MORSE
Health Technology Assessment Group

Roche

Predicting long term survival using non-parametric bayesian methods: the melanoma case

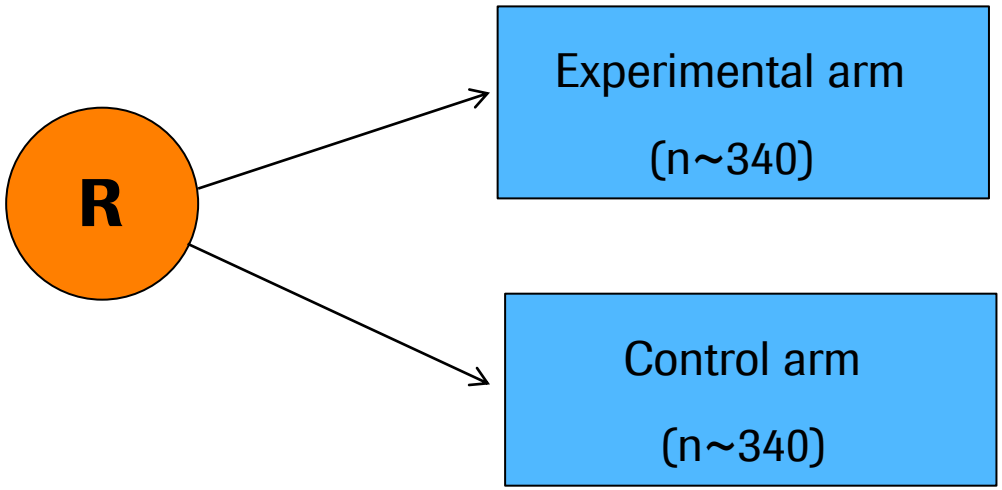
Yovanna Castro
Pierre Ducournau

BBS - EFSPI 2015 – June 23, 2015

Melanoma

- Type of skin cancer
- Less common than other skin cancers
- More dangerous if it is not treated early
- Causes 75% of deaths related to skin cancer

Clinical trial

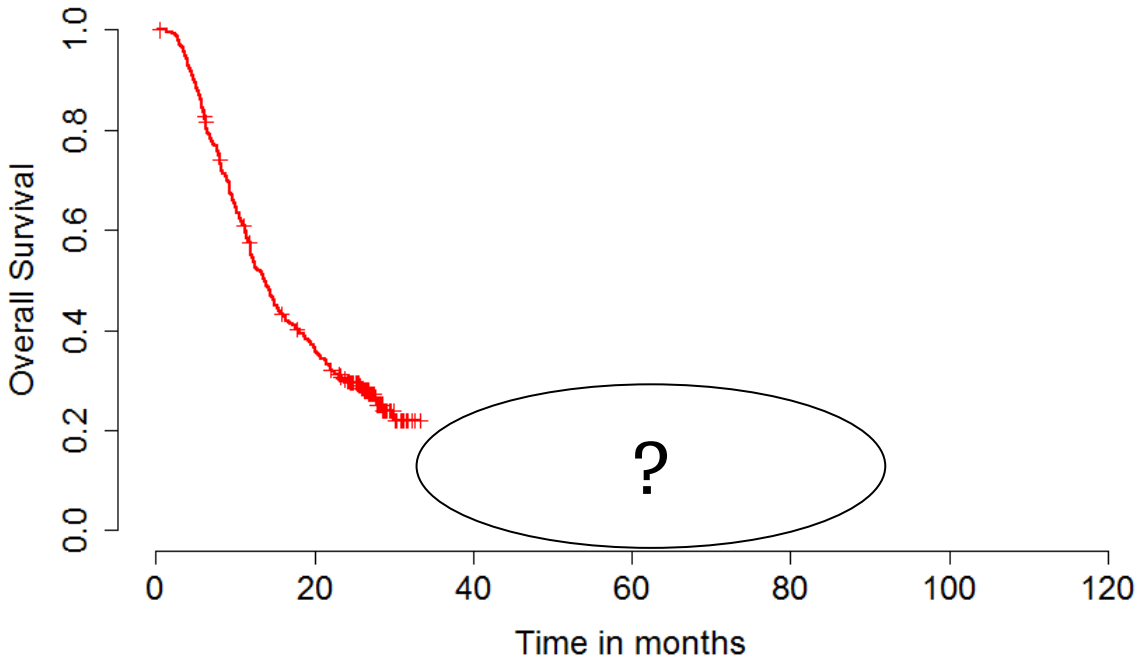


For the purpose of this application:

- Consider overall survival endpoint. Focus on active treatment arm due to high percentage of “crossover” after early data cut
- 94% of patients in trial were stage IV → 5-year survival rates of 15%-20%

A key question in Health Technology Assessment is:

How to extrapolate survival data from a clinical trial?



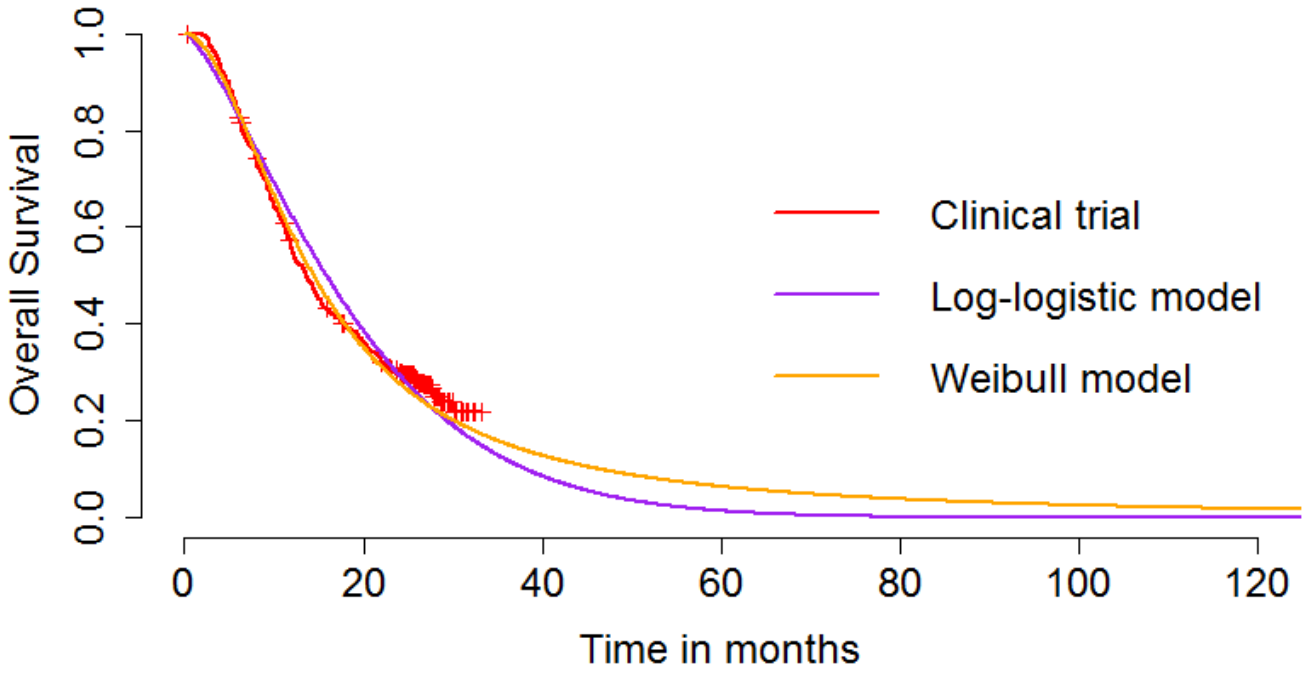
Characteristics of a clinical trial data

Ideal Conditions

- Randomization
- Blinding
- Clean database

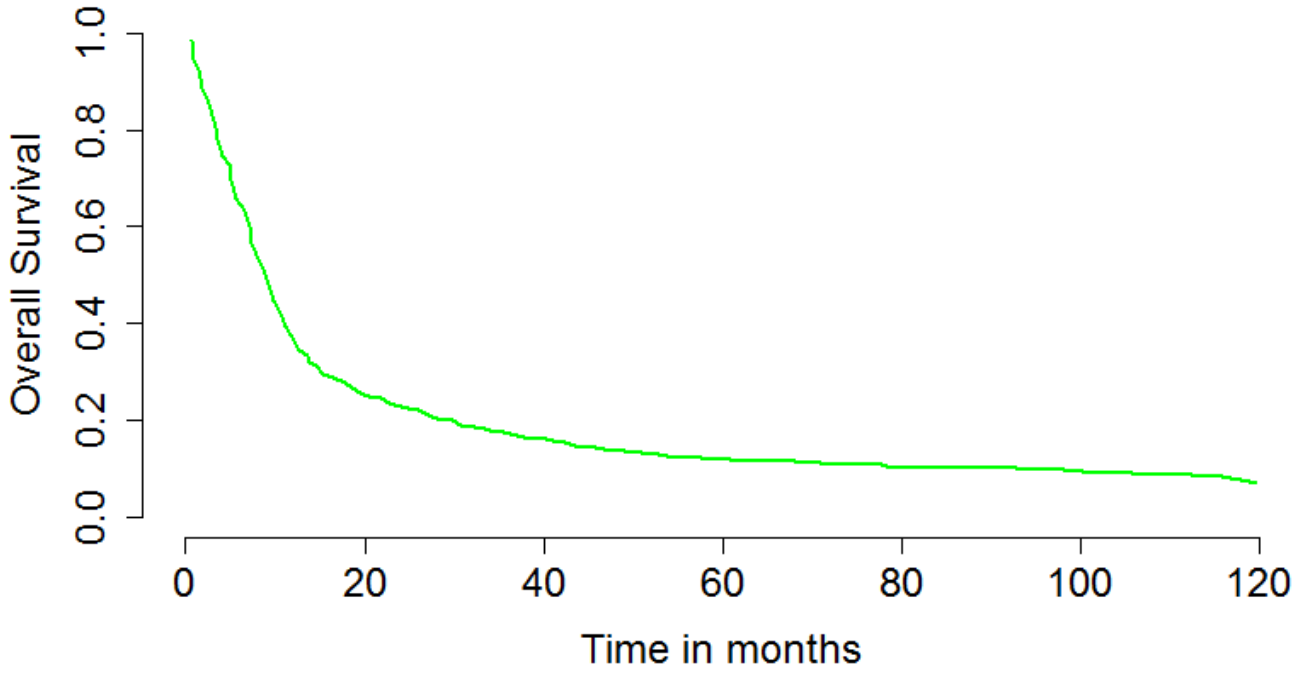
- May not reflect real practice
- Limited follow up

One way to answer is to apply parametric extrapolation



We should assess plausibility of our extrapolations. Latimer (2013).

In fact we can consider registry data:



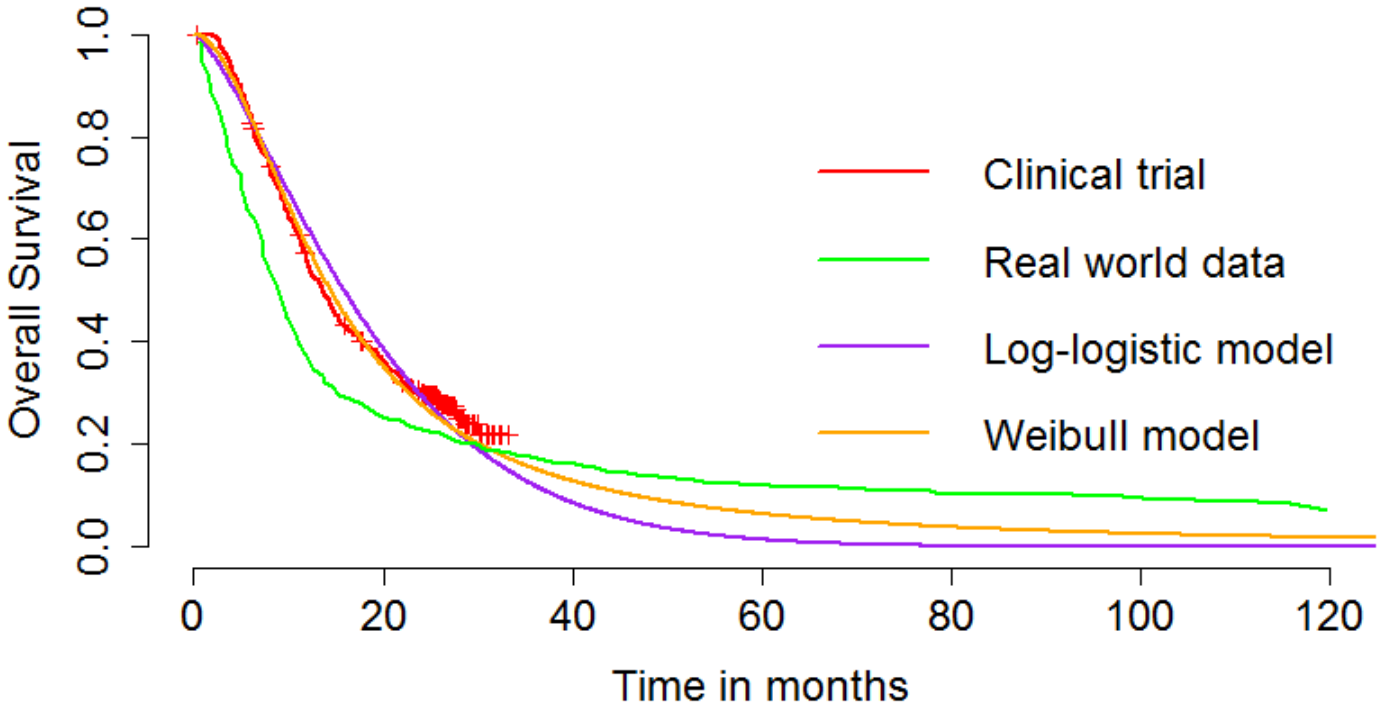
Patients with at least 5 years of follow up from a registry published in Xing et al (2010).

Characteristics of a real world data

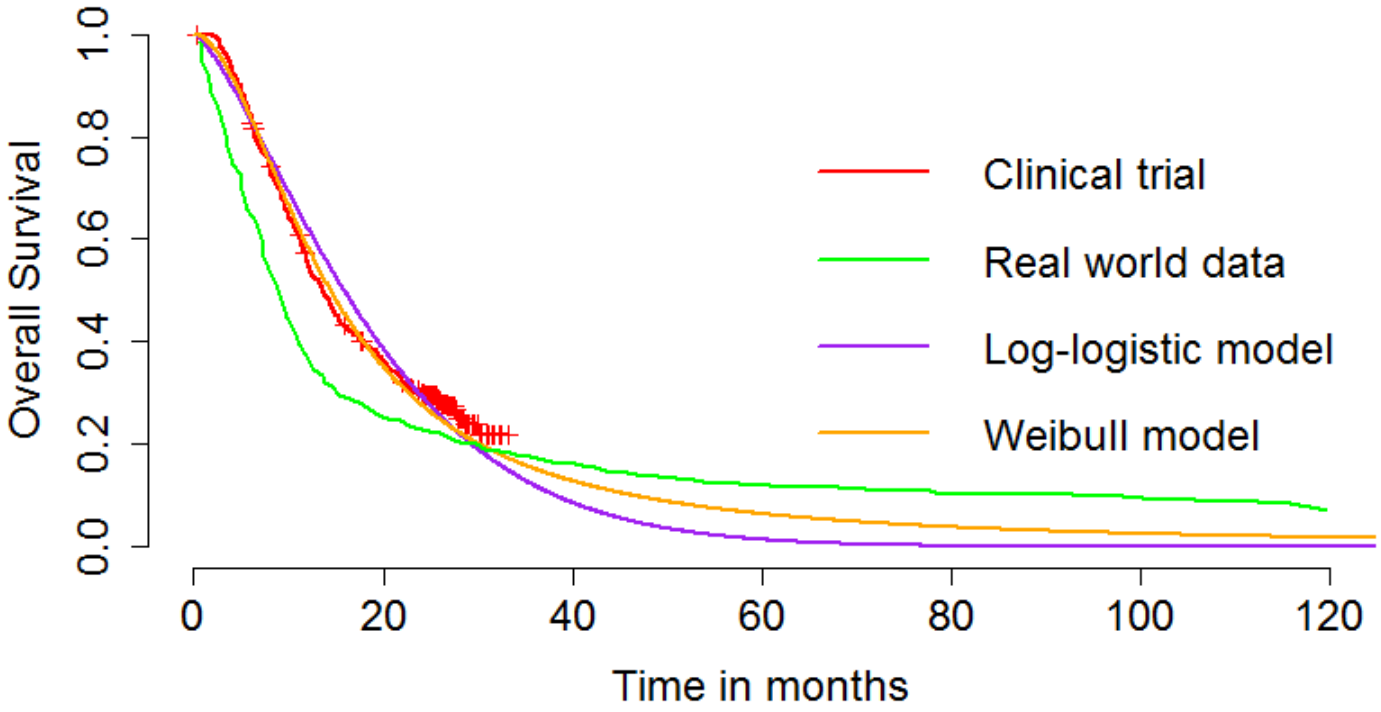
- It may reflect clinical practice
- Longer follow up

- May be limited to one country or one region
- Incomplete information about patients

What happen when we compare our parametric extrapolation with the real world data



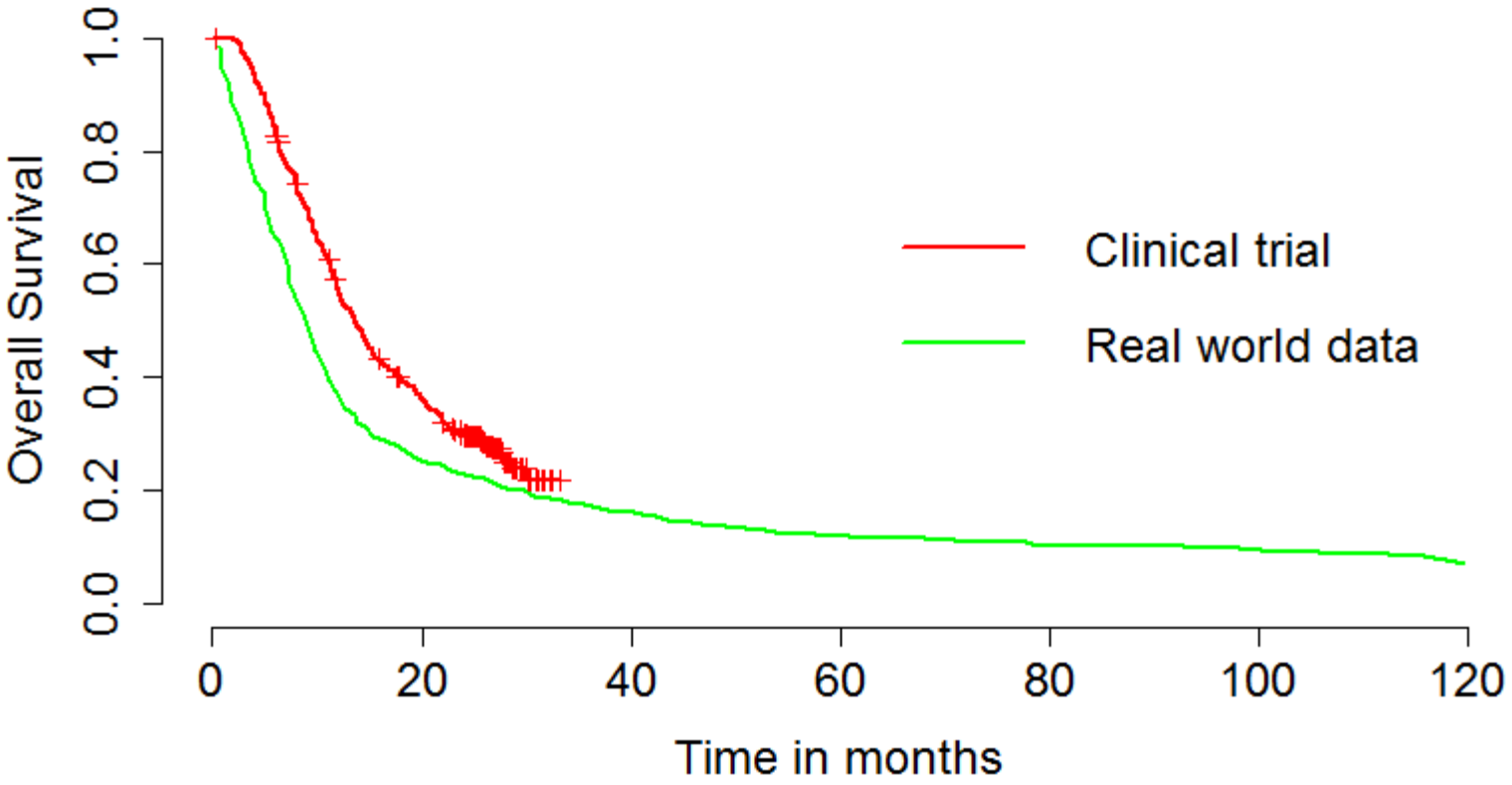
What happen when we compare our parametric extrapolation with the real world data



The problem is all the parametric extrapolations we perform lead to a heavy underestimation of survival rate

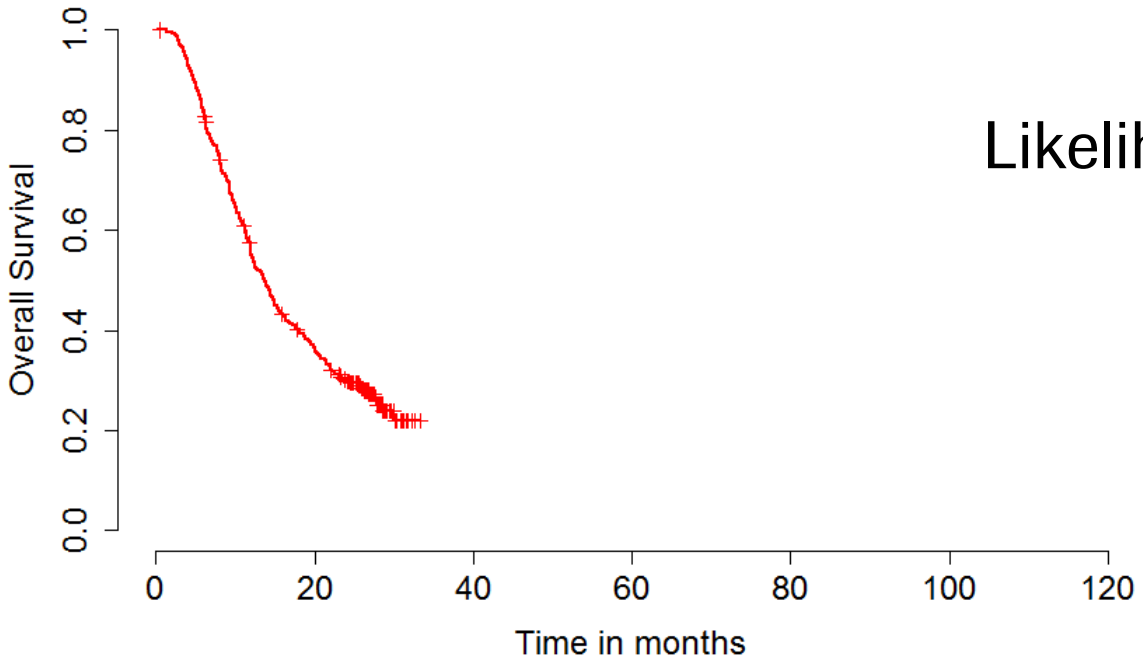
Another option is:

Combine the two sources of information we have



The clinical trial data has:

- “Short” follow up relatively to the time horizon considered in the health economics models
- “A lot” of censored observations specially in the tail

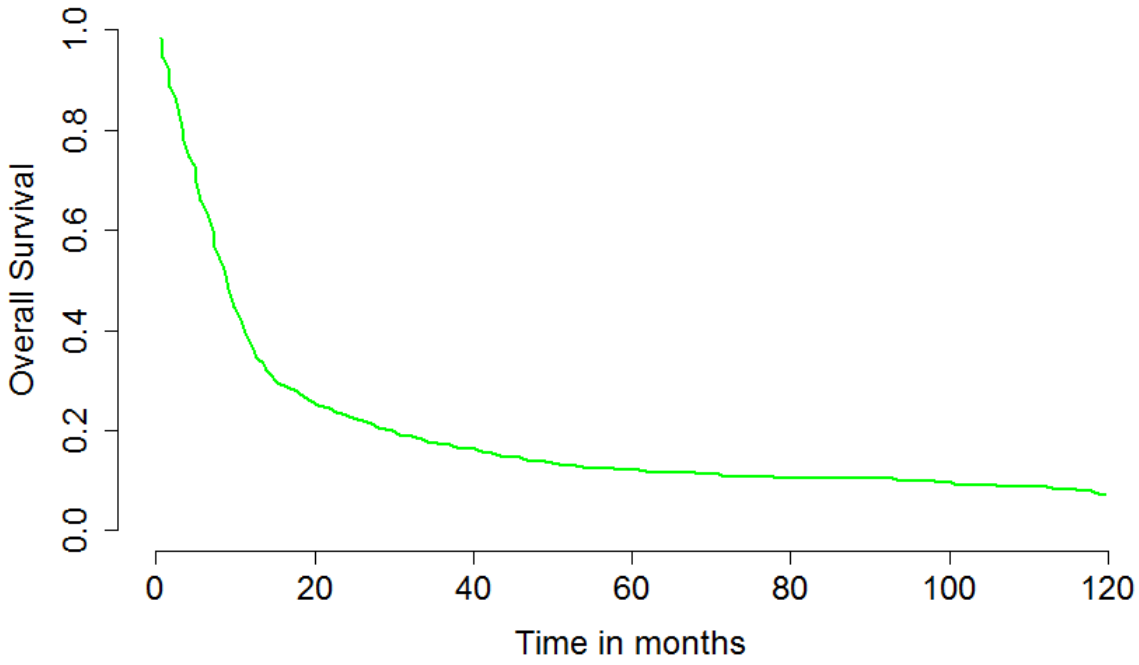


Likelihood

We have some previous knowledge:

- Real world data
- Longer follow up clinical trial

Prior



We can combine them using Bayesian estimation

$$\text{Posterior} \propto \text{prior} * \text{likelihood}$$

Prior = observational data

Likelihood = available (trial) data

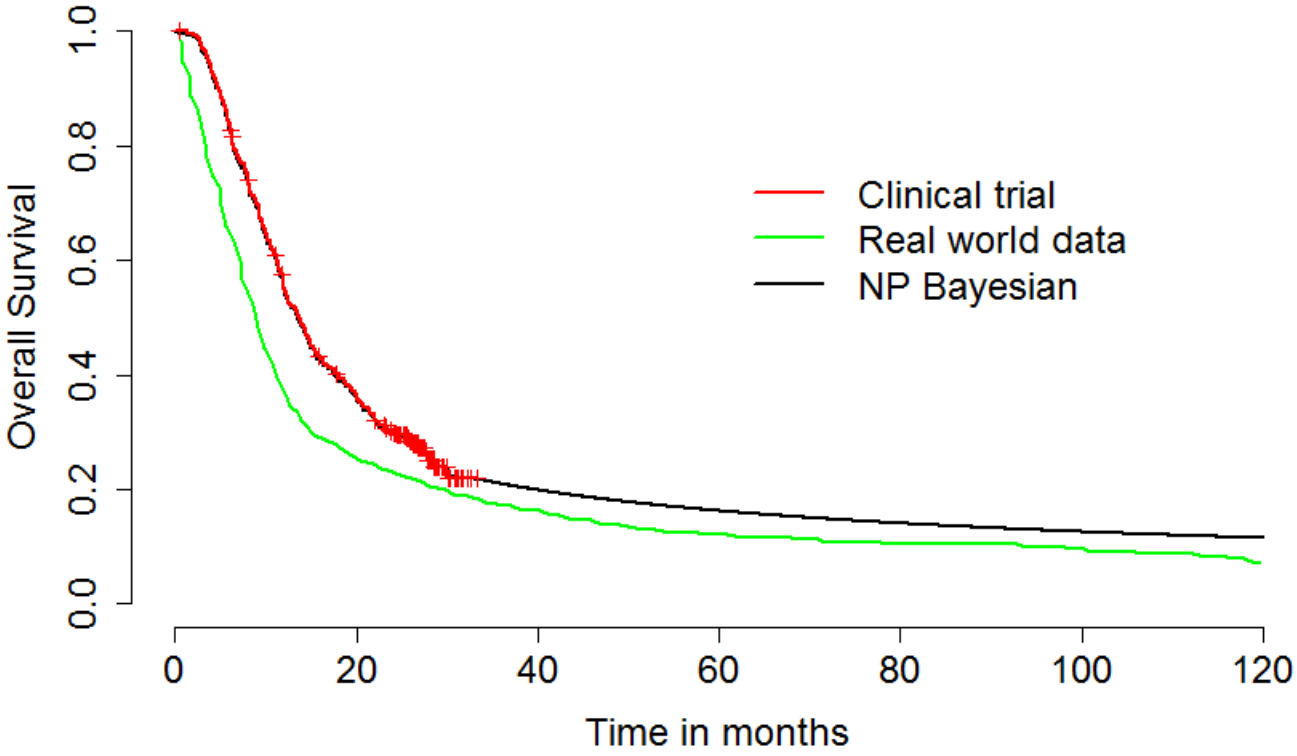
We use a Bayesian nonparametric estimation

- The prior is based on a Dirichlet process.
- For survival analysis previous work based on Dirichlet processes was proposed by Ferguson and Phadia (1979) and Susarla and Van Ryzin (1976).
- We assume the survival function follows a Dirichlet distribution with certain parameter.
- The form of the $S(t)=cS_0(t)$
- $S_0(t)$ is our prior guess at the survival function
- c is a measure of how much weight we put on our prior guess (larger value of c lead to smoother function)

Non parametric Bayesian estimator

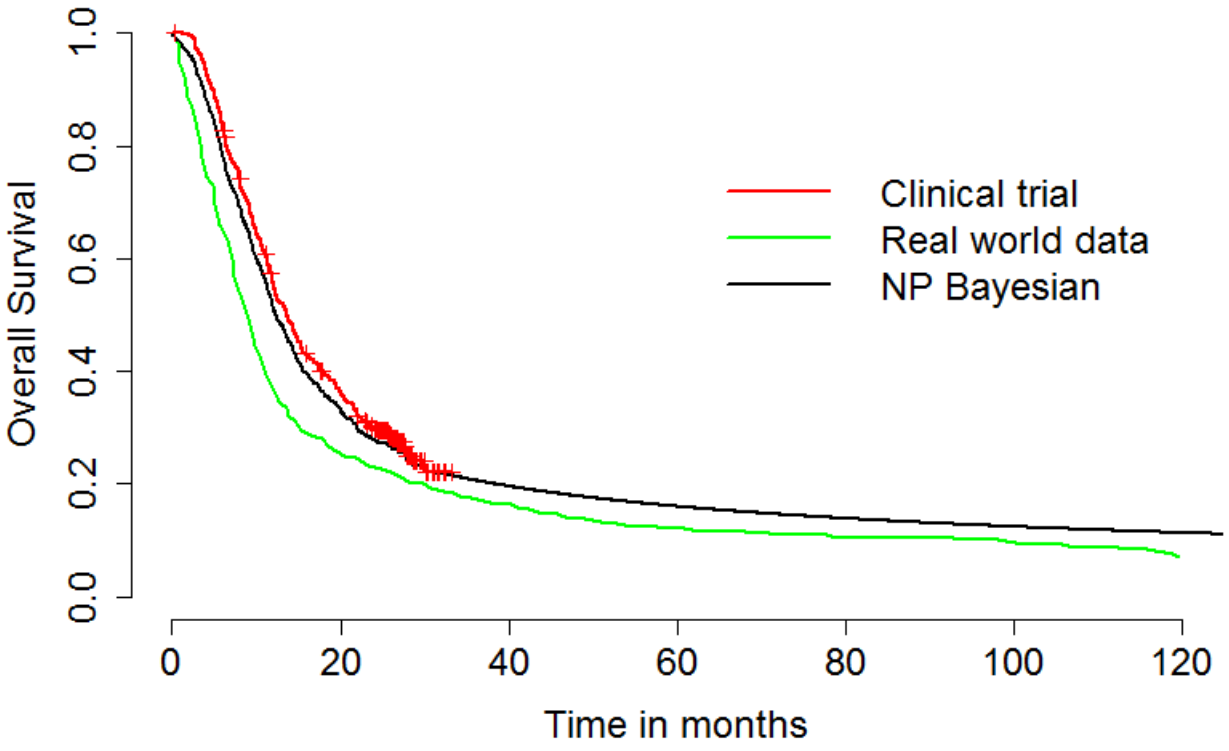
- Continuous function between two event times
- Coincides with the Kaplan Meier estimation for big sample size
- Is driven by the prior information for small sample size
- Takes into account the censoring and the event times

Nonparametric Bayesian estimation c=10



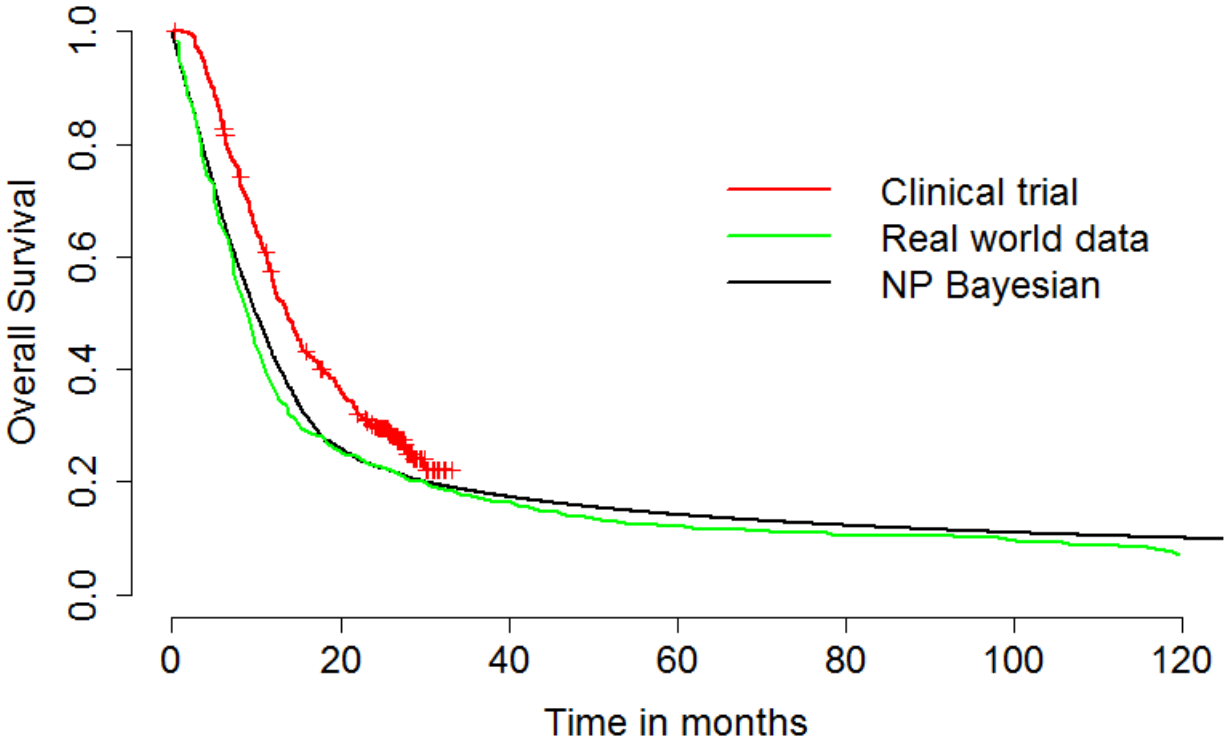
It overlaps with Kaplan Meier estimate while there is clinical trial available, when c equal to 10

Nonparametric Bayesian estimation c=100



Slightly under the Kaplan Meier from the clinical trial when c is equal to 100

Nonparametric Bayesian estimation c=1000



It overlaps with the Kaplan Meier from the real world data when c is equal to 1000

How to extrapolate survival data from a clinical trial?

- Combining clinical trial data with real world data
- This is possible in the Bayesian framework
- Several sensitivity analyses should be carried out

Some advantages of the Bayesian nonparametric estimation

- It is defined for all the time points (not only for the follow up trial)
- It allows combination between prior information and clinical trial data
- If we assume a Dirichlet process $S_0(t)$ is an exponential distribution
- Assuming a squared error loss function we have a conjugate prior, therefore we have a close form solution for the posterior distribution.

Statistical background

Using a squared-error loss function:

$$L(S, \hat{S}) = \int_0^\infty \left[\hat{S}(t) - S(t)\right]^2 dw(t),$$

where $w(t)$ is a weight function.

There are two classes of prior distribution that lead to a closed form estimates of the survival functions.

- Prior distribution for the survival function.
- Prior distribution for the cumulative hazard function

Prior distribution for the survival function

- Assuming survival function is sampled from a Dirichlet process with a parameter function α.
- $\alpha[t,\infty) = cS_0(t)$ where $S_0(t)$ is our prior guess at the survival function and c is a measure on how much weight to put on our prior guess.
- $\alpha(0,\infty) = cS_0(0)$
- Prior mean is given by: $E[S(t)] = \frac{\alpha[t,\infty)}{\alpha[0,\infty)} = \frac{cS_0(t)}{cS_0(0)} = S_0(t)$
- $S_0(t) = \exp(rt)$

The Bayesian nonparametric estimation:

Given the fact that is a conjugate prior the posterior distribution, the parameter α^* is given by:

$$\alpha^*\big((a,b)\big) = \alpha\big((a,b)\big) + \sum_{j=1}^{n} I\left[\delta_j > 0, a < T_j < b\right)$$

n distinct events times

The Bayesian nonparametric estimation:

Assuming M distinct times (censored or uncensored)

The bayes estimator of the survival function is given by:

$$\tilde{S}_D(t) = \frac{\alpha[t,\infty) + Y_{i+1}}{\alpha[0,\infty) + n} \prod_{k=1}^{i} \frac{\alpha[t_k,\infty) + Y_{k+1} + \lambda_k}{\alpha[t_k,\infty)}$$

At time i, Y_i is the number of individuals at risk, and λ_i is the number of censored observations.

For large n the bayes estimator reduces to a Kaplan Meier estimator.

For small sample size the prior will dominate.

How to assess uncertainty?

- How to sample from $\alpha^*\big((a,b)\big)$?

$$\alpha^*\big((a,b)\big) = \alpha\big((a,b)\big) + \sum_{j=1}^{n} I\left[\delta_j > 0, a < T_j < b\right)$$

- The posterior distribution is a Dirichlet

To assess uncertainty (work in progress):

So from Wikipedia we have:

- Using of Gamma-distributed random variables (y_i) one can sample a random vector from Dirichlet distribution

$$Gamma(\alpha_i, 1) = \frac{{y_i}^{\alpha_i - 1} e^{-y_i}}{\Gamma(\alpha_i)}$$

- Then

$$x_i = \frac{y_i}{\sum_{j=1}^{K} y_j}$$

x_i is a sample from a Dirichlet distribution

Take home messages

- In economic evaluations we are interested to assess long term outcomes
- The plausibility of the results should be also considered
- The non-parametric bayesian estimator provides a very natural way to combine two sources of information
- We can decide how much weight we put in our prior knowledge
- This approach is specially useful when patients in the control arm have switch to the experimental arm

References

- Ibrahim, J. G., Ming-Hui C., and Debajyoti S. *Bayesian survival analysis*. John Wiley & Sons, Ltd, 2005.
- Klein, J. and Moeschberger, M. (2003). *Survival analysis techniques for censored and truncated data*, Springer, New York.
- Latimer, N. R. (2013) Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. *Medical Decision Making*, 743-754.
- Xing Y., et al. (2010) Conditional survival estimates improve over time for patients with advanced melanoma. *Cancer,* 116(9), 2234-2241.
- Wikipedia: http://en.wikipedia.org/wiki/Dirichlet_distribution
- Wikipedia: http://en.wikipedia.org/wiki/Melanoma

Doing now what patients need next