Statistics as a weapon for improving women's health

Kathrine Frey Frøslie, biostatistician, PhD

Oslo Centre for Biostatistics and Epidemiology

Norwegian Resource Centre for Women's Health

UiO : University of Oslo

Oslo University Hospital

Overview

Background on women's health:
- **Obesity**
- **Blood glucose**
- **Pregnancy**

The weapon: Functional data analysis

Results & health summary

Not convinced yet? FDA can also handle this: _____

Take-home message

Obesity

BMI < 18.5	BMI 18.5 – 25	BMI 25 – 30	BMI > 30
Underweight	Normal weight	Overweight	Obese

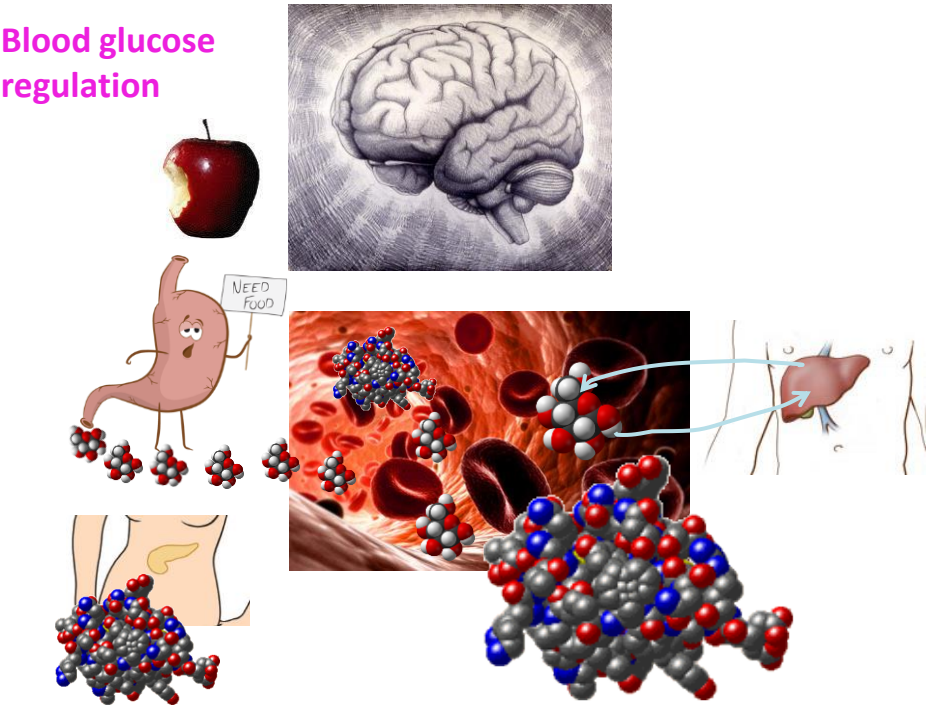
Systemic low-grade inflammation

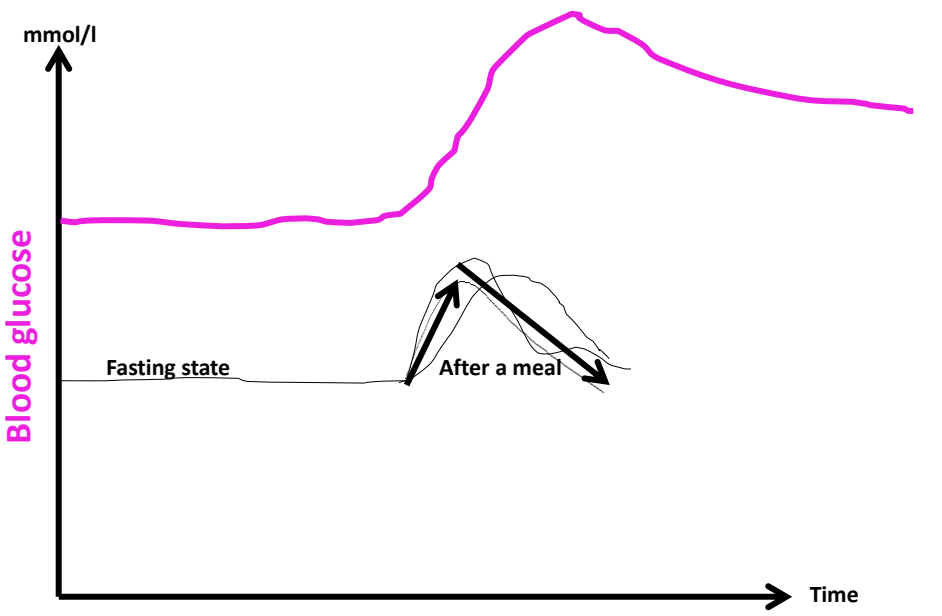
Insulin resistance

Obesity in pregnancy

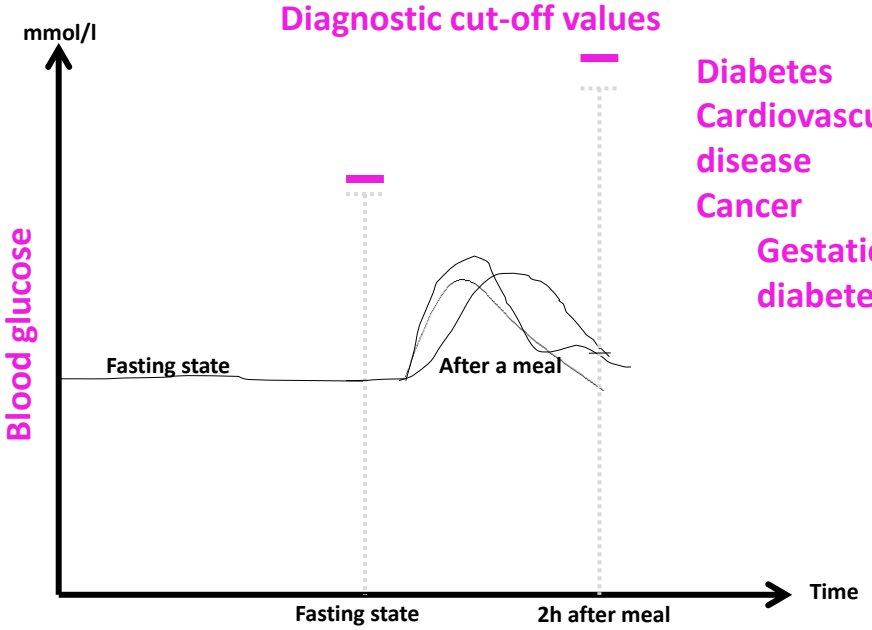
	Short-term risk	Long-term risk
Mother	Big baby Complications	
Child	Diabetes Cardiovascular diseases Obesity	

Blood glucose regulation



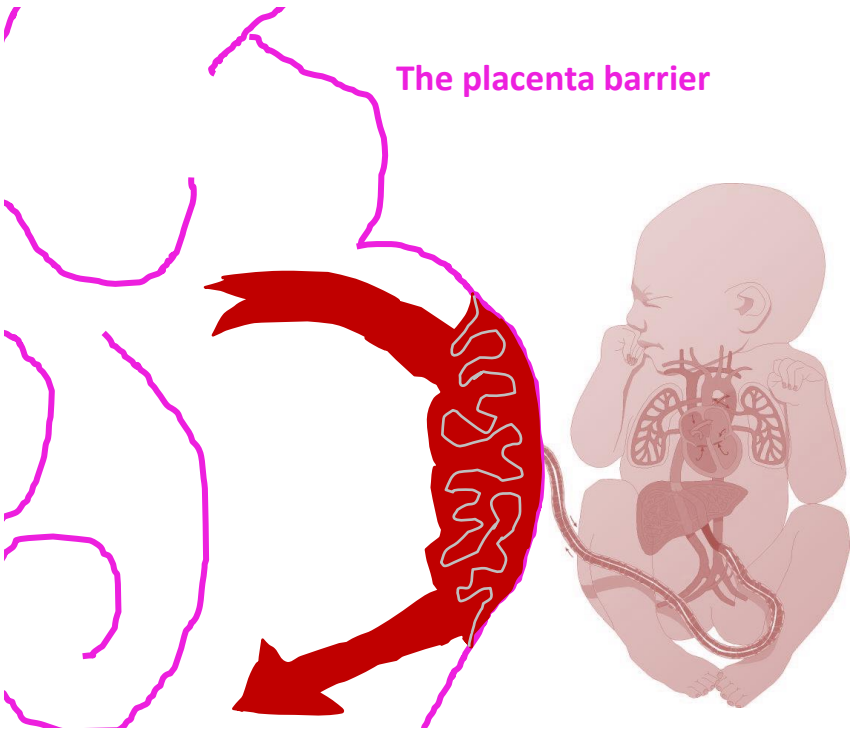


Diagnostic cut-off values

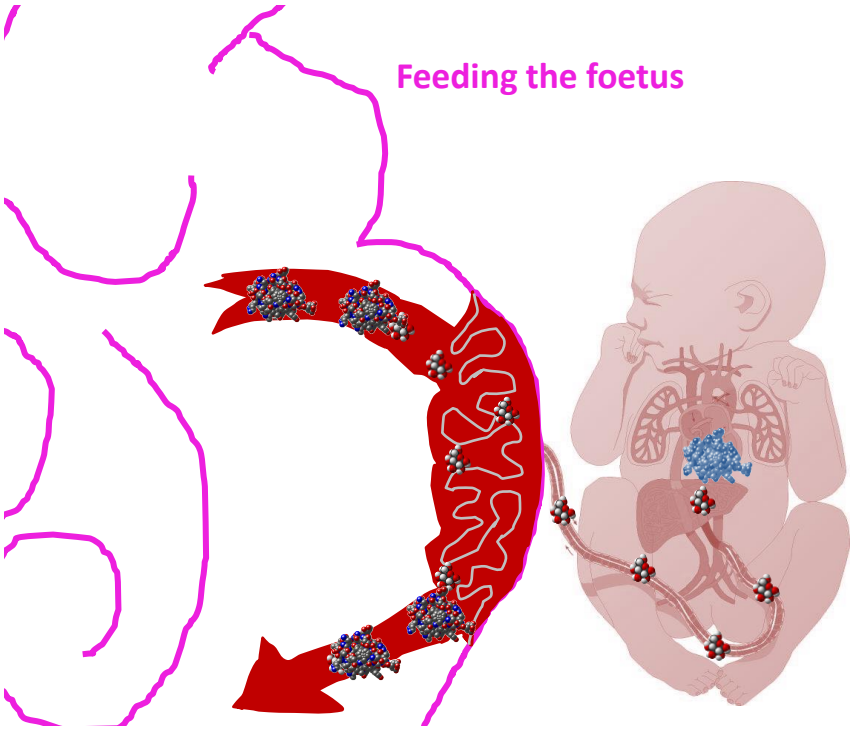


Diabetes
Cardiovascular disease
Cancer
Gestational diabetes

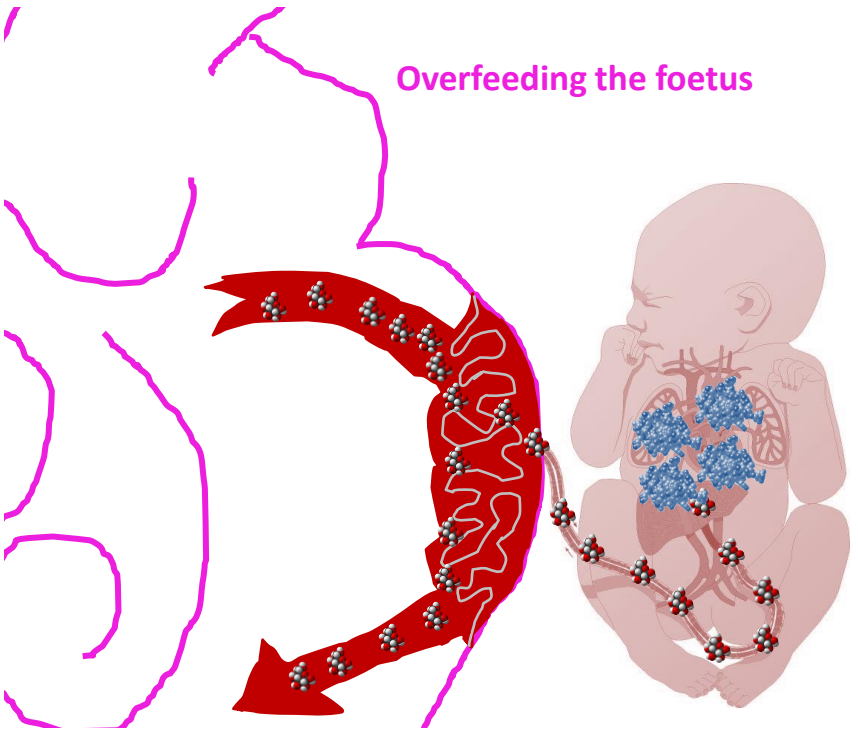
The placenta barrier

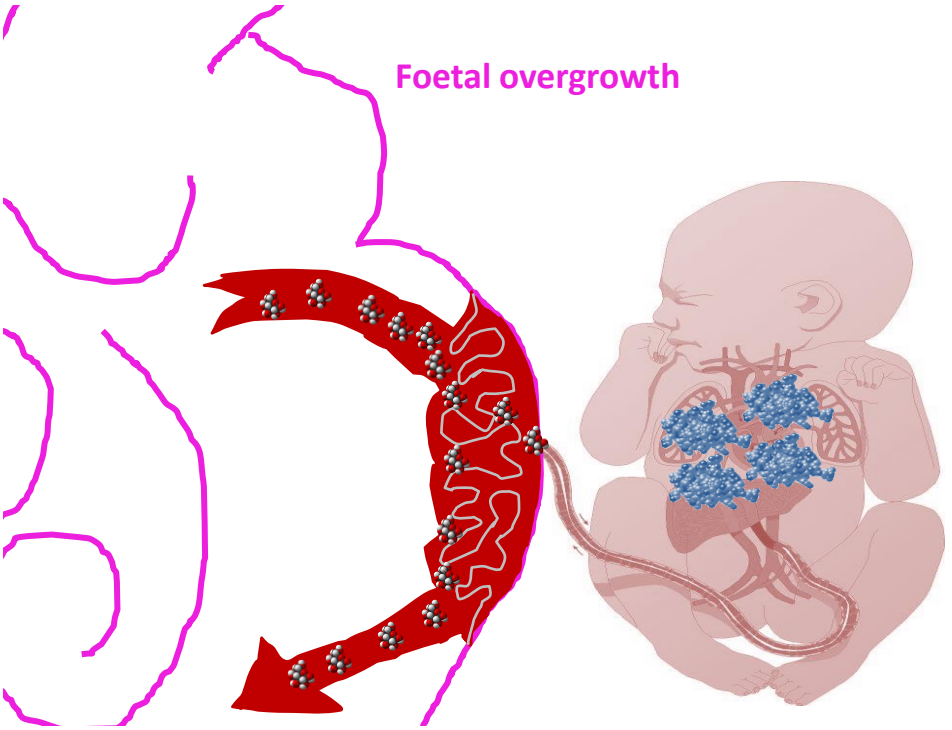


Feeding the foetus



Overfeeding the foetus





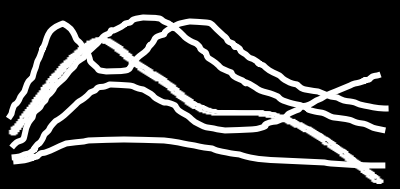
«High» blood glucose in pregnancy

	Short-term risk	Long-term risk
Mother	Big baby Complications	
Child	Diabetes Cardiovascular diseases	

The weapon: Functional data analysis (FDA)

- ✓ **Functional data: Fitting curves**
- ✓ **Characterising curves:**
  Mean and principal component analysis
- ✓ **Placing FDA in the landscape of statistical analyses**
- ✓ **Women's health: The STORK study**
- ✓ **FDA can also handle this:**
  Derivatives
  Alignment

Functional data: Fitting curves

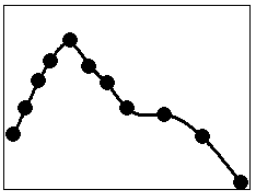


The basics

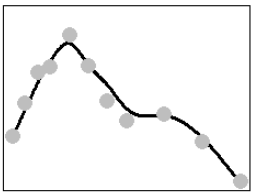
An underlying, continuous process

Time (t)

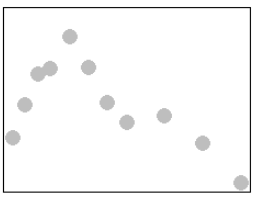
…measured at discrete time points



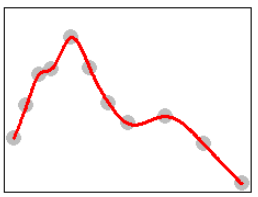
,with measurement error



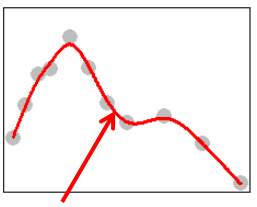
Hence, these are the measurements



…that we use to fit a curve

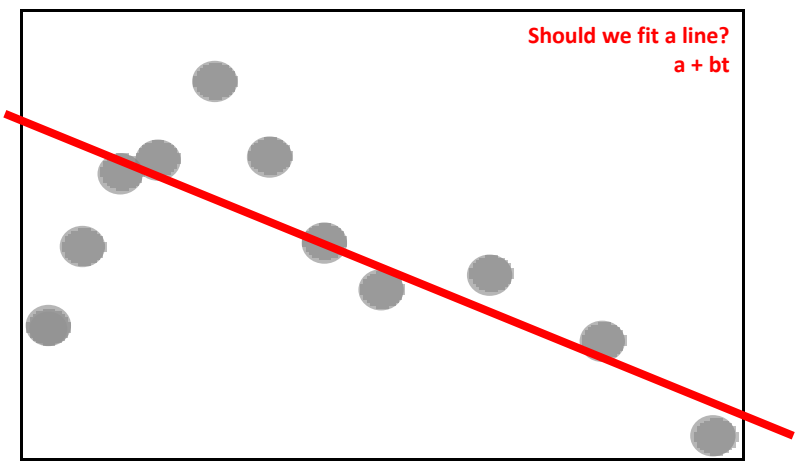


…smoothed, due to the measurement error



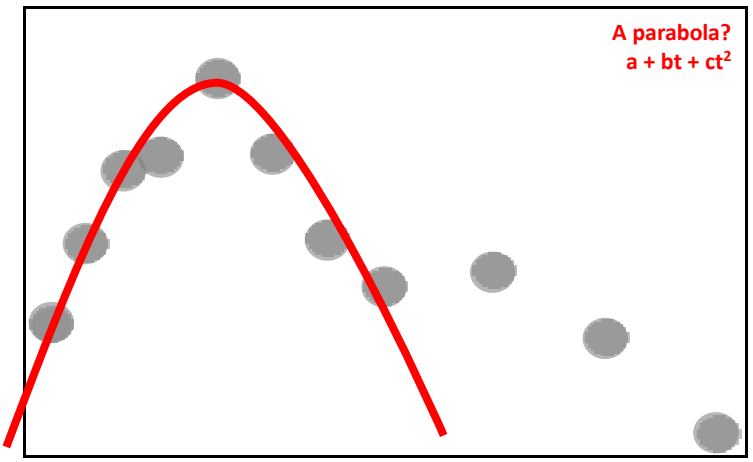
The basic unit in the analysis

How to construct a curve (the basic unit) from the data?

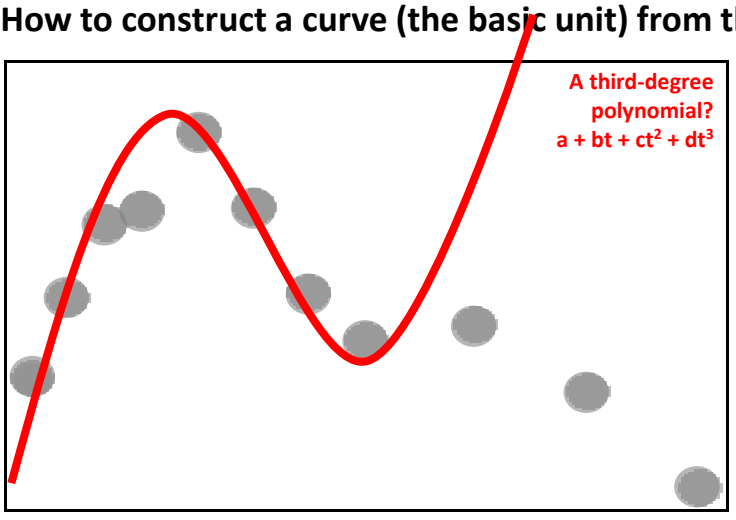


Time (t)

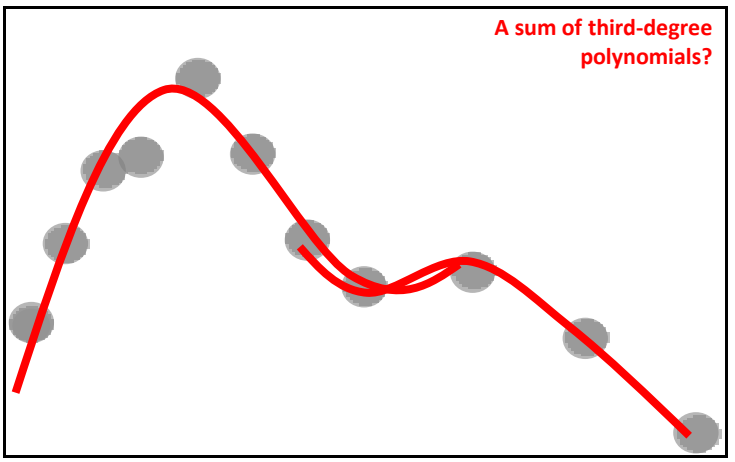
How to construct a curve (the basic unit) from the data?



How to construct a curve (the basic unit) from the data?

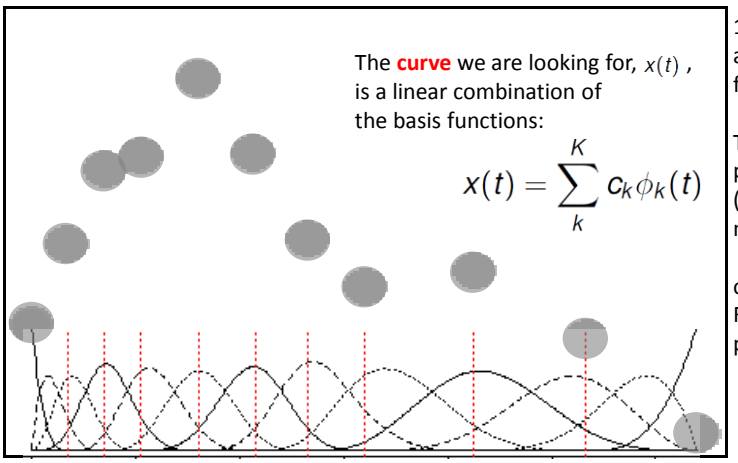


How to construct a curve (the basic unit) from the data?



Time (t)

How to construct a curve (the basic unit) from the data?



The **curve** we are looking for, $x(t)$, is a linear combination of the basis functions:

$$x(t) = \sum_{k}^{K} c_k \phi_k(t)$$

1. Choose a set of basis functions $\phi_k(t)$:

Third-degree polynomials (= B-splines) for non-periodic data

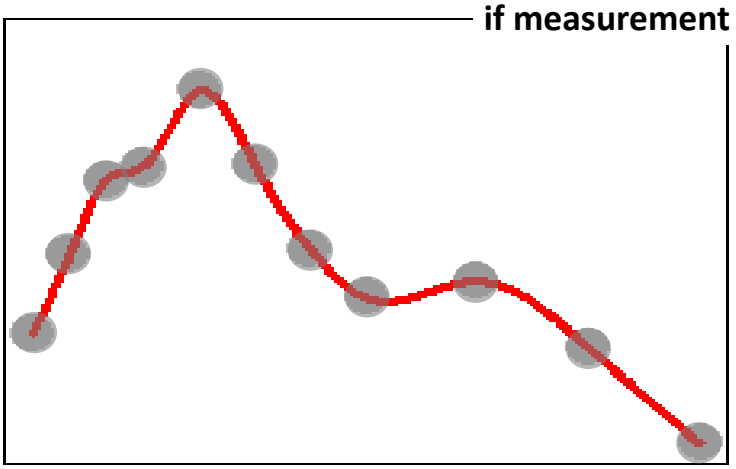
or
Fourier basis for periodic data

The best linear combination of the basis functions (the curve that fits the points best) is found by minimising the least squares expression

$$\mathrm{SSE}(x|\mathbf{y}) = [\mathbf{y} - x(\mathbf{t})]'[\mathbf{y} - x(\mathbf{t})]$$

Best curve: overfitted curve (too curvy)

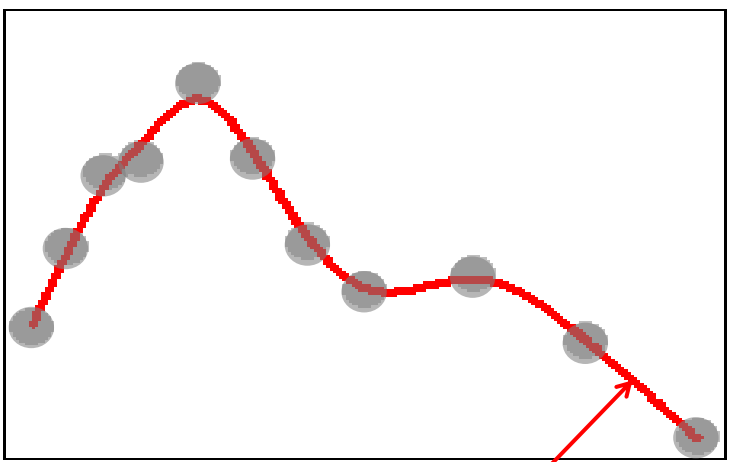
if measurement error



We want to restrict the total amount of curvature, and minimize these least squares instead:

$$\mathrm{PENSSE}_\lambda(x|\mathbf{y}) = [\mathbf{y} - x(\mathbf{t})]'[\mathbf{y} - x(\mathbf{t})] + \lambda\,\mathrm{PEN}(x)$$

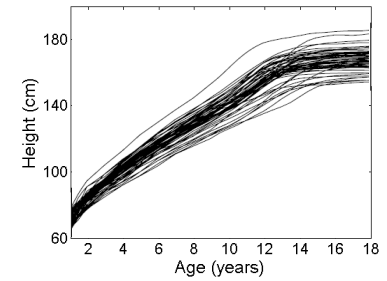
Smoothed curve (less curvy)

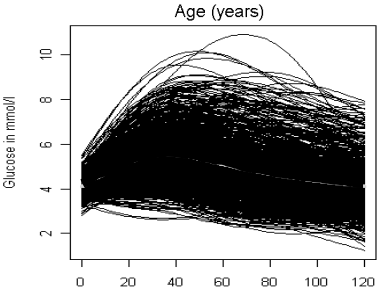


The basic unit in FDA
From now on, we only use curves
We never look back!

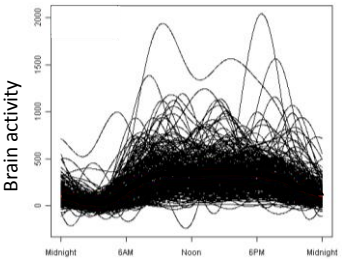
Examples of functional data

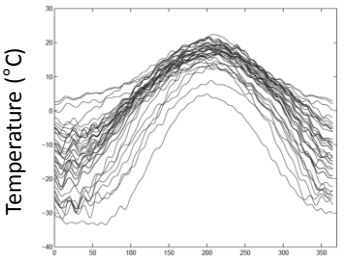
Non-periodic patterns





Periodic patterns





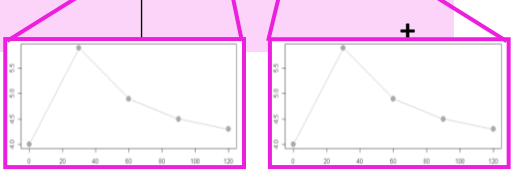
Day no. during the year

STORK@Oslo University Hospital

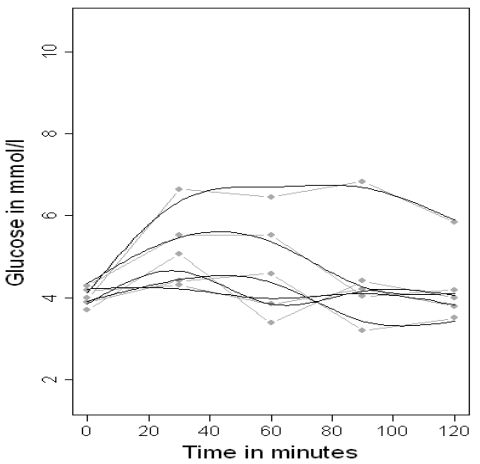
BIG babies and Complications

n=1031 healthy, pregnant women

Pregnancy week	14-16	30-32	Birth
Background/demographics	+		
Anthropometry (BMI)	+	+	+
Glucose challenge (2-hour OGTT)	+	+	
Inflammation	+	+	
Birth data			+



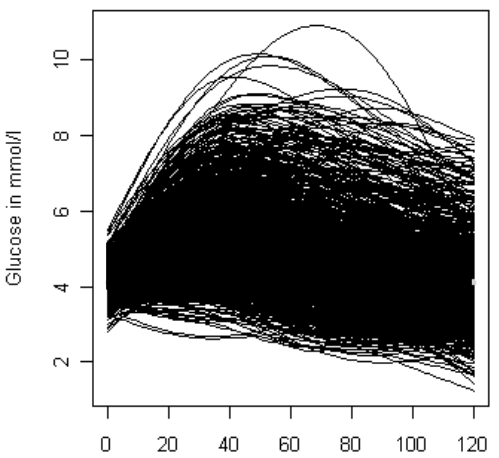
Glucose measurements in early pregancy



Curves as basic units:
Must be fitted

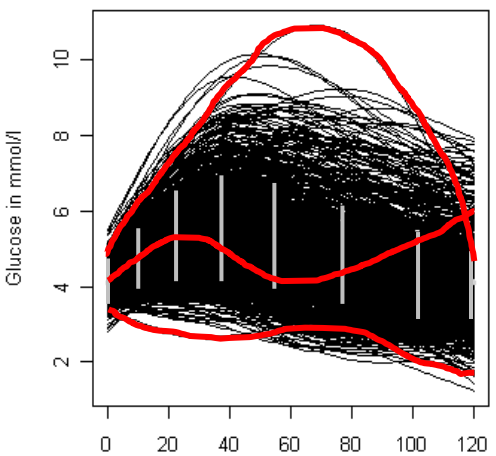
Frøslie KF, Røislien J, Qvigstad E, Godang K, Voldner N, Bollerslev J, Henriksen T, Veierød MB.
Shape information from glucose curves: Functional data analysis compared with traditional summary measures.
BMC Medical Research Methodology 2013;13:6

Glucose curves in early pregancy



Characterising curves

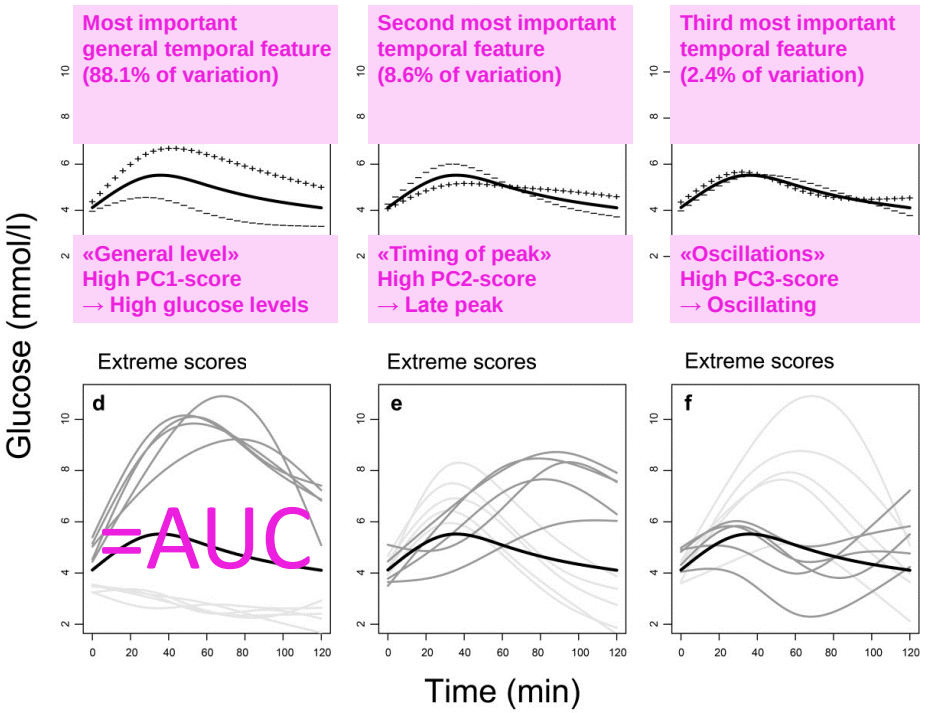
Interesting features of curves

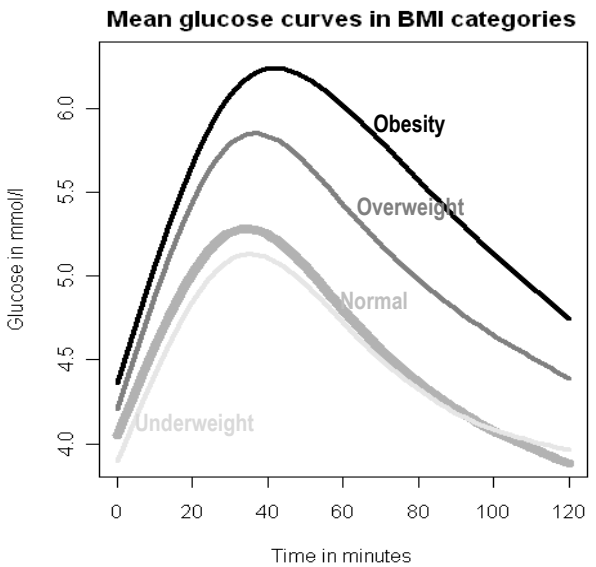


Mean curve: OK.

Variation: At single timepoints?
Even more important: temporal variation

Functional principal component analysis

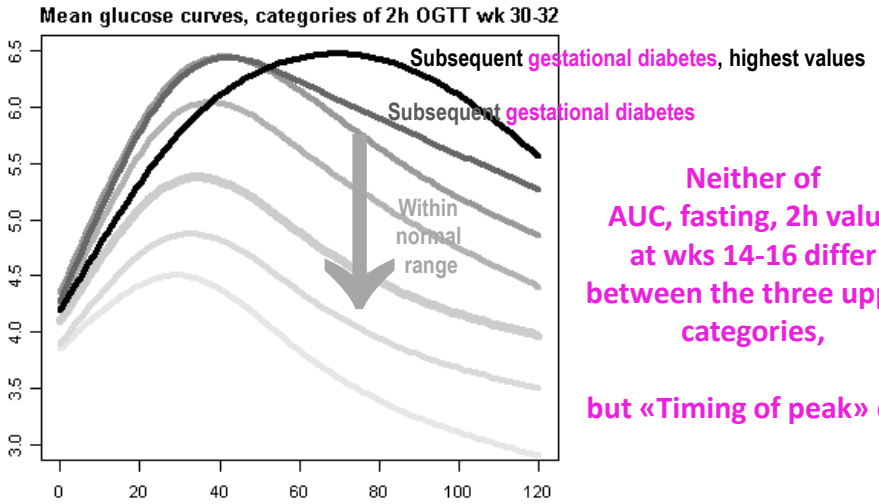




AUC

would also be fine here

Glucose curves, early pregnancy → 2h glucose categories, late pregnancy



Neither of AUC, fasting, 2h value at wks 14-16 differ between the three upper categories,

but «Timing of peak» do.

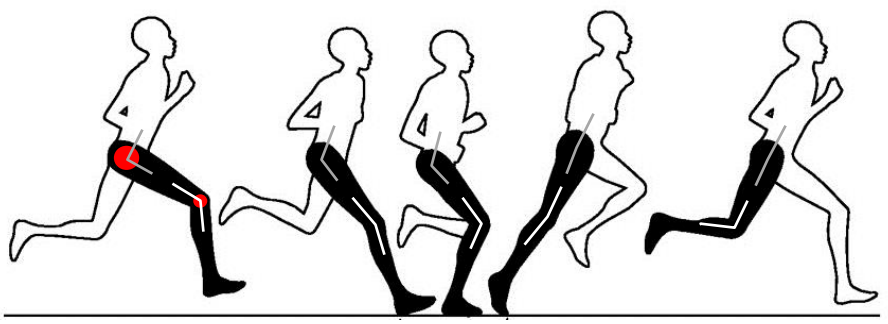
Placing FDA in the landscape of statistical analyses

Placing FDA in the landscape of statistical analyses

Types of data (basic units in the analysis)

Methods	Categorical data	Continuous data	Censored data	Functional data
Descriptive statistics	Proportions Tables	Mean or median SD or quartiles	Kaplan-Meier plot Survival tables	Mean curve Principal components
Comparisons of groups	Cross-tables Pearson χ^2 test	T-test or Mann-Whitney ANOVA or Kruskal-Wallis	Log-rank test	Functional ANOVA, permutation F-tests
Simple regression models	Logistic regression	Linear regression	Poisson-regression Cox-regression	Functional regression
Regression models for repeated data/ hierarchical structure	Multilevel/mixed models			

Examples of functional data

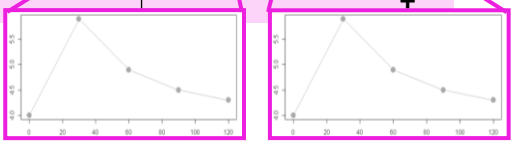


STORK@Oslo University Hospital

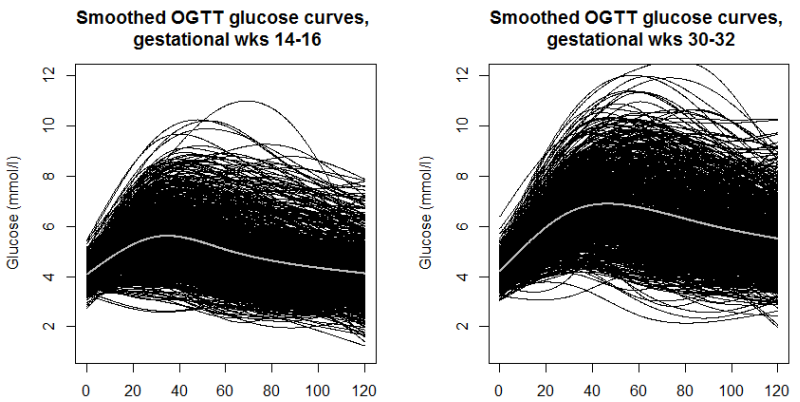
BIG babies and Complications

n=1031 healthy, pregnant women

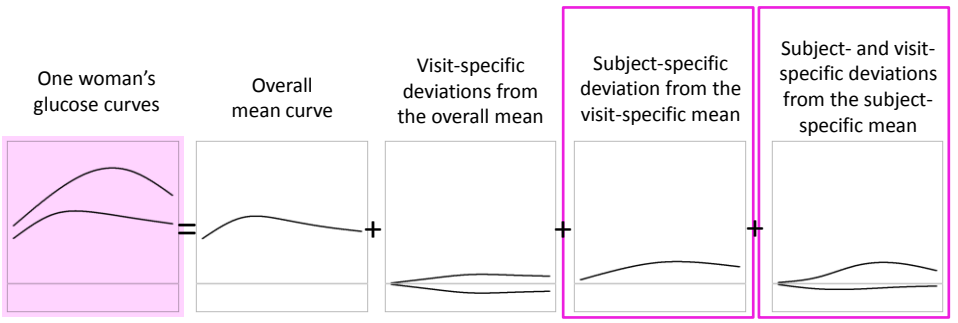
Pregnancy week	14-16	30-32	Birth
Background/demographics	+		
Anthropometry (BMI)	+	+	+
Glucose challenge (2-hour OGTT)	+	+	
Inflammation	+	+	
Birth data			+



Repeated glucose curves during pregancy: Multilevel functional data analysis

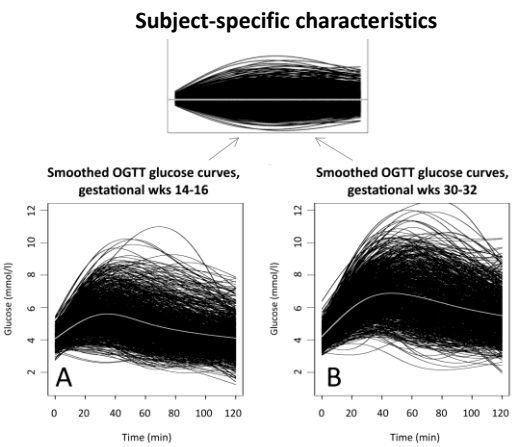


Frøslie KF, Røislien J, Qvigstad E, Godang K, Bollerslev J, Henriksen T, Veierød MB.
Shape information in repeated glucose curves during pregnancy provided significant physiological information for neonatal outcomes. PLoS One 2014;9:e90798

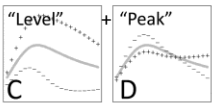


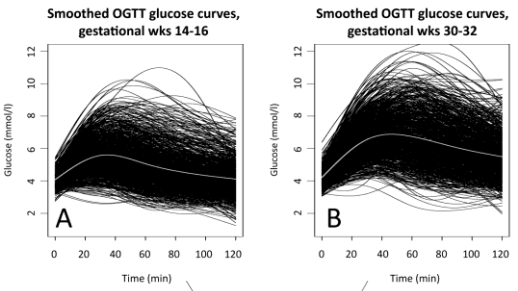
$$\gamma_{iv}(t) = \mu(t) + \eta_v(t) + X_i(t) + U_{iv}(t)$$

Glucose curves = Fixed effects curves + **Random effects curves**

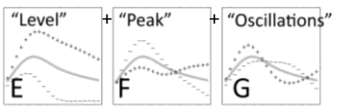


Subject-specific characteristics

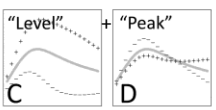


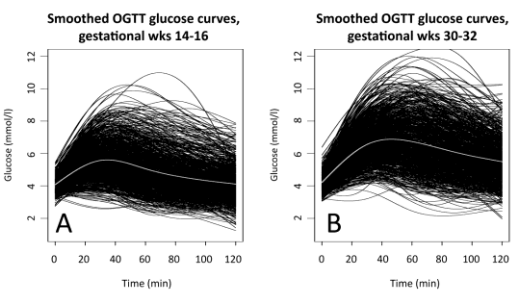


Subject- and visit-specific characteristics



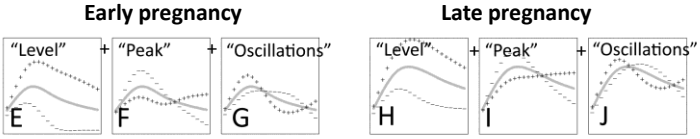
Subject-specific characteristics



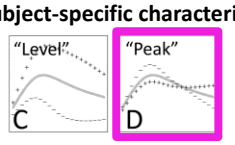


Subject- and visit-specific characteristics

Early pregnancy **Late pregnancy**



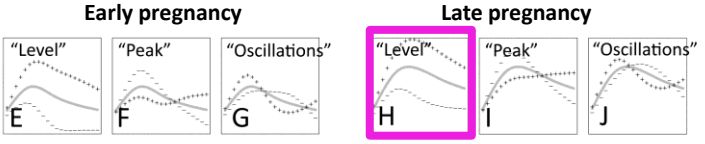
Subject-specific characteristics



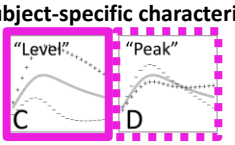
→ **Birth weight**

Subject- and visit-specific characteristics

Early pregnancy **Late pregnancy**



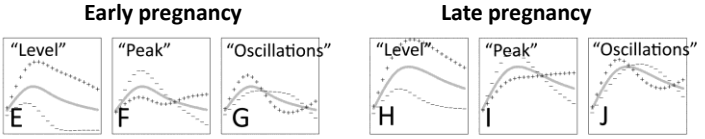
Subject-specific characteristics



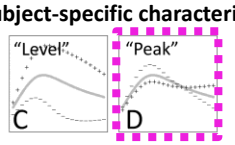
→ **Neonatal % fat**

Subject- and visit-specific characteristics

Early pregnancy **Late pregnancy**



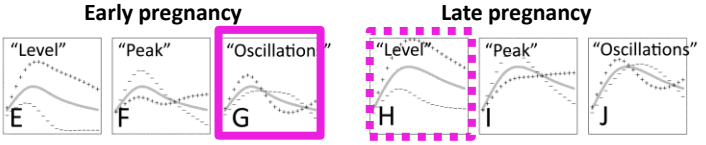
Subject-specific characteristics



→ **Neonatal insulin**

Subject- and visit-specific characteristics

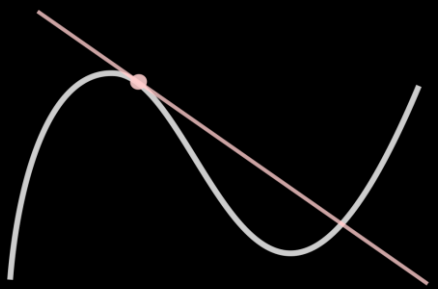
Early pregnancy **Late pregnancy**



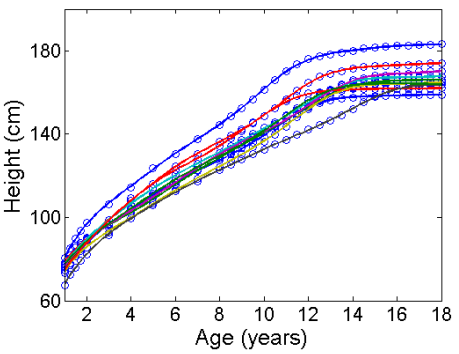
Summary, health

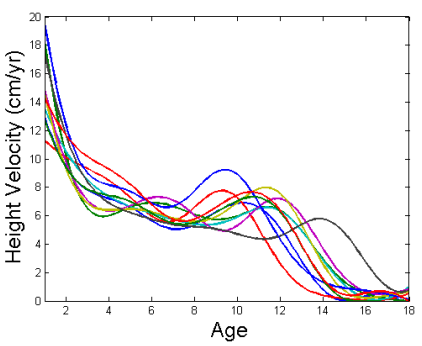
- ✓ Women's curve shapes contain important information.
- ✓ Glucose levels rise, and the glucose peak lags, as pregnancy proceeds
- ✓ A late glucose peak in early pregnancy is a risk factor for subsequent gestational diabetes
- ✓ Glucose curve characteristics influence neonatal outcomes

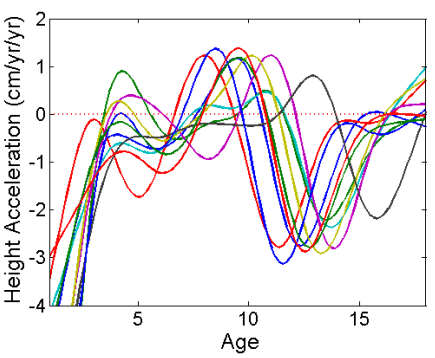
Functional data: Derivatives



FDA: assuming continuity and smoothness opens for analysis of derivatives







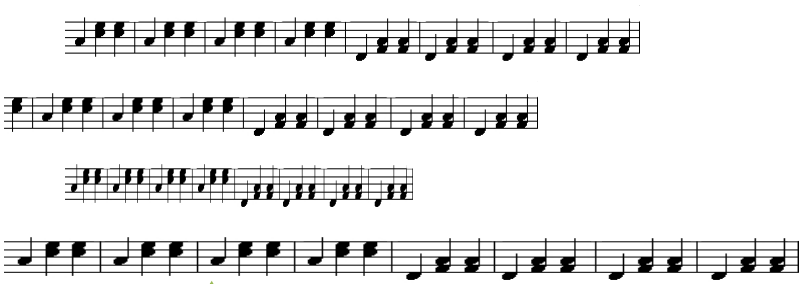
Data alignment

Data alignment

D a t a a l i g n m e n t

D a t a a l i gnm e n t

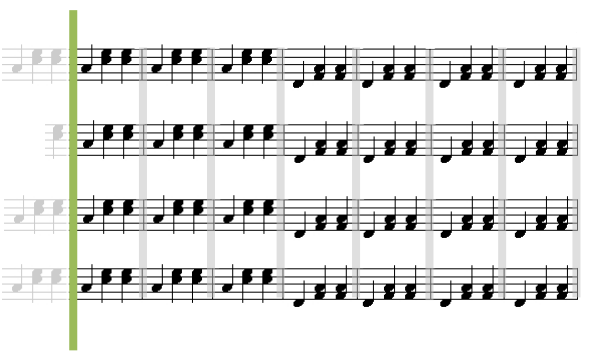
Take this waltz



With its very own breath
of brandy and Death

37/47

Take this waltz

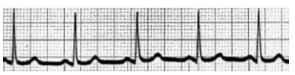


Alignment: What and why?

"**Alignment**" = "**Registration**" = "**Warping**"

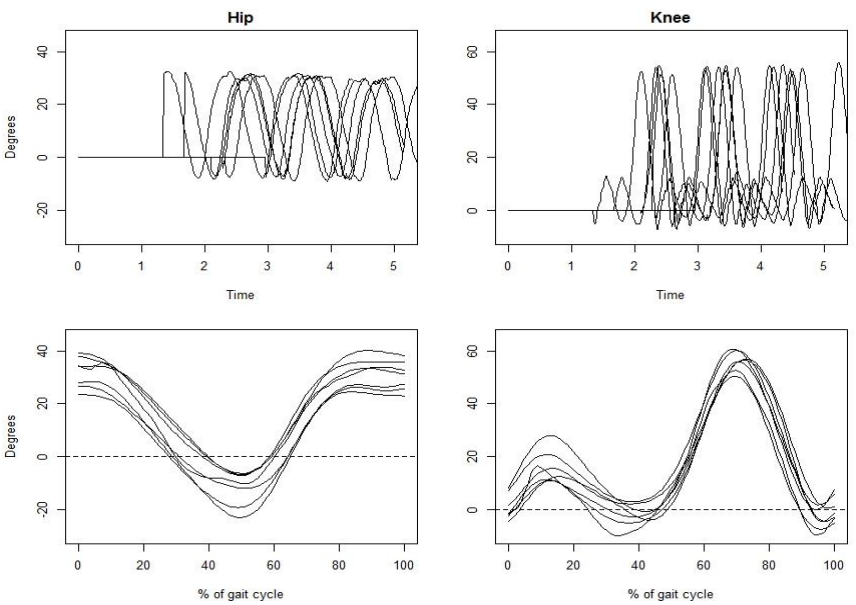
Important features are observed across units – but at different times (or space location).

Cf. Rigid **metric of physical time** versus **musical or biological rhythms**.



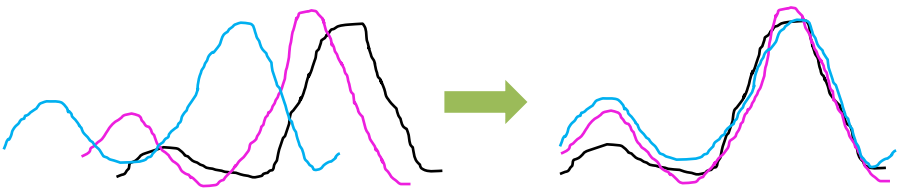
Misalignment blur any subsequent analyses, including descriptive statistics, and acts as a confounder.

Example: Recording gait cycles. Unaligned vs **aligned**



Alignment: Landmark-based

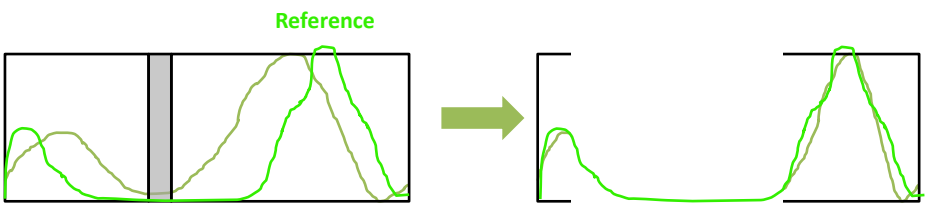
A) Left-right alignment ("Shift")



B) Stretching the axis

Alignment: Continuous registration

Alternative to landmark-based alignment:
Continuous registration
Non-linear transformation of t, "warping functions",
often solved by iterative procedures



Take-home message

- ✓ Curve shapes may contain important information. Use FDA to analyse them.
- ✓ For analyses of other basic units than curves, check out object-oriented data analysis

References

- ✓ Ramsay JO. www.functionaldata.org
- ✓ Ramsay JO, Silverman BW. Functional Data Analysis, 2nd ed. *Springer,* 2005.
- ✓ Marron JS, Alonso AM. Overview of object oriented data analysis. Biom J. 2014;5:732-53.
- ✓ Frøslie KF, Røislien J, Qvigstad E, Godang K, Voldner N, Bollerslev J, Henriksen T, Veierød MB. *Shape information from glucose curves: Functional data analysis compared with traditional summary measures.* BMC Medical Research Methodology 2013;13:6
- ✓ Frøslie KF, Røislien J, Qvigstad E, Godang K, Bollerslev J, Henriksen T, Veierød MB. *Shape information in repeated glucose curves during pregnancy provided significant physiological information for neonatal outcomes.* PLoS One 2014;9:e90798

Tools

- ✓ R (package `fda`)
- ✓ Matlab

Thanks to

Marit B Veierød, Jo Røislien, Tore Henriksen, Petter Laake, co-authors,co-workers at OCBE and OUS,
the 1031 women in the STORK study