Analysis of Male Elephants' Reproductive Success with Age: What Do the Data Tell Us?

Eric Vance

Saturday, June 18, 2005

1 Previous Data Decisions and Assumptions

- Since no calves were matched before 1977, all analyses will be done for the (conception) years 1977-2000.
- Because we do not know the ages of elephants "SM1" and "SM2", they will be removed from the data set, thus removing 1 matched calf. (Matched calves = 130.)
- We assume that, after possibly accounting for the specific **Years** a bull has been an adult and the bull's individual **Virility**, each genotyped bull has the same chance of siring a calf, given his **Age**.
- We also assume that the chance of siring a calf is equivalent to the chance of being matched with a calf in our data set. We can infer the rate of siring (all) calves from the rate of being assigned paternity to one of the genotyped calves.
- We assume that a bull sires calves at a rate μ_{ijk} dependent on his $\mathbf{Age}_j$, and possibly dependent on the $\mathbf{Year}_k$ and the bull's individual $\mathbf{Virility}_i$. The number of calves sired by a bull in one year has a Poisson distribution, $c_{ijk} \sim Poisson(\mu_{ijk})$.
- A bull's reproductive success **RS** at an age is defined by his rate of siring calves which is inferred from the actual number of calves sired at that **Age**.

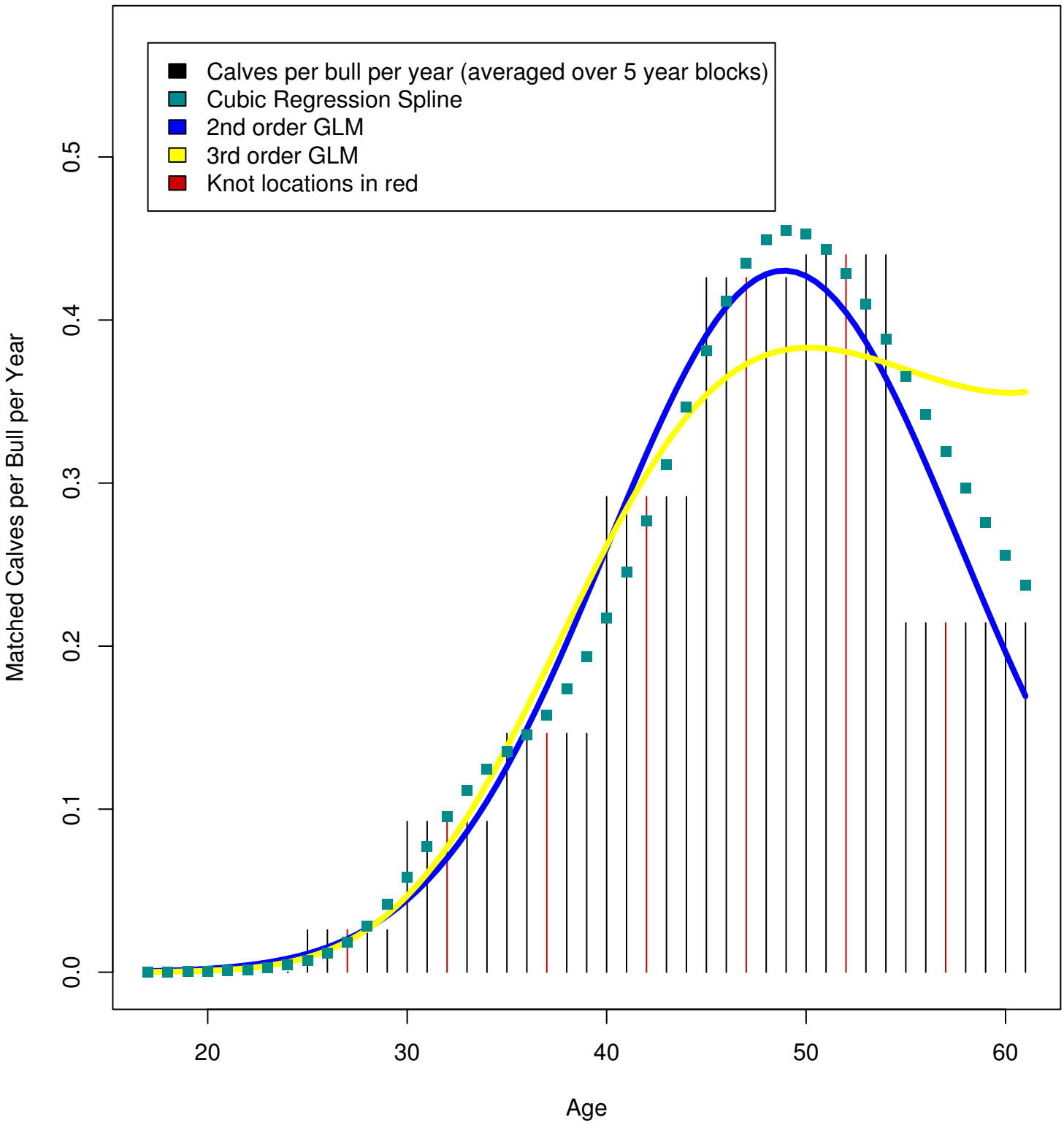
2 Preliminary Questions to be Answered

2.1 To spline or not to spline

- Should the number of calves sired between, for example, ages 17-30 affect the shape of the **RS** curve between, say, ages 50-60? If one assumes a parametric relationship between **Age** and **RS**, then data points at one extreme of the curve will influence the shape of the curve at the other extreme since they're all part of the same equation relating **RS** to **Age**.
- We have already assumed the Poisson model. If we further assume that the (log) rate, μ_j, of siring calves at $\mathbf{Age}_j$ is a quadratic function of **Age** (one that decrease at very old ages),

$$log(\mu_j) = \beta_0 + \beta_1 \, \mathbf{Age}_j + \beta_2 \, \mathbf{Age}_j^2,$$

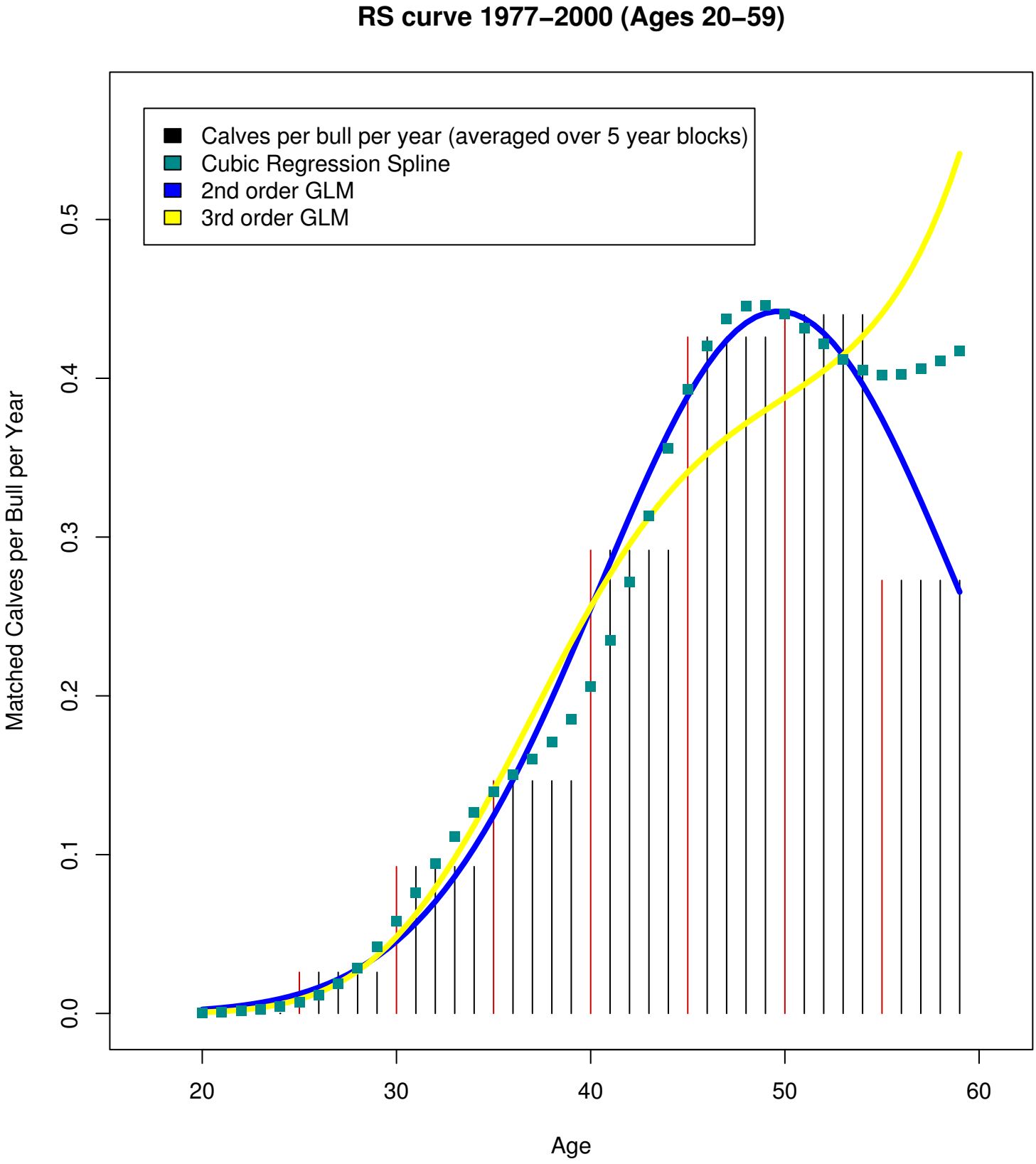
  then **RS** will be a normal curve! That means that the expected number of calves bulls sire in a year will be normally distributed over **Age** (with a peak somewhere around 50, and a SD $\approx$ 10 years).
- If you don't believe (or don't want to be forced to believe) that the specific relationship between **RS** and **Age** is the same across all ages, then you should use splines. Cubic Regression splines fit a smooth curve to individual sections of the data (between knots 17-21, 22-26, ..., 57-61), and then the algorithm forces each curve to line up (to be continuous) and smooth at each of the knots.
- Using splines means that what happens at ages 17-25 doesn't impact the shape of the RS curve between ages 56-61, and therefore it is a reasonable method to use for this data set. However, splines do not change the shape of the **RS** curve much (as can be seen in the first two figures).



- The quadratic GLM fit (in blue) forces **RS** to be normally distributed over **Age**.
- The cubic polynomial GLM fit (in yellow) predicts that **RS** will increase after age 60.
- The spline looks the most like the data. It looks like the best fit.

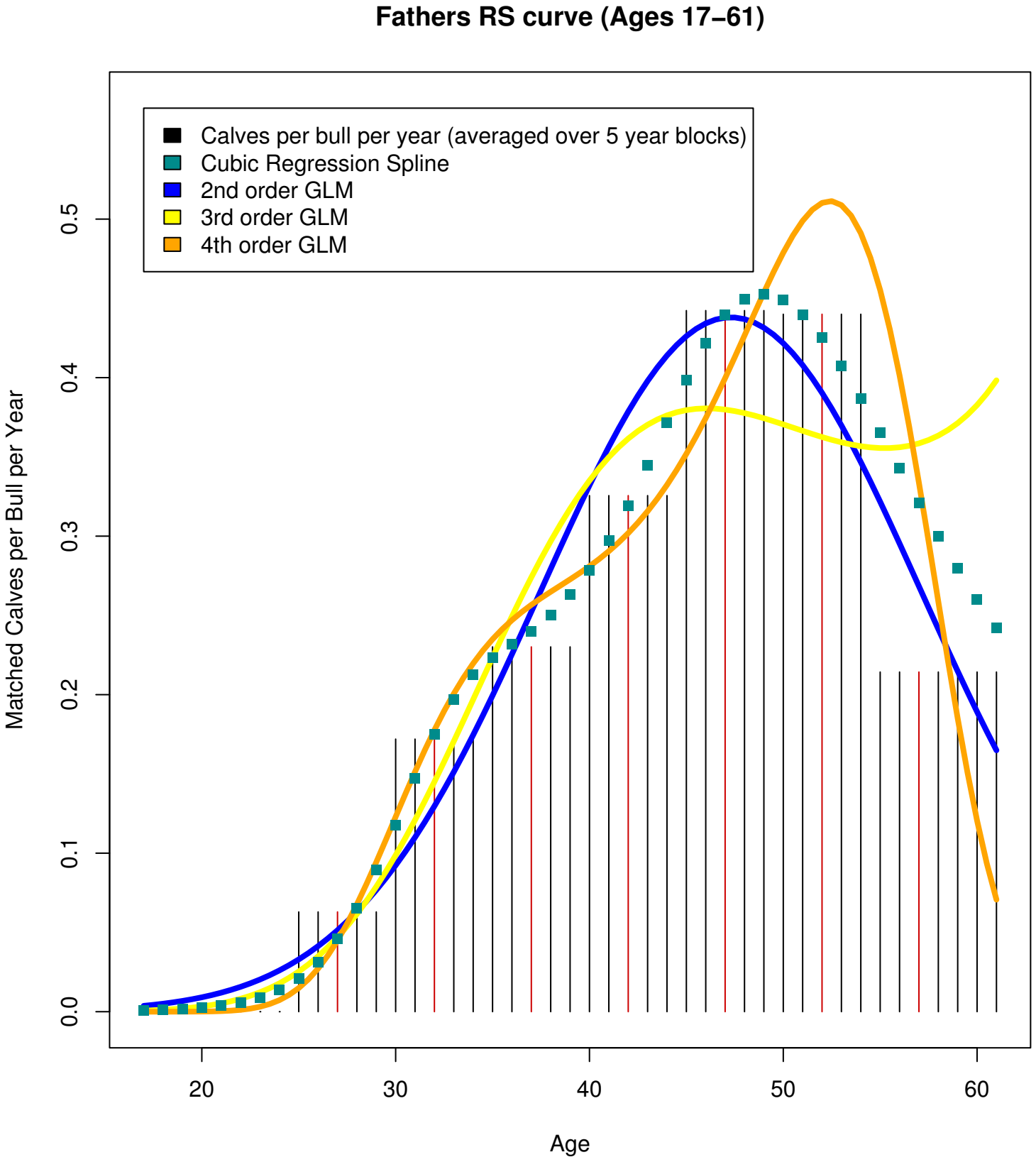
2.2 Sensitivity to extreme Age values

- What would the **RS** curve look like if no data were collected on the very oldest elephants at age 60 and 61? How reliable is the decline in **RS** in old age as seen in the previous figure?
- The spline fit shows no real decline after age 50. The 2nd order fit shows decline. The 3rd order fit shows an accelerated increase in **RS** after age 55.

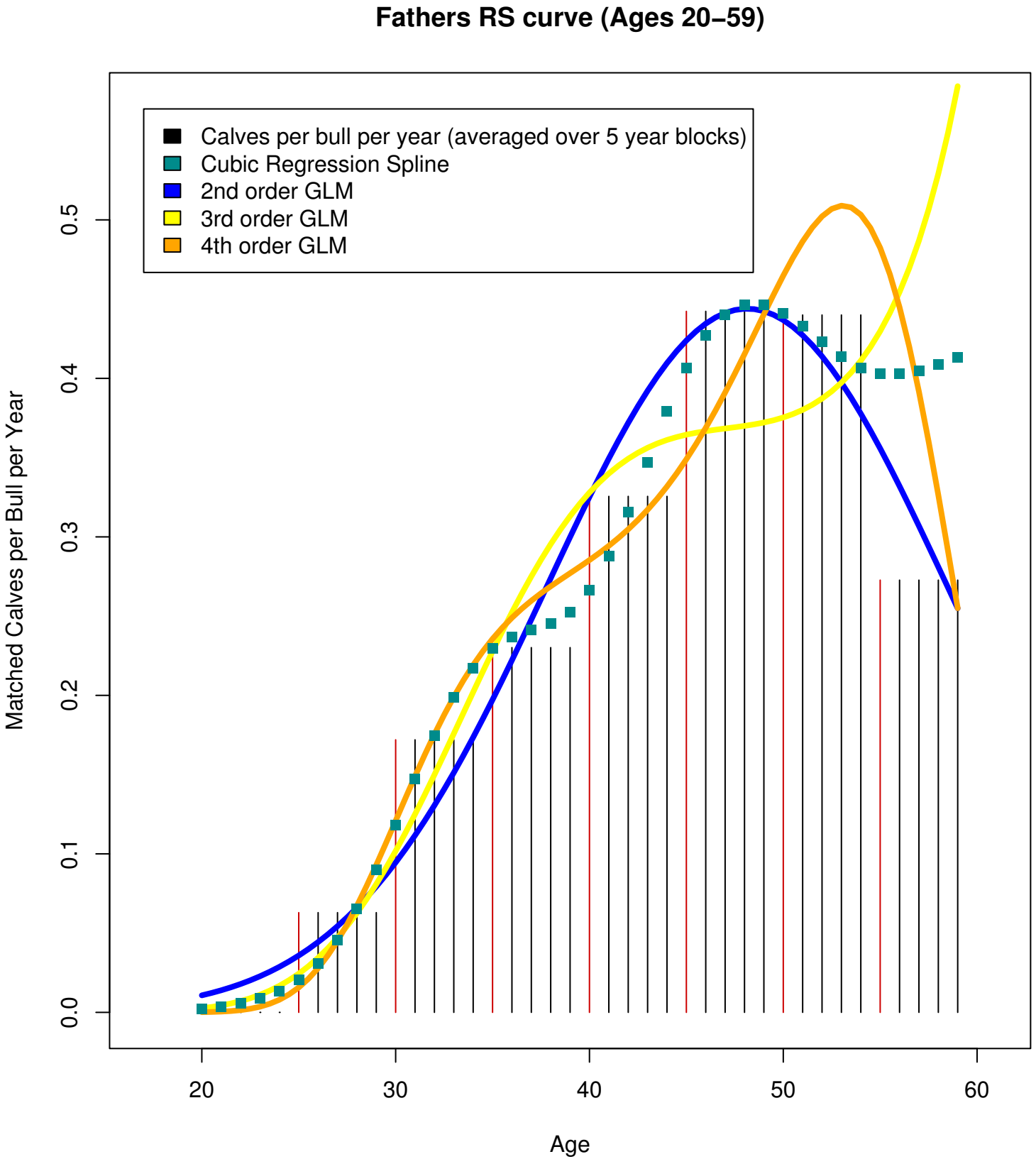


2.3 RS Curve for Fathers only

- How does the **RS** curve look for only the bulls (**Fathers** n=37) which have a paternity match? The **RS** looks very nearly the same as when all genotyped bulls (n=112) are included.
- In the previous plots the 4th order polynomial curve was not shown as these fits looked bad. This time, however, the 4th order polynomial looks like a reasonable fit.

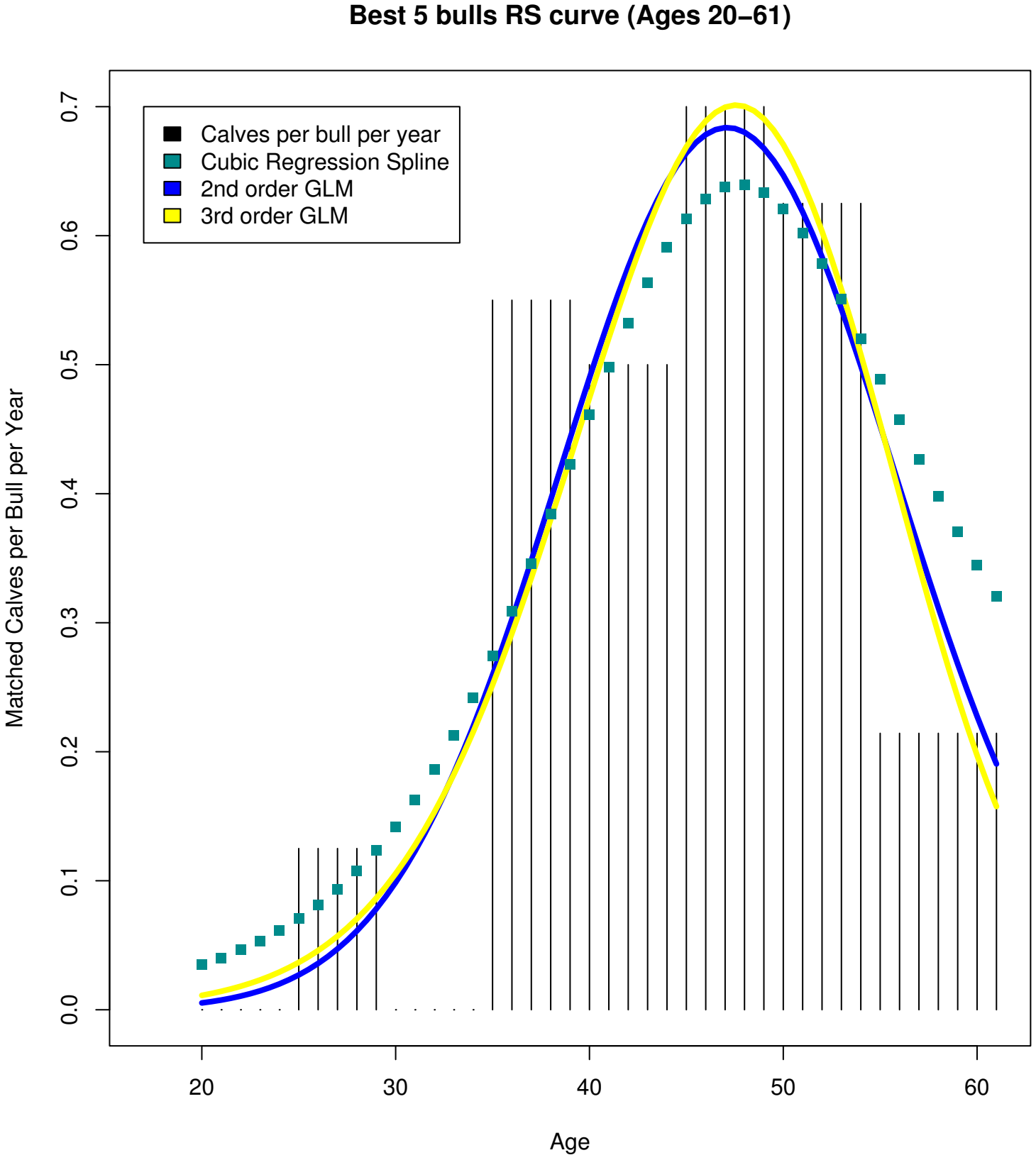


- How sensitive is the inference about the very old **Fathers** to the oldest three ages (60, 60, 61)?
- Only the 3rd order polynomial is unreasonable. The spline works well again.

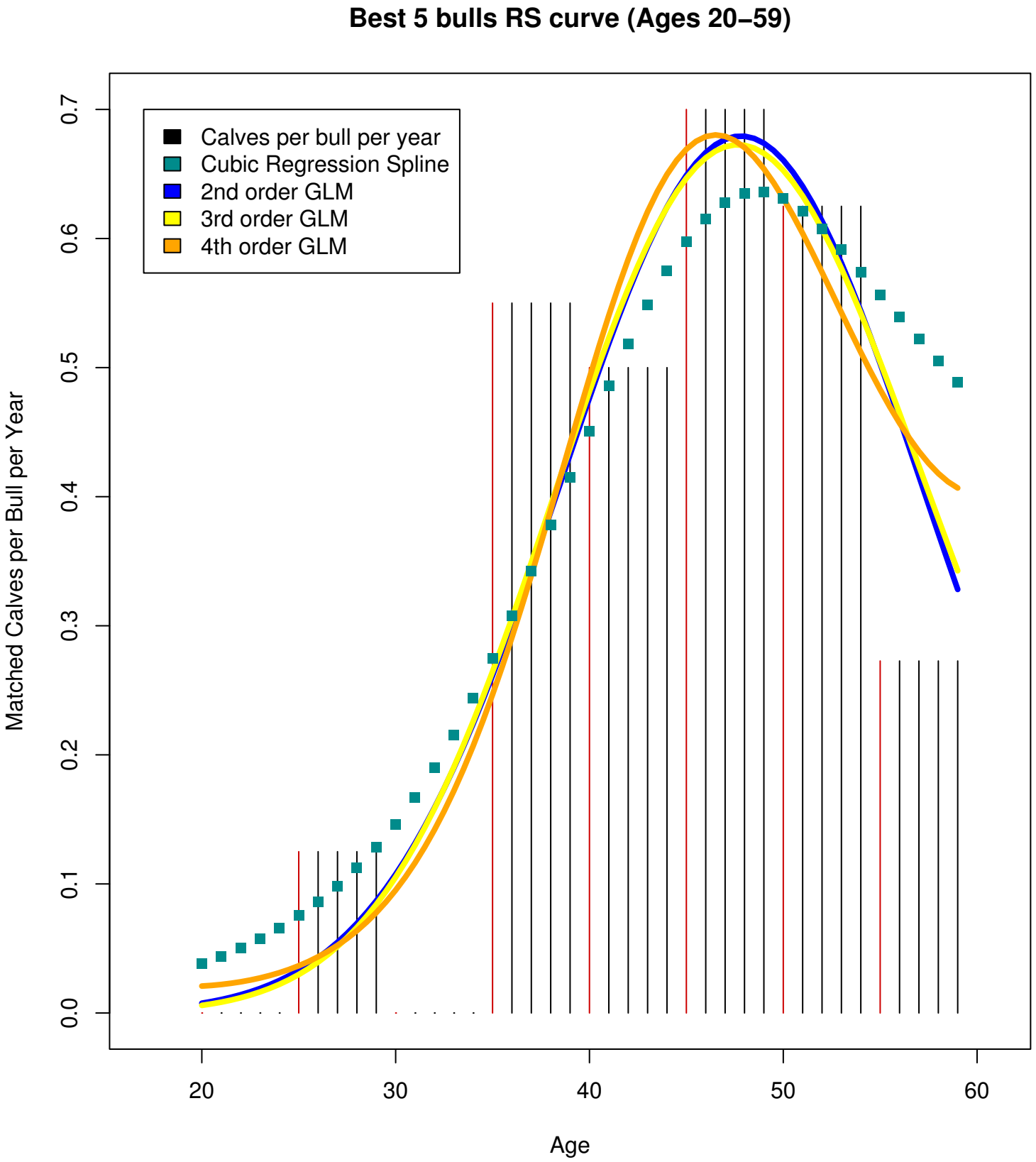


2.4 The 5 Best Male elephants

- How does the **RS** curve look for the five most successful bulls (in terms of total number (n=51) of calves sired)? The **RS** looks very nearly the same as for all the bulls and for all the **Fathers**.
- The Spline, 2nd, 3rd, and 4th order fits are nearly identical. Perhaps **RS** *is* distributed normally throughout the **Ages**. (Perhaps the sum/mixture of the individual bull's **RS** curves is Gaussian.)

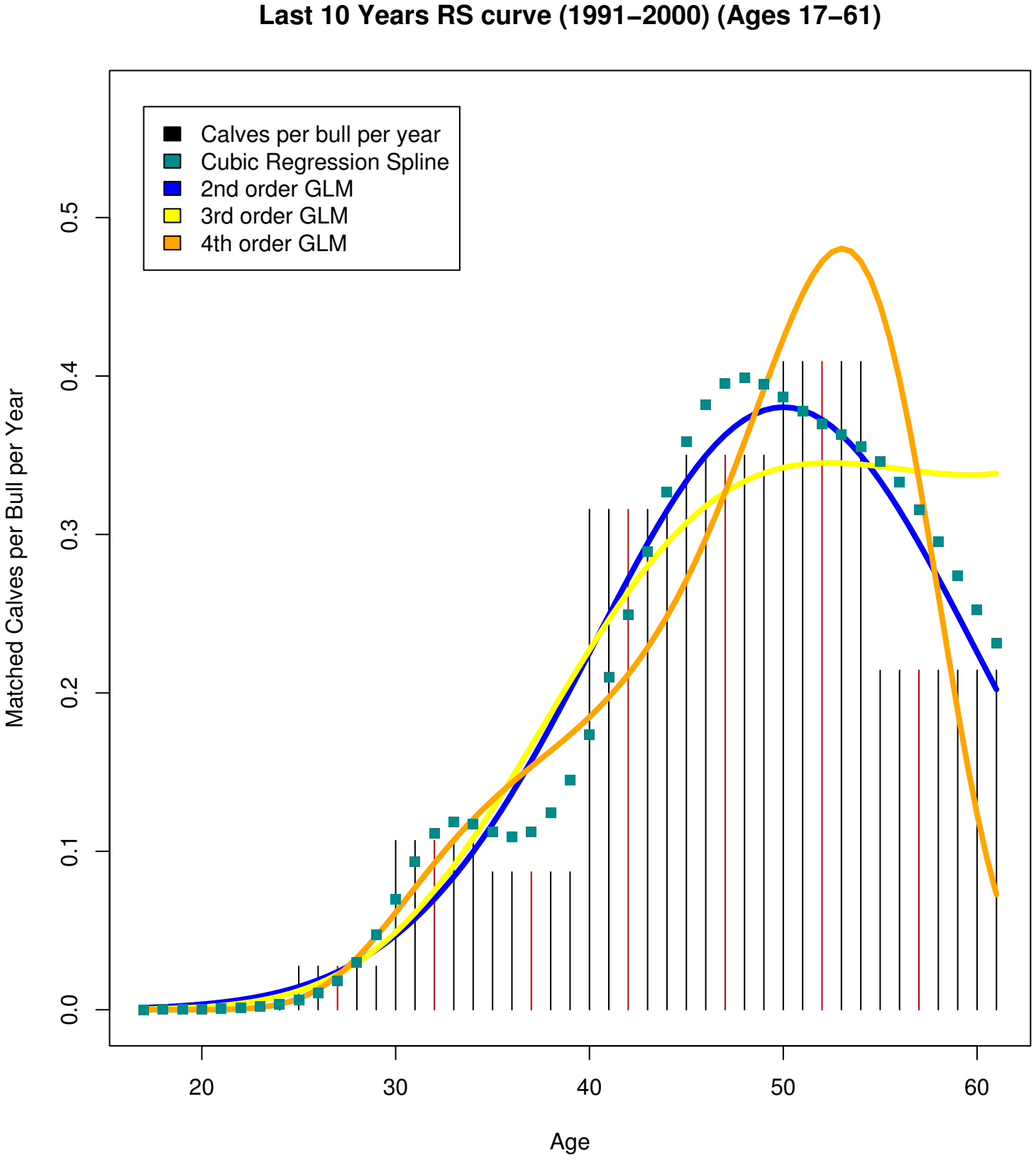


- How sensitive is the inference on the **RS** decline of the **5 best** elephants to the three data points of the oldest three ages (**Age** = 60, 60, 61)?
- The Spline flattens out, but a decline is still detected. All the fits look reasonable and similar.

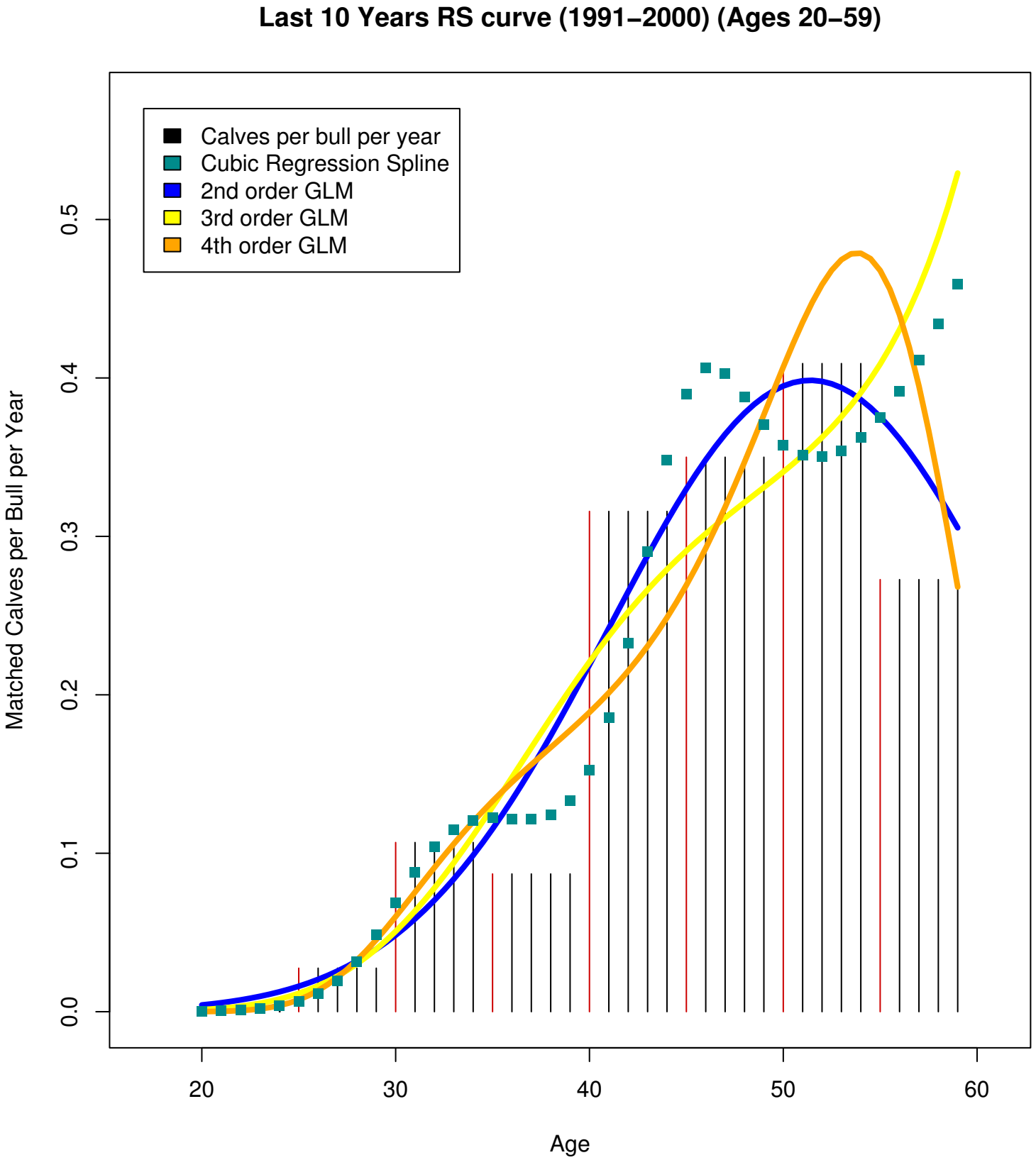


2.5 The Last 10 Years

- How does the **RS** curve look for the most recent (best quality?) 10 years of data, 1991-2000? The **RS** curve looks similar to when all the years (1977-2000) are included. The Spline looks the best.
- From 1991-2000, 83 calves were matched to 33 bulls. From 1977-1990, 57 calves were matched to 15 bulls.

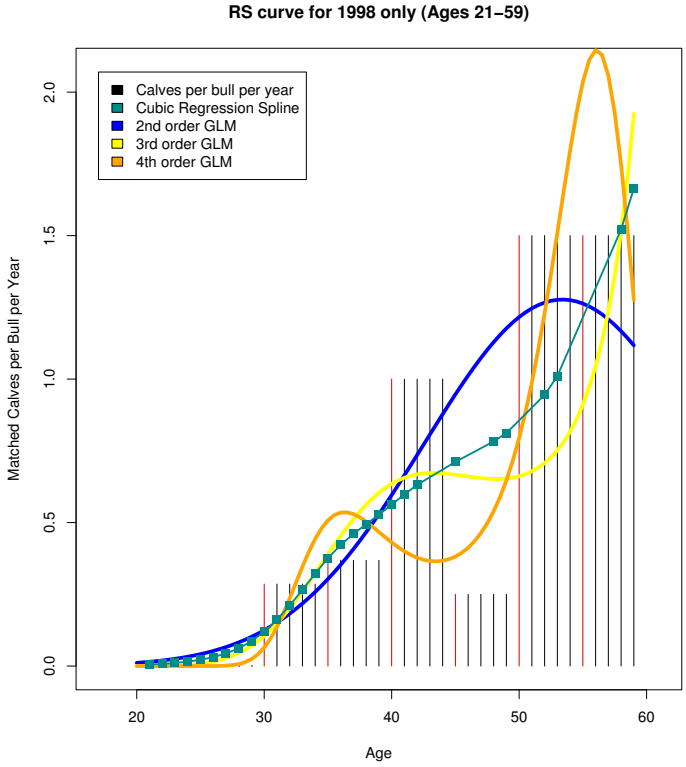


- How sensitive is the inference on the **RS** decline for the last 10 years to the three data points of the oldest three ages (**Age** = 60, 60, 61)?
- The Spline and 3rd order fits keep increasing for old **Ages**. It is clear that our inferences are very sensitive to those two very old bulls (comprising 3 years of no matched calves).



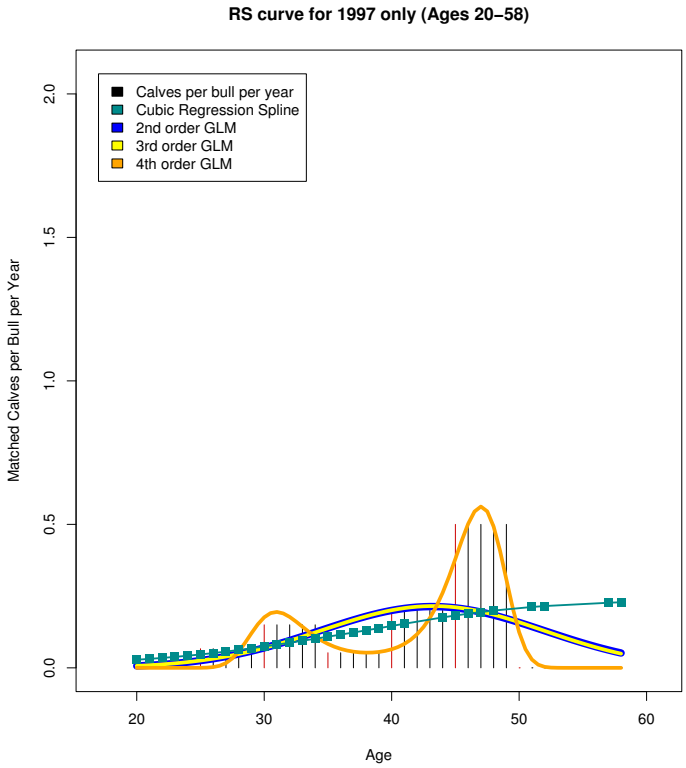
2.6 1998: A Huge Year

- How does the **RS** curve look in a single **Year**? 1998 was a very big year as far as matching calves (n=25 to 19 bulls). Nearly one-fifth of all the calves (n=130) in the data set were conceived in 1998.



1997: A Typical Year?

- In 1997, 8 calves were matched to 7 bulls. None of the methods (especially this spline) does a very good job of fitting a curve to such little data. Yet, the shape of the **RS** curve remains roughly similar.



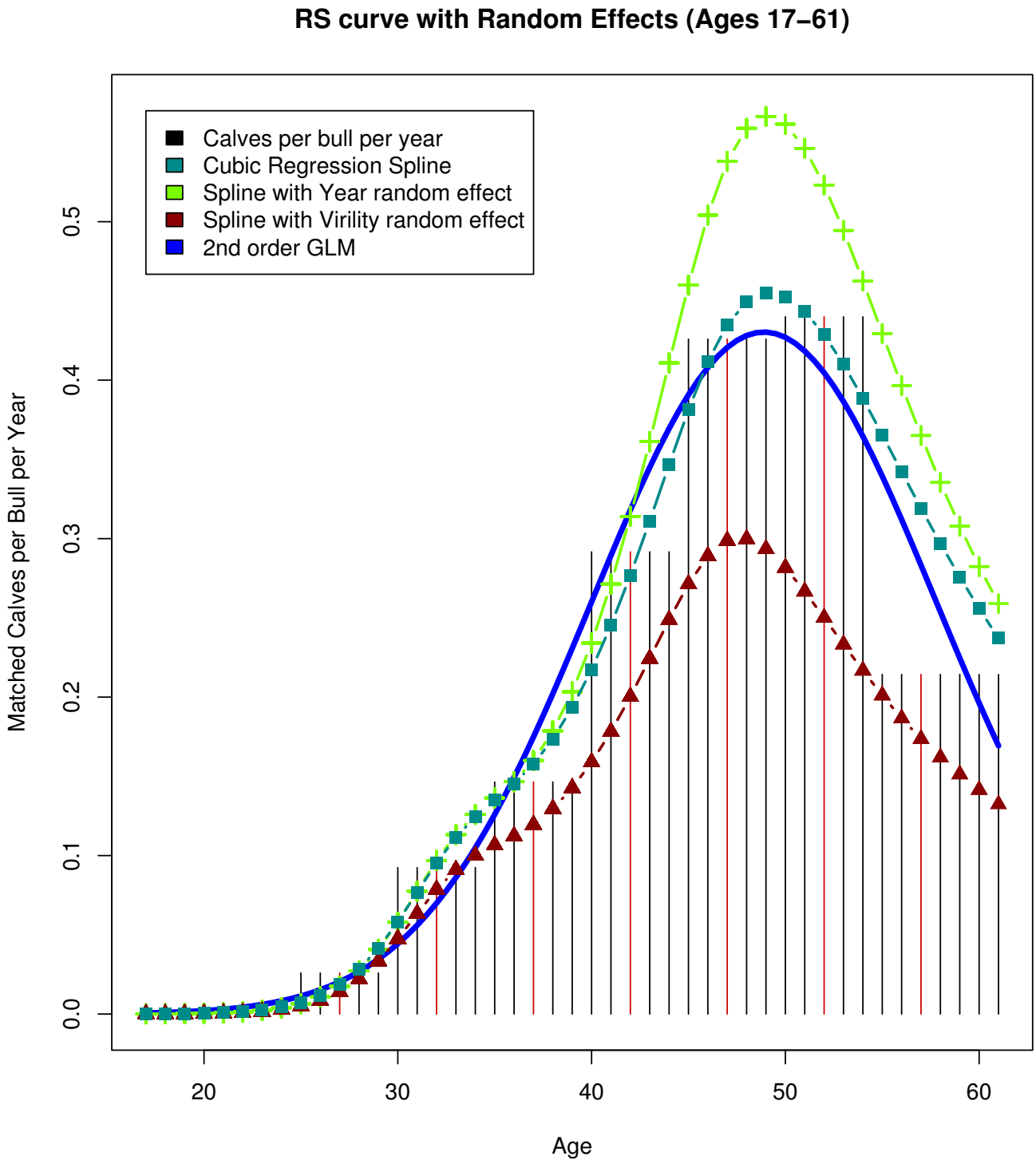
2.7 Random effects?

- Including a random effect for each $\mathbf{Year}_k$ in the model means that we think the **RS** curve for each year has the same shape, but is multiplied by a constant, $e^{\mathbf{Year}_k}$. This lessens the overall impact of a really big year (or bad years) by "assigning" some of the $\mathbf{RS}_{jk}$ to the $\mathbf{Year}_k$, and not to the $\mathbf{Age}_j$.

$$log(\mu_{jk}) = \beta_0 + \mathbf{Y}_k + \beta_1\,\mathbf{Age}_j + \beta_2\,\mathbf{Age}_j^2 \quad \Longrightarrow \quad \hat{\mathbf{RS}}_{jk} = \hat{\mu}_{jk} = e^{\hat{\mathbf{Y}}_k}\, e^{\hat{\beta}_0 + \hat{\beta}_1\,\mathbf{Age}_j + \hat{\beta}_2\,\mathbf{Age}_j^2}$$

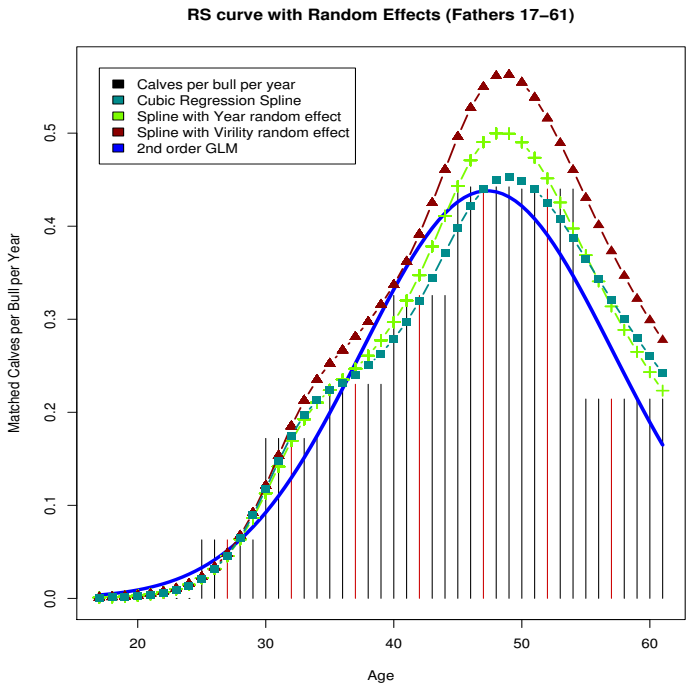
- Similarly, we might think that each bull has an individual **Virility** that inflates or deflates his **RS**.

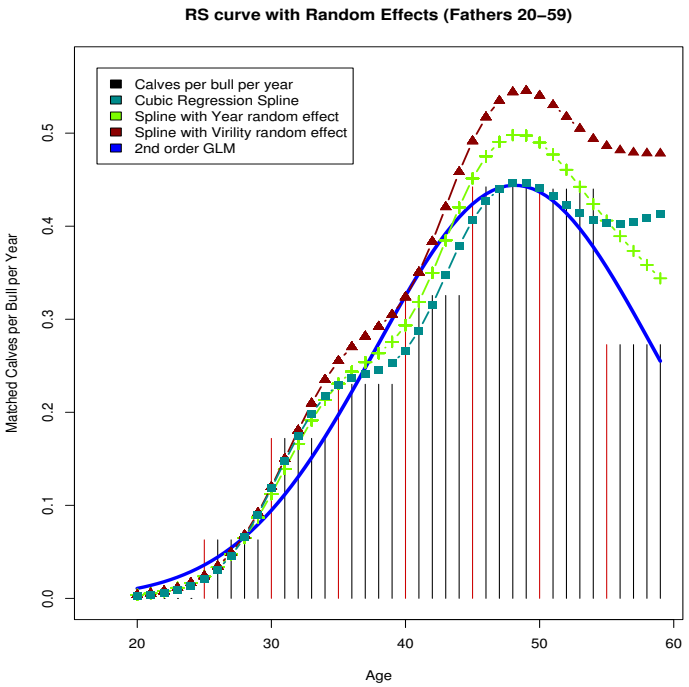
$$log(\mu_{ij}) = \beta_0 + \mathbf{V}_i + \beta_1\,\mathbf{Age}_j + \beta_2\,\mathbf{Age}_j^2 \quad \Longrightarrow \quad \hat{\mathbf{RS}}_{ij} = \hat{\mu}_{ij} = e^{\hat{\mathbf{V}}_i}\, e^{\hat{\beta}_0 + \hat{\beta}_1\,\mathbf{Age}_j + \hat{\beta}_2\,\mathbf{Age}_j^2}$$

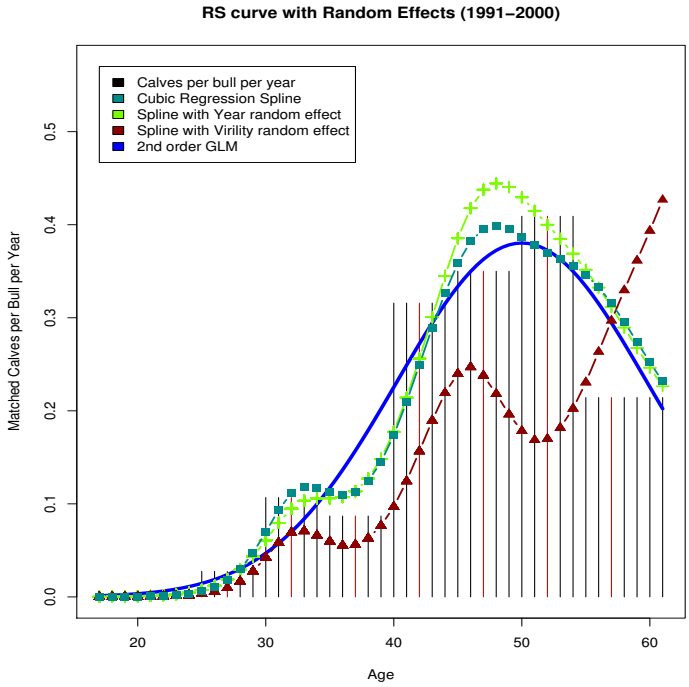


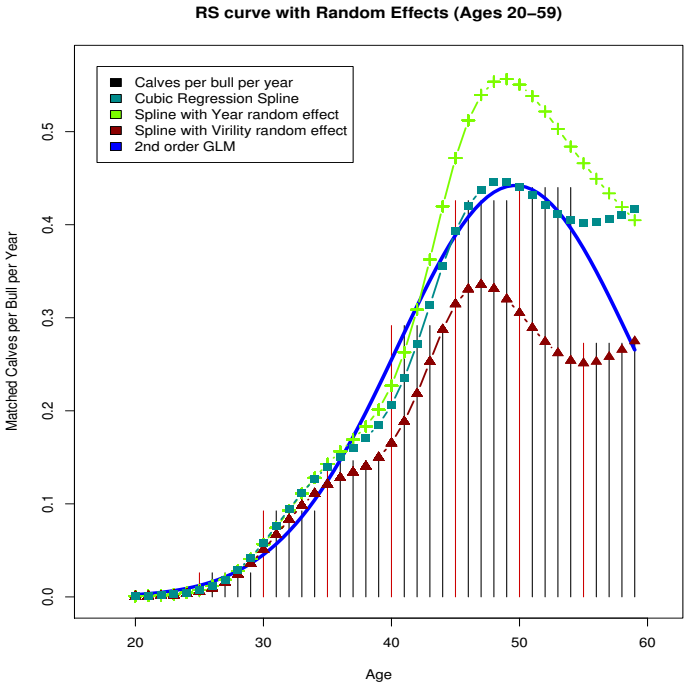
- These **RS** curves all have a similar shape. **Virility** deflates the average curve. **Year** inflates it.

- For **Fathers** the curves are similar. The young, low-virility, non-fathers no longer deflate **RS**.
- When only the **Fathers** aged 20-59 are considered, both the non-random effect Spline and the **Virility** spline show no decline in **RS** with **Age**. However, when **Year** is included in the model, the **Year** Spline shows that **RS** does decrease with **Age**, even when the three oldest ages are removed.
- For the last 10 years, 1991-2000, adding **Virility** to the model makes the curve crazy. The **Year** Spline and the non-random effect Spline both seem reasonable fits for the last 10 years of data.
- When the three oldest data points are removed from consideration, the **Year** Spline still indicates that **RS** decreases with **Age**. One reason for this is that in 1998 an unusually high percentage of calves were matched to older bulls. The **Year** Spline model attributes some of this observed **RS** for the old elephants to the fact that 1998 was a very good year for matching calves. Therefore the **Year** Spline model downweights this big year. It also downweights the importance of the lack of calves to the 60+ year-old elephants as these "failures" occurred in years with relatively few matched calves.









3 Conclusions

⋆1⋆ Splines allow for a flexible relationship between **Age** and **Reproductive Success** and do not force too many parametric assumptions. Splines give reasonable fits to the data and should be used.

⋆2⋆ There are very few data for the very old bulls. Therefore it is difficult to make inferences on what **RS** should be at old **Ages**. Ideally I would be able to quantify the uncertainty in the curve at **Ages** 55+, but instead I will just say that it is quite uncertain. What I have done is test how sensitive our estimates of the curves are to the three data points at ages 60, 60, and 61 by removing them and refitting the **RS** curves. The **Year** Spline gives similar results with and without these points. The other curves (except the normal curve) predict that **RS** stays flat or even increases in old **Age**.

⋆3⋆ The **RS** curve estimated from the data for only the bulls that have sired a calf (**Fathers** n=37), is very similar to the **RS** estimated by considering all the bulls (n=112). Again, only the **Year** Spline is robust against the three oldest ages.

⋆4⋆ The **RS** curve for the five most prolific bulls looks similar to the curves for all bulls and for all **Fathers**. The splines with random effects (see Appendix) are also very well-behaved. If the best bulls had an **RS** curve with a different shape, we would be worried about our model and inferences, but everything seems to be okay. The distribution of **RS** across **Age** looks surprisingly Gaussian.

⋆5⋆ The majority of the calves (83/130) were matched in the last ten years (1991-2000). Only the **Year** Spline gives a reasonable fit for this data when the three oldest ages are removed (see Appendix). Otherwise, the **RS** curve looks similar as that for the entire data set.

⋆6⋆ 1998 was an extremely successful year for matching calves to the genotyped bulls. Many calves were matched to older bulls. Thus the **RS** curve increases indefinitely with **Age**. In contrast, 1997 was a fairly typical year. Eight calves were matched to 7 fathers, but none of those fathers were over age 50. Chance variation in the number of calves born will exist among **Years**, yet there may also be some bias in the number of matched calves year-to-year as a result of the non-random sampling of the bulls (and perhaps the mothers and calves too?). Because years can be wildly different, the **RS** for a bull in a given year can be due to factors specific to the **Year** itself and not to the bull's **Age**. Including **Year** in the model accounts for the year-specific effects is therefore a good idea.

⋆7⋆ It is likely that there exists some sort of (unmeasurable) "virility" factor distributed among the population of male elephants. This factor implies that elephants with high virility will be more successful in siring calves than elephants with low virility. In terms of the Poisson model (which models the rate at which bulls sire calves), each bull will have an individual multiplier to the overall age-specific rate of siring calves. However, the many bulls that have never sired a calf could conceivably have a very wide range of negative **Virility** (possibly approaching $-\infty$). This leads to unstable estimates of **Virility** and therefore a loss of interpretability since the average virility must be zero. In addition, the **Virility** Splines do not always give reasonable estimates for the **RS** curve. I do not think we should include **Virility** in the model.

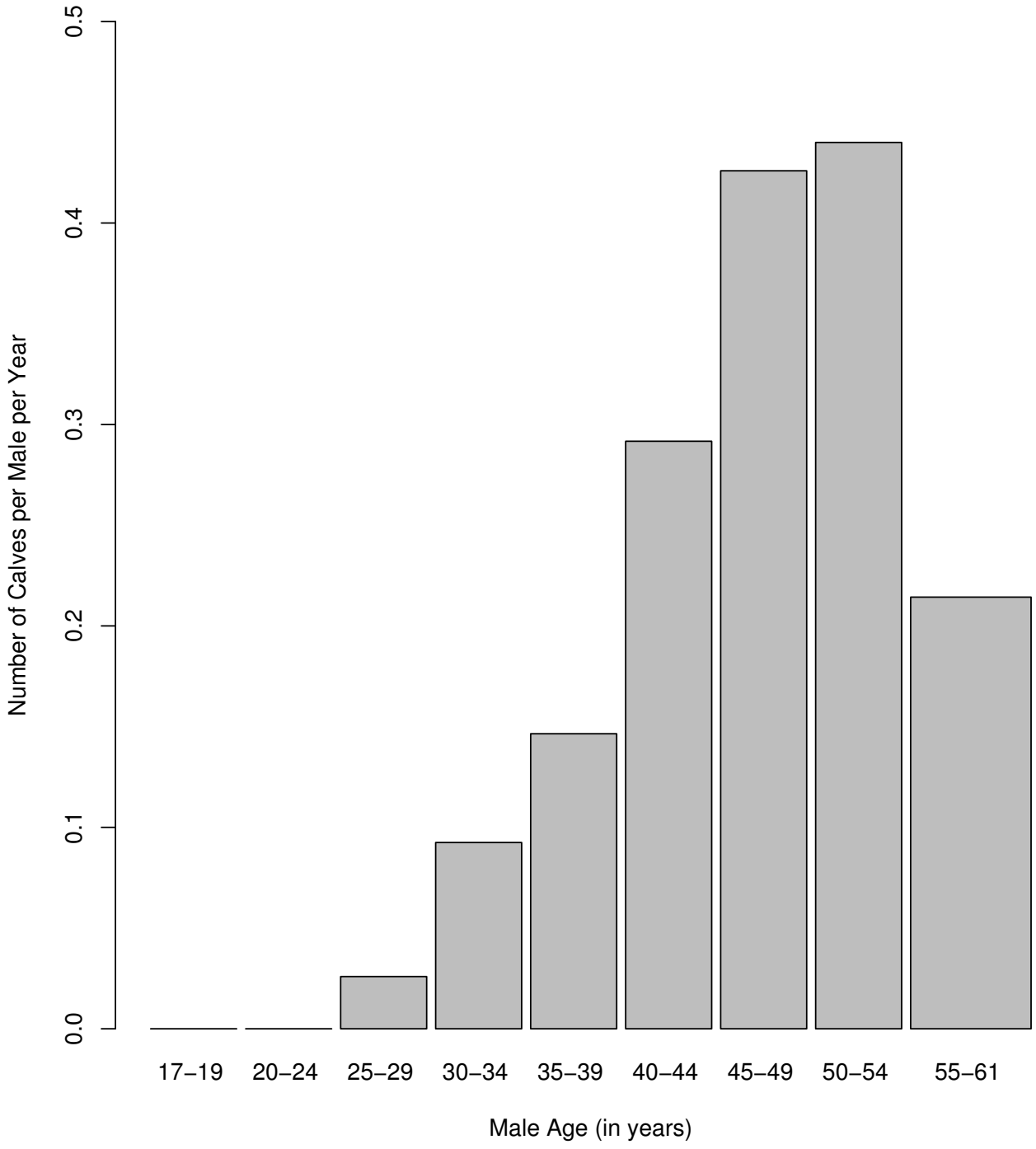
⋆⋆⋆ On the other hand, the **Year** random effect does aid our model. In every scenario I have tested, the **Year** Spline produces reasonable and robust estimates of the **RS** curve. Therefore, I propose that the **Year** Spline is used to analyze the relationship between **Age** and **Reproductive Success**.

4 Analysis for Nature paper

Age and Reproductive Success

The 112 genotyped males in our sample ranged in age from 17 to 61. The youngest male for whom we documented paternity was 26 years old at the time of conception while the oldest was 59 years old. Reproductive success, as measured by the number of calves sired per male at a given age, increases steadily from the mid-twenties until a peak between ages 45-55 (Fig. 1a). Reproductive success then declines in the oldest age class to levels comparable with a male in his early-forties.

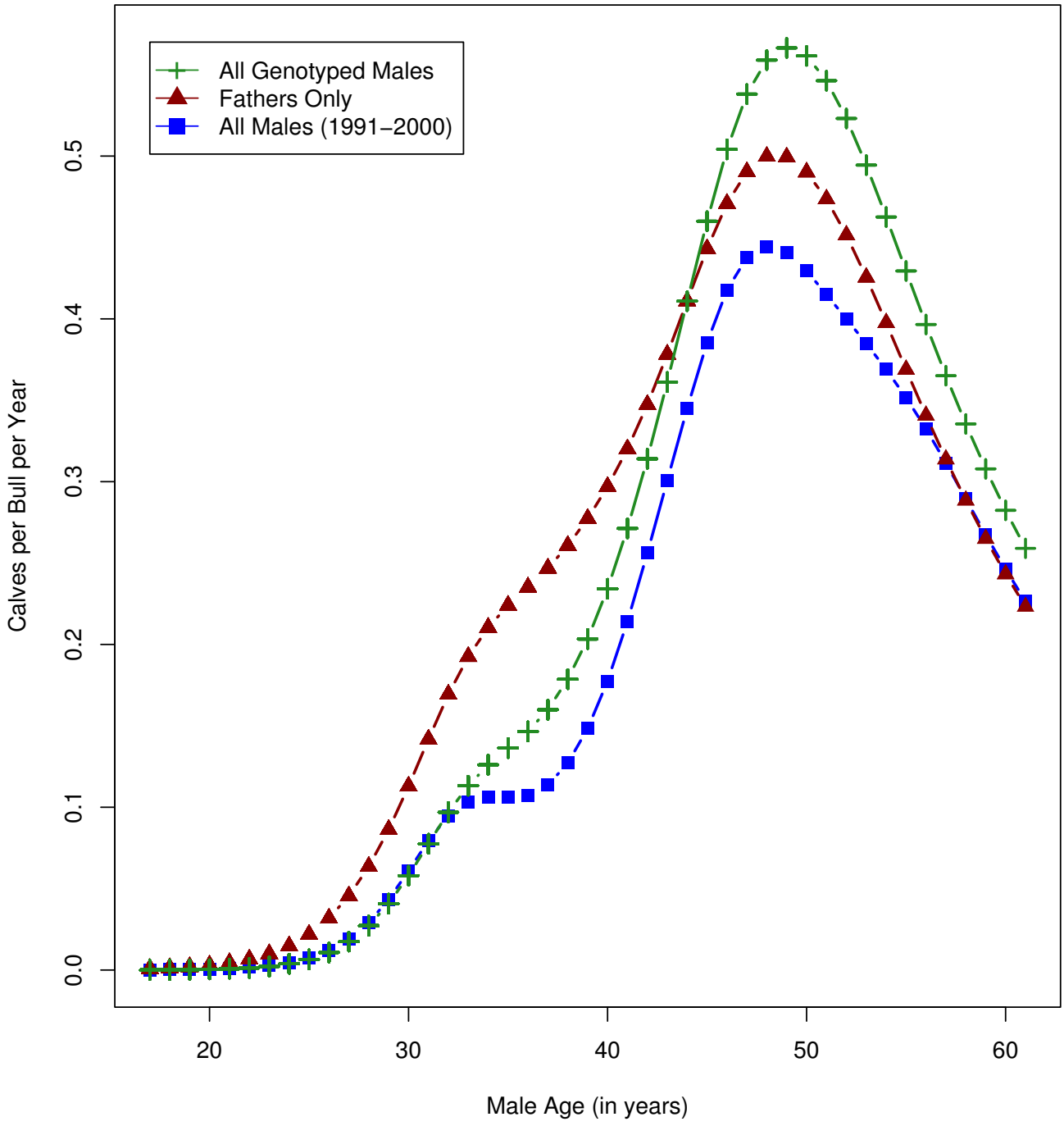
Fig. 1a: Number of Calves per Male per Year (1977–2000)



- (2nd paragraph)

We modeled the rate at which a male sires calves as a function of the male's age and the year in which the calf was conceived. The reproductive success rates were estimated using the data on the entire sample of 112 genotyped males for the duration of the study (1977-2000), using just the 37 males we identified as fathers, and using only the last ten years of data (1991-2000) (Fig. 1b). In each case reproductive success peaks in the late-forties to early-fifties, then declines to the rates of the early-forties.

Fig. 1b: Rate of Reproductive Success with Age



(note: the Fig. 1b curves could be overlaid onto the Fig. 1a bars)

• (3rd paragraph) There are a few inaccuracies in this paragraph, specifically the claim that "older males sire the majority of calves", as 50% of all calves in our data set were sired by bulls younger than 40. I started rewriting the paragraph but then got sidetracked into an analysis of the five males with the highest RS rate. These "studliest five" are different from the "Best 5" who happened to have sired the most total calves (actually 4 of 5 are the same). Of course, generally, the older a bull is the more calves he will have sired in his lifetime.

• Last note: I think the truncated Gaussian shape of the RS curves is fascinating. What we are seeing is that males' RS peaks at age 50. Then they die quickly (because of their teeth!) before their RS declines too much and they become useless for the survival of the species. The RS curve predicts that a 70 year-old would be as reproductively successful as a 30 year-old, if not for the fact that the 70 year-old is dead.

5 Appendix