Usage of interval censored estimations for surrogate endpoints in case of a large portion of missing data

Sergei Kulikov, Irina Tischenko,
National Center for Hematology,
Moscow, Russia

Motives

- There are incorrect attempts to extrapolate the tasks, methods, results, interpretation from controlled clinical studies to analysis of registry population data.
- Expanding usage of surrogate endpoints and overoptimistic expectation from its analysis based on population data.
- Sensitivity and bias of classical survival estimates to quality of data, missings, irregularity of evaluation of surrogate events like response.
- Availability of new modern methodology for analyzing interval censored data.

Data sets used

1. 508 CML patients with 2005-2006 as years of diagnosis from study EUTOS-OSP (retrospective population, Russian register)
   Median age at diagnosis was 49.3 years, range from 18 to 82, 47.6% were men, 6.7% in AC,BC phase, 29.3% at high risk by Sokal
2. 200 CML patients of PBS EUTOS study (2009-2012 as years of diagnosis, prospective population, Russian register)
   Median age at diagnosis was 50.4 years, range from 16 to 82, 50.8% were men, 6.0% in AC,BC phase, 31.7% at high risk by Sokal.
3. Simulation model data (cure fraction exponential model)

Surrogate endpoints and surrogate events

Advantages

- Short duration of study
- Lower sample size (less censoring)
- Less troubles with heterogeneity, competing risks, fragility

Troubles

Most “events” of interest are no real events:

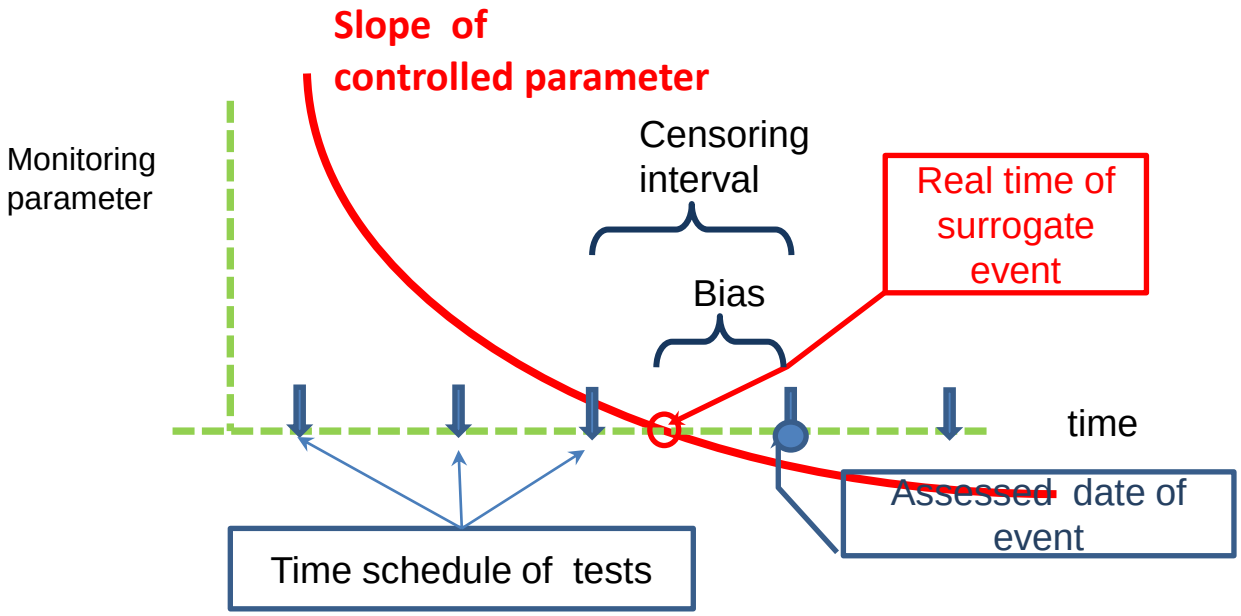
- Any kind of response is no sharp change of status , it is not the event
- Resistance is not an event - it is the status of knowledge about the patient status
- Treatment failures are subjective events with low accurate time assessment
- Combined end points = “basket” of random values of different - nature, weight distributions, factors, etc
- Low precision of time measures
- Extrapolated usage out of frame of controlled clinical studies

More troubles with surrogate events

- Measured by intervals
  - When the intervals are too large , the event may not be captured at all
  - Problem for comparisons between groups: if the intervals are not the same along time, and moreover not the same at all between groups
- Estimations based on time assessment of event are always biased
  - Positive surrogate events rate (responses, remissions) are always underestimated (biased to the right)
  - Survival without negative surrogate events rate (relapse, loss of response) is overestimated (biased to the right)
- Complicate analysis if event and risk are of different signs (f.e response - death)

Interval “nature” of the surrogate events

Surrogate endpoints are based on periodical measurement or clinical evaluation so are always interval censored
Surrogate event is not event



Irregular time schedule and missings are the source of noise in measurements and errors in estimates,
The discretization error (noise) exists even for strong regular and uniform time schedule.

The measurement accuracy of real and surrogate events

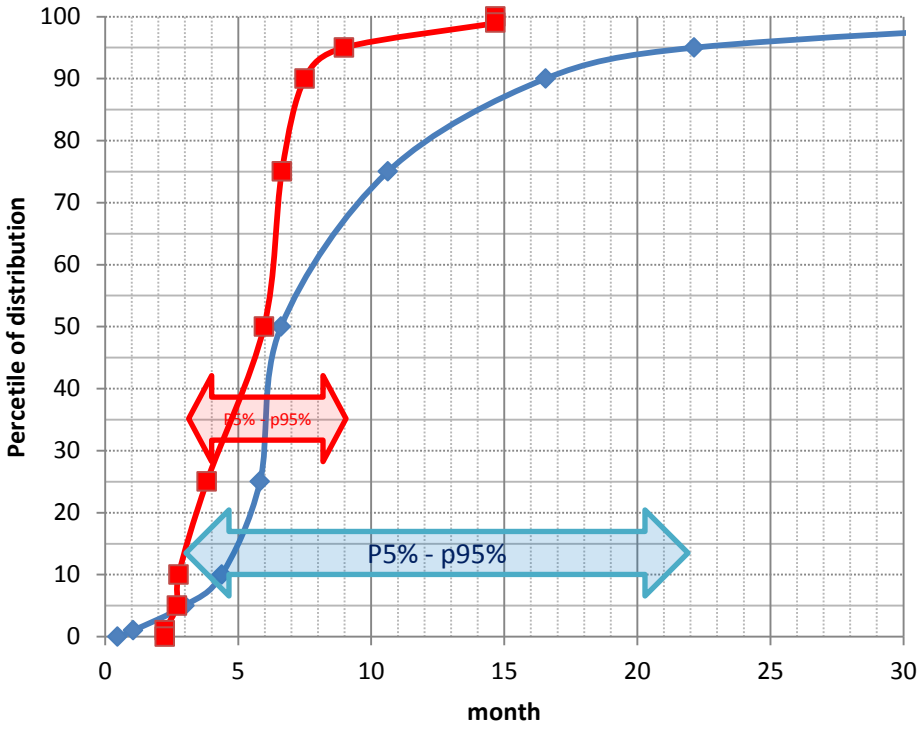
- Life duration = 1 day, (accuracy 10^{-3})
- Time to progression = 10-100 days, (accuracy 10^{-2})
- Time to response = 100-300 days, (accuracy 0,5-1)

surrogate/real events accuracy ratio is up to 1/1000

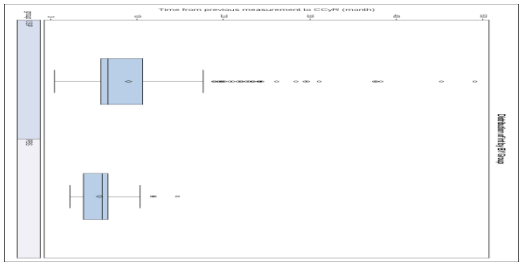
- Surrogate events are noisy measures not only because of errors in monitoring parameter but also due to errors in time assessed

Time interval between cytogenetic tests which includes CCyR is the measure of event evaluation precision

Distributions of interval duration



	OSP	PBS
n	368	94
mean	8.99	5.64
median	6.61	5.97
range (5-95%)	19.15	6.28
StDev	6.44	2.21



Time of CCyR assessment in OSP study is 3 times more “noisy” than in PBS

Accuracy of surrogate events and statistical power of conclusions

- Surrogate events are measured with noise
- So statistical power of ANY conclusion based on surrogate events is much less then for exact real events

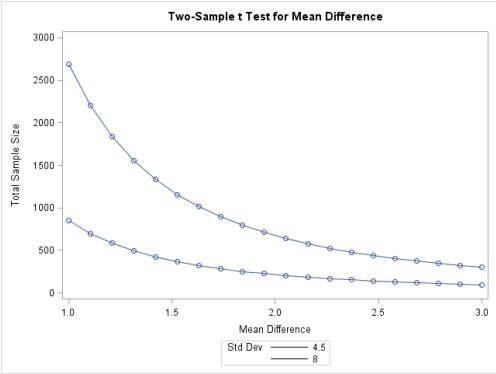
Example: Endpoint - time to CCyR

Mean =8m,
StdDev (population + error) = 8m,
StdDev of error = 6.5m
StdDev of "pure time"
(population) = 4.5m

Results of sample calculation for hypothetical two arm clinical trials

Power=0.9,
Target difference in means =1,2,3 months,
StdDev =4.5, 8 months

Output of SAS procedure POWER:

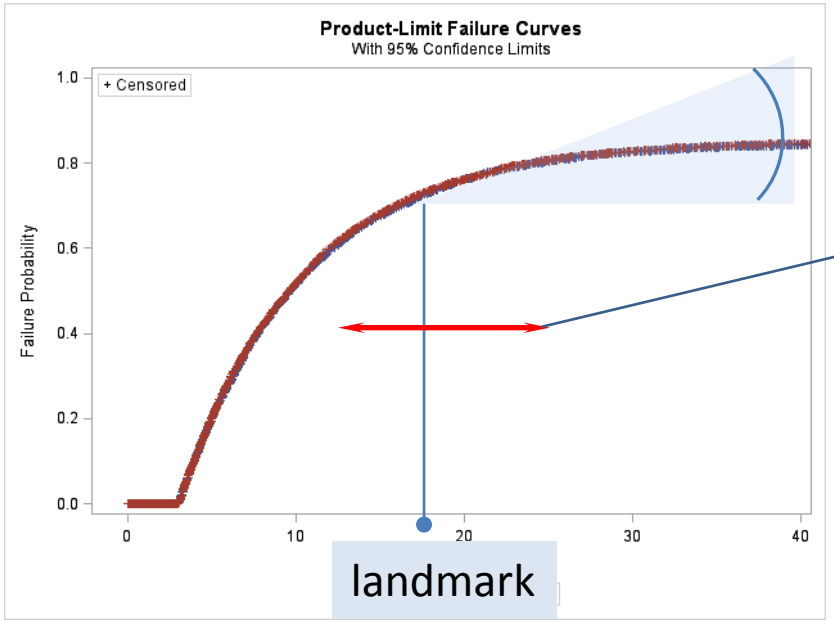


Mean Diff	N1 (Stdev = 4.5)	N2 (Stdev = 8)	Ratio N2/N1
1	854	2692	3.2
2	215	675	3.1
3	97	301	3.1

Hypothetically, we can decrease the needed sample size up to 3 times if the controlled parameter could be measured more frequent (every day)

Frequencies are also “noisy”

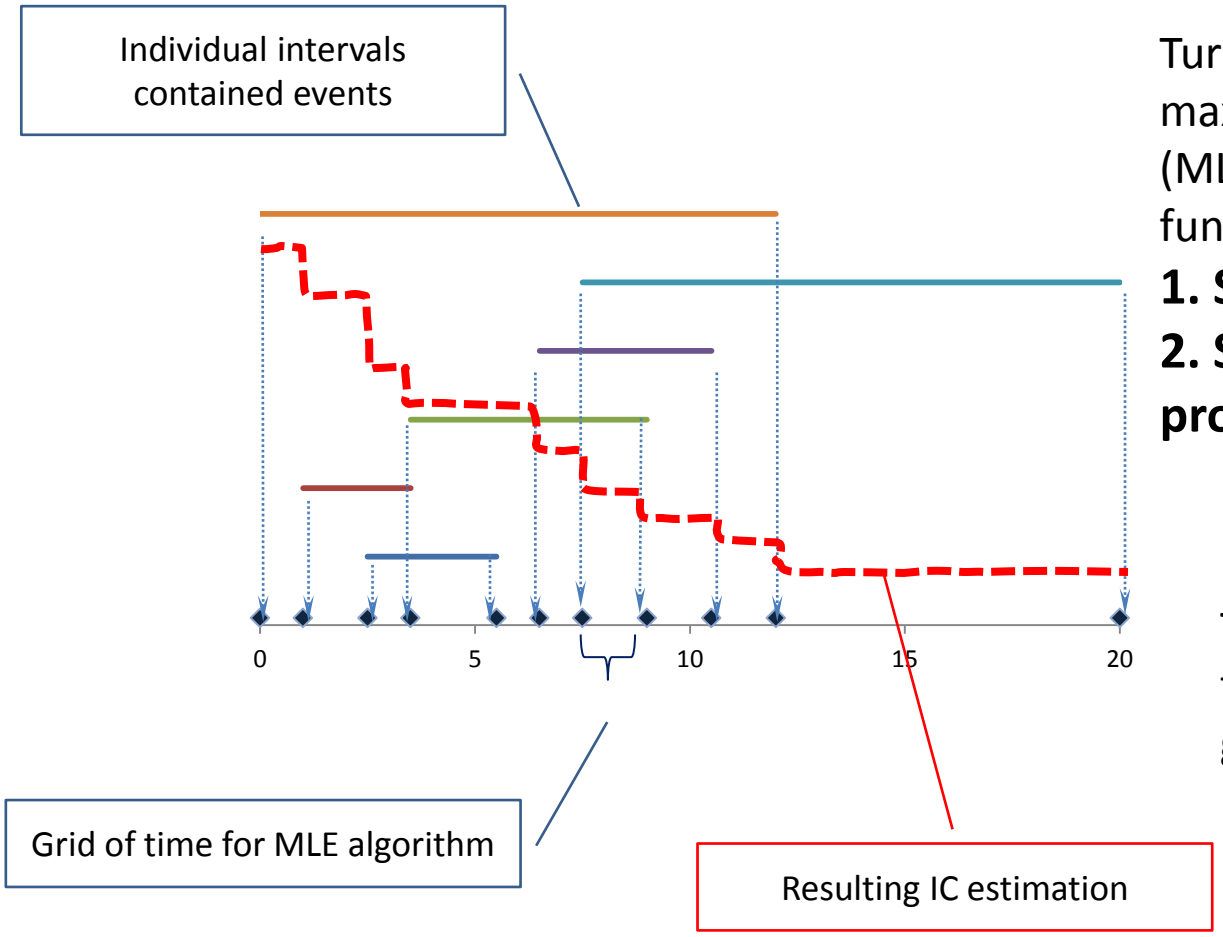
- Time uncertainty is converted into count uncertainty, the coefficient is tangent of slope angle of respond rate in landmark (control time point).
- Not more then 1/5.



Tan(α)

Event with interval censoring around the landmark is the source of errors in counting

Survival estimates based on interval-censored data



Turnbull's algorithm for maximum likelihood estimator (MLE) of step-wise survival function is realized in

1. SAS macro %EMICM

2. SAS ICLIFETEST procedure (in ver. 6.4)

The generalized log-rank test of Sun is used for group comparison

Comparison of two Russian CML population studies (OSP, PBS Eutos)

Two CML studies (data sets):

1. **OSP** - 2005-2006 years, retrospective
2. **PBS** - 2009-2012 years, prospective

Findings:

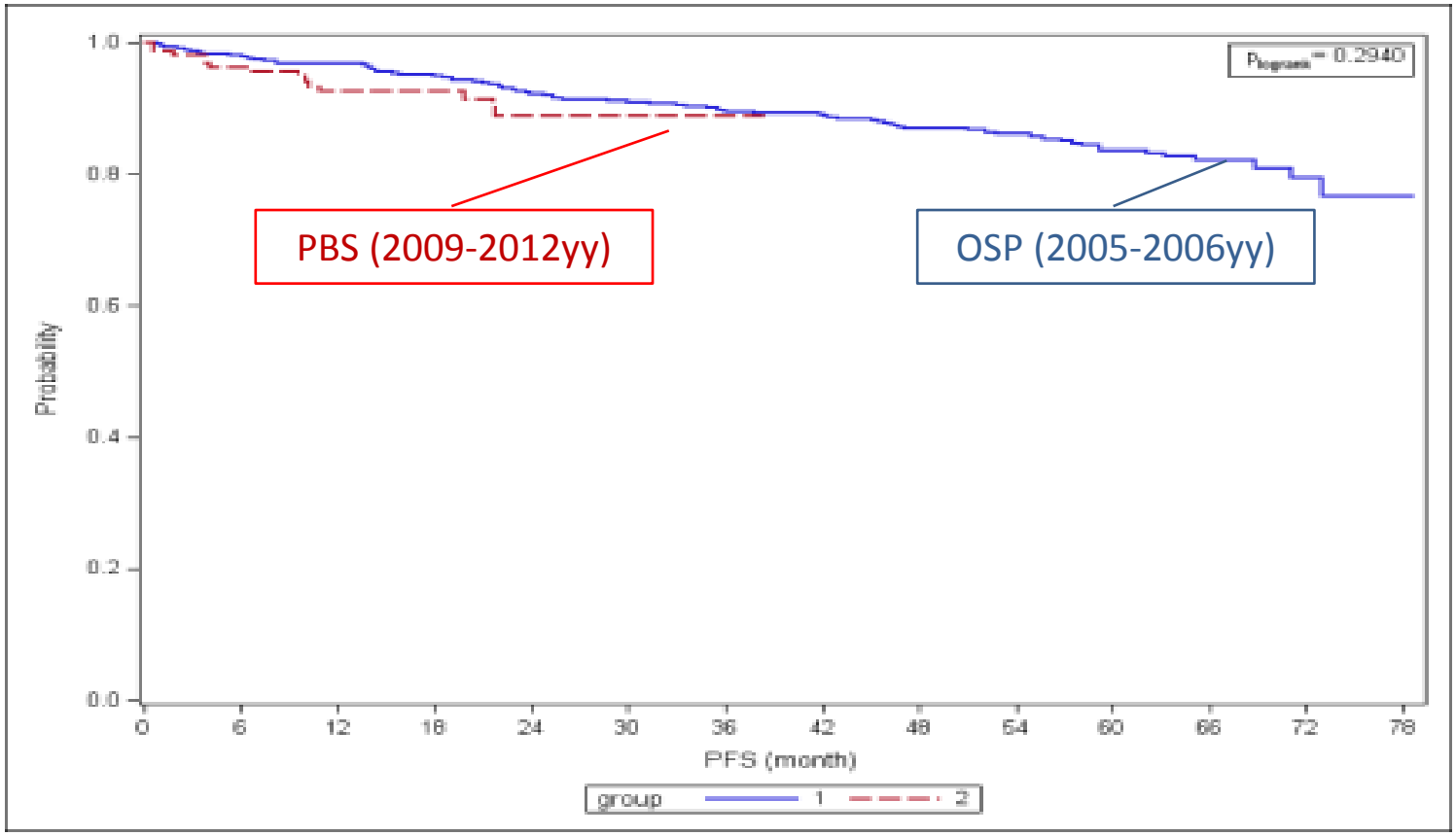
Long term results of studies are almost identical although the response rates are essentially different if calculated by standard survival methods.

Question:

Are this differences real? Should they be interpreted?
Or are these the results of instrumental errors and different biases of estimates?

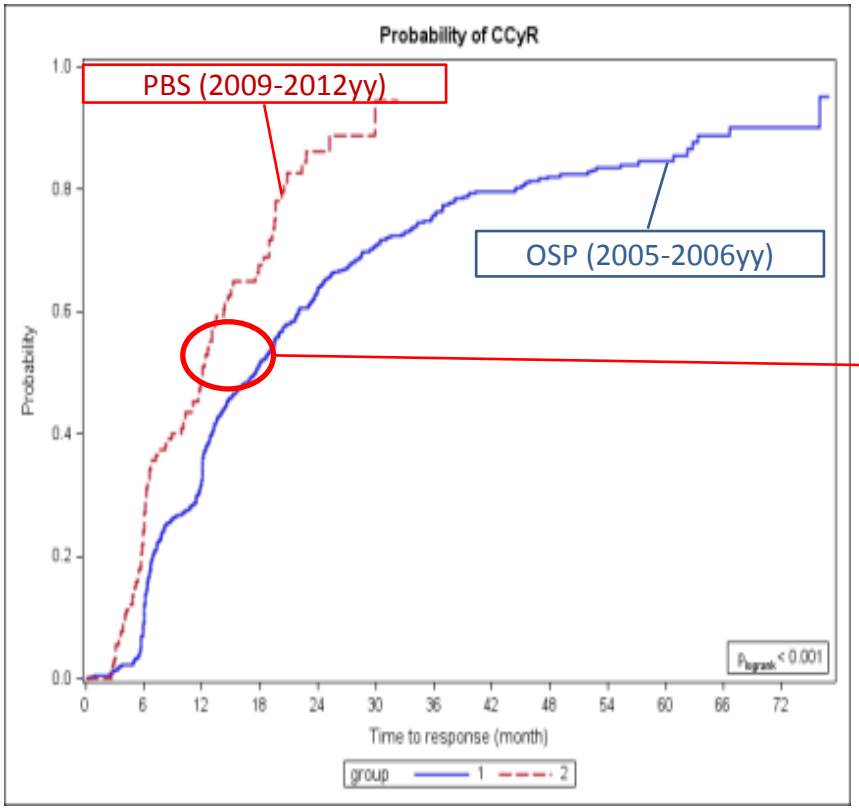
The progression free survival are similar in two studies

KM estimations of PFS in OSP and PBS sets.

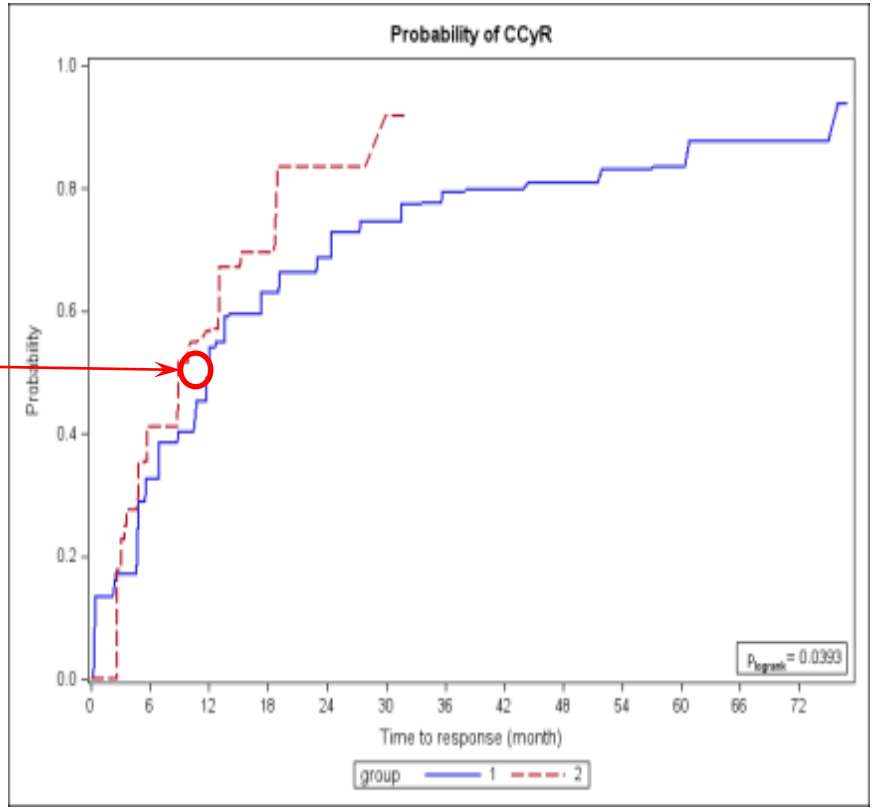


Are the CCyR rates really different in the studies ?

KM estimations of CCyR in OSP and PBS sets.



Interval Censor Estimations (ICE) of CCyR in OSP and PBS sets.



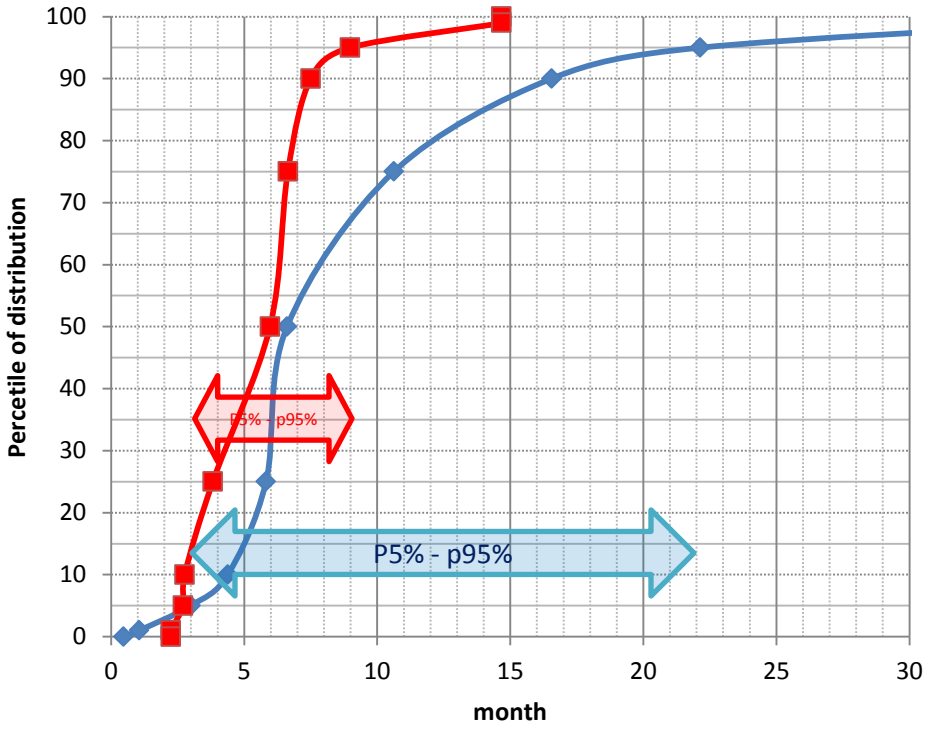
Conclusion:

Difference in CCyR rates disappears if ICE is used.

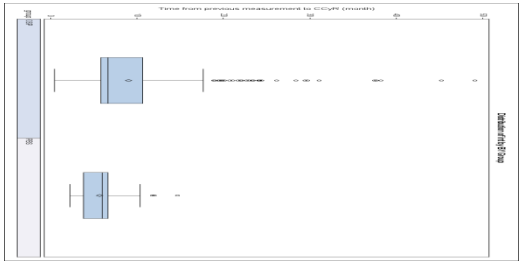
The source of false “discover” is different intervals of cytogenetic evaluations

Different mean time intervals between cytogenetic tests is the reason of different bias of KM estimations of CCyRs

distributions



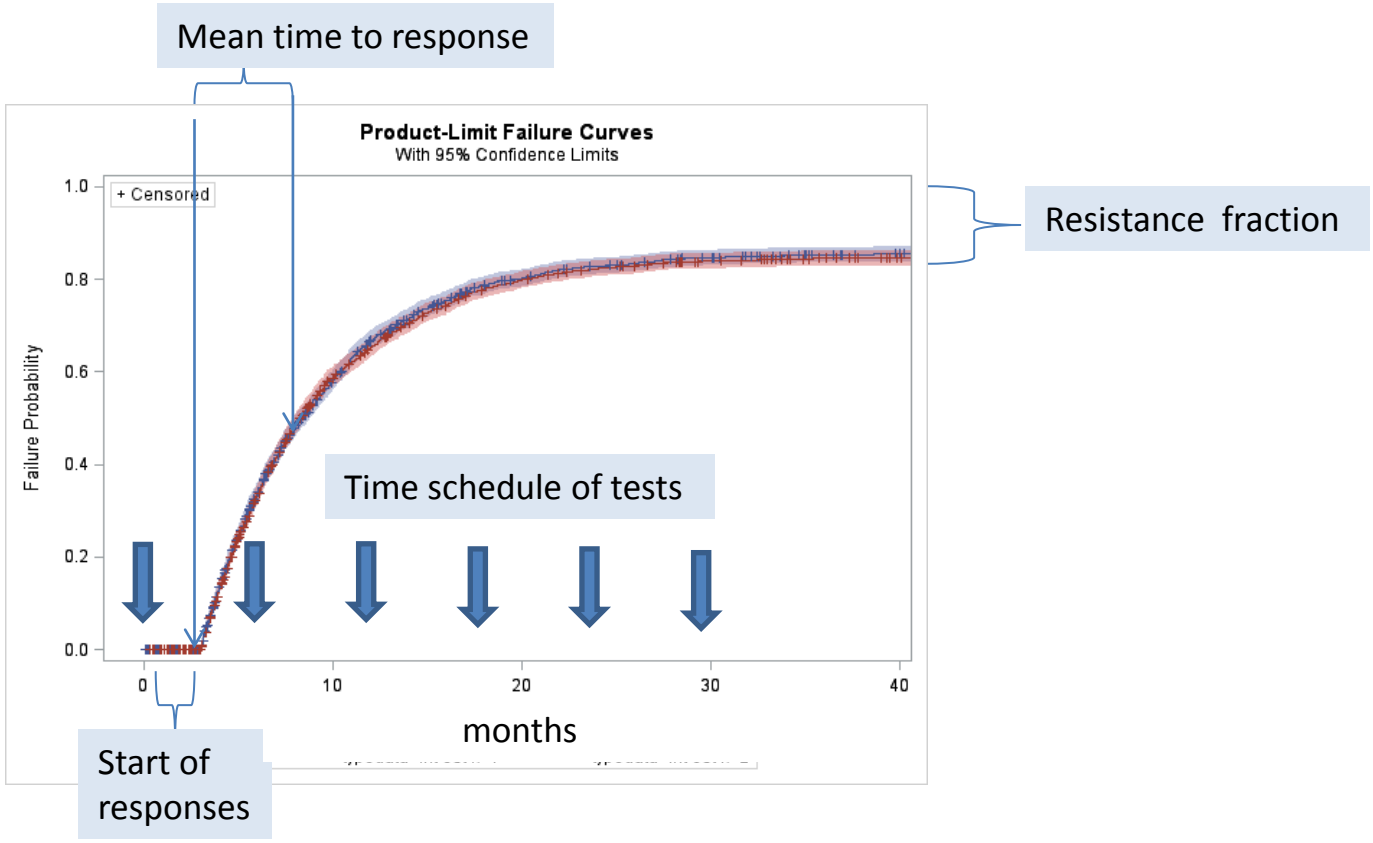
	OSP	PBS
n	368	94
mean	8.99	5.64
median	6.61	5.97
range (5-95%)	19.15	6.28
StDev	6.44	2.21



Time of CCyR assessment in OSP study is 3 times more “noisy” then in PBS because of missings

Simulation model

- Time to response generated from mixture of two exponential distributions: for cure (responder) & non-cure (resistance) fractions.
- Interval of response tests: (equal intervals + random delay) * probability of missing

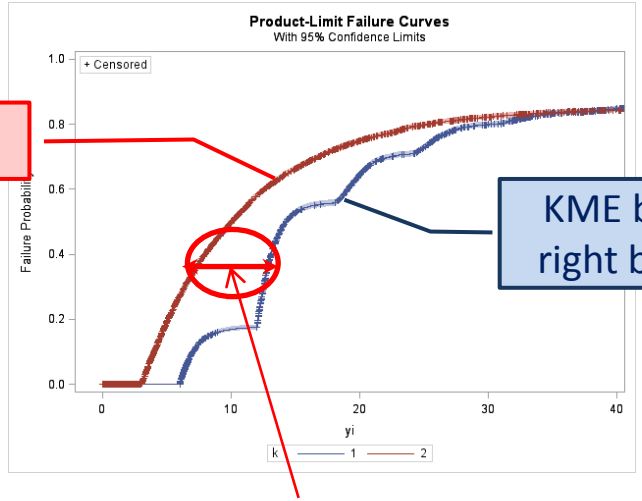


Simulation results. Bias of KM estimates of CCyR

Model parameters:

1. Resistance fraction = 15%
2. Mean time to response
   Responders = 8 m
   Non-responders = inf.
3. Test intervals = 6m
4. Mean delay = 1m
5. Probability of missing = 0.5
6. Sample volume = 10000

True time to CCyR



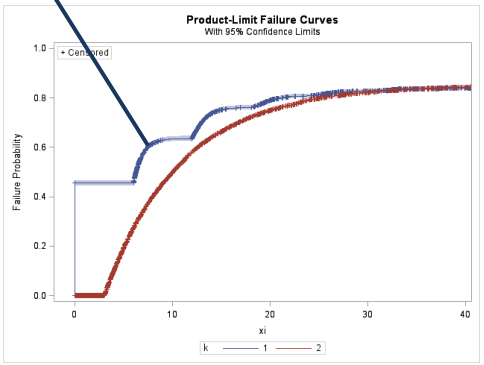
KME based on right boundary

Bias of median time = 7 mounts

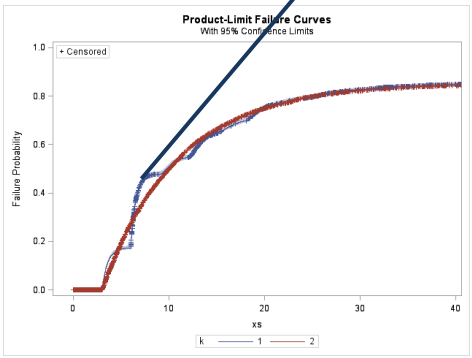
If KM estimation is used:

1. Right boundary = big bias
2. Left boundary, mid point are not applicable

left boundary



Mid point



Simulation results. IC versus KM estimates of CCyR

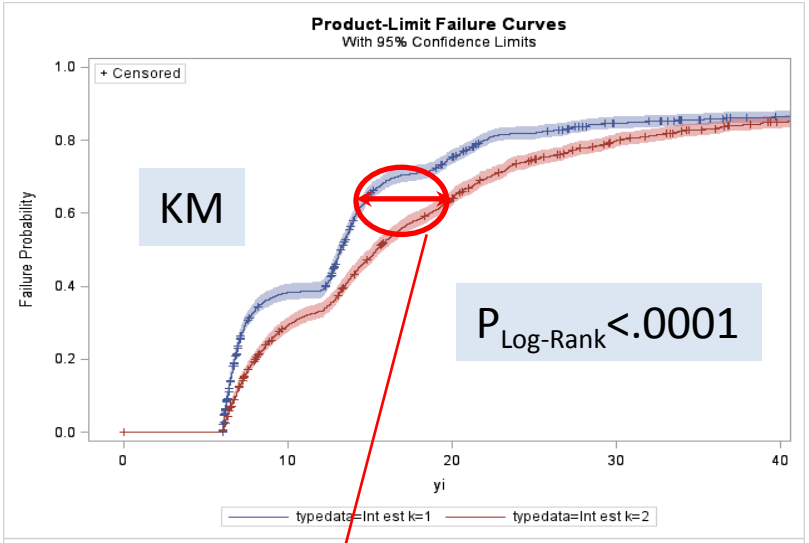
Model parameters:

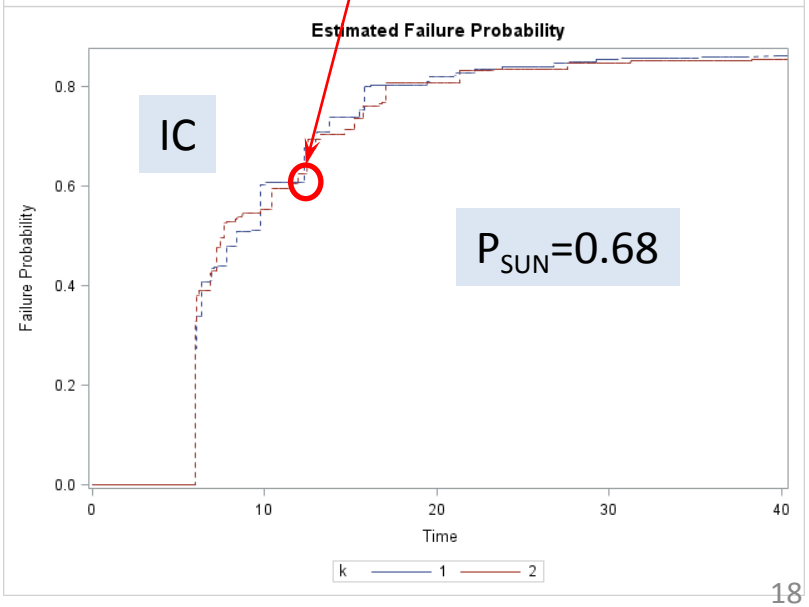
1. Resistance fraction = 15%
2. Mean time to response
   - Responders = 8 m
   - Non-responders = inf.
3. Test intervals = 6m
4. Mean delay = 1m
6. Sample volume = 2000
5. Probability of missing
   - 1 group = **0.05** (low missing)
   - 2 group = **0.25** (high missing)

Same for both groups

IC vs KM estimation:

1. KME show difference in groups which no exists
2. ICEs in groups are near the same as postulated in model





More advantages of estimations based on interval censorships

- You can use additional information for interval building
- F.e. FMR can not occur before CcyR -> cytogenetic test can be used for left limit for molecular responses
- Molecular tests can be used for right limits calculation of cytogenetic responses

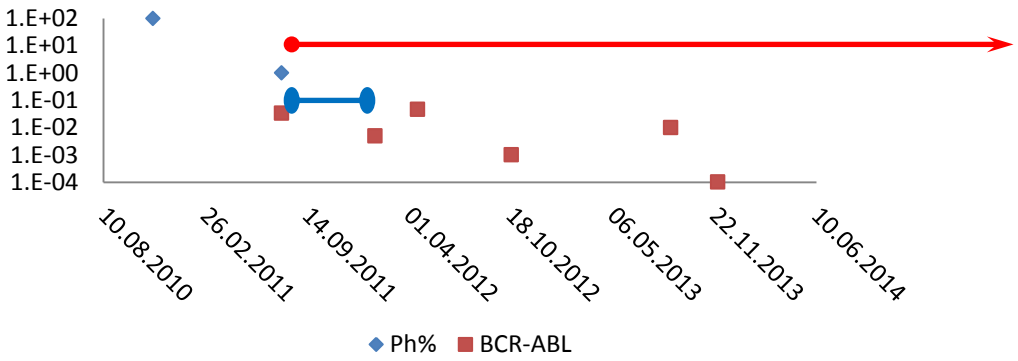
Common conclusion - non direct tests, evaluations, additional logistic info can be put in date interval calculation for each surrogate event and this leads to less biased statistical estimation of target event rate

Algorithms for CCyR and MMR interval assessment by usage both cytogenetic (Ph%) and molecular (BCR-ABL) tests results

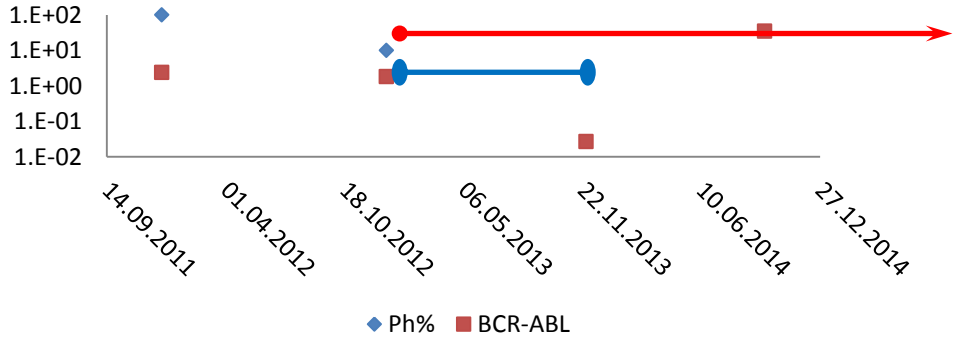
Complete Cytogenetic Response date interval:
Left boundary =Last data of (Ph%>0 or if no - BCR-ABL>1)
Right boundary =First data of (Ph%=0 or if no - BCR-ABL<0.1)

Major Molecular Response date interval:
Left boundary =Last data of (BCR-ABL>0 .01 or if no – Ph=min)
Right boundary =First data of (BCR-ABL<0 .01)

How it works (real cases)
Examples of reduction of CCyR interval

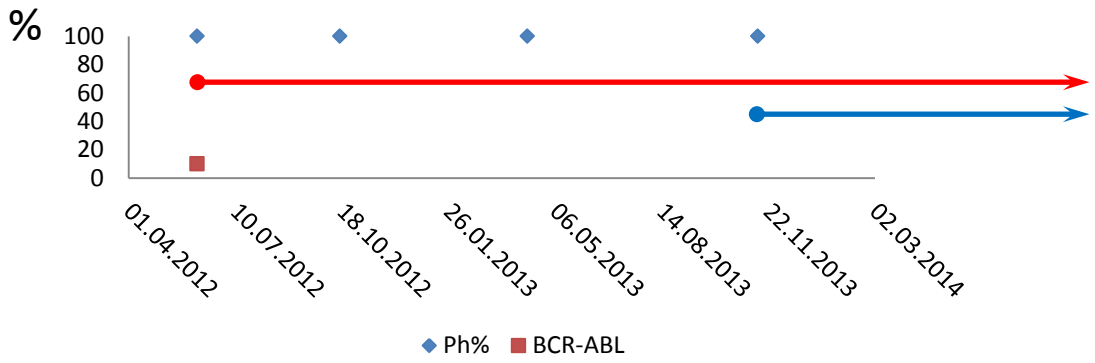


Before adjusting (only CY tests)
(26/07/11,∞)
After (Cy+M tests)
(26/07/11, 25/01/12)

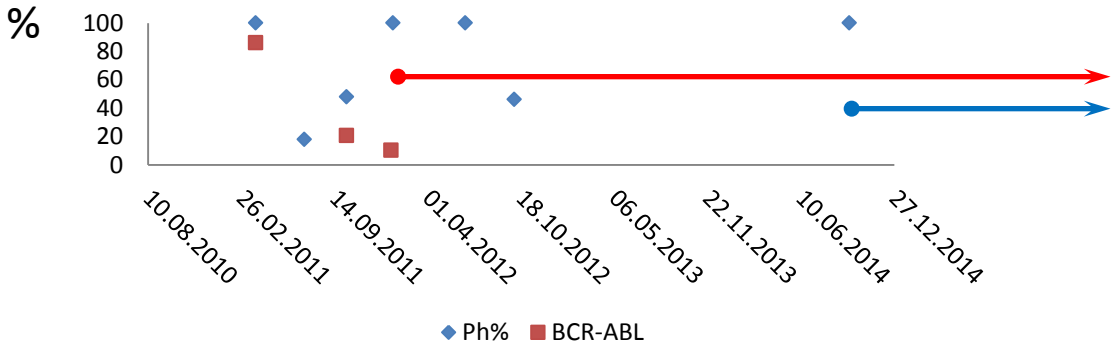


Before adjusting (only CY tests)
(27/12/12,∞)
After (Cy+M tests)
(27/12/12, 28/11/13)

How it works (real cases)
Examples of reduction of MMR interval



Before adjusting (only M tests)
(04/06/12,∞)
After (Cy+M tests)
(12/11/13,∞)



Before adjusting (only M tests)
(13/01/14,∞)
After (Cy+M tests)
(22/09/14,∞)

Conclusions

- Proceed with caution surrogate event statistics in population and non-controlled studies.
- Do not interpret absolute (not comparative) value of estimates of surrogate events.
- Random errors in time measurement of surrogates leads to loss in statistical power and this errors can compensate only by reduction of evaluation intervals.
- The bias of surrogate's time estimation can be partially compensate by interval censoring technique.
- Use interval censoring technique as more robust and less biased then classical right censoring estimators .
- Combine information from different laboratory and clinical measurements to justify censor intervals.

References

1. SAS Institute, Inc. SAS/STAT ® User's Guide, SAS OnlineDoc® 9.4, Cary, NC: SAS Institute, Inc.
2. So Y., Johnston G. Kim S.H.: // Analyzing Interval-Censored Survival Data with SAS® Software. , SAS Global Forum 2010.
3. Guo C., So Y., G. Johnston,:// Analyzing Interval-Censored Data with the ICLIFETEST Procedure, Paper SAS279-2014, SAS Institute Inc.
4. Singh R., Totawattage D.:// The Statistical Analysis of Interval-Censored Failure Time Data with Applications, Open Journal of Statistics, 2013, 3, 155-166, http://dx.doi.org/10.4236/ojs.2013.32017
5. Panageas K., Ben-Porat L., Dickler M.,Chapman P., Schrag D. //When you look matters: the effect of assessment schedule on progression-free survival . J Nat Cancer Inst, 2007; 99: 428 – 32 .
6. Chen D., Sun J., Peace K. //Interval-Censored Time-to-Event Data: Methods and Applications, 2012, Chapman & Hall/CRC Biostatistics Series

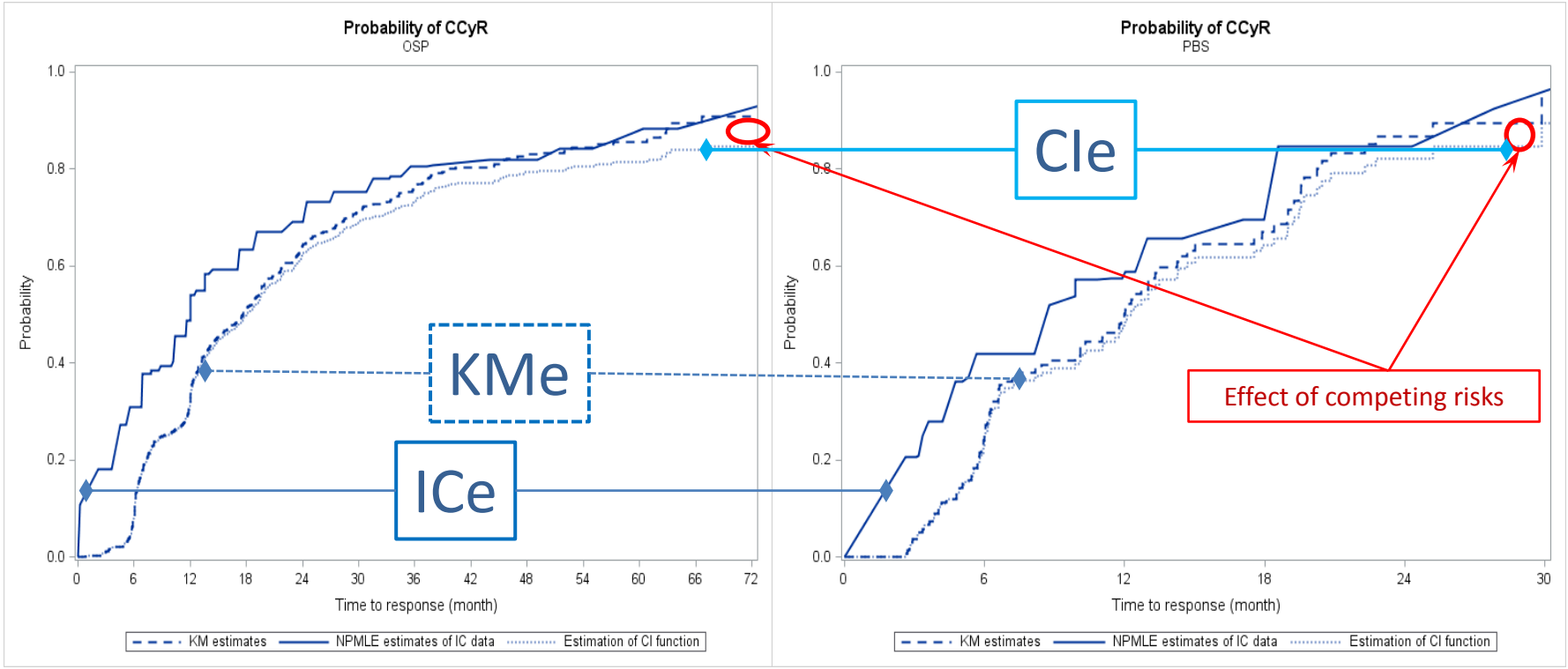
Acknowledgements:

- Turkina A.
- Lasareva O.
- Chelysheva E.
- Parovichnikova E.
- Rusinov M.
- Chabaeva Ju.
- Guilhot J.

Comparison of KM, IC and CI estimators

OSP data (high level of missing)

PBS data (low level of missing)



Conclusion:

KM and CI are close because of low level of competing risk events (less then 10%)