



Royal Veterinary College
University of London

Improving the fit of non-linear regression models in SNP genotype association studies

Geoff Pollott

Senior Lecturer in Bioinformatics and Genetics

Motivation

- Many quantitative traits are measured repeatedly over time on the same animals. (For example, liveweight with age)
- Genetic analyses of such traits often involve fitting a linear mixed model, including a term to account for the change in the trait over time
- Such models can suffer from three major drawbacks
 - the best-fit overall relationship between the trait and age may be difficult to linearise,
 - the variance of the trait may vary with age.
 - the estimates of fixed effects may vary with age (SNP genotype effects)

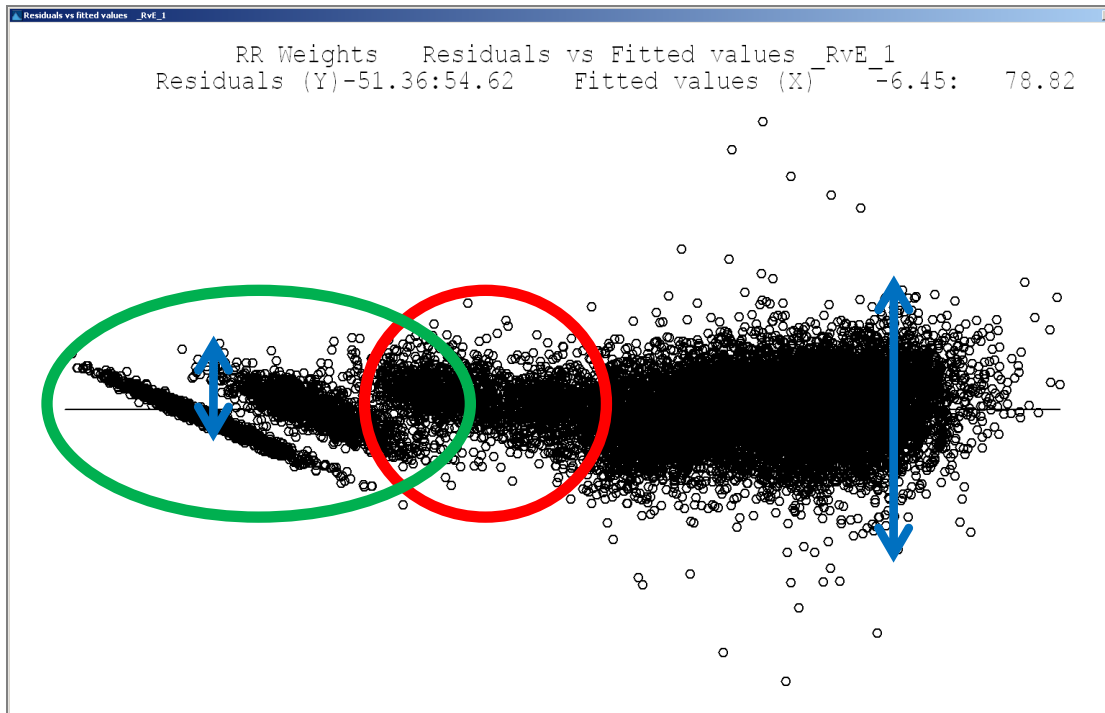
Two examples

- 1) Sheep growth from birth to maturity – genetic parameter estimation
- 2) Heifer growth to first mating – leptin SNP effects
Clempson et al. (2011) *Animal* **5**:1335-1343.

Common approach

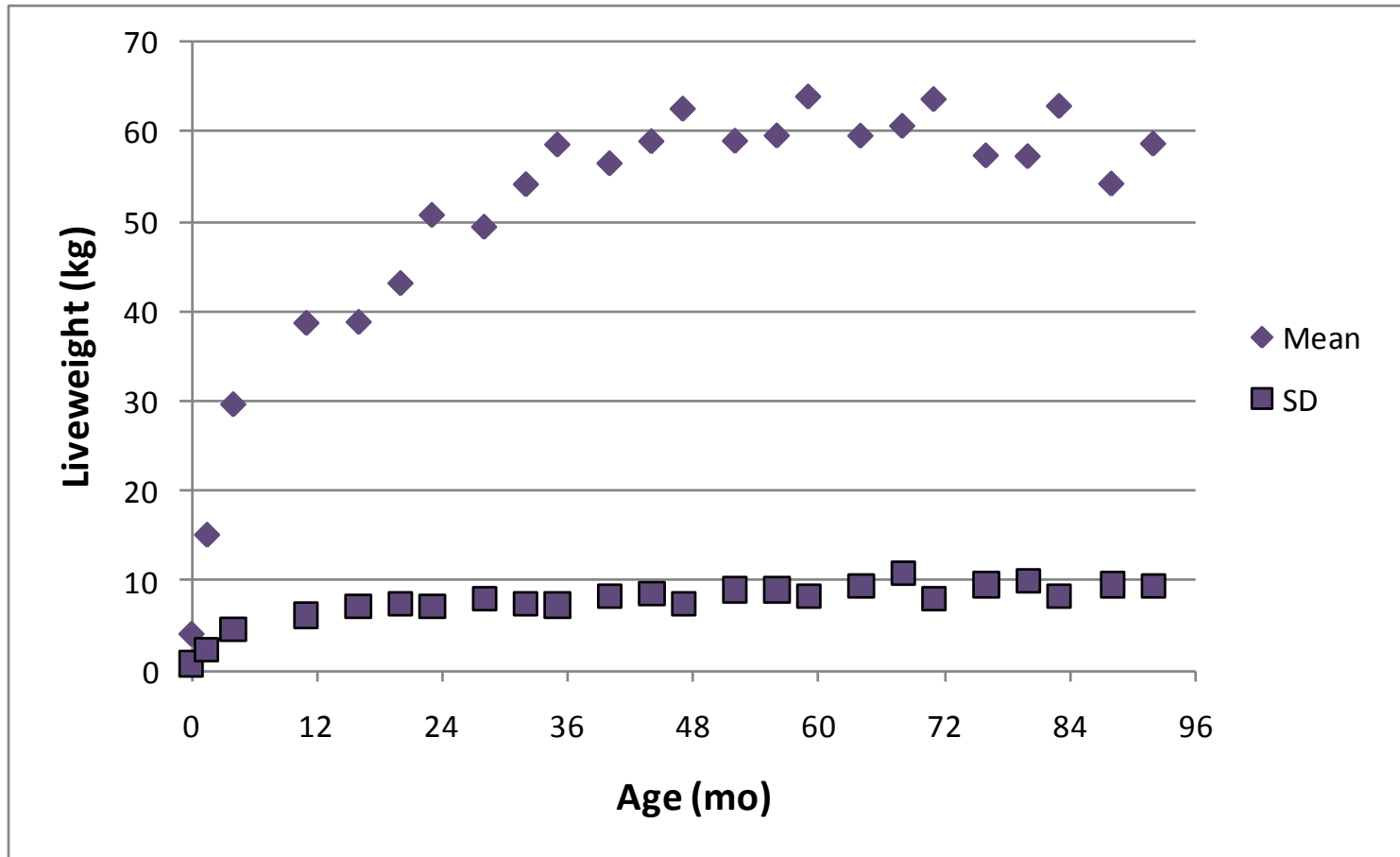
- Mixed model including term for the change in weight with age
- Polynomial used because difficult to linearise appropriate growth model

Residuals - mixed model - polynomial - sheep



- Poor fit of polynomial
- Residual variance differs with weight (age)
- 'Correction' factors unsuitable for all ages

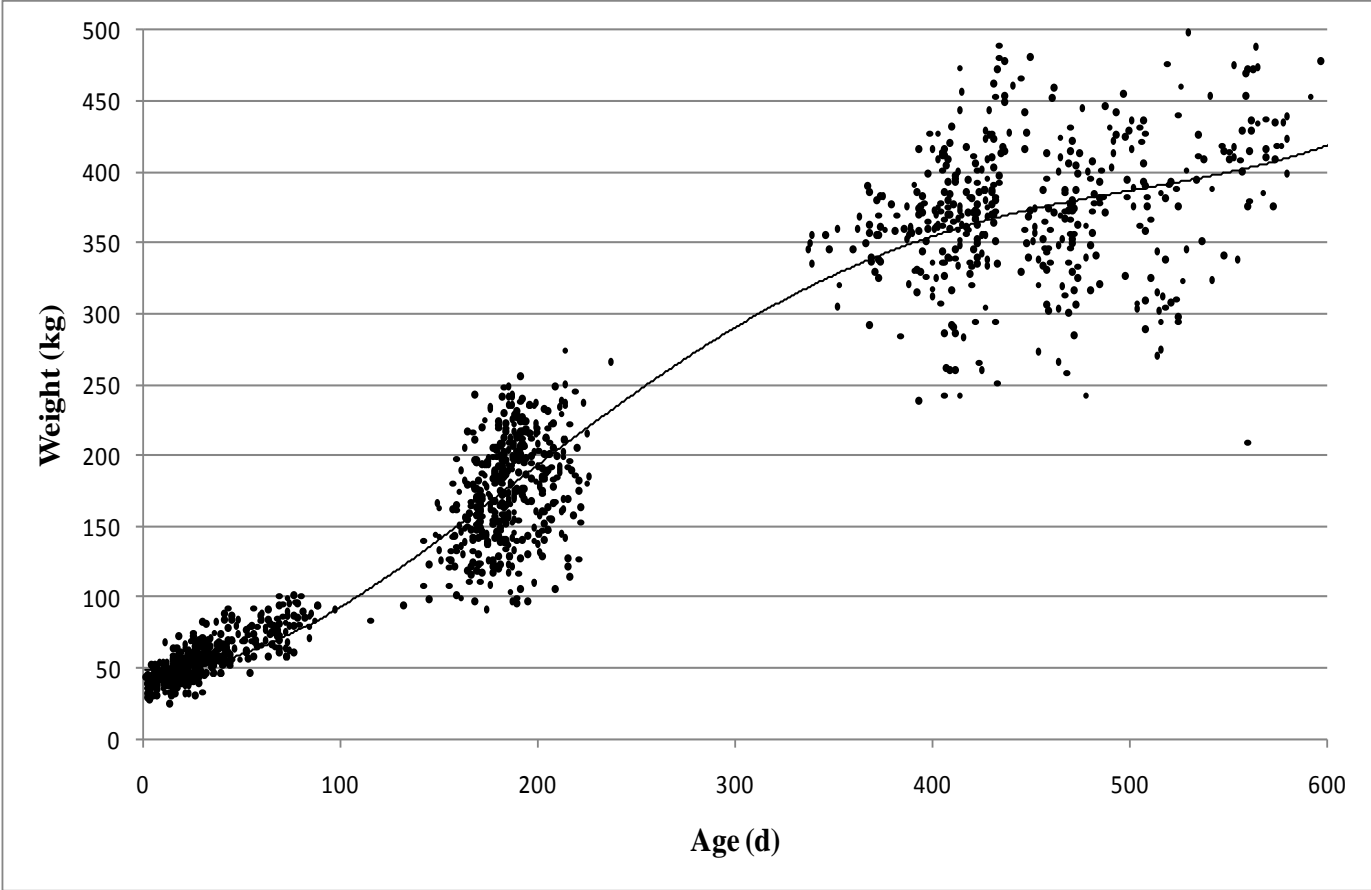
Mean and SD of liveweight by age - sheep



Effect of birth type on liveweight at different ages (kg) - Sheep

	Birth	6 wks	16 wks	8mo	36mo	40mo	45mo
Twin	3.88	14.59	28.51	30.26	49.62	50.46	57.89
Single	4.57	17.74	33.76	32.20	51.20	52.86	58.60

Plot of raw data - Heifers



Proposed solution

Adjust data

For each liveweight measurement:

Subtract a suitable mean

Divide by a suitable SD

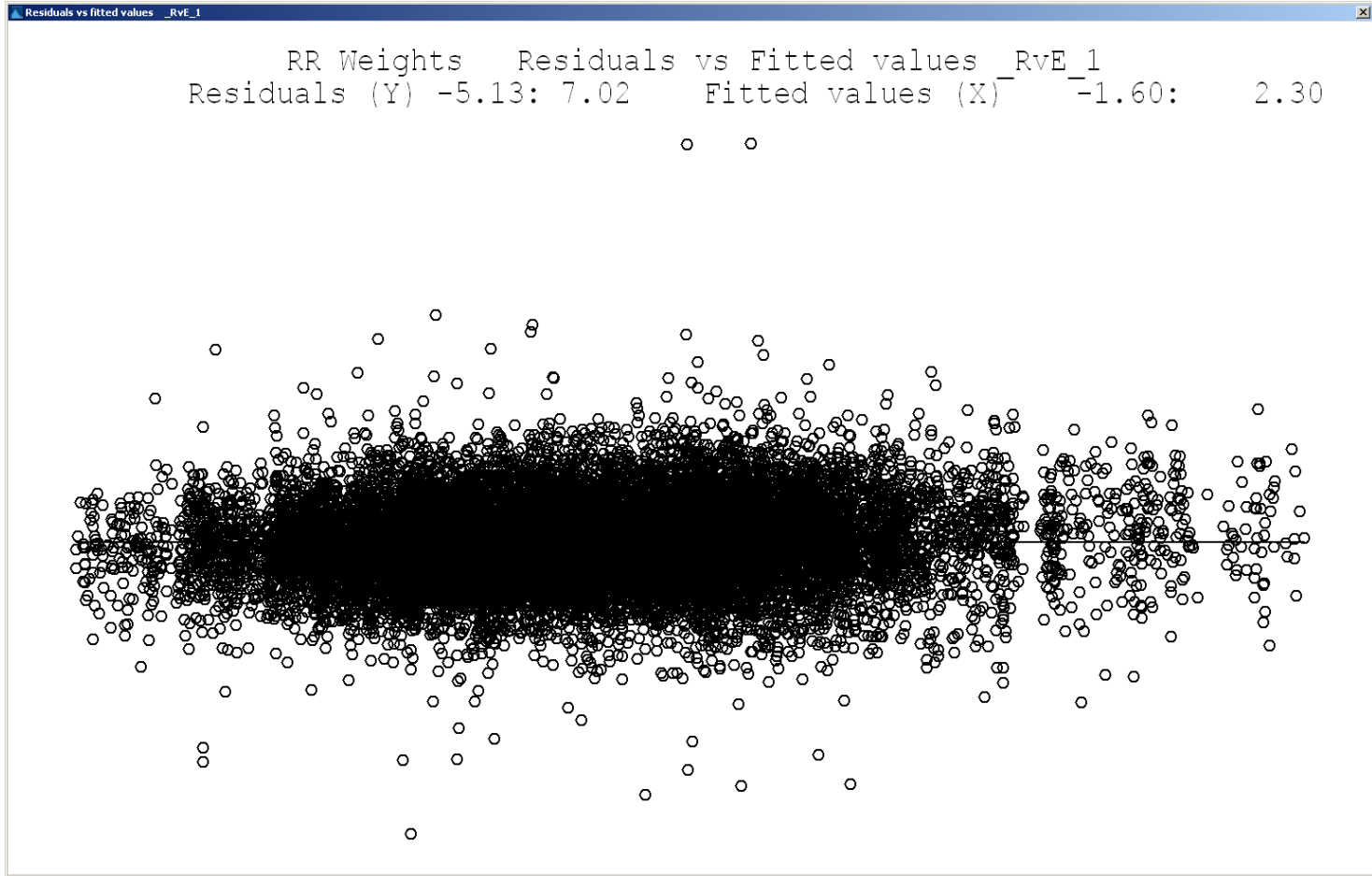
For each weight calculate adjusted weight as:

$$(\text{weight}_{\text{BWt}} - \text{mean}_{\text{BWt}}) / \text{SD}_{\text{BWt}}$$

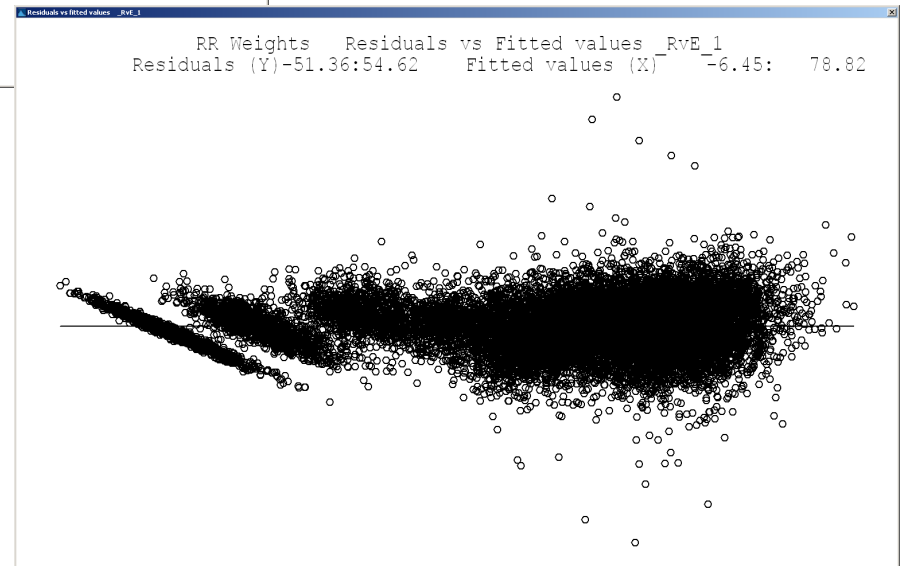
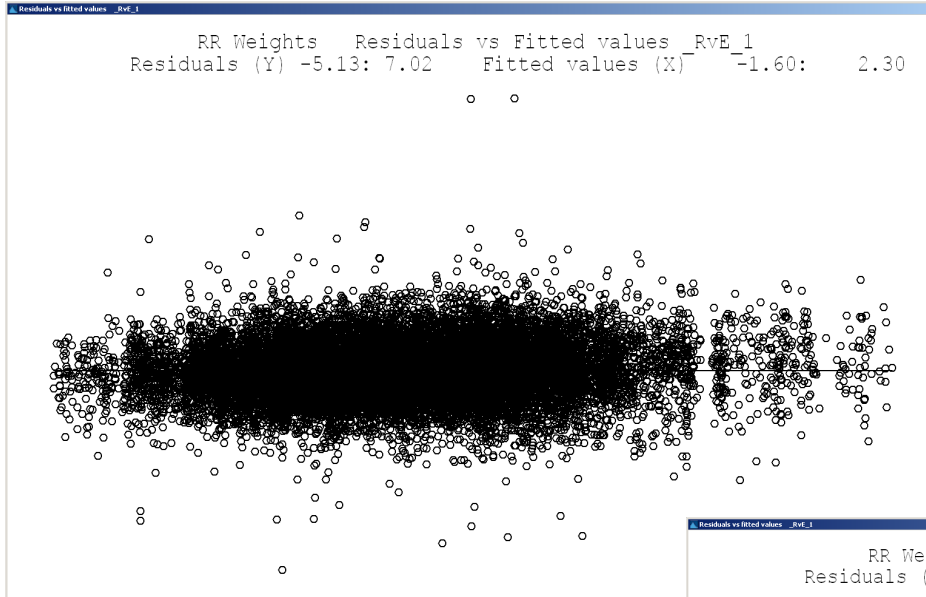
using sheep birth weight (BWt) as an example

Now analyse adjusted weights with appropriate mixed model

Residuals from analysis of adjusted data - sheep



Comparison with original analysis - sheep



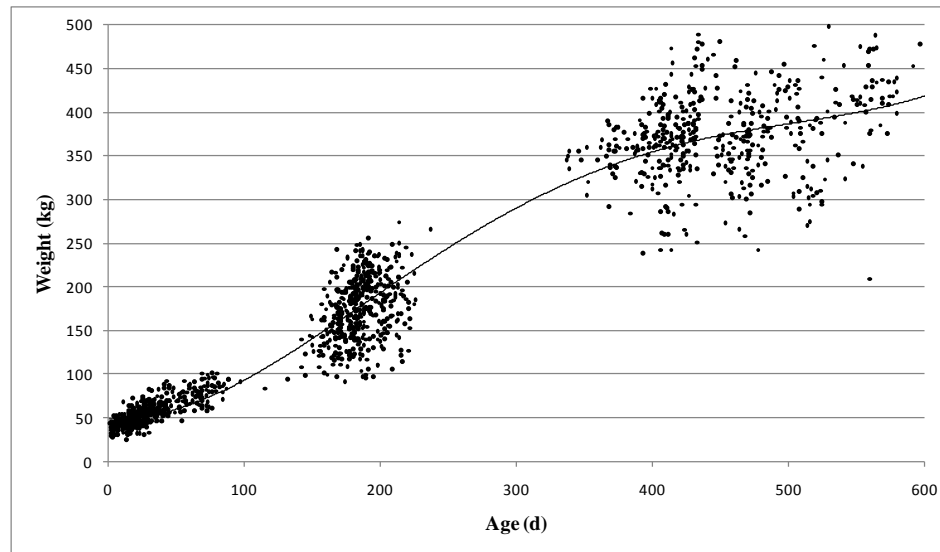
ANOVA summary for sheep liveweight (kg)

	Original	Mean-adjusted	Mean/SD-adjusted
Birth type	***	***	***
Year of birth	***	***	***
Dam age	***	***	***
Age of animal	***	*	NS
Animal variance component ¹	8.74	8.98	0.216
Ewe variance component	6.36	7.43	0.133
Error variance	49.61	31.36	0.587
<i>Calculated values</i>			
Phenotypic variance	64.71	47.77	0.936
Heritability	0.14 ± 0.02	0.19 ± 0.03	0.23 ± 0.03
Repeatability	0.23 ± 0.01	0.34 ± 0.01	0.37 ± 0.01
Variance of residuals	46.63	29.24	26.90 ²

¹ All variances are in kg² except for the Mean/SD-adjusted data, which is unitless.

² In kg² after conversion of all residuals back to the original scale.

What are the appropriate means and SD to adjust the heifer data?



- Calculate mean and SD for each contiguous run of 50 animals along the age scale
- Find appropriate relationship with age
- Adjust each liveweight by its age-related mean and SD

ANOVA for heifer liveweight (kg)

	Original	Mean-adjusted	Mean/SD-adjusted
Herd/year/season effect	***	***	***
SNP genotype effect	NS	**	*
Age of animal	***	***	**
Animal variance component ¹	79.7	78.6	0.166
Error variance	692.2	692.0	0.525
<i>Calculated values</i>			
Phenotypic variance	771.9	770.6	0.691
Heritability	0.10 ± 0.03	0.10 ± 0.03	0.24 ± 0.04
Variance of residuals	623	623	405 ²

¹ All variances are in kg² except for the age/SD-adjusted data, which is unitless.

² In kg² after conversion of all residuals back to the original scale.

SNP effects under different adjustments

Age	Original	Mean/SD-adjusted ¹
<i>Age 50 d</i>		
Homozygote 1	62.3 ^a	66.6 ^a
Heterozygote	60.6 ^a	65.8 ^a
Homozygote 2	54.9 ^a	64.1
<i>Age 500 d</i>		
Homozygote 1	392.4 ^a	398.7 ^a
Heterozygote	390.6 ^a	396.0 ^a
Homozygote 2	385.0 ^a	390.3

¹Standard errors were calculated from the means predicted from the mixed model analyses and adjusted using the age and SD for 50d and 500d, as appropriate.

^a Means within a column and age grouping with the same superscript were not significantly different ($P > 0.05$).

Discussion points

Suitable mean and SD to use

Genome-wide association study - this approach should pick up more significant SNPs

Alternatives - 10 error variances
- allow fixed effects to vary over age

SNP effects under different adjustments - 2

Age	Original	Mean/SD-adjusted ¹	Mean/SD-adjusted age within SNP ¹
<i>Age 50 d</i>			
Homozygote 1	62.3 ^a	66.6 ^a	67.6 ^a
Heterozygote	60.6 ^a	65.8 ^a	64.7
Homozygote 2	54.9 ^a	64.1	65.1 ^a
<i>Age 500 d</i>			
Homozygote 1	392.4 ^a	398.7 ^a	399.5 ^a
Heterozygote	390.6 ^a	396.0 ^a	396.9 ^a
Homozygote 2	385.0 ^a	390.3	391.1 ^a

¹Standard errors were calculated from the means predicted from the mixed model analyses and adjusted using the age and SD for 50d and 500d, as appropriate.

^a Means within a column and age grouping with the same superscript were not significantly different ($P > 0.05$).

Conclusions

Using mean and SD adjusted data in mixed model analyses can overcome problems of heterogeneous variance and fixed effect estimates that change with age

Effects such as SNP genotypes can achieve significance due to reduction in residual variance

Allowing fixed effects to vary with age may pick up SNP effects unseen in other analyses

Thank you for your attention