

What is nonlinear trajectory modelling with splines?

Ahmed Elhakeem, University of Bristol

<https://github.com/aelhak/NCRM2023>



Analysis of repeated measures / trajectories

Investigate long-term development patterns

Enables characterization of key features important for understanding development dynamics e.g., timing of peak growth

Improve understanding of the influences on health trajectories across life

Identify adverse changes in trajectories and when changes might occur

Inform not only the need, but also timing, for monitoring or intervention efforts

Linear mixed effects (LME) model

Also known as random-effects, multilevel, or hierarchical models

Estimates a population-average trajectory (line) as 'fixed effects' as well as the variation of individual trajectories around this average as 'random effects'

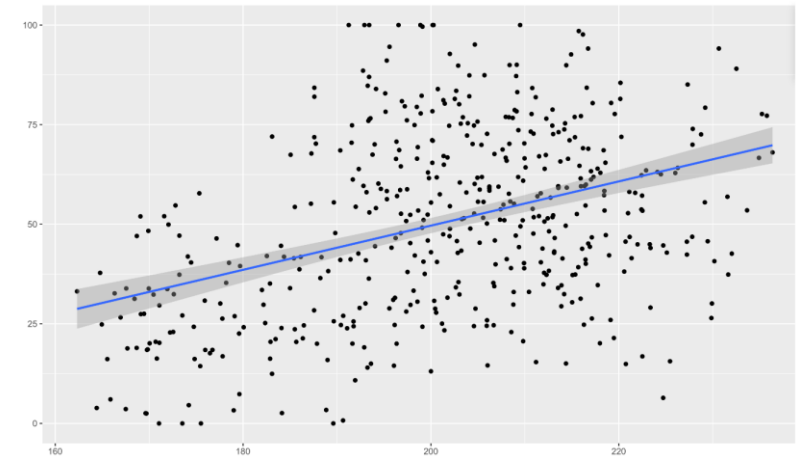
Repeated outcome is evaluated by a linear combination of the fixed and random effects

Assumes *linear change* in the outcome with increasing time

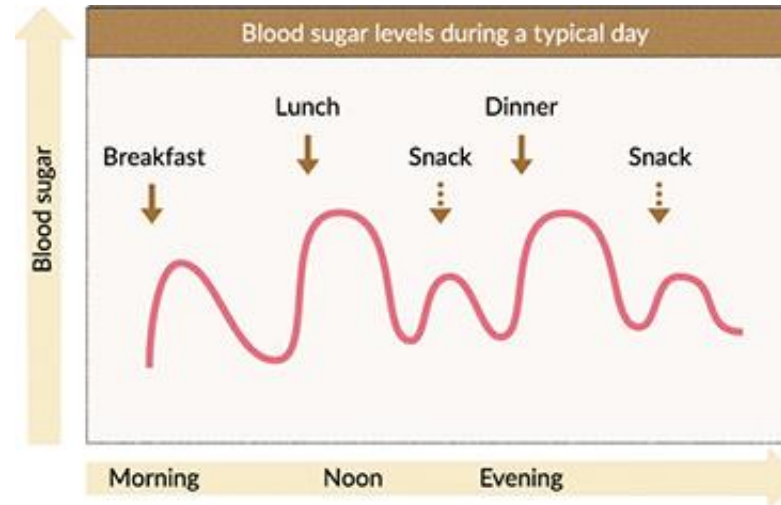
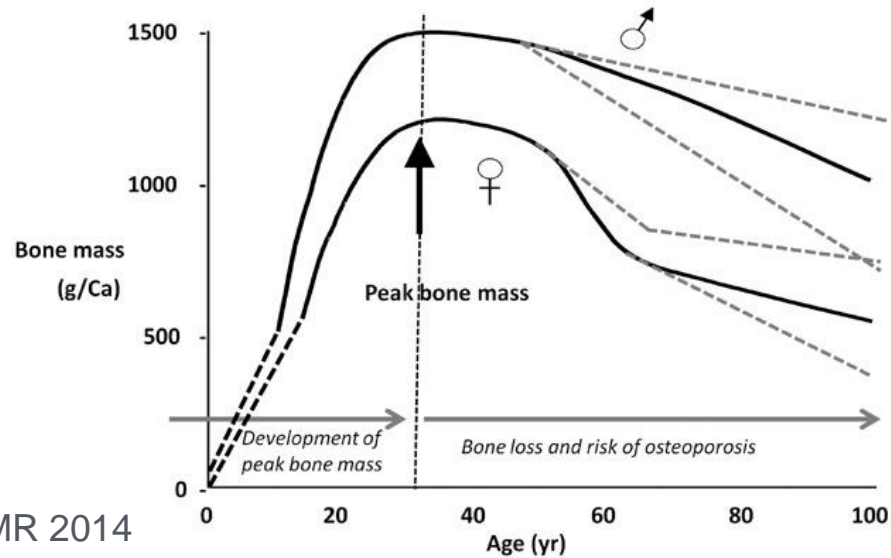
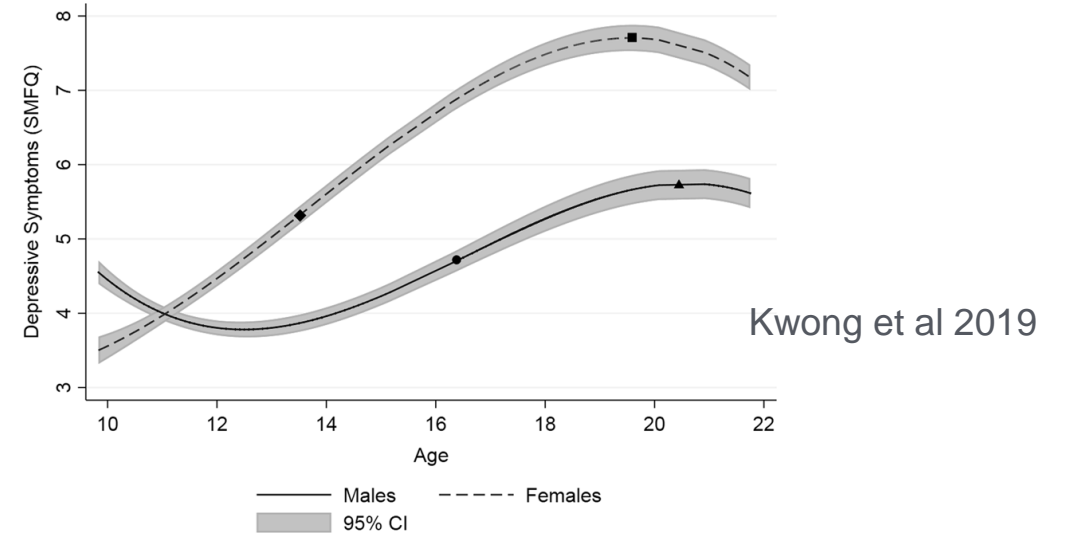
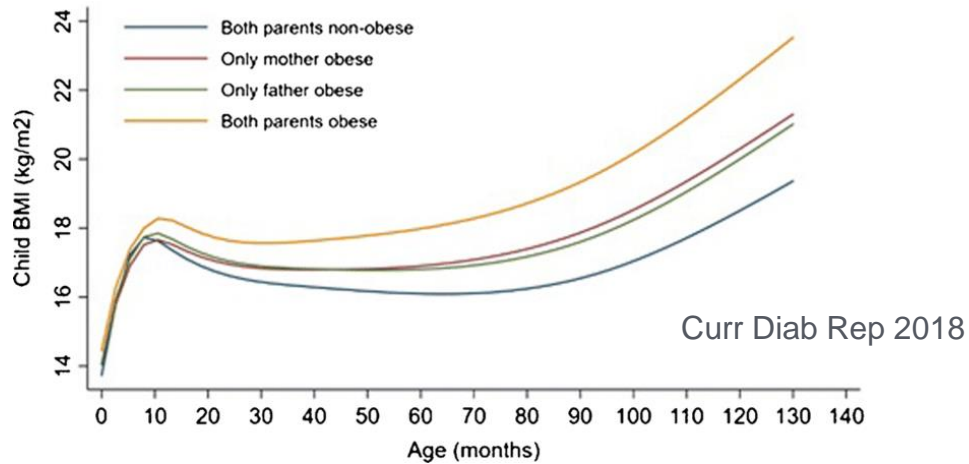
$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2), \quad i.i.d.$$

$$\beta_{0i} = \beta_0 + u_{0i}, \quad u_{0i} \sim N(0, \sigma_0^2), \quad i.i.d.$$

$$\beta_{1i} = \beta_1 + u_{1i}, \quad u_{1i} \sim N(0, \sigma_1^2), \quad i.i.d.$$



In many cases / applications change is not linear



<https://www.otsuka.co.jp/en/health-and-illness/glycemic-index/glucose-level/>

Beyond linear change

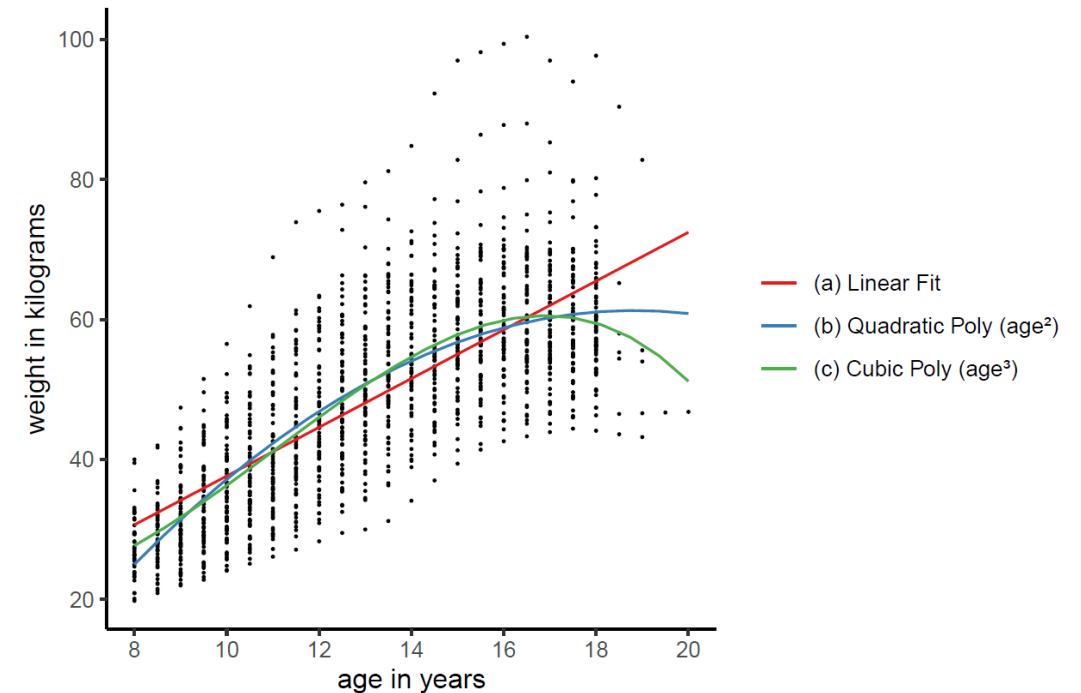
Nonlinear age changes can be incorporated into LME models by including linear combinations of nonlinear terms for age in the model

Standard approach has been to use conventional polynomial functions to produce nonlinear curves

Simpler polynomials give few curve shapes and more complex polynomials can fit badly at the ends and produce artefactual turns in the curve

A more flexible alternative to modelling complex patterns of change is to use spline functions

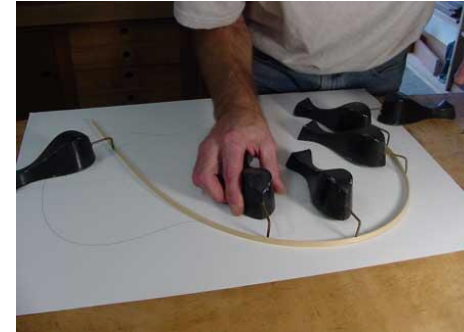
Limitations of using conventional polynomials to approximate a nonlinear growth trajectory



Splines

Physical crafting tools: flexible thin strip of wood / metal used to draft smooth curves e.g., for boats

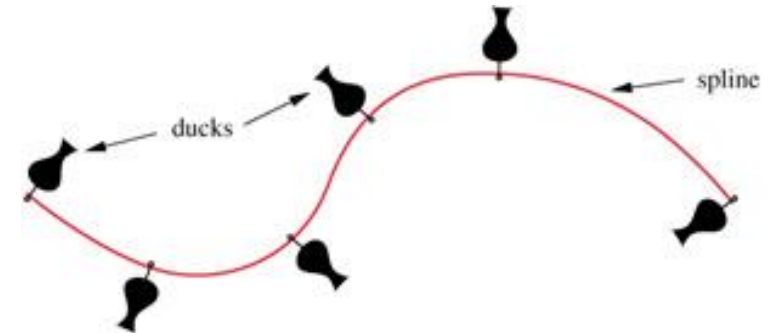
Strips would bend according to number and position of weights & take the shape



<http://www.alatown.com/spline-history-architecture/>

In statistics, spline functions can be used to mathematically reproduce flexible shapes

A spline function is a set of piecewise polynomials that are smoothly joined (i.e., constrained to be continuous) at turning points (called knots)



Splines

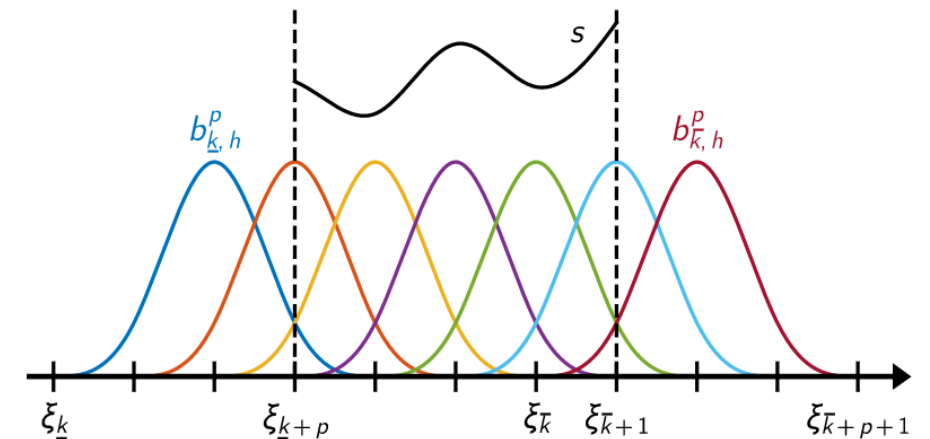
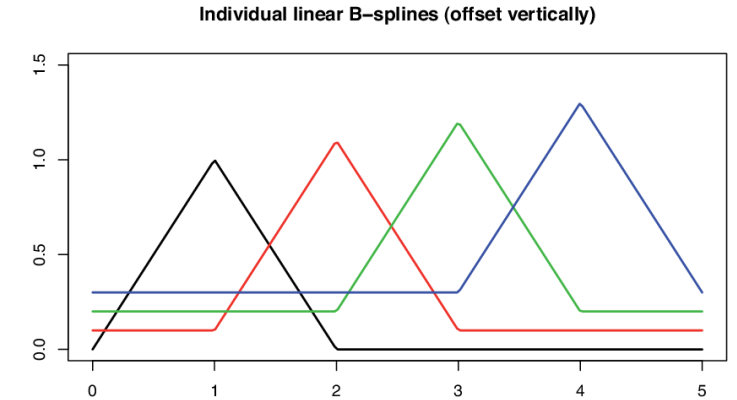
Splines are formed using spline basis functions (the dimensional space containing each element from a set of local polynomials)

- Created (usually) using basis splines (B-splines) in R: this is an alternative way for expressing splines that is more convenient than piecewise polynomials, and is mathematically equivalent

Polynomial degree and method of knot placement is generally what distinguishes between different spline functions, e.g.,

- degree 1 for linear spline
- degree 3 for cubic spline

Regression spline: functions that use fewer knots than observations – number and location of knots must be specified by the user



Uniform cubic B-splines forming a cubic spline.

<https://bsplines.org/>

Linear spline

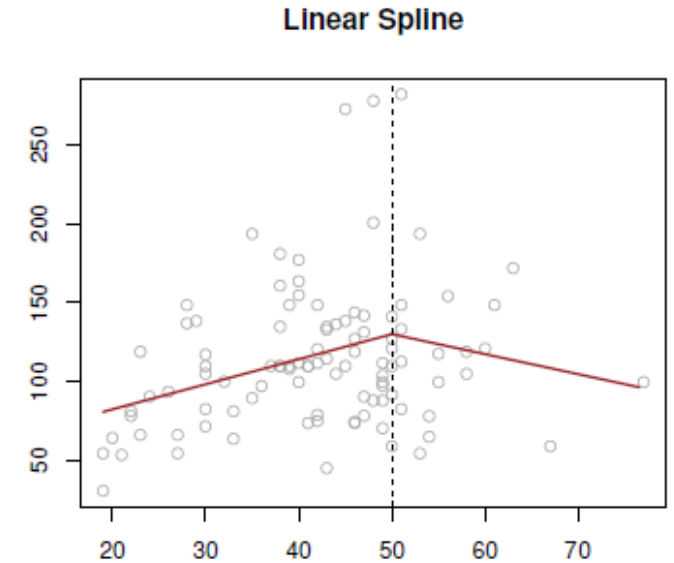
Connected straight lines with a change in slope at each knot

Can be used in LME to describe growth as a series of connected lines (linear slopes) joined at knot points

Linear spline function for age (20–70 years) with one knot point at age 50 years produces 2 different linear slopes of the repeated measure (e.g., weight): 20 to ≤ 50 , and 50 to ≤ 70 years

Linear growth rate is described over age windows (1st derivative is not continuous at the knots)

There can be an abrupt change from positive to negative growth – curve shape is unrealistic



$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \sum_{k=1}^K b_{ki}(t_{ij} - \xi_k)_+ + \varepsilon_{ij}$$

$$(t_{ij} - \xi_k)_+ = \begin{cases} 0 & t_{ij} < \xi_k \\ (t_{ij} - \xi_k) & t_{ij} \geq \xi_k \end{cases}$$

Cubic and restricted cubic (natural) spline

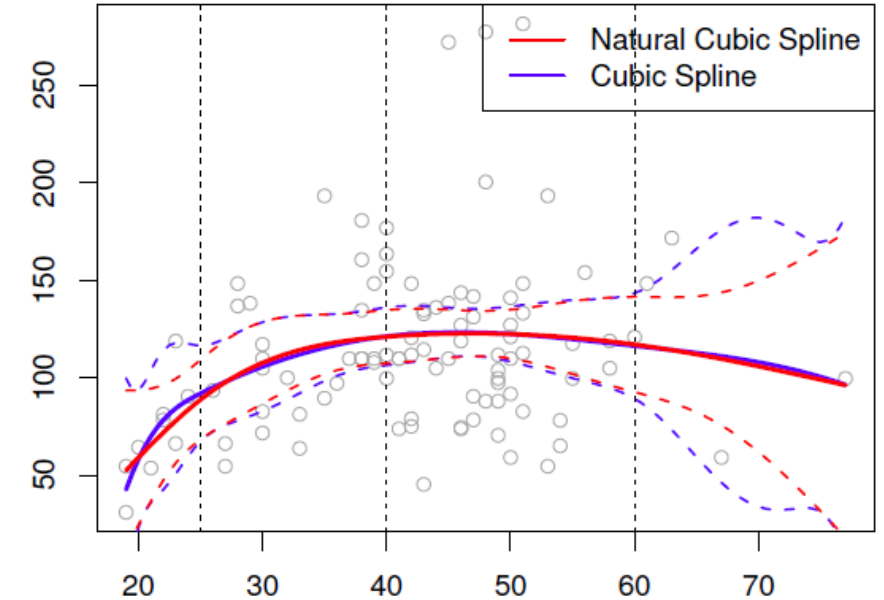
A cubic spline is set of connected cubic polynomials with slope and continuity constraints at each knot – continuity constraints -> more parsimonious for complex shapes than linear spline with many knots

Cubic splines have continuous 1st (velocity) and 2nd (acceleration) derivatives) -> possible to differentiate the growth curve to get a velocity curve, and identify the nature of the peak

A natural spline is a cubic spline with additional constraints of linearity before (usually) the first knot and after the last knot

Linearity constraint allows trajectory to be less erratic at the tails: more reliable than linear splines and (unrestricted) cubic splines

Natural spline models growth by cubic polynomials between internal knots, and with a line before the 1st knot and after the last knot



$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \sum_{k=1}^{K-2} b_{ki} (t_{ij} - \xi_k)_+^3 + \varepsilon_{ij}$$

$$(t_{ij} - \xi_k)_+^3 = (t_{ij} - \xi_k)_+^3 - (t_{ij} - \xi_{K-1})_+^3 \frac{\xi_K - \xi_k}{\xi_K - \xi_{K-1}} + (t_{ij} - \xi_K)_+^3 \frac{\xi_{K-1} - \xi_k}{\xi_K - \xi_{K-1}}, k = 1, 2, \dots, K-2$$

Choosing number and location of knots

Flexibility/smoothness of regression splines is determined by the number and position of knots

A small number of knots (~ 3 to 5 knots) provides a good fit to most patterns however, with many repeat measurements more knots may be required

For natural splines, knot position less important than number (both important for linear splines)

Common approaches to knot placement:

- quantile placement: knots at equal centiles (of age distribution) or at Harrell's centiles
- equidistant placement: equally spaced knots over the range of age
- fit a smoothing curve & select knot position based on this (e.g., `geom_smooth()` / FPs)
- placing the knots at the mean age of data collection / where data are available
- subject knowledge: e.g., to test sensitive periods

Model comparison / selection

Model comparison and selection can be done by a combination of informal checking and formally testing difference between fitted models

Models with different knots can be compared informally e.g., by inspecting plots from competing models and avoiding those showing unrealistic trajectory shapes, and by examining residuals for the fitted models

Valid comparison between nested models with different knot points can be done using standard likelihood ratio tests under either ML or REML estimation

Models with non-nested mean structures can be compared informally; valid comparison using information criteria (e.g, BIC) can be done provided ML estimation is used

- A model with $k-1$ evenly spaced knots will not generally be nested within a model with k evenly spaced knots, same for models using evenly-spaced quantile placement of knots

Problems with regression splines

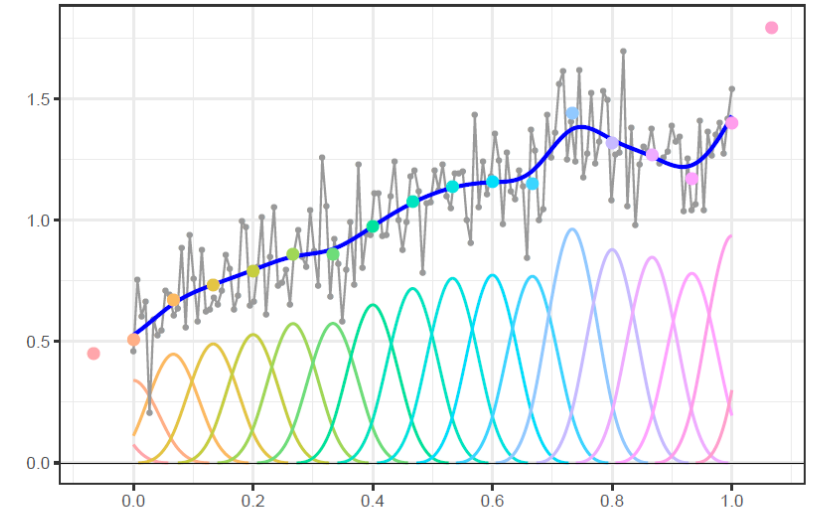
- Decision on how many knots to use and where to place these knots – must be made by the user
- Choices regarding number of knots (and where they are located) can have a large effect on the fit
 - with too few B-splines, the resulting spline curve is too sensitive to exact knot locations, and not flexible enough
 - Using many B-splines removes these problems, but estimated spline curve overfits data and has too much uncertainty
- No general consensus on model comparison and selection

Penalized regression splines

- Regression splines that include an added penalty on the basis coefficients to control the smoothness of the spline function
- Can be used to model nonlinear change in a repeated outcome within a generalised additive (mixed) model framework – this is a generalized linear (mixed) model with a linear predictor involving a sum of smooth function(s) of one or more covariates (**mgcv** package in R)
- Estimation is carried out by penalized regression methods (penalized likelihood maximisation) and the appropriate degree of smoothness for the spline function is estimated from the data
- User does not need to specify the number of knots, instead, input amounts to choosing a maximum spline basis dimension / complexity (and method for estimating the smoothing parameter) – smoothing parameter penalizes the basis coefficients to avoid overfitting

P-splines

- A type of penalized regression spline
- P-splines combine evenly spaced B-splines with a discrete difference penalty applied directly on their coefficients to control the wiggleness of the spline function
- Easy to set up and use
- Flexible – any order of B-spline basis can be combined with any order of penalty e.g., basis order 2 gives cubic P-splines and penalty order 3 gives a penalty built on 3rd differences
- Can be especially useful for MCMC Bayesian inference
- Cubic P-splines useful for modelling biological growth



The core idea of P-splines: a sum of B-spline basis functions, with gradually changing heights.

JOPS book

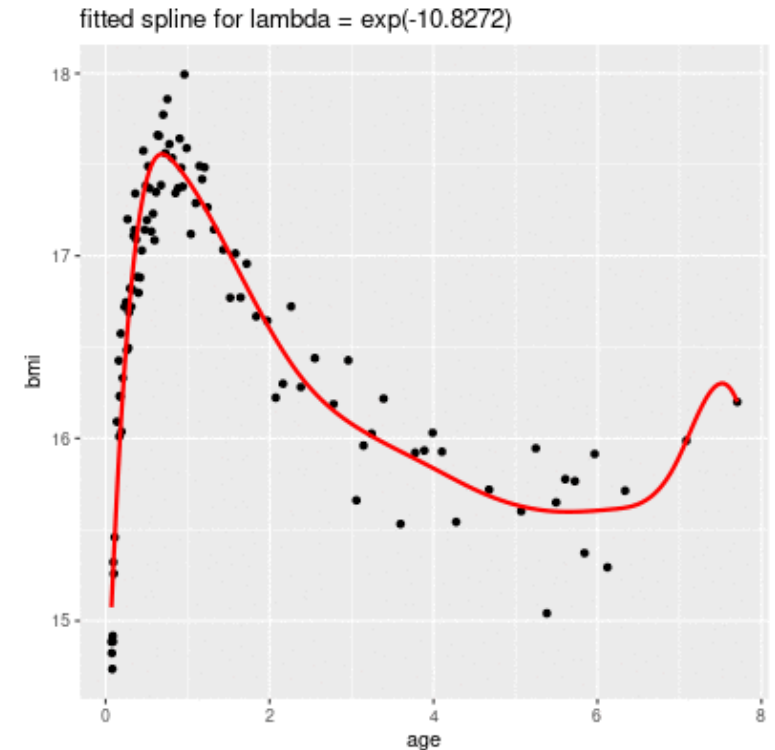
Choosing basis dimension (k)

- The modeller needs to decide roughly how large a basis dimension is fairly certain to provide just adequate flexibility and use that
 - Keep k a little larger than it is believed could be reasonably necessary
- Provided k is large enough that the basis is slightly more flexible than we expect to need (i.e., do not make basis dimension too restrictively small), then the exact choice of k (or location of knots) should have only a small effect on the fitted model
 - caveat: making k too large often changes the flexibility of the fitted curve – i.e., model fit tends to retain some sensitivity to choice of k
- The problem of estimating the smoothness of the curve (effective degrees of freedom) is now a problem of estimating the smoothing parameter
- Informal checking that the chosen basis dimension is appropriate is easy (if needed) e.g., using `gam.check()` for residual randomization test, or smoothing residual with increased k value

Smoothing parameter (λ)

- Controls trade-off between smoothness of the spline function and fidelity to the data to avoid overfitting (by constraining the B-spline coefficients so that they are less variable)
 - $\lambda = 0 \rightarrow$ unpenalized regression spline
 - $\lambda \rightarrow \infty$ straight line estimate
- P-splines penalize (some particular order of) differences between adjacent basis
- Computationally challenging to estimate – can be estimated by REML (recommended), generalised cross-validation or other approaches
- Full model estimated by maximizing penalized log-likelihood. Log-likelihood is a measure of fit of the model, while the penalty helps avoid fitting overly complex smooths

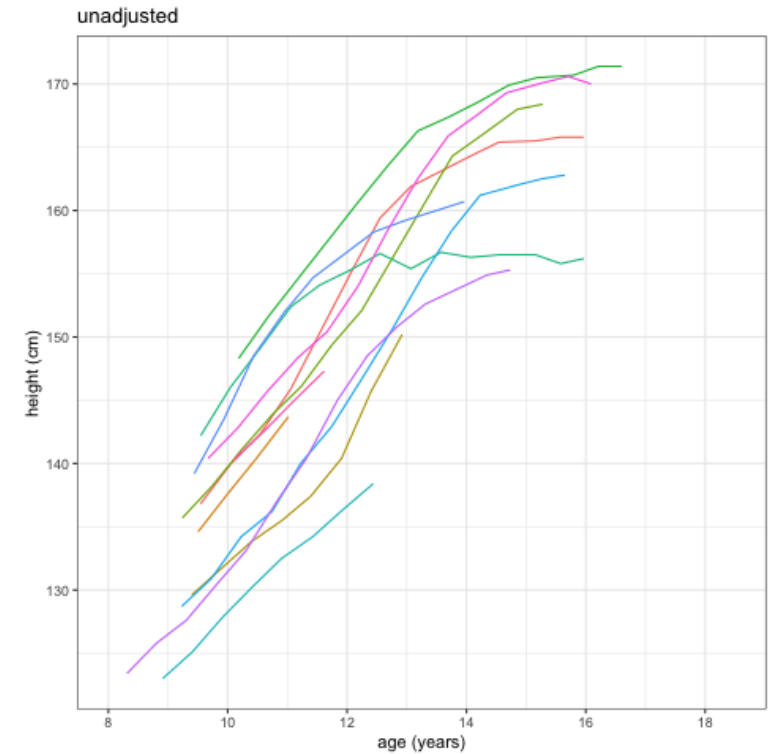
animation of P-splines fit for varying λ



<https://github.com/ZheyuanLi/P-splines-workshop>

SITAR

- So far, we have considered linear mixed effects models that allow terms (spline functions) to describe a nonlinear trajectory with age
- SITAR (Super Imposition by Translation and Rotation) is a shape invariant nonlinear mixed effects model
- A shape invariant model fits just one curve, and then tailors it to match each individual (subject-specific) curve
- A nonlinear model is nonlinear in the coefficients
- Parameters of nonlinear models reflect the model's specific purpose – in the case of SITAR, it was designed to model pubertal growth (in height)
- SITAR simplifies estimation of subject-specific timing and intensity of peak pubertal growth



SITAR

- Population curve fitted as natural cubic spline of y on age – so the data define curve shape
- Individual curves are all versions of the mean curve
- Mean curve is shifted and scaled by 3 random effects (size, timing and intensity) to fit any individual curve
- SITAR models growth on both the x and y axes – allows differences in developmental age to be modelled
- Assumes growth velocity is the same for all individuals post-puberty

$$y_{ij} = \alpha_0 + \alpha_i + h\left(\frac{t - \beta_0 - \beta_i}{\exp(-\gamma_0 - \gamma_i)}\right) + \varepsilon_{ij}$$

y_{ij} = the outcome measurement for individual i at age j

$h(t)$ = natural cubic spline curve

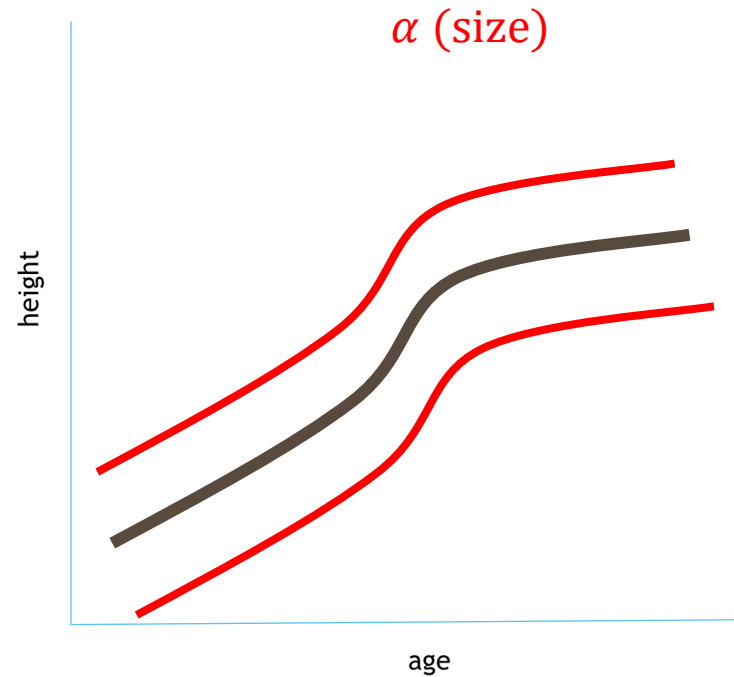
$\alpha_0, \beta_0, \gamma_0$ = fixed effects

$\alpha_i, \beta_i, \gamma_i$ = random effects for i th individual – describe size (α_i), timing (β_i), intensity (γ_i) of individual i 's growth relative to the mean growth curve

ε_{ij} = independent normally distributed errors

SITAR random effects

α_i adjusts for the differences in y and geometrically reflects individual shifts up or down (translation) in the mean curve



Subject 1 taller than peers

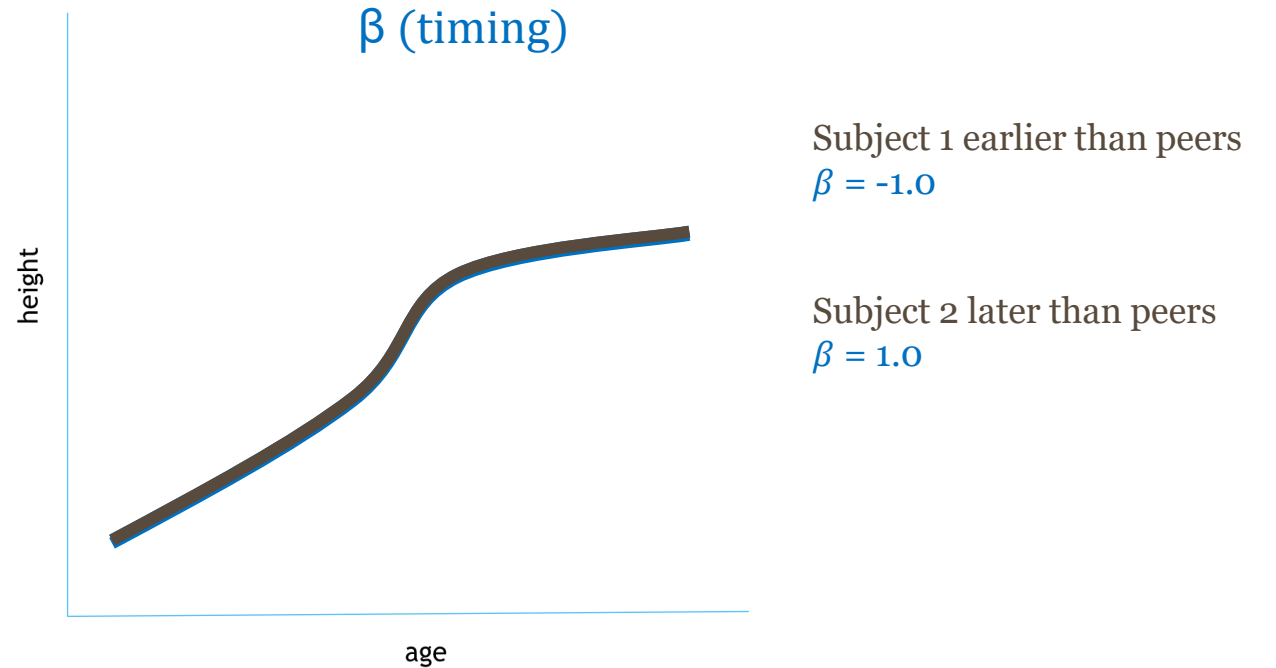
$\alpha = 1.0$

Subject 2 shorter than peers

$\alpha = -1.0$

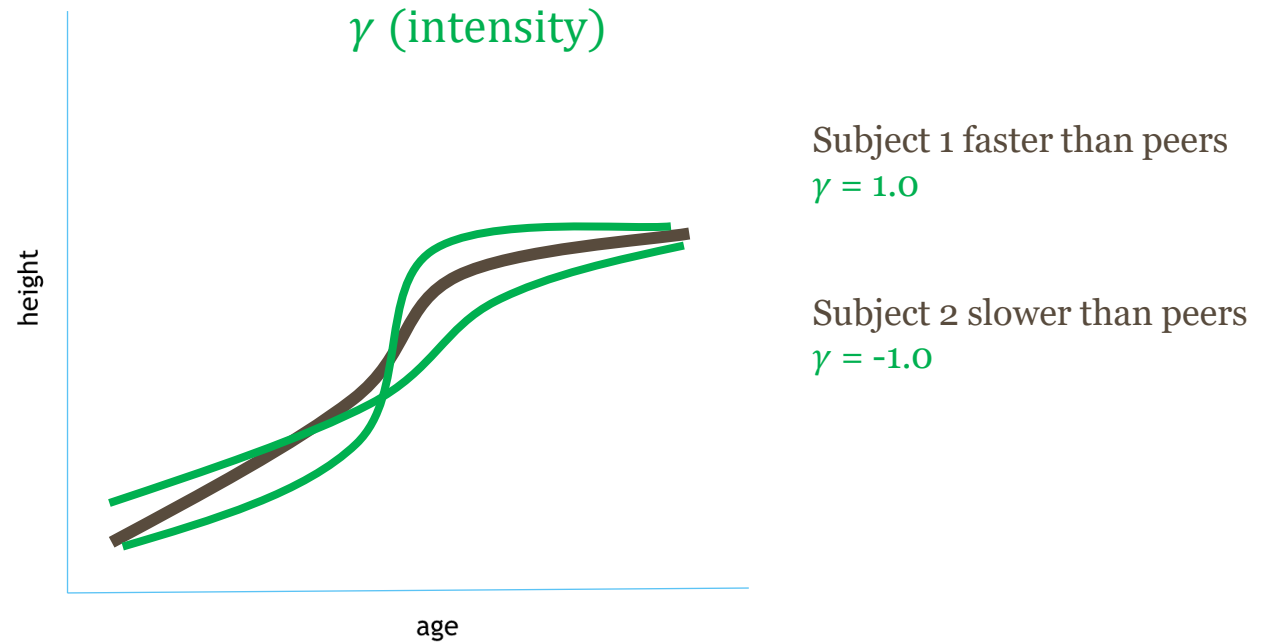
SITAR random effects

β_i adjusts for differences in the timing of peak growth in y and geometrically reflects left to right shifts (translation) in the mean curve



SITAR random effects

γ_i adjusts for the duration of the growth spurt and geometrically corresponds to shrinking or stretching of the age scale and rotating the curve



R Code

Linear splines

<https://www.rdocumentation.org/packages/lspine/versions/1.0-0/topics/lspine>

Set up using the `bs()` function from the `splines` package

For convenient parametrisation of spline coefficients, use `lspine` package which is a wrapper for `bs()`

Coefficients are expressed as the linear slopes of consecutive spline segments (default), or as the slope change at consecutive knots

To use in a linear mixed effects model, replace the age variable with a function for age (same for both `lmer()` from the `lme4` package and `lme()` from `nlme` package)

`lspine` contains three functions:

`lspine(x, knots)` # to manually specify knots

`qlspline(x, q)` # quantile knot placement (at equal quantiles (input a number) or specified quantiles (input a vector of locations) of x

`elspline(x, n)` # equidistant placement – equal intervals of x

Natural splines

<https://www.rdocumentation.org/packages/splines/versions/3.6.2/topics/ns>

Set up using the `ns` function from the **splines** package

To use in a linear mixed effects model, replace the age variable with a function for age (same for both `lmer()` from the **lme4** package and `lme()` from **nlme** package)

```
ns(  
  
  x, # predictor column name (age)  
  
  df, # number for degrees of freedom, used for  
      quantile knot placement (at equal quantiles).  
      The number of knots is df-1  
  
  knots, # can be used instead of df for manual  
          knot placement by supplying a vector for the  
          location of knots  
  
)
```


P-splines

<https://www.rdocumentation.org/packages/mgcv/versions/1.8-42/topics/gam>
<https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/s.html>
<https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/smooth.terms.html>

Can be fitted using the `gam()` function from the `mgcv` package;

```
gam(  
  formula, # model of the form y ~ s() (other smooth options include te, ti, t2)  
  data, # data frame / list containing response variable and covariates  
  method # smoothing parameter estimation method (option include REML, ML, and GCV)  
)
```

```
s(  
  ..., # list of variables to transform  
  k, # basis dimension (default=10 for 1 variable)  
  bs, # smoothing spline to use (ps for P-spline)  
  m, # order of basis/penalty (e.g., c(2, 2)) for cubic P-spline with 2nd order penalty)  
  by, # numeric/factor variable to multiply/replicate the smooth  
  xt, # extra information required to set up a particular basis  
  data # name of dataset containing x and y variables  
)
```

P-splines

<https://www.rdocumentation.org/packages/mgcv/versions/1.8-42/topics/gam>
<https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/s.html>
<https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/smooth.terms.html>

Random intercept – random (linear) slope
P-spline model in **gam()**

```
gam(  
  y ~ s(age, bs = 'ps') +  
  s(id, by = "re") +  
  s(age, id, bs = "re"),  
  data = dat  
)
```

gam() ok for simple random effects

For more complex random effects (i.e., random P-splines) or large number of random effects, use **gamm()** function from the **mgcv** (or use the **gamm4** package)

gamm() is more efficient with random effects: it uses **gam()** to specify the smooth function, converts the smooths to a random effects parameterization, and calls **lme()** from **nlme** to fit the re-parameterized model

However, it still struggles with large datasets – ~1000 individuals

gamm() syntax:

```
gamm(  
  formula, # gam() formula  
  random, # optional random effects  
  Data # dataset  
)
```

Random P-spline model in **gamm()**

```
gamm(  
  y ~ s(age, bs = 'ps') +  
  s(age, id, bs = "fs", xt = "ps"),  
  data = dat  
)
```

SITAR

<https://rdr.io/cran/sitar/man/sitar.html>

- Using the **sitar** package in R (based on **nlme** package): data in long format
- Common practice is to fit SITAR models with varying number of knots placed at equal quantiles of the age distribution for the spline curve
- Log transforming x and/or y often leads to better fit
- Best fitting SITAR model selected based on smallest BIC (or AIC) and inspecting the fitted population curves, and residuals
- Customizable similar to **nlme** (e.g., within-subject correlations)

```
sitar(  
  x, # vector of ages.  
  y, # vector of measurements  
  id, # factor of subject identifiers  
  data, # data frame containing variables x, y and id  
  df, # degrees of freedom for restricted cubic regression spline  
  knots, # vector of values for knots  
  fixed = NULL, # a, b, c fixed effects (default to those in random)  
  random = "a + b + c", # a, b, c random effects  
  pdDiag = FALSE, # unstructured variance-covariance matrix  
  correlation = NULL, # within-subject correlation structure  
  weights = NULL, # object or one-sided formula describing the  
    within-subject heteroscedasticity structure  
  method = "ML", # estimation method  
  control = nlmeControl(msMaxIter = 100, returnObject = TRUE),  
)
```

References / further reading

- Perperoglou, A., Sauerbrei, W., Abrahamowicz, M. et al. A review of spline function procedures in R. BMC Med Res Methodol 19, 46 (2019).
- Elhakeem A, Hughes RA, Tilling K, et al. Using linear and natural cubic splines, SITAR, and latent trajectory models to characterise nonlinear longitudinal growth trajectories in cohort studies. BMC Med Res Methodol. 2022;22(1):68.
- Wood SN. Generalized additive models an introduction with R. 2nd Edition 2017.
- Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ. 2019;7:e6876.
- Eilers, P., & Marx, B. (2021). Practical Smoothing: The Joys of P-splines. Cambridge: Cambridge University Press. doi:10.1017/9781108610247
- Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. Int J Epidemiol. 2010;39(6):1558-1566.
- Analysis of Repeated Measures Short Course. Bristol Medical School.
<https://www.bristol.ac.uk/medical-school/study/short-courses/courses/analysis-repeated-measures/>

**Please remember to complete
your feedback form**

