Event projection: quantify uncertainty and manage expectations of broader teams

Kaspar Rufibach
Department of Biostatistics, Roche Basel
Methods, Collaboration & Outreach Group
Basel, 28th April 2016

Generated 2016-04-24 at 21:13:16.

Agenda

1. Challenge
2. How to predict analysis timepoint in general?
3. Hybrid estimate of S: Kaplan-Meier with Exponential tail
4. Predicted analysis timepoints
5. Uncertainty in event prediction
6. Small proportion of events
7. Remarks

Agenda

Challenge

Clinical trial with **time-to-event** outcome:

- Portion of patients accrued, some patients still to be accrued.
- Interim or final analysis after **pre-specified number of events**.

Questions:

1. **When** does this number of events happen?
2. Can we assign a **confidence interval** around that timepoint?
3. How to manage the potentially large uncertainty in clinical teams (and higher up)?

Implications:

- Resource allocation for running the trial.
- Gating of other trials, e.g. early phase combinations.
- Keep investors happy.

Example

Phase 3 trial with time-to-event primary endpoint (PFS).

All $n = 1202$ patients accrued: FPI 2010-04-06, LPI 2012-11-06.

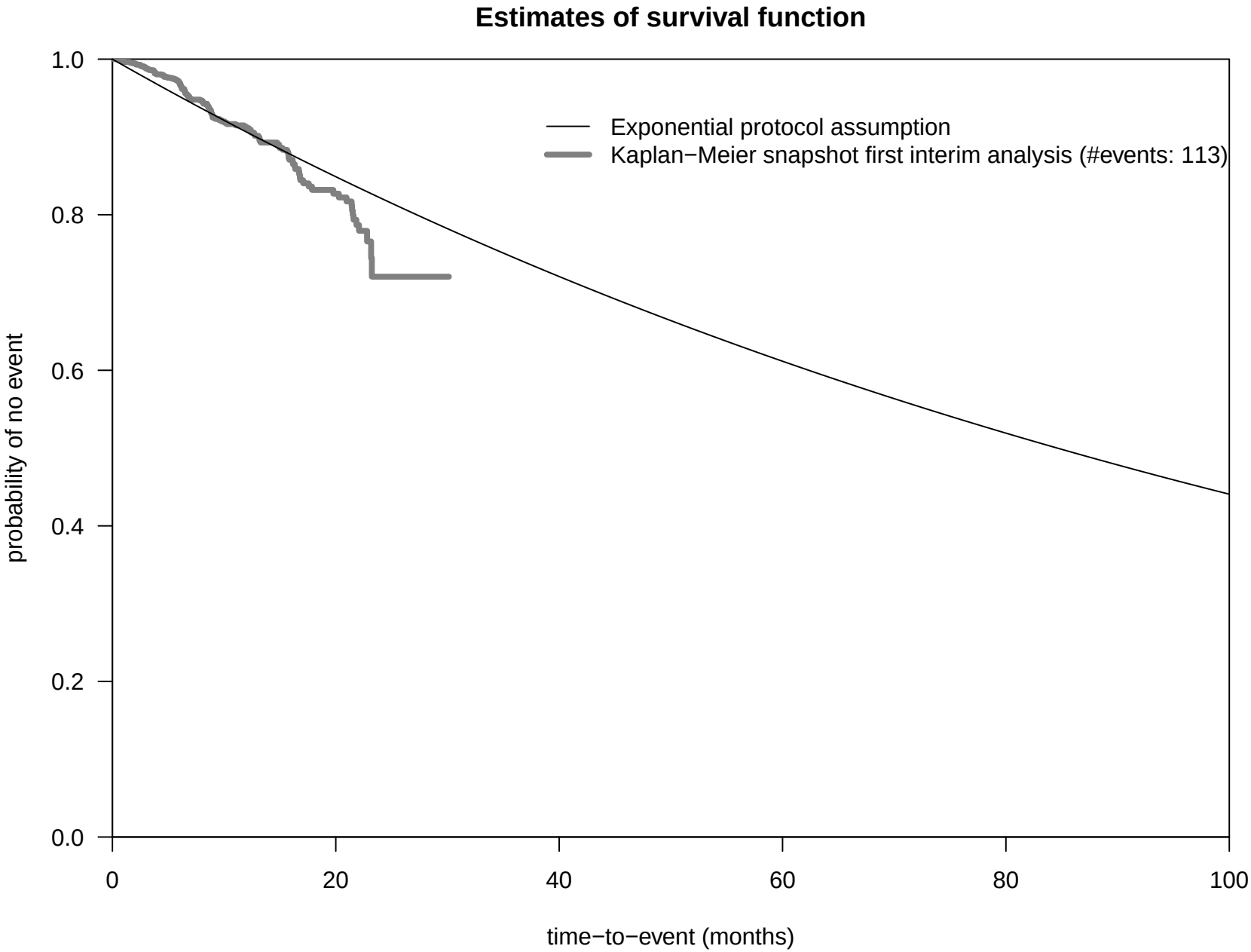
First interim analysis:

- Clinical cutoff date: 2012-11-21.
- #events: **113**.

Based on this interim data $\Rightarrow$ predict time when we see

- **248** events: second interim,
- **370** events: final analysis.

Kaplan-Meier estimate based on snapshot



Agenda

Expected number of events at T

S: chosen or estimated survival function.

Expected number of events m at T:

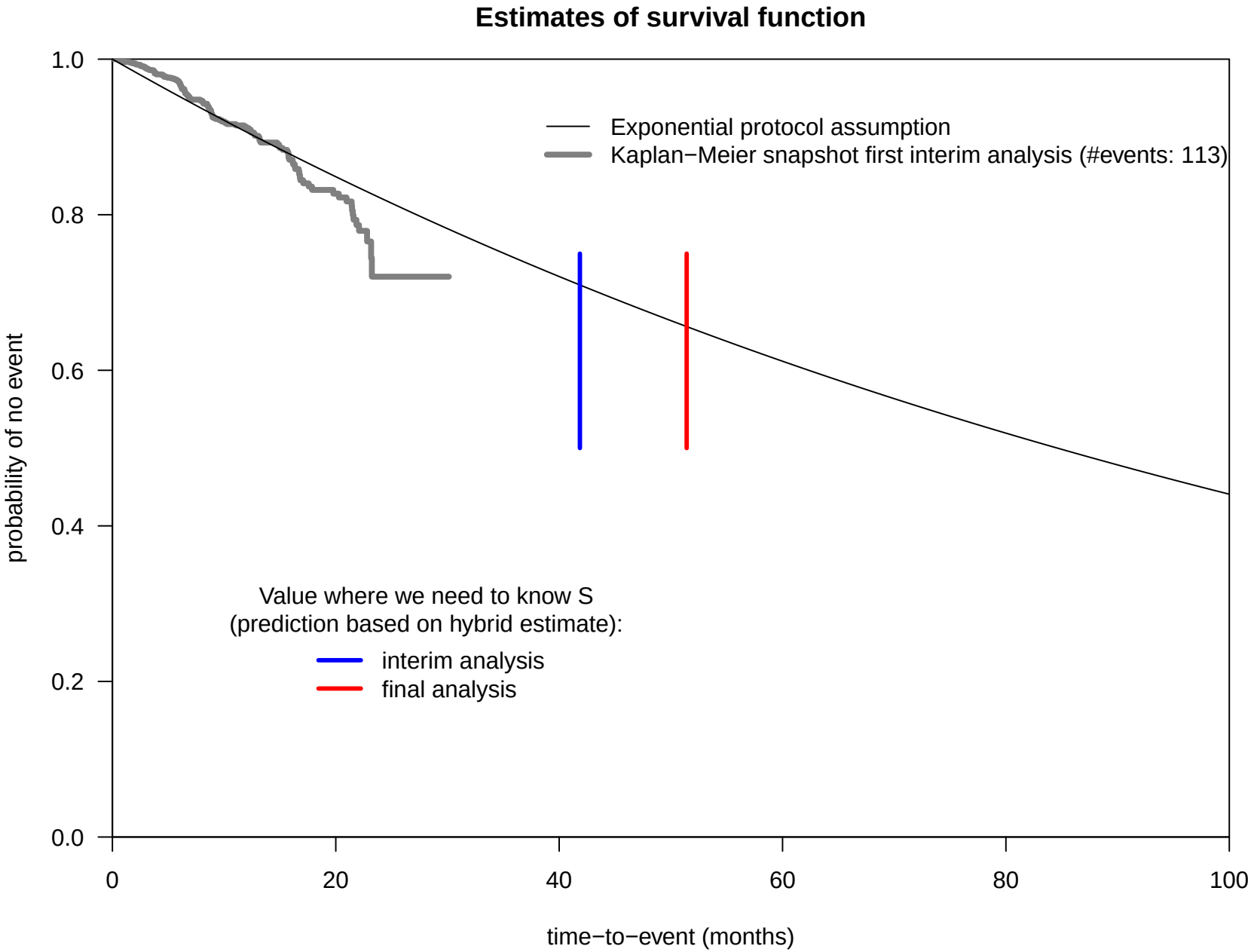
$$
\begin{aligned}
m(S) &= \sum_{h \in \mathcal{I}_1} 1 + \sum_{j \in \mathcal{I}_2} P\Big(\text{event in } (t_j, T] | \text{no event up to } t_j\Big) + \sum_{j \in \mathcal{I}_3} P\Big(\text{event in } (a_i, T]\Big) \\
&= n + \sum_{j \in \mathcal{I}_2} \frac{S(t_j) - S(T)}{S(t_j)} 1\{T > t_j\} + \sum_{j \in \mathcal{I}_3} \Big(1 - S(T - a_i)\Big) 1\{T > a_i\}.
\end{aligned}
$$

Sets of patients:

- $\mathcal{I}_1$: patients who already had event, n.
- $\mathcal{I}_2$: patients censored at t_j.
- $\mathcal{I}_3$: patients to be recruited at a_i.

Need to have value for S at least **up to** T.

Kaplan-Meier for current data



How to estimate S?

Nonparametrically via Kaplan-Meier or kernel estimate:

+ No or few assumptions about true $S \Rightarrow$ unbiased.
+ Can account for the "large steps" due to schedule of assessment.
- High variability, especially in tail.
- No extrapolation beyond last event / censoring time.

Fully **parametric**, e.g. Exponential or Weibull:

+ Efficient, if assumption is true.
+ Accurate estimate of tail, if assumption is true.
+ Can estimate S beyond where we have data.
- Biased if assumption not true.
- Not able to capture particular features of oncological time-to-event data (inspection intervals).

Hybrid approach: Use nonparametric where we have data, complement with parametric tail.

Agenda

Hybrid estimate

Proposed in Fang and Su (2011). Recipe:

1. Choose total number of change points k.
2. Estimate piecewise Exponential hazard with k change points.
3. Test for "significance" of change points.

Borrows strength of nonparametric and parametric approach.

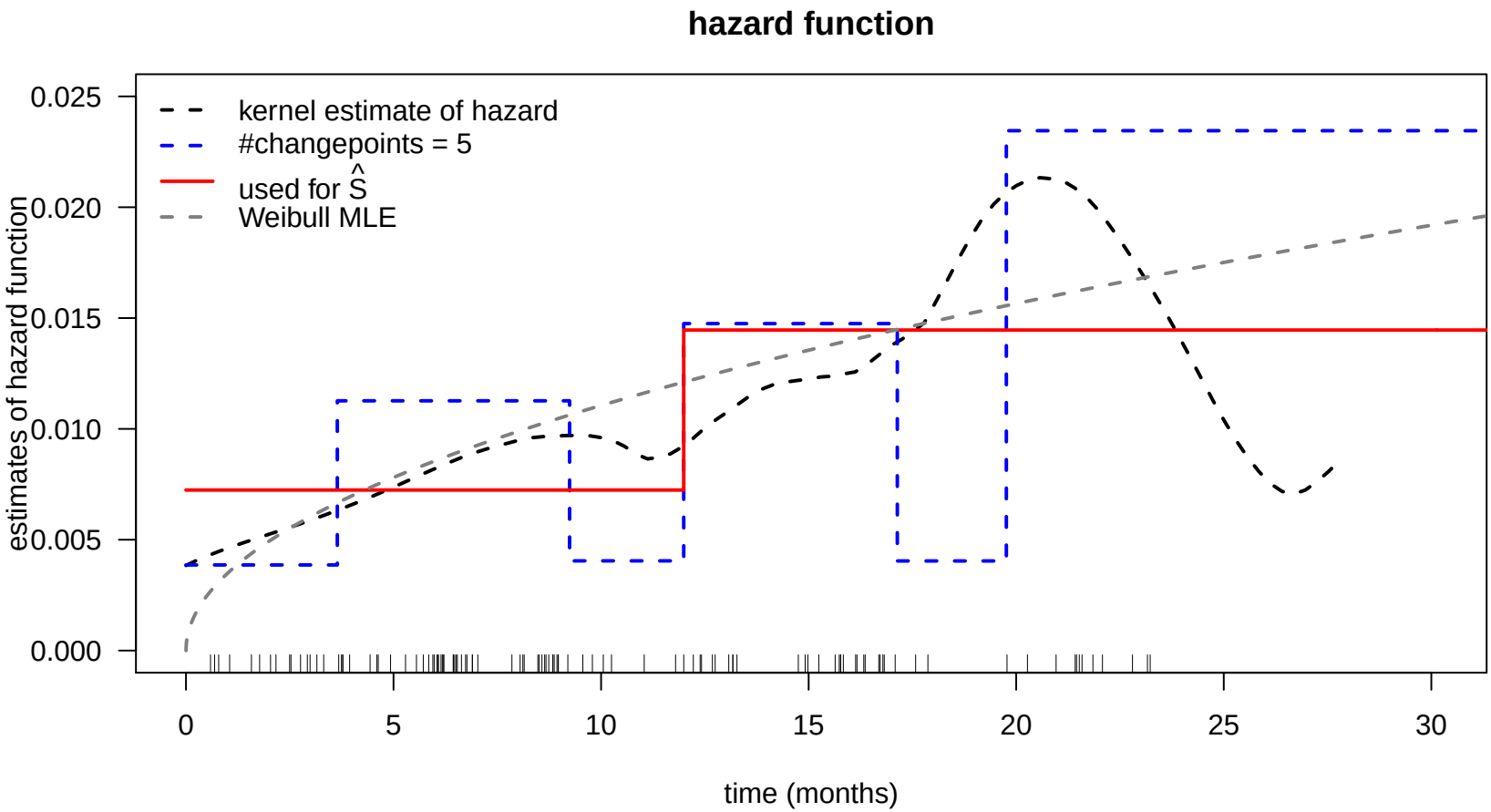
Allows to capture "big steps" due to schedule of assessments.

Change point:

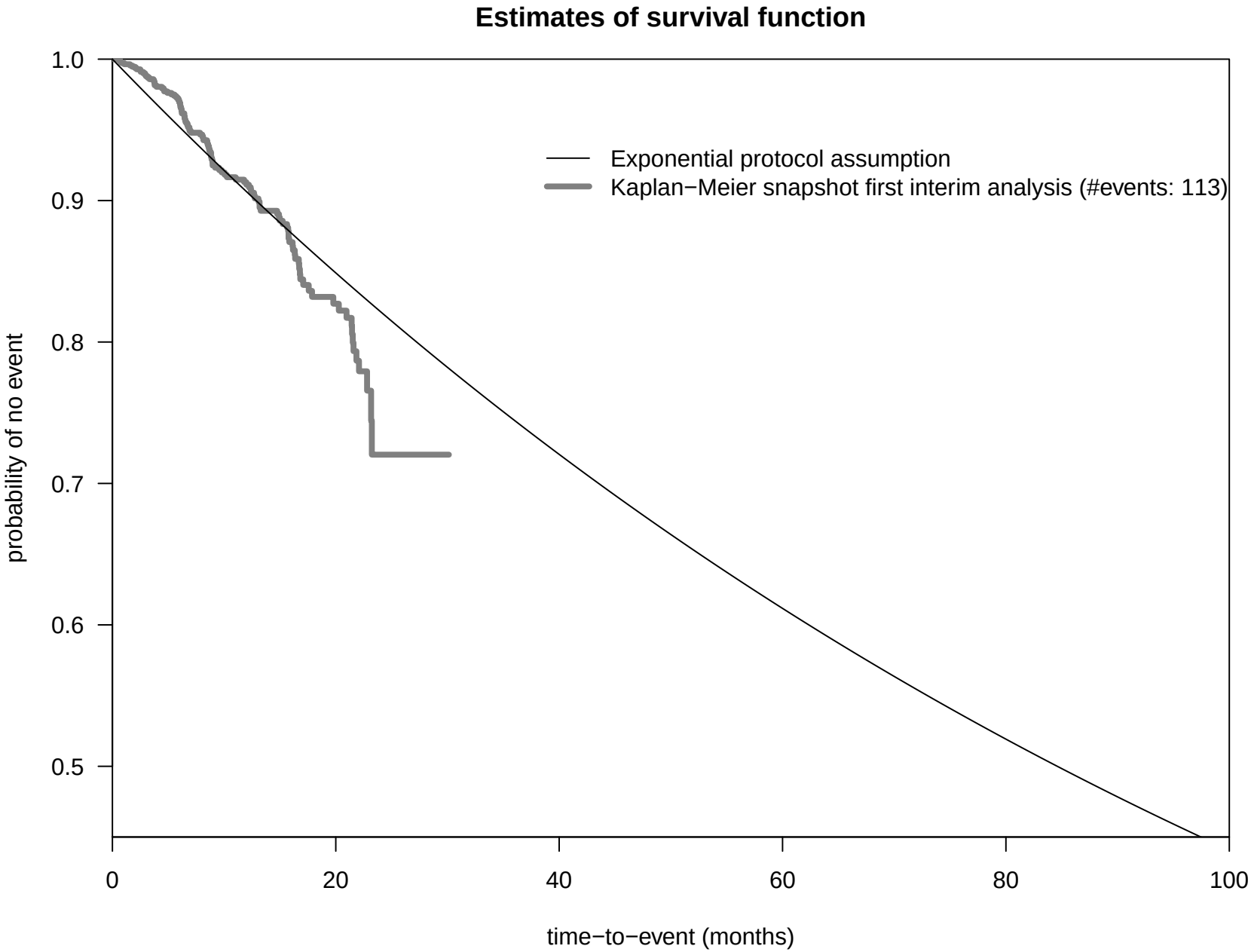
- Sequential test according to Goodman et al. (2011).
- To avoid **overfitting** correct α according to Lan and DeMets (1983).
- No change point: pure Exponential fit.
- ≥ 1 change point: Kaplan-Meier prior to **selected** change point, Exponential tail fit beyond selected change point.

Alternatively: Bagiella and Heitjan (2001), Ying et al. (2004), Ying and Heitjan (2008), Di et al. (2016).

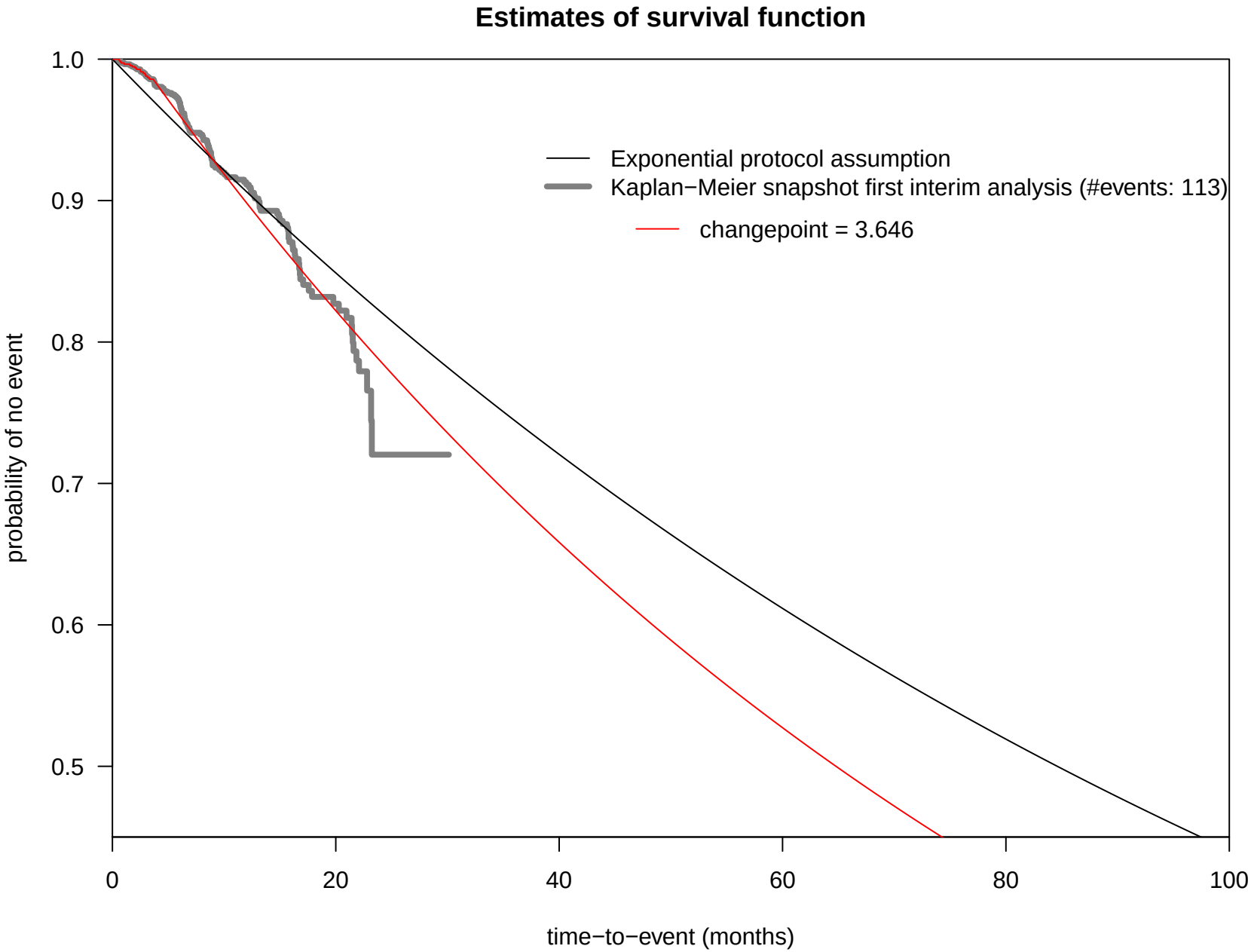
Apply hybrid approach



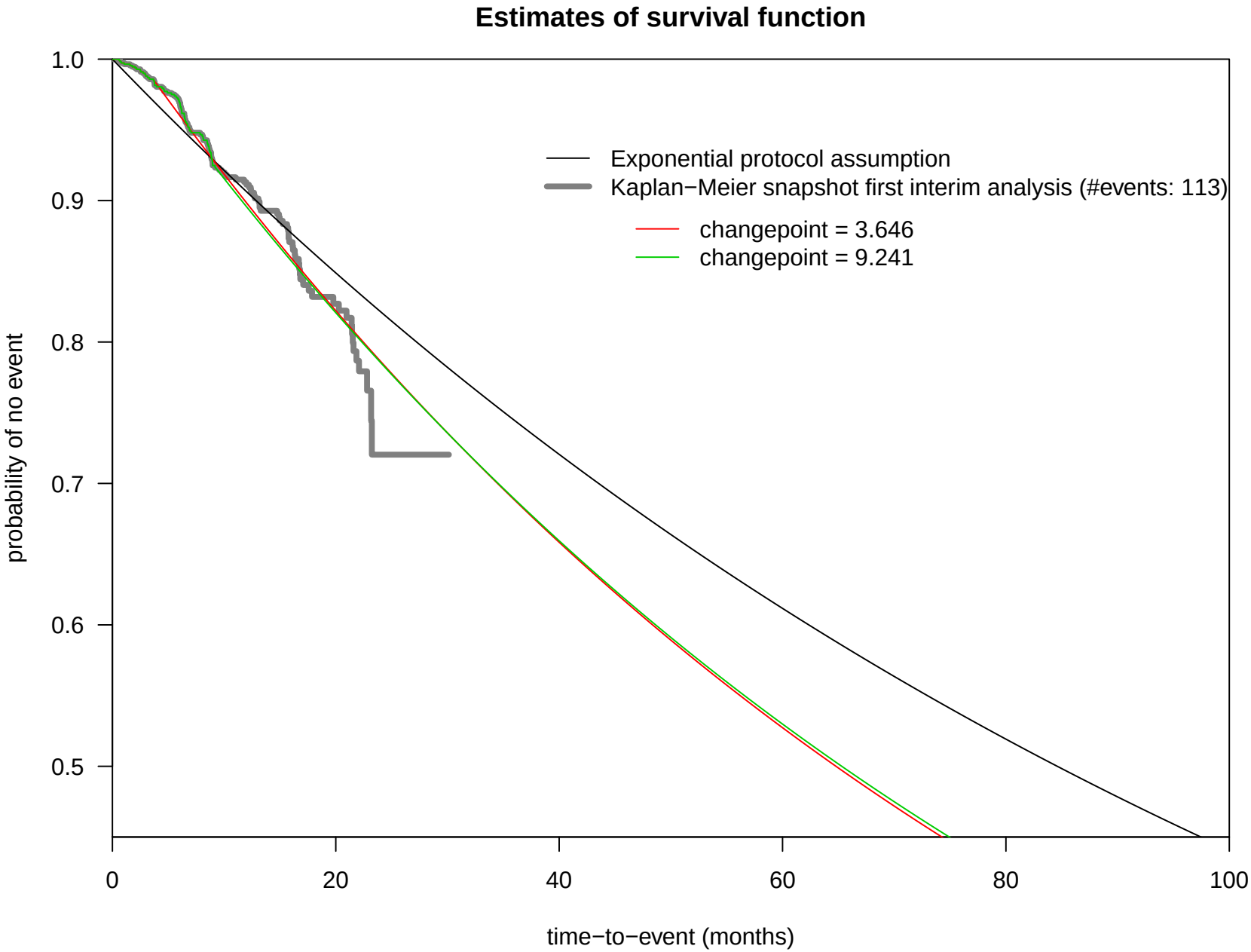
Kaplan-Meier for current data



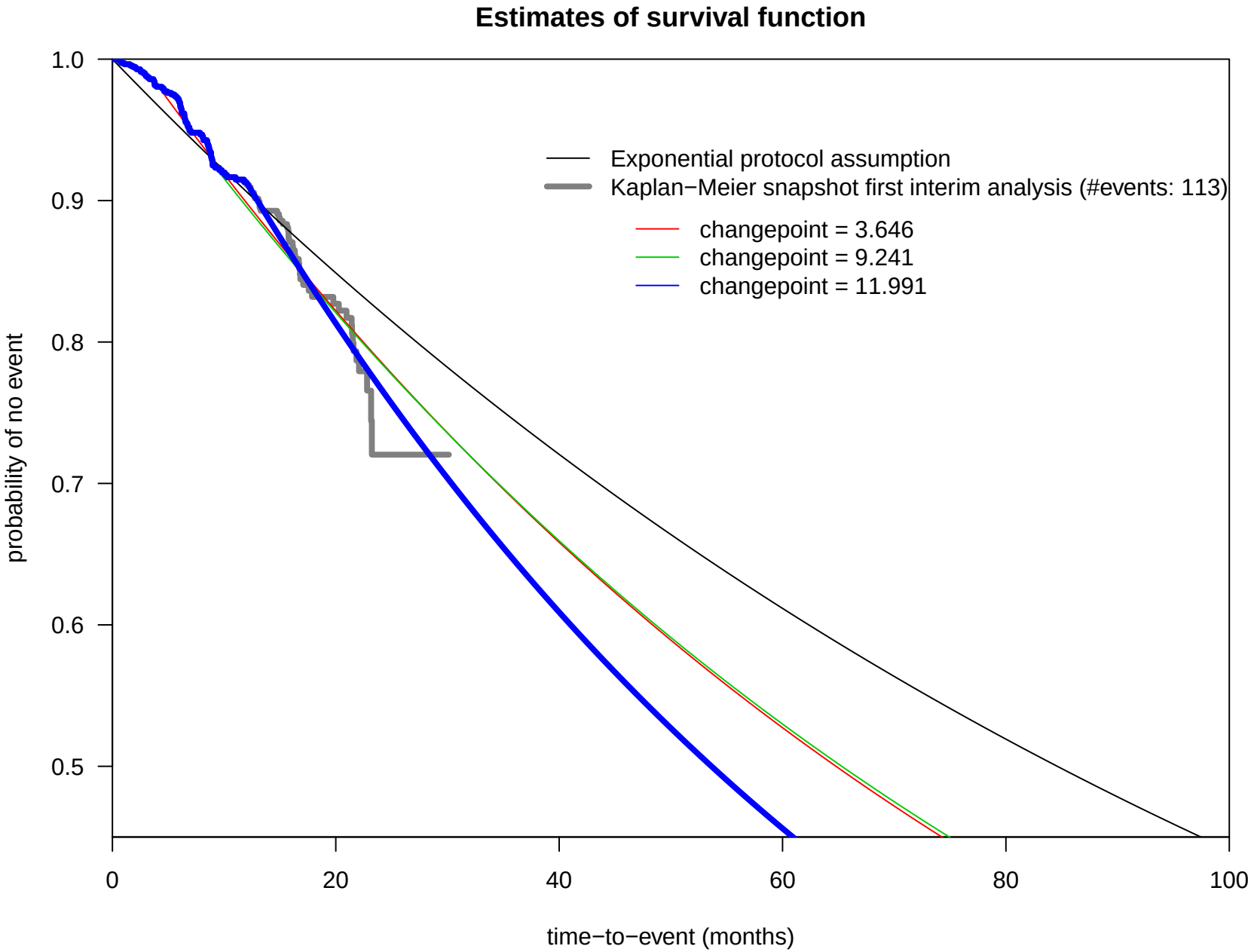
Kaplan-Meier for current data



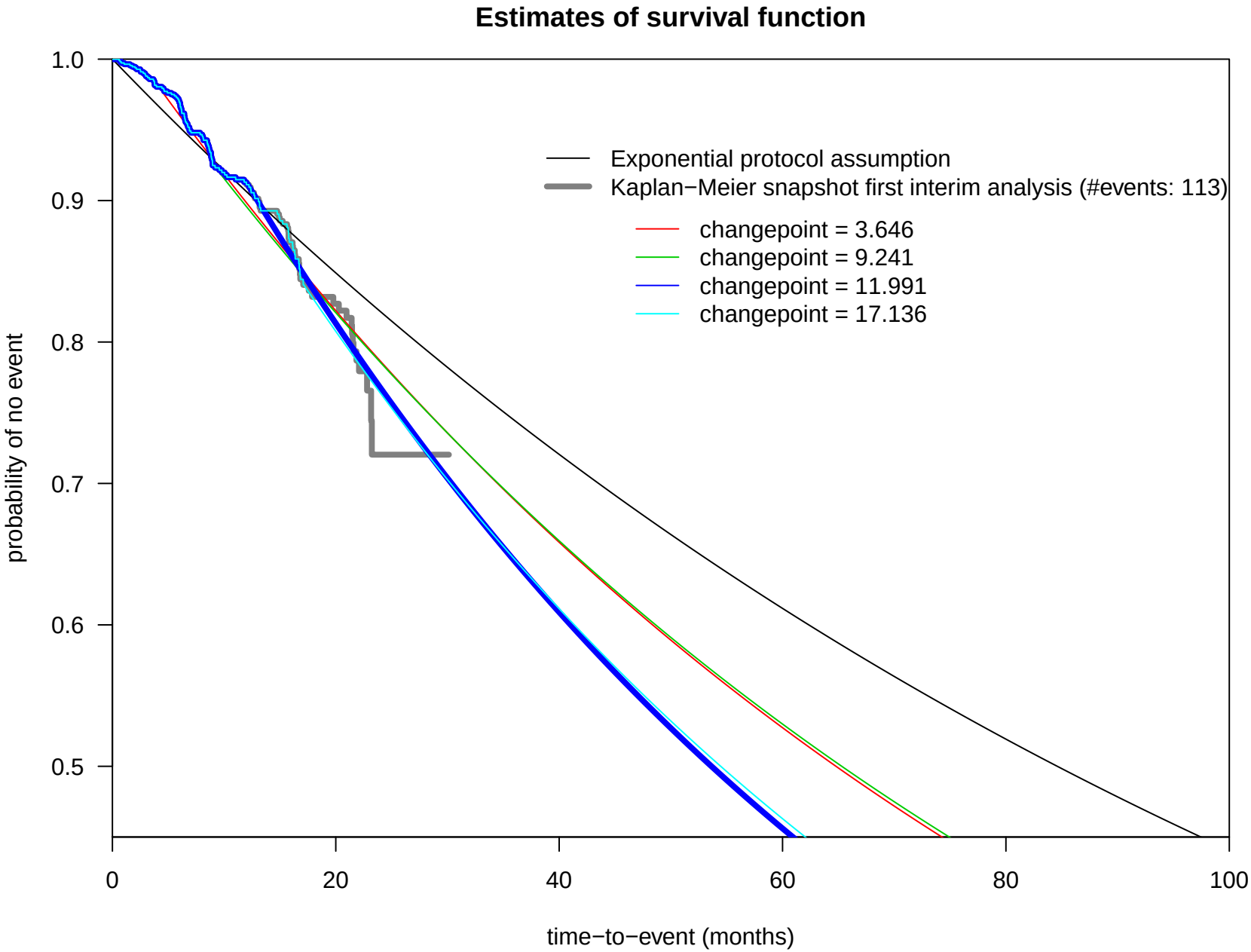
Kaplan-Meier for current data



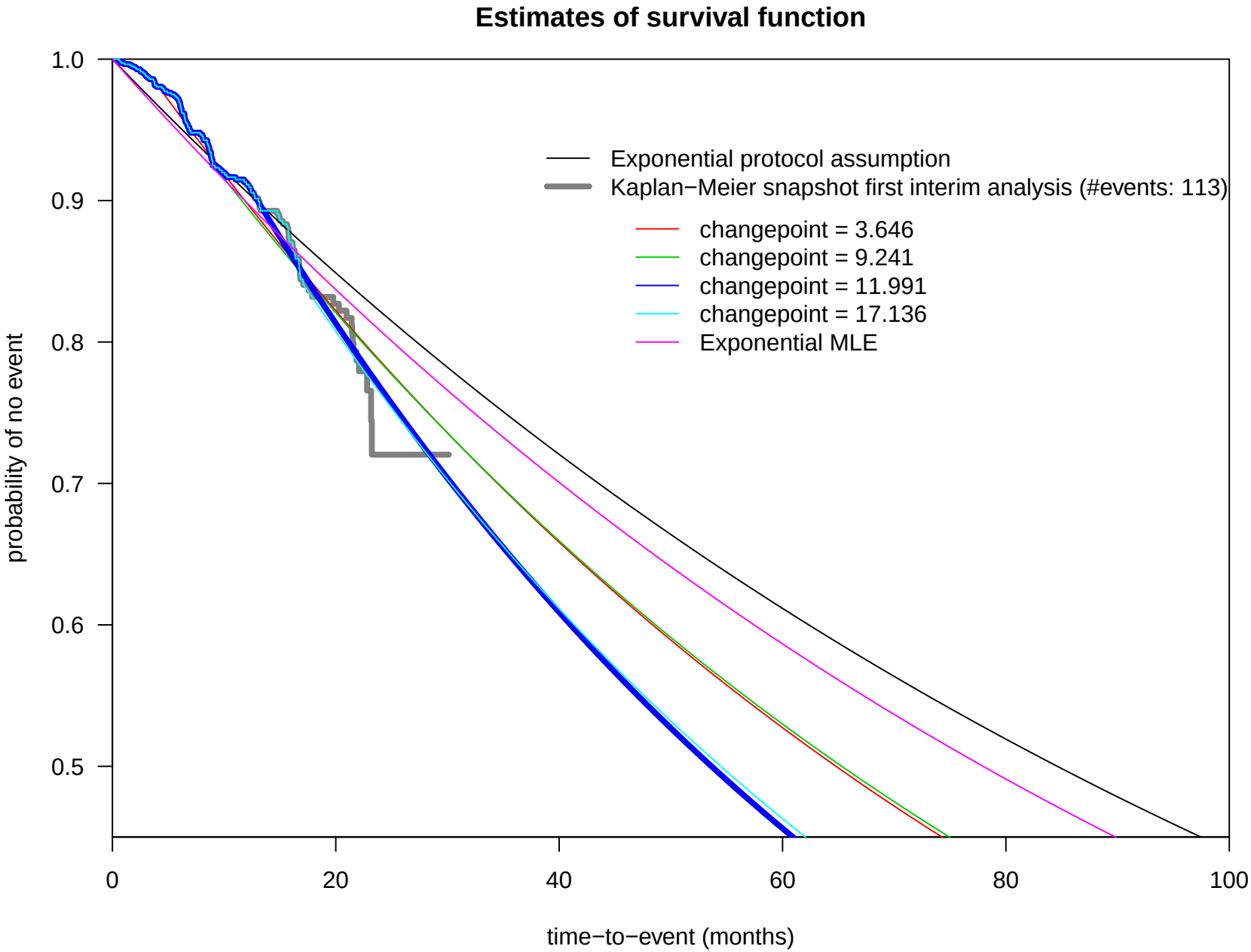
Kaplan-Meier for current data



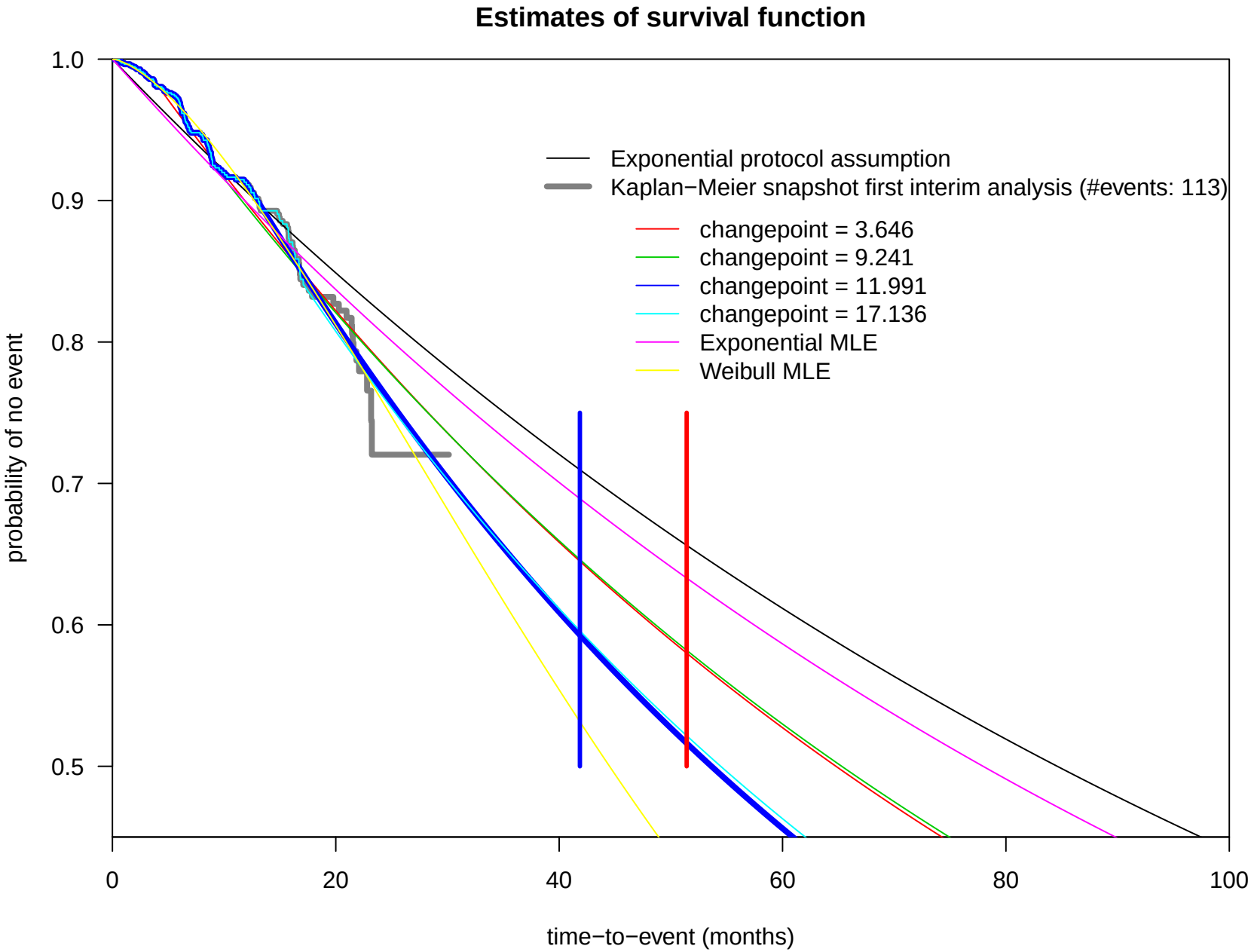
Kaplan-Meier for current data



Kaplan-Meier for current data

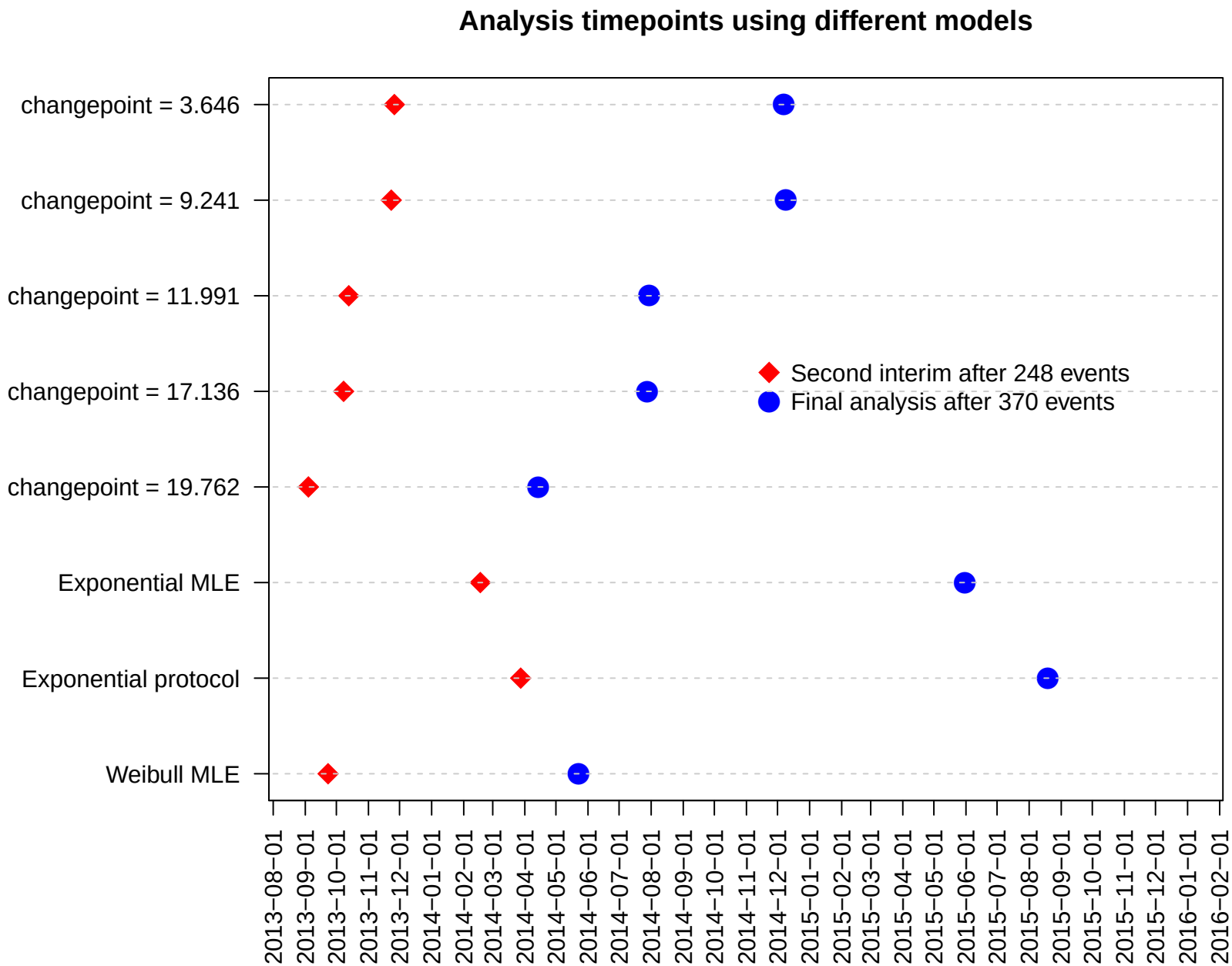


Kaplan-Meier for current data

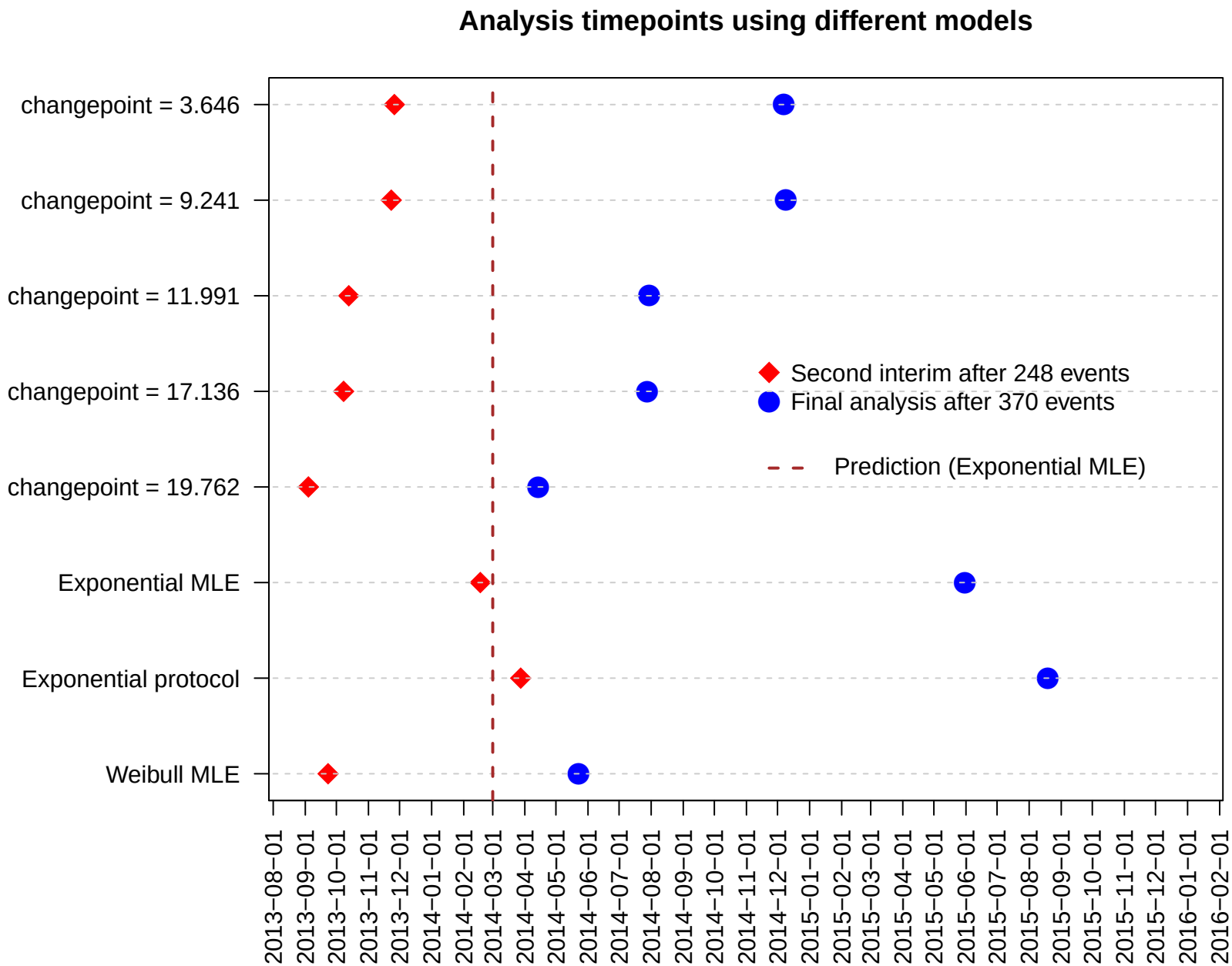


Agenda

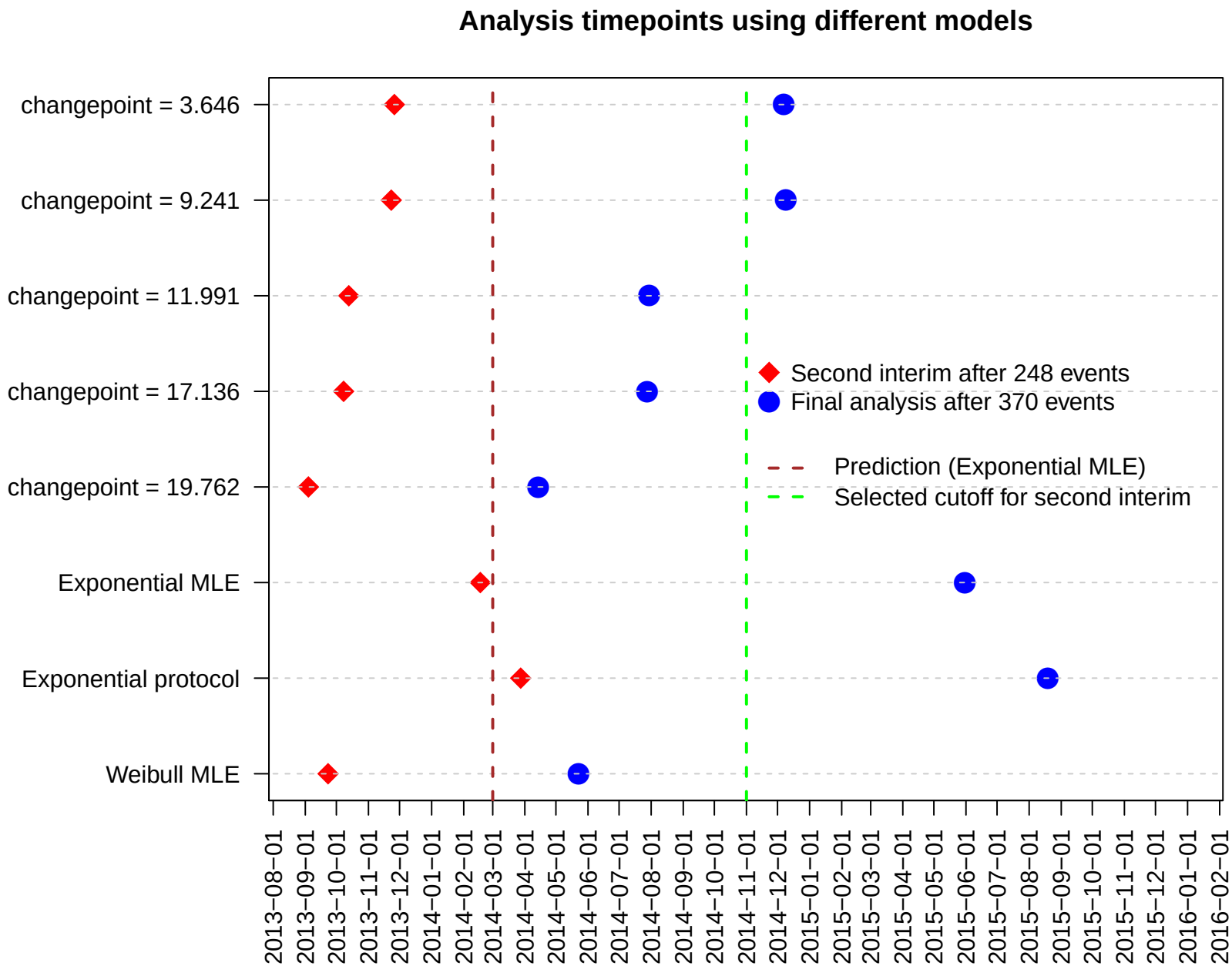
Predicted analysis timepoints



Predicted analysis timepoints



Predicted analysis timepoints



Shift

Communicated prediction: Exponential MLE $\Rightarrow$ synthesis of protocol assumption and prediction based on hybrid estimate.

Prediction of second interim analysis:

- Data: 113 events in 1202 patients.
- Clinical cutoff: 2012-11-21.
- Initial prediction: 2014-03-01.
- Actual cutoff: 2014-11-01.

Shift of cutoff: 8.0 months.

8.0 months!

p.s.: 8.0 months earlier would certainly have been ok, but 8.0 months later...

Agenda

Uncertainty in event prediction

Recall formula:

$$m(S) \quad = \quad n \; + \; \sum_{j \in \mathcal{I}_2} \frac{S(t_j) - S(T)}{S(t_j)} 1\{T > t_j\} \; + \; \sum_{j \in \mathcal{I}_3} \Big(1 - S(T - a_i)\Big) 1\{T > a_i\}.$$

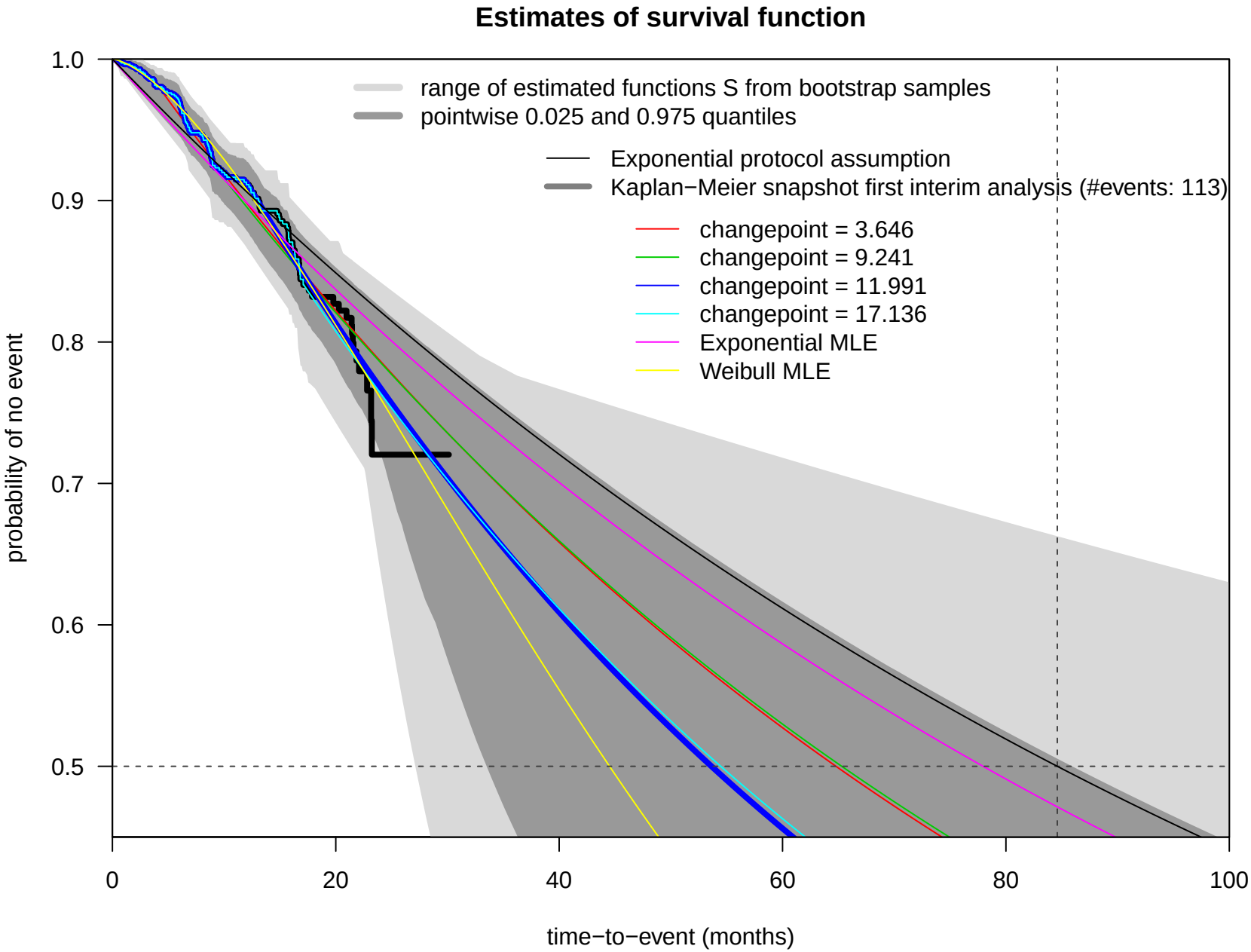
Trial was fully recruited when prediction made $\Rightarrow$ uncertainty = **Sampling error** in estimation of S.

Proposal:

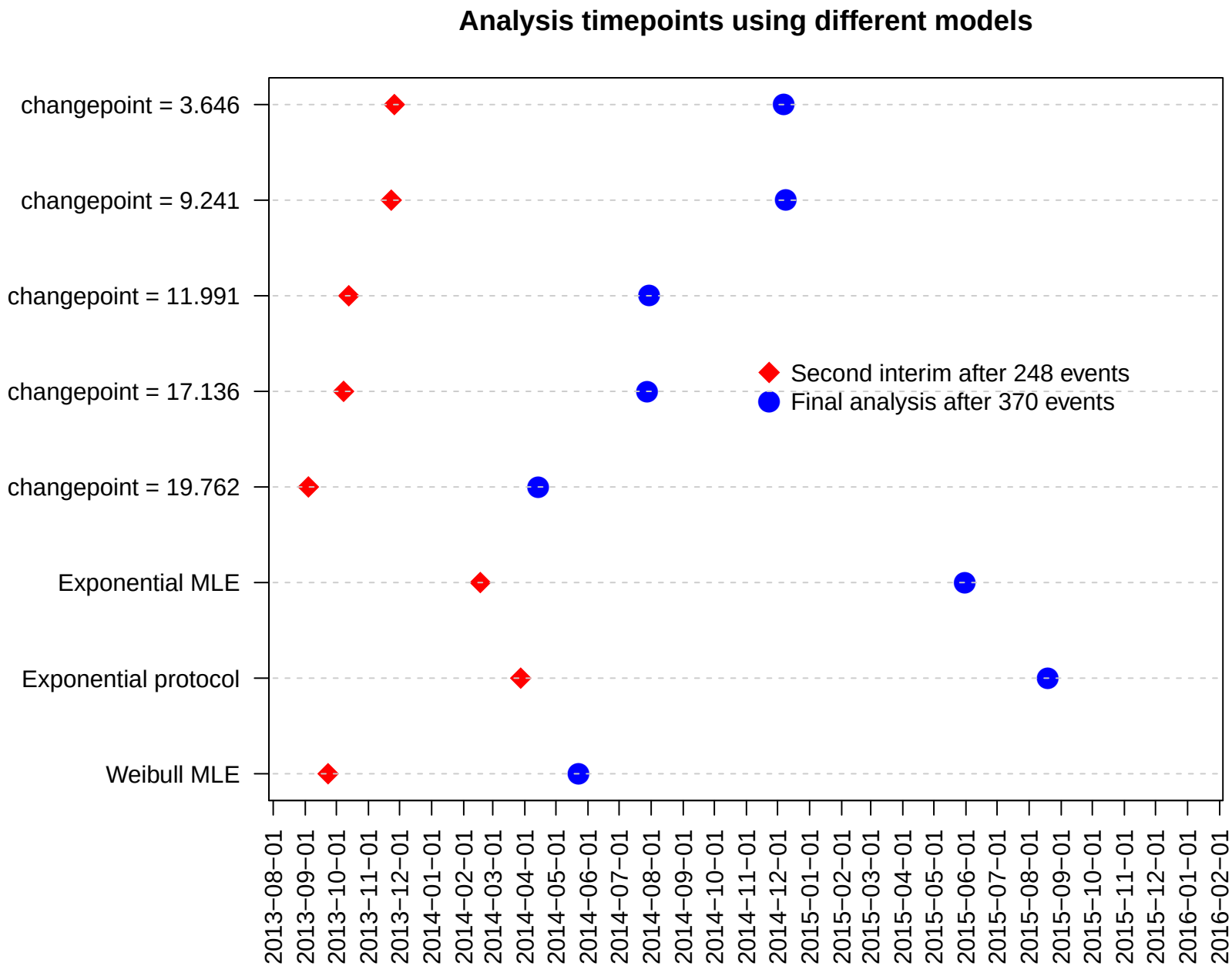
- **Bootstrap** time-to-event data: sample with replacement from (X_i, δ_i).
- Re-estimate S via hybrid Exponential model for each bootstrap sample, choose change point based on sequential test $\Rightarrow$ S estimated **fully automatic**.
- Compute analysis timepoint in each sample.
- Compute quantiles of these analysis timepoints $\Rightarrow$ bootstrap percentile confidence interval for analysis timepoint.

Validity of bootstrap for right-censored data: Efron (1981), see also Akritas (1986).

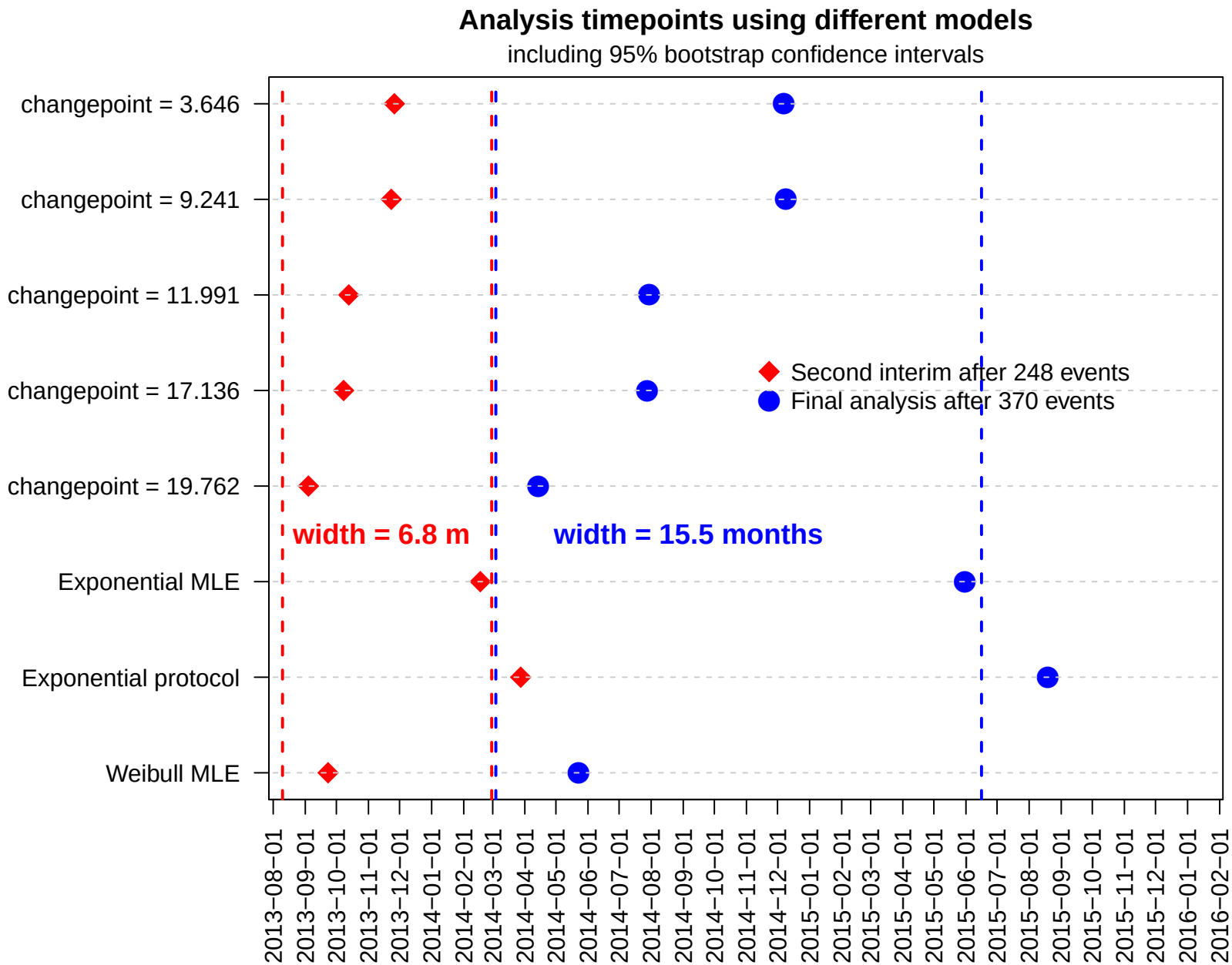
Bootstrapped survival function estimates



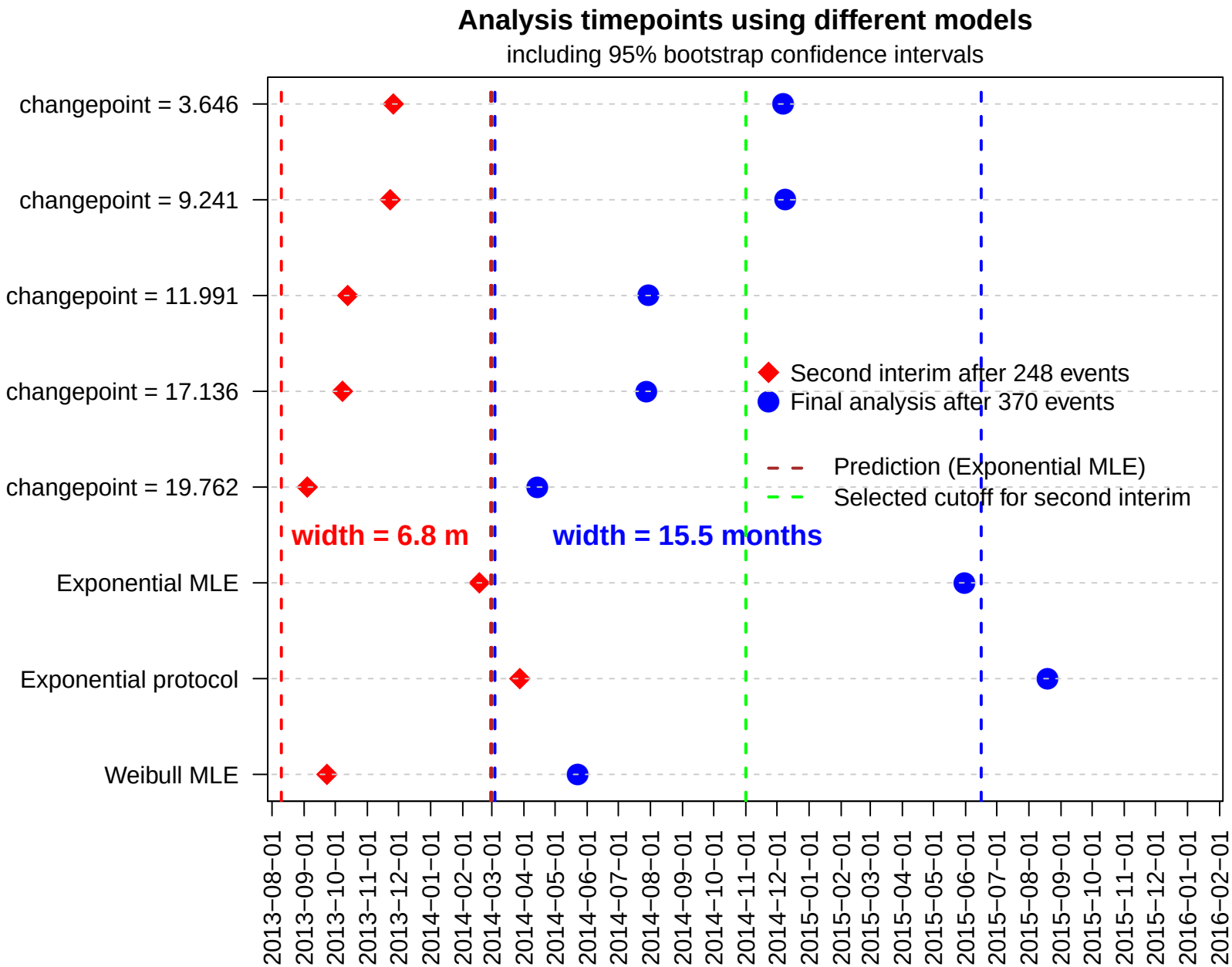
Analysis timepoints



Analysis timepoints

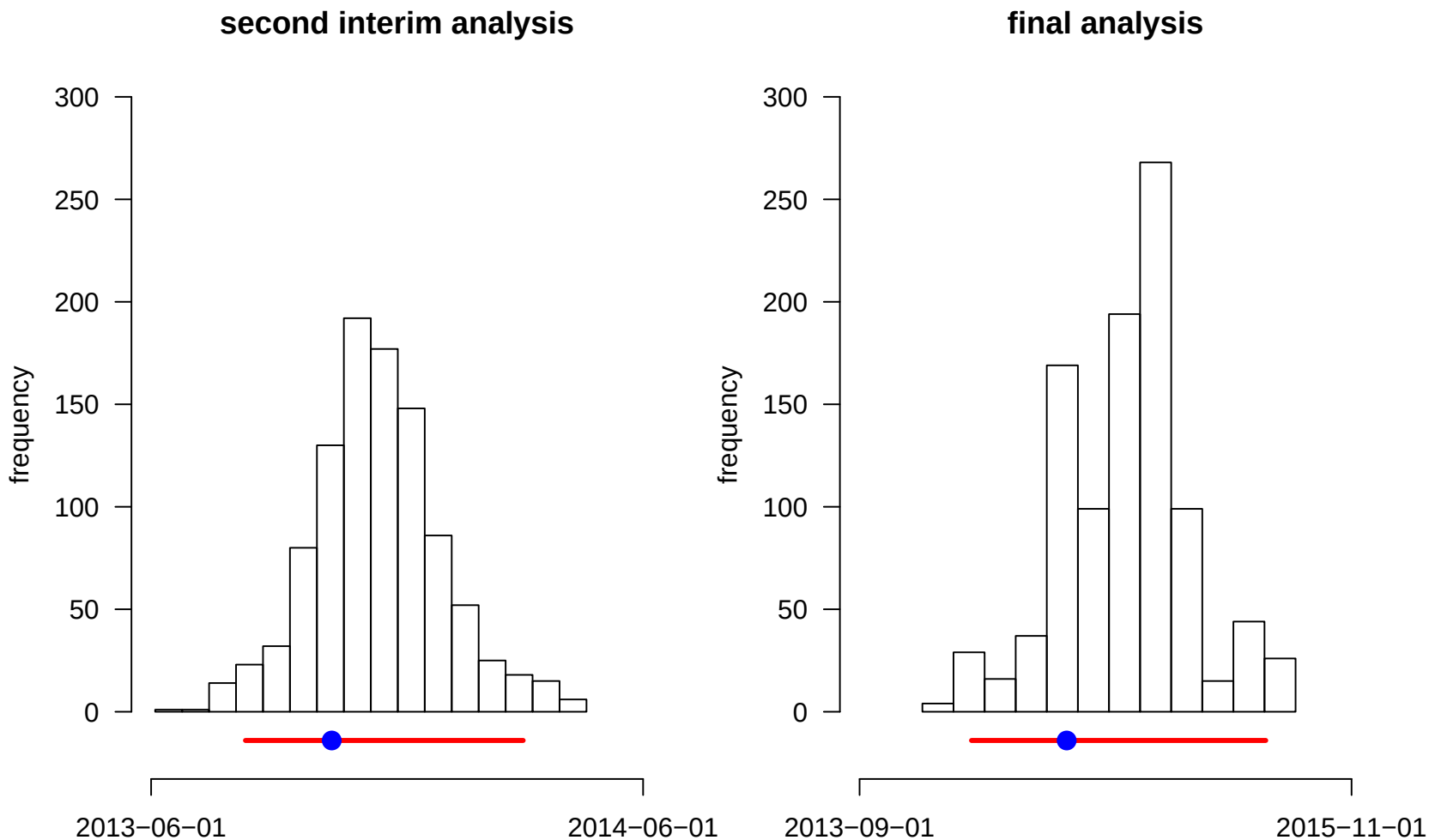


Analysis timepoints



Bootstrap distributions

Histograms of analysis timepoints, based on 1000 bootstrap samples
Blue dot: point estimate based on hybrid estimate. CCOD 2012−11−21, 113 events.



Confidence intervals

Statistically sound.

Difficult to communicate to Big Boss. Initially, we did not communicate confidence interval, only gave point estimate.

Statistician:

> *Your interim analysis cutoff will be reached between 2013-08-10 and 2014-03-04, with 95% confidence.*

Confidence intervals

Statistically sound.

Difficult to communicate to Big Boss. Initially, we did not communicate confidence interval, only gave point estimate.

Statistician:

> *Your interim analysis cutoff will be reached between 2013-08-10 and 2014-03-04, with 95% confidence.*

Big boss:

> *Fine.*

Confidence intervals

Statistically sound.

Difficult to communicate to Big Boss. Initially, we did not communicate confidence interval, only gave point estimate.

Statistician:

> *Your interim analysis cutoff will be reached between 2013-08-10 and 2014-03-04, with 95% confidence.*

Big boss:

> *Fine.* **Give me the date**.

Agenda

Small proportion of events - large variability

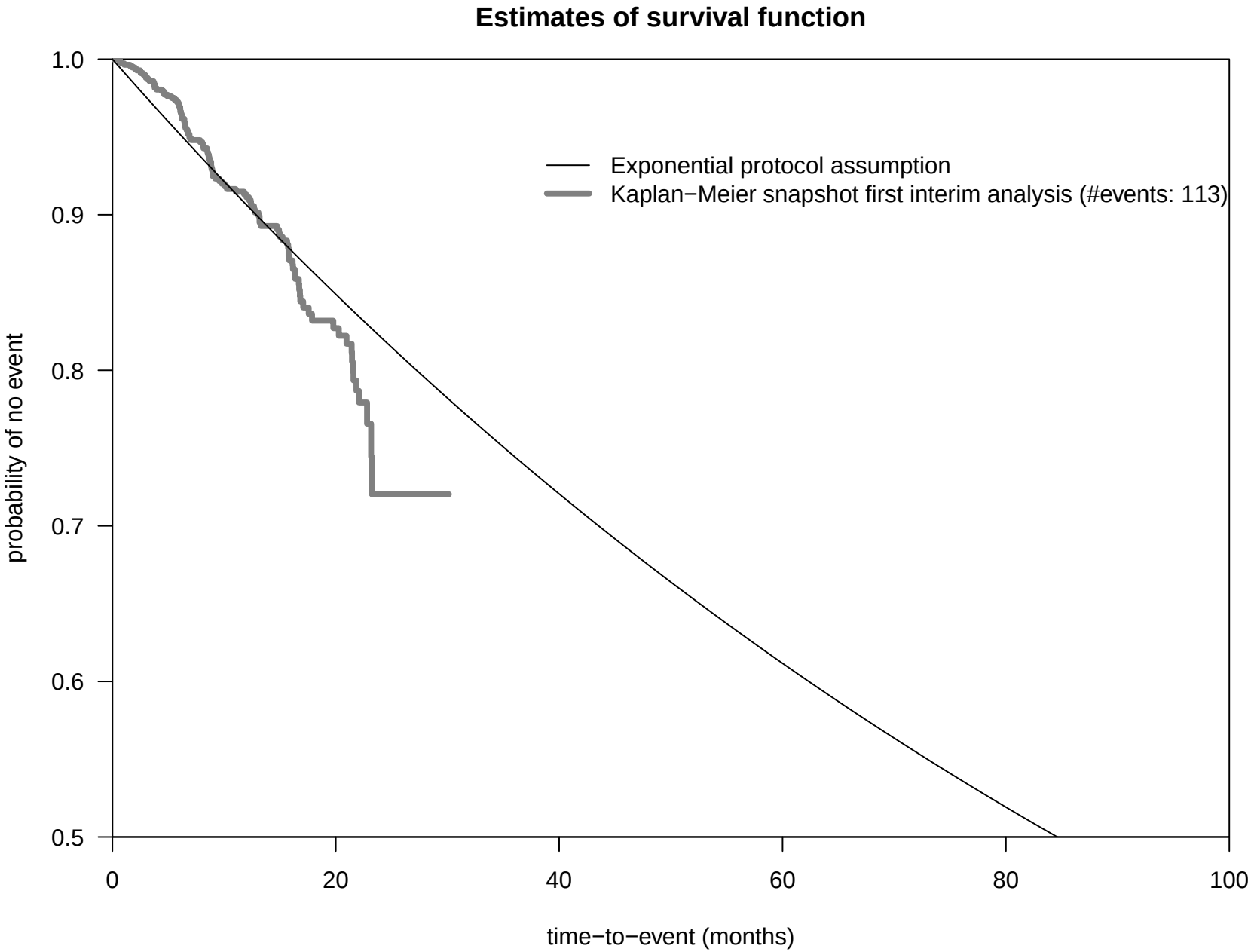
Approx. 1y after initial prediction.

- Data cleaning milestone.
- Based on 193 events became clear we need to shift cutoff.

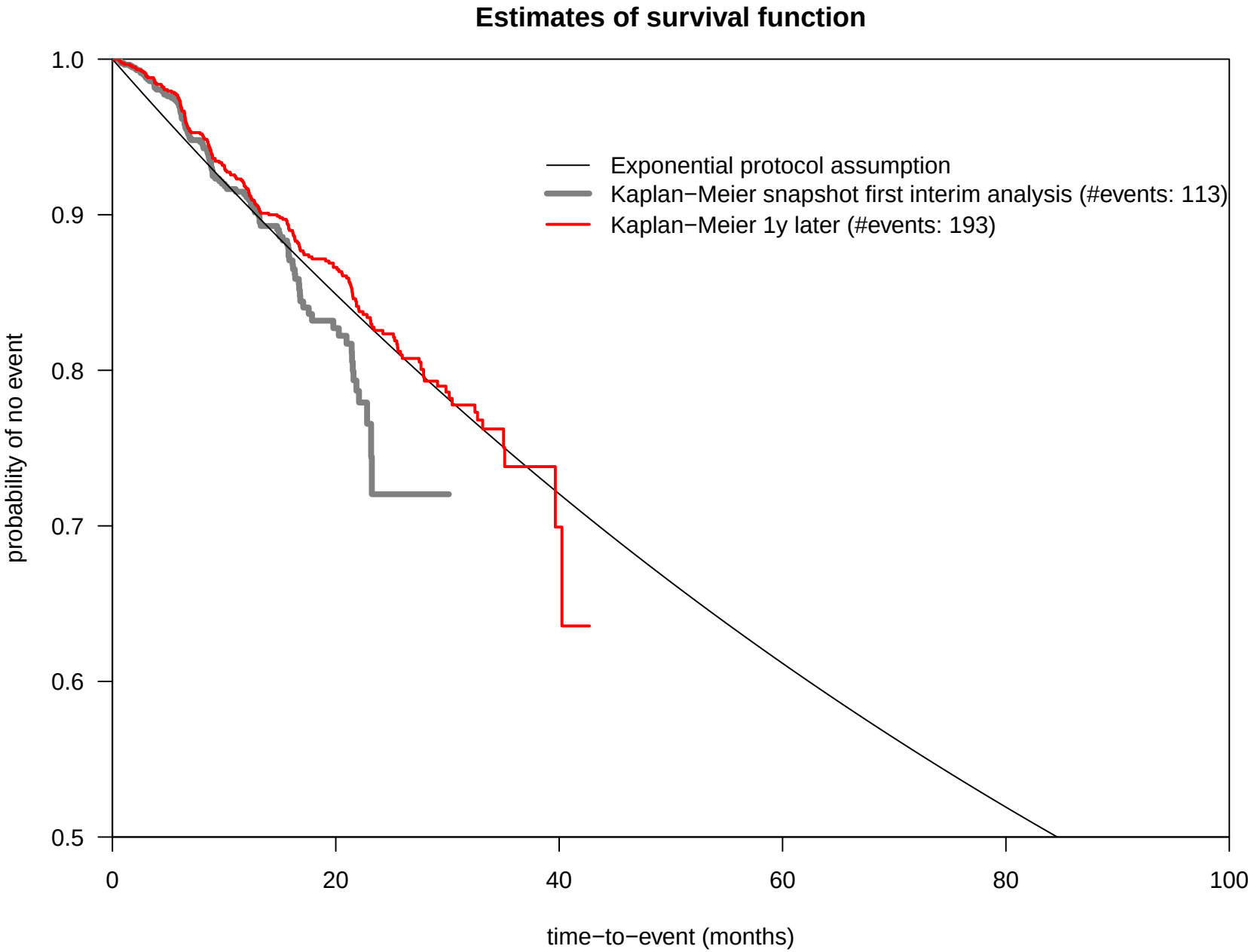
"Root-cause" analysis for 8.0 months shift.

Statistics team explained uncertainty around predictions $\Rightarrow$ estimates of S quite different based on two snapshots.

Variability in Kaplan-Meier estimate over time



Variability in Kaplan-Meier estimate over time



Big Boss comes back

Big Boss:

> *Seems you had more events in relation to amount of follow-up one year ago? How good was your cleaning back then anyway?*

Maybe we had all events, but tumor assessment follow-up for those event-free was not comprehensive?

Compare:

- Data Snapshot 1 of first interim analysis: 3892 tumor assessments.
- Snapshot 2 one year later, **cut back to first interim analysis cutoff**: 3895 tumor assessments.
- Proportion of tumor assessments available in snapshot 1 if Snapshot 2 is considered "full dataset": 3892 / 3895 = 99.923%.

Message: we had everything back then. Difference in estimates

- either due to variability,
- or **early** events?

Agenda

Remarks

- Hybrid estimate of $S \Rightarrow$ provides estimated value of S beyond last data point.
- Confidence intervals wide.
- In fact used prediction based on Exponential MLE. Could simply have used that standard error.
- Typically: uncertainty driven by #events. But here: need to predict event time for those **who had no event yet**. Uncertainty driven by proportion of events at prediction (113) to patients (1202)? More research needed?
- Sources of variability not considered: drop-out.
- Not discussed, but essential: do such projections only on **cleaned** data. Otherwise, you likely have most of the events, but no comprehensive follow-up.
- Communication is key. But difficult.

Technical comments

- Once k **is selected** $\Rightarrow$ automatic method to do event projection $\Rightarrow$ **unambiguity**.
- Dependence on k $\Rightarrow$ sensitivity analysis.
- Point estimate might not lie in center of bootstrap confidence interval. Use alternative function of bootstrap sample (mode, median).
- R package **eventTrack**, available internally. Everything implemented from scratch $\Rightarrow$ likely not optimally efficient.
- Confidence interval: 1202 patients and 1000 bootstrap samples: takes about 3h. Depends on T, k.

Thank you for your attention.

References

- Akritas, M. (1986). Bootstrapping the Kaplan-Meier Estimator. *J. Amer. Statist. Assoc.* **81** 1032–1038.
- Bagiella, E. and Heitjan, D. F. (2001). Predicting analysis times in randomized clinical trials. *Stat. Med.* **20** 2055–2063.
- Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat. Med.* **19** 1141–1164.
- Di, J., Wang, D., Brashear, H. R., Dragalin, V. and Krams, M. (2016). Continuous event monitoring via a bayesian predictive approach. *Pharmaceutical Statistics* **15** 109–122.
- Efron, B. (1981). Censored Data and the Bootstrap. *J. Amer. Statist. Assoc.* **76** 312–319.
- Fang, L. and Su, Z. (2011). A hybrid approach to predicting events in clinical trials with time-to-event outcomes. *Contemp Clin Trials* **32** 755–759.

References

- Goodman, M. S., Li, Y. and Tiwari, R. C. (2011). Detecting multiple change points in piecewise constant hazard functions. *J Appl Stat* **38** 2523–2532.
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis*. 2nd ed. Springer-Verlag.
- Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70** 659–663.
- Yao, Y. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Comm. Statist. Theory Methods* **15** 2455–2466.
- Ying, G. S. and Heitjan, D. F. (2008). Weibull prediction of event times in clinical trials. *Pharm. Stat.* **7** 107–120.
- Ying, G. S., Heitjan, D. F. and Chen, T. T. (2004). Nonparametric prediction of event times in randomized clinical trials. *Clin Trials* **1** 352–361.

Backup slides.

Piecewise constant hazard

Piecewise constant hazard $\Rightarrow$ piecewise Exponential density or survival function.

"Bridges" parametric and nonparametric model.

Assume for a **given** number of change points k, hazard function h, and $0 = \tau_0 < \tau_1 < \ldots < \tau_{k+1} = \infty$:

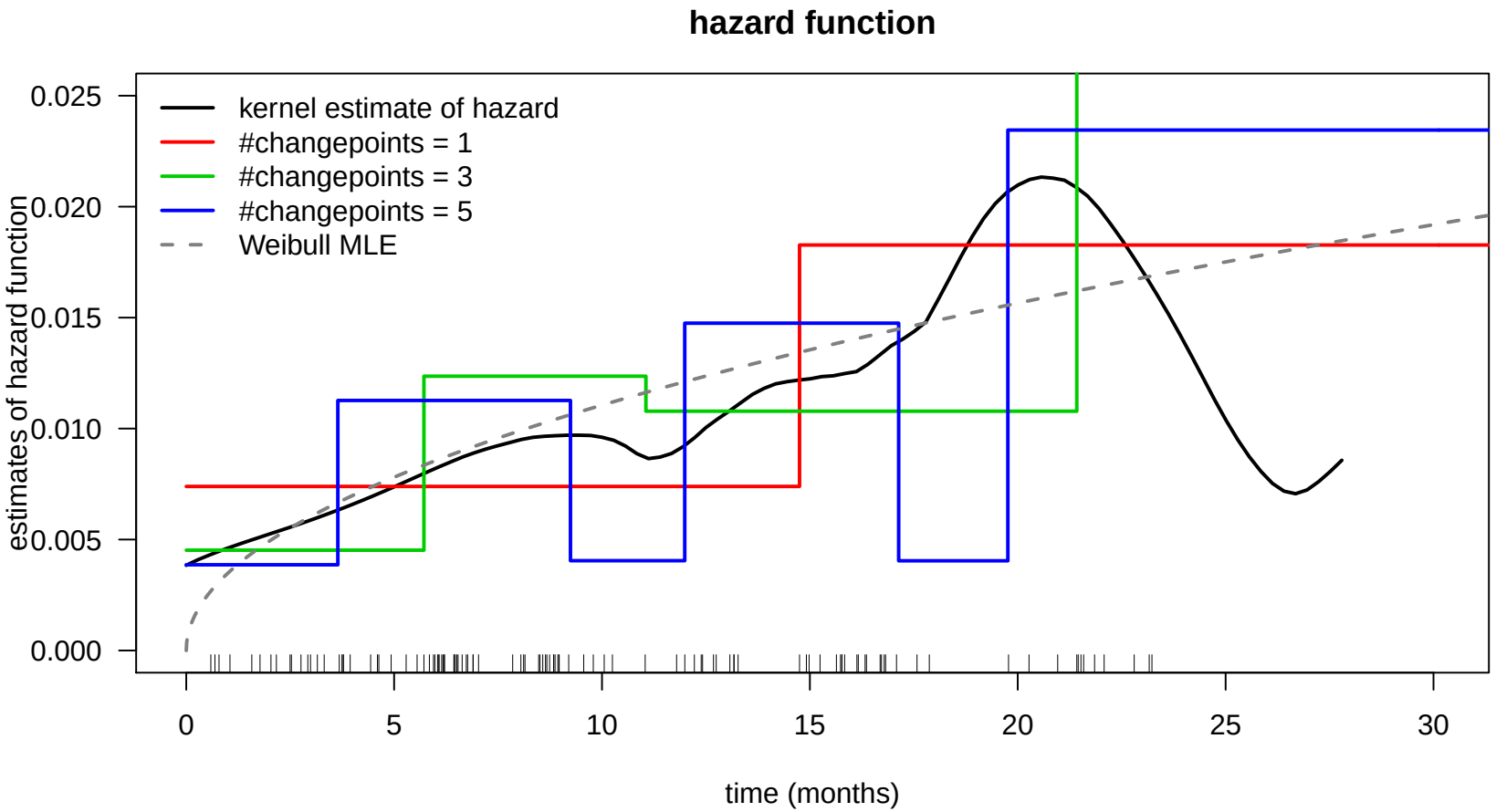
$$h(t) \quad = \quad \begin{cases} \lambda_1 & t \in [0, \tau_1) \\ \lambda_2 & t \in [\tau_1, \tau_2) \\ \vdots & \\ \lambda_{k+1} & t \geq \tau_k. \end{cases}$$

Survival function computed from hazard function:

$$S(t) \quad = \quad \exp\Big(-\int h(t)dt\Big).$$

Estimate $(\lambda_1, \ldots, \lambda_{k+1}, \tau_1, \ldots, \tau_k)$ incl. standard errors using **maximum likelihood** (backup for details).

Piecewise constant hazard



Maximum likelihood for censored data

General log-likelihood function for censored data:

- We observe $(T_i, \delta_i) = (\min\{X_i, C_i\}, \delta_i), i = \ldots, n$,
- h_λ is the hazard,
- S_λ the survival function, both depending on λ.

Then (see e.g. Klein and Moeschberger (2003)):

$$\begin{aligned} \log(L(\lambda)) &= \log\Big(\prod_{i=1}^{n} h_\lambda(T_i)^{\delta_i} S_\lambda(T_i)\Big) \\ &= \sum_{i=1}^{n} \delta_i \log h_\lambda(T_i) + \sum_{i=1}^{n} \log S_\lambda(T_i). \end{aligned}$$

For a **constant** hazard: $h_\lambda(t) = \lambda$, $S(t) = \exp(-\int \lambda dt) = \exp(-\lambda t)$. So:

$$\log(L(\lambda)) = \sum_{i=1}^{n} \delta_i \log \lambda - \lambda \sum_{i=1}^{n} T_i.$$

Taking derivative w.r.t. to λ and setting equal to 0 yields $\widehat{\lambda} = \sum_{i=1}^{n} \delta_i / \sum_{i=1}^{n} T_i$.

MLE in a piecewise model for the hazard with censored data

Generalize to piecewise function $\Rightarrow$ again, general recipe.

Piecewise Exponential: see e.g. references in Goodman et al. (2011). We get:

$$\begin{aligned}\log L(\lambda_1, \ldots, \lambda_{k+1}, \tau_1, \ldots, \tau_k) &= \\ &= \sum_{j=1}^{k+1} \{\#\text{events in } (\tau_{j-1}, \tau_j]\} \log \lambda_j - \sum_{i=1}^{n} \sum_{j=1}^{k+1} \lambda_j \Big(\min\{T_i, \tau_j\} - \min\{T_i, \tau_{j-1}\} \Big).\end{aligned}$$

Closer look at $\min\{T_i, \tau_j\} - \min\{T_i, \tau_{j-1}\}$:

$$\begin{aligned}\min\{T_i, \tau_j\} - \min\{T_i, \tau_{j-1}\} &= \begin{cases} 0 & \text{if } T_i < \tau_{j-1} \\ T_i - \tau_{j-1} & \text{if } T_i \in [\tau_{j-1}, \tau_j] \\ \tau_j - \tau_{j-1} & \text{if } T_i > \tau_j. \end{cases} \\ &= (\min\{T_i, \tau_j\} - \tau_{j-1}) 1\{T_i > \tau_{j-1}\}.\end{aligned}$$

Observation i contributes to estimation in interval $(\tau_{j-1}, \tau_j]$ the observation time in that interval.

MLE in a piecewise model for the hazard with censored data

Assume $\tau_1, \ldots, \tau_k$ as **fixed**.

Take derivative of $\log L(\lambda_1, \ldots, \lambda_{k+1}, \tau_1, \ldots, \tau_k)$ w.r.t. to $\lambda_1, \ldots, \lambda_{k+1}$. This gives:

$$\widehat{\lambda}_j \quad = \quad \frac{\#\text{events in } (\tau_{j-1}, \tau_j]}{\sum_{i=1}^{n} (\min\{T_i, \tau_j\} - \tau_{j-1}) 1\{T_i > \tau_{j-1}\}}.$$

"Usual" MLE for Exponential data, just in interval $(\tau_{j-1}, \tau_j]$.

Plug $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{k+1}$ into $\log L(\lambda_1, \ldots, \lambda_{k+1}, \tau_1, \ldots, \tau_k)$, maximize numerically over $\tau_1, \ldots, \tau_k$ to get estimates $\widehat{\tau}_1, \ldots, \widehat{\tau}_k$.

Profile-likelihood! Justified since **asymptotically**, $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{k+1}$ and $\widehat{\tau}_1, \ldots, \widehat{\tau}_k$ are **independent**, i.e. asymptotically, estimates are the same irrespective of whether we maximize separately for τ's and λ's, or jointly, see Yao (1986).

Get standard errors of estimates based on **standard maximum likelihood** theory. At least I do so.

Sequential test

Piecewise constant hazard estimate with $k = 5$ change points:

$$\widehat{h}(t) = \begin{cases} 0.00386 & t \in [0, 3.6) \\ 0.01127 & t \in [3.6, 9.2) \\ 0.00405 & t \in [9.2, 12.0) \\ 0.01475 & t \in [12.0, 17.1) \\ 0.00404 & t \in [17.1, 19.8) \\ 0.02345 & t \geq 19.8 \end{cases}$$

Choose **global** $\alpha = 0.05$.

H_0	α_k	χ^2-quantile	X_W	p-value	reject H_0	established change point
$\lambda_1 - \lambda_2$	0.0250000	5.02	15.52	< 0.0001	1	3.6
$\lambda_2 - \lambda_3$	0.0125000	6.24	9.90	0.0017	1	9.2
$\lambda_3 - \lambda_4$	0.0062500	7.48	10.91	0.00095	1	12.0
$\lambda_4 - \lambda_5$	0.0031250	8.73	7.19	0.0073	0	
$\lambda_5 - \lambda_6$	0.0015625	10.00	6.98	0.0083	0	

Final hybrid estimate of S

$\widehat{\tau}$: last established change point.

Estimate piecewise constant hazard assuming $k = 1$ with change point $\widehat{\tau}$. Then:

$$\begin{aligned}
\widehat{S}(t) &= \begin{cases} \widehat{S}_{\text{KM}}(t) & 0 \leq t \leq \widehat{\tau} \\ \widehat{S}_{\text{KM}}(t) \cdot \widehat{S}_{\text{exponential beyond } \widehat{\tau}} & t > \widehat{\tau} \end{cases} \\
&= \begin{cases} \widehat{S}_{\text{KM}}(t) & 0 \leq t \leq \widehat{\tau} \\ \widehat{S}_{\text{KM}}(t) \cdot \exp\Big(-\widehat{\lambda}(t - \tau)\Big) & t > \widehat{\tau}. \end{cases}
\end{aligned}$$

Plug $\widehat{S}$ in generic formulas to compute m.

Confidence intervals width decreases with more events

Question: if we get more events - does width of confidence intervals decrease?

Setup: so far, used data with CCOD of 2012-11-21. Cut data back to CCOD of 2011-11-22 and 2012-05-21.

Results for prediction of second analysis, targeted #events = 248:

	CCOD = 2011/11/22	CCOD = 2012/05/21	CCOD = 2012/11/21
#events at CCOD	29	68	113
predicted timepoint	2014-05-14	2014-05-16	2013-10-13
95% confidence interval	[2013/03/14, 2015/02/15]	[2013/05/26, 2014/01/04]	[2013/08/10, 2014/03/04]
width of CI (months)	23.1	7.3	6.8

Dates get "earlier" with later cutoff date $\Rightarrow$ initial underreporting of events.

Confidence intervals width decreases with more events

Question: if we get more events - does width of confidence intervals decrease?

Setup: so far, used data with CCOD of 2012-11-21. Cut data back to CCOD of 2011-11-22 and 2012-05-21.

Results for prediction of final analysis, targeted #events = 370:

	CCOD = 2011/11/22	CCOD = 2012/05/21	CCOD = 2012/11/21
#events at CCOD	29	68	113
predicted timepoint	2015-10-18	2015-11-30	2014-07-31
95% confidence interval	[2013/12/09, 2018/06/13]	[2014/01/21, 2016/06/02]	[2014/02/28, 2015/06/16]
width of CI (months)	54.1	28.4	15.5

Dates get "earlier" with later cutoff date $\Rightarrow$ initial underreporting of events.

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 3.2.2 (2015-08-14)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: reporttools / xtable / fitdistrplus / MASS / eventTrack / muhaz / survival

This document was generated on 2016-04-24 at 21:13:16.