Development of new Austrian height and weight references

Andreas Gleiß, Michael Schemper
Center for Medical Statistics, Informatics and Intelligent Systems
Gabriele Häusler
Dept. of Paediatrics and Adolescent Medicine
MEDIZINISCHE UNIVERSITÄT WIEN

Introduction

Reference curves used in Austria:

- First Zurich longitudinal study (Prader et al, 1989)
  - 274 children (longitudinal) from birth years 1954 to 1956
  - middle and upper class, mostly not breastfeeded
- Germany: Kromeyer-Hauschild et al (Mon.schr.KHK, 2001)
- WHO: de Onis et al (Bulletin WHO, 2007)
- Germany: Rosario et al (Ann.Hum.Biol., 2011)

Girls' height at 10 years (cm)	3rd perc	**50th perc**	97th perc
Prader et al (1989)	126.2	**138.2**	148.7
Kromeyer-Hauschild et al (2001)	128.9	**141.2**	153.4
de Onis et al (2007)	126.6	**138.6**	150.7
Rosario et al (2011)	128.6	**141.3**	154.1

Introduction

Clinical use of reference curves in paediatrics:

- monitor a child's somatic development
- decision making in diagnostic work-up for shortness
- based on Standard Deviation Scores (SDS)
- height and body proportions

Data (1/4)

- targeted sample size: ≥ 12,000
- according to age- and sex-specific population proportions in 9 provinces (Statistik Austria, 2009)
- Austrian Federal Ministry of Education, Arts and Culture
- External health care system consulting company
- 117 kindergartens, primary schools, secondary schools, and vocational schools
- November 2009 to June 2011
- sampling & statistics financed by Merck Ges.m.b.H.

Data (2/4)

- sampling whole classes
  - higher sample size
  - outside planned age range of 4 to <19 years
- 15,368 recruited
- exclude 67 (0.44%), e.g.: plaster cast, trisomy 21
- 15,301 eligible
- 14,544 of 4 to <19 years (7,415 boys and 7,129 girls)
- + 442 half year outside
- cross-sectional!

Data (3/4)

- measurements:
  - height
  - sitting height
  - weight

Data (4/4)

- measurements:
  - height
  - sitting height
  - weight
- derived quantities:
  - subischial leg length = height – sitting height
  - ratio of sitting height / leg length
  - BMI = weight / height^2
- all variables measured completely

GAMLSS Method (1/9)

- **G**eneralized **A**dditive **M**odels for **L**ocation, **S**cale and **S**hape
  - Rigby & Stasinopoulos (Stat. Med., 2004)
  - Rigby & Stasinopoulos (Appl. Statist., 2005)
  - Stasinopoulos & Rigby (J. Stat. Software, 2007): R-package
- GAMLSS model has two components:
  a) parametric distribution at each value of independent var.
  b) model dependence of distributional parameters on independent variable

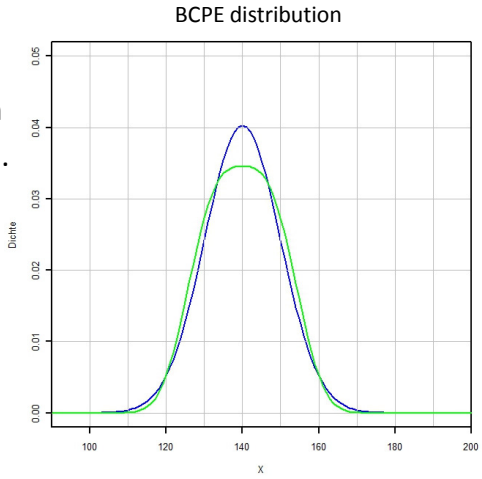
GAMLSS Method (2/9)

a) at each age parametric distribution:

- normal, log-normal, …
- more general, e.g.:
  - Box-Cox t-distribution
  - **B**ox-**C**ox **P**ower **E**xpon. (BCPE)

→skewness

→kurtosis



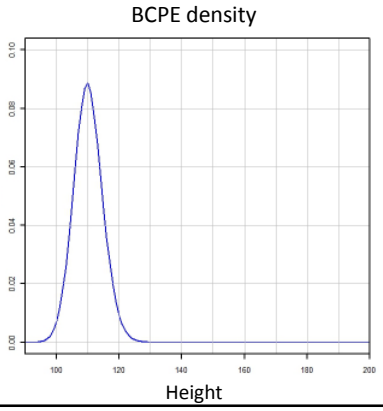
GAMLSS Method (3/9)

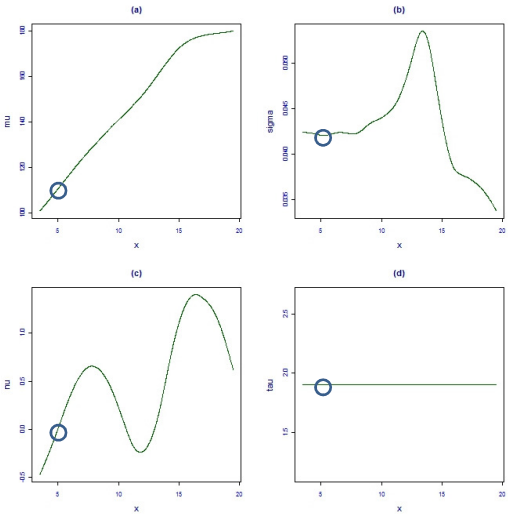
b) distributional parameters depending on age:

- linear, polynomial
- smoothing functions:
  - fractional polynomials
  - **cubic splines**

GAMLSS Method (4/9)

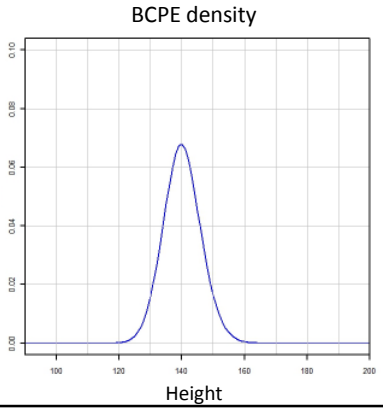
- Example: height (boys)
- 5 years

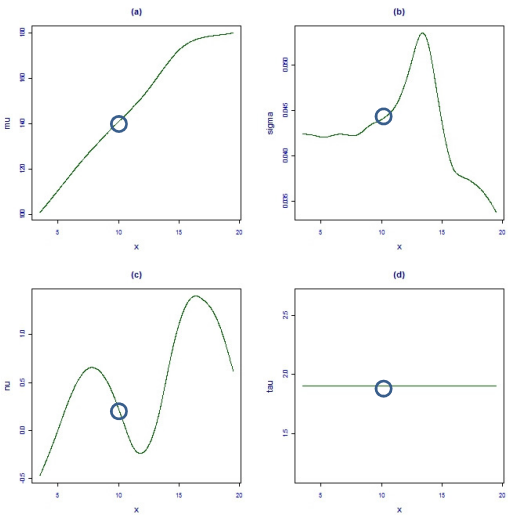




GAMLSS Method (5/9)

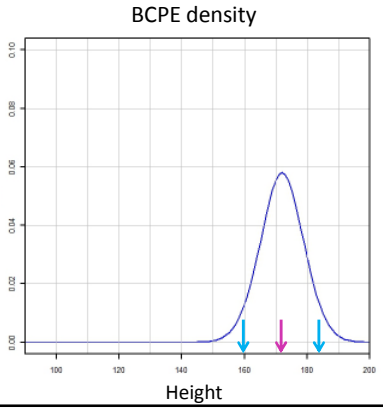
- Example: height (boys)
- 10 years

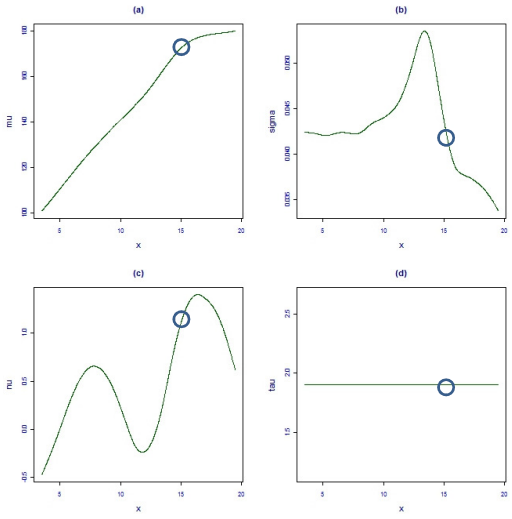




GAMLSS Method (6/9)

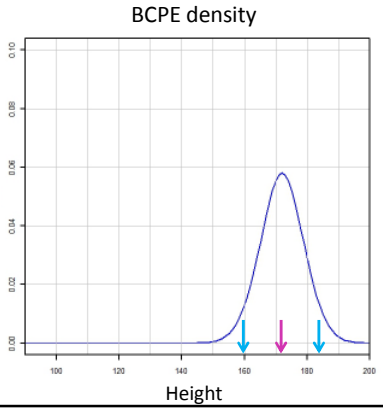
- Example: height (boys)
- 15 years

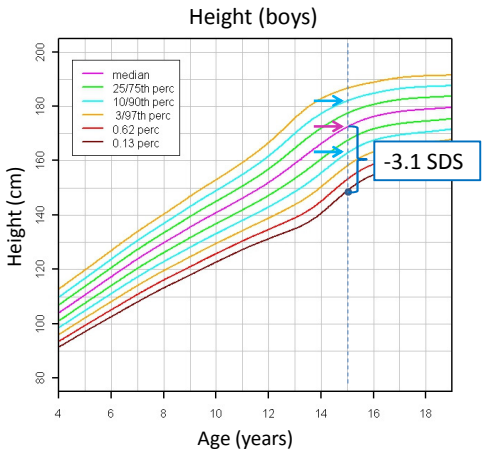




GAMLSS Method (6/9)

- Example: height (boys)
- 15 years





GAMLSS Method (7/9)

- estimation of Standard Deviation Scores (SDS) for concrete measurement *X*:

$$Z = \begin{cases} \frac{1}{\sigma\nu}\left[\left(\frac{x}{\mu}\right)^{\nu} - 1\right] & \text{if } \nu \neq 0 \\ \frac{1}{\sigma}\log\left(\frac{x}{\mu}\right) & \text{if } \nu = 0 \end{cases}$$

Rigby & Stasinopoulos (Stat. Med., 2004)

$$\text{SDS} = \Phi^{-1}\left(\left(1 + sign(z)\, G_{1/\tau}\left(\frac{1}{2}\left|Z/c\right|^{\tau}\right)\right)/2\right)$$

Φ standard normal distribution function

$G_{1/\tau}$ standard Gamma distribution function with parameter $1/\tau$

$$c = \left[2^{-2/\tau}\Gamma(1/\tau)\left(\Gamma(3/\tau)\right)^{-1}\right]^{1/2}$$

Γ Gamma function.

GAMLSS Method (8/9)

- further corrections needed:
  - correct for truncation of distribution function (see Appendix A in Rigby & Stasinopoulos, 2004)
  - correct for extreme observations removed before model estimation

$$SDS_{corr} = \Phi^{-1}\left(\left(\Phi(SDS)\cdot(n - o_l - o_u) + o_l\right)/n\right)$$

n: original # observations
o_l: # omitted at lower end
o_u: # omitted at upper end

e.g. for $o_l = o_u = 2$: SDS_{corr} = -2.5 for *SDS* = -2.84 (n=500)
or *SDS* = -2.52 (n=5000)

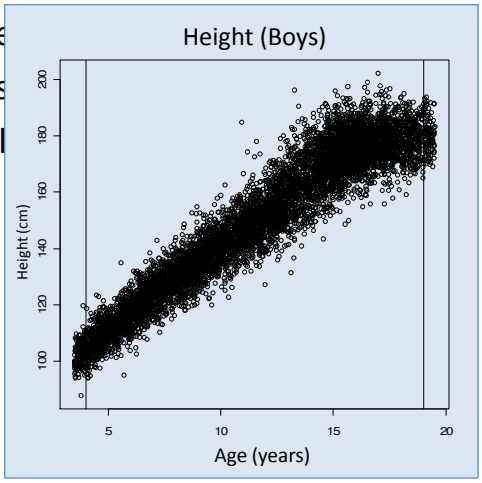
GAMLSS Method (9/9)

- separate model for each sex
- separate model for each outcome variable
- LMS-method by Cole & Green (Stat.Med., 1992)
  = special case of GAMLSS
- Borghi et al. (Stat.Med., 2006):
  evaluation of methods for WHO
- de Onis et al (Bulletin WHO, 2007):
  GAMLSS with BCPE and cubic splines for estimating WHO growth curves

Modelling (1/4)

Step 1: **data preparation**

- clarify unplausible value
- remove uneligible cases
- weight, BMI: subtract cl



Modelling (2/4)

Step 2: **starting model**

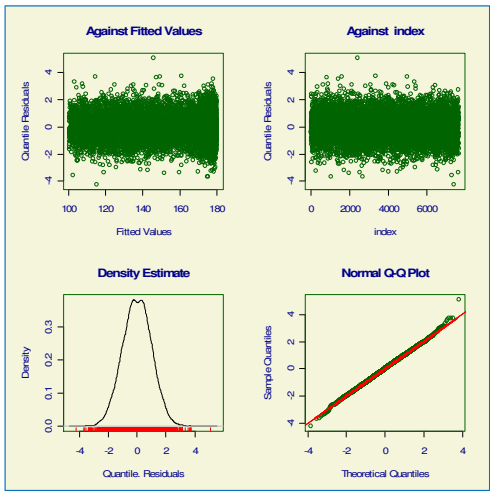
- mean and variability: 10 d.f.
- skewness and kurtosis: linear

→ remove extreme values (residuals outside +/-4 SD)

Modelling (2/4)

Step 2: **starting model**

```
Residuals: < -4 or > +4
========================
     4290     14352
 5.076757 -4.223742
         nr groesse     alter
4290    4384   184.8 10.929423
14352 14707    95.0  5.696547
```



Modelling (3/4)

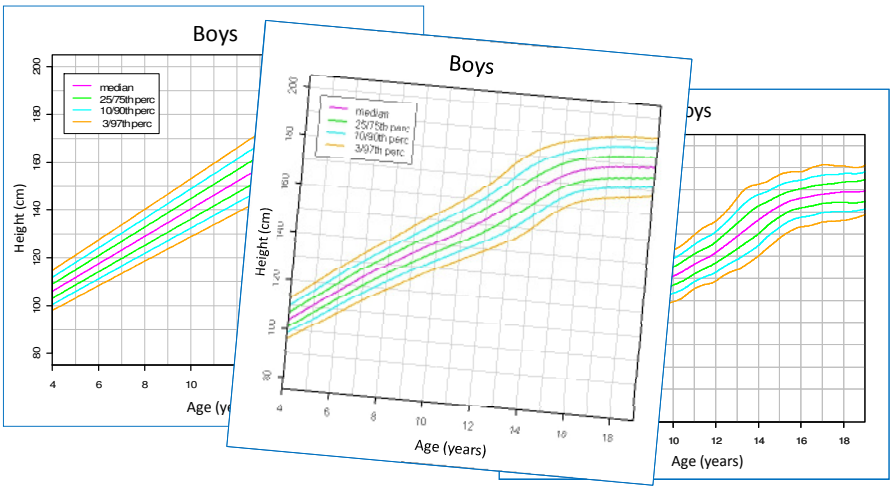
Step 3: **interim model**

- based on reduced data set
- optimization:
  - degrees of freedom for cubic splines
  - age exponent
  - using R-function `find.hyper()`
- criterion = (Generalized) Akaike Information Criterion
  = $k \cdot p - 2 \log(L)$

k = penalty, p = # parameters, L = likelihood

Modelling (3/4)

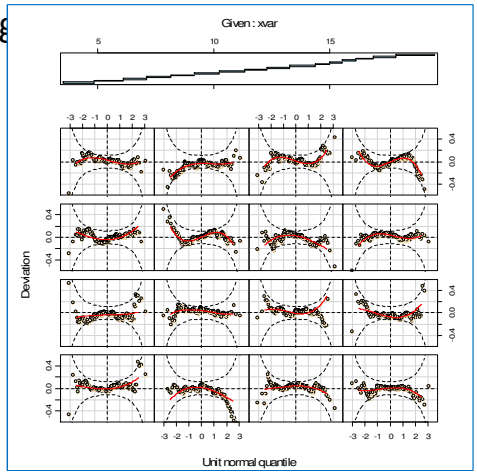
Step 3: **interim model**



Modelling (3/4)

Step 3: **interim model**

- diagnostic Tools (cf. Borg
  - residual plots
  - worm plots
    (van Buuren & Fredriks,
    Stat.Med., 2001)



Modelling (3/4)

Step 3: **interim model**

- diagnostic Tools (cf. Borghi et al, Stat. Med., 2006):
  - residual plots
  - worm plots
    (van Buuren & Fredriks,
    Stat.Med., 2001)
  - observed vs. fitted proportions below centiles
- remove further extreme values if necessary
- re-run optimization

Modelling (4/4)

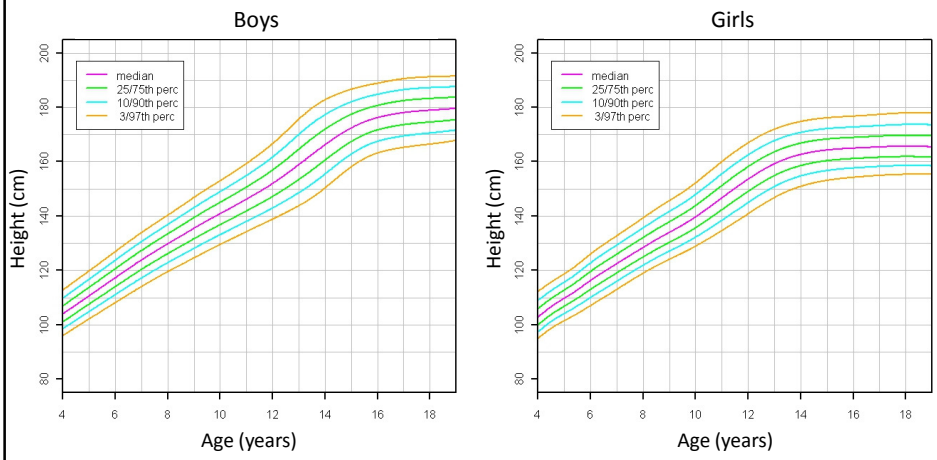
Step 4: **final model**

- consultation with paediatrician
- adapt model if necessary:
  - kurtosis more rigid (e.g. height for boys)
  - BMI generally more rigid:
    higher penalty *k* in GAIC optimization criterion

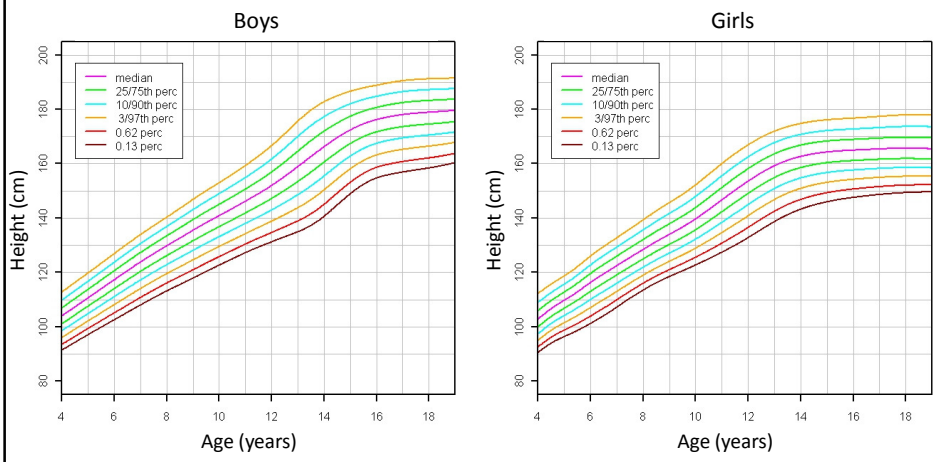
Modelling (4/4)

Boys	outliers removed	penalty *k* in GAIC	d.f. μ	d.f. σ	d.f. ν	d.f. τ	Age Expon.
Height	1+1	2	8.45	8.23	3.49	-	1.36
Weight	2+0	2	10.71	7.33	0.10	3.40	1.10
BMI	6+0	3	6.31	4.13	2.97	2.63	0.04
Sit. height	4+2	2	9.01	8.29	5.10	-	1.00
Leg length	3+2	2	8.79	7.53	4.33	-	1.01
Ratio SH/LL	8+11	2	8.19	2.80	0.79	0.17	1.00

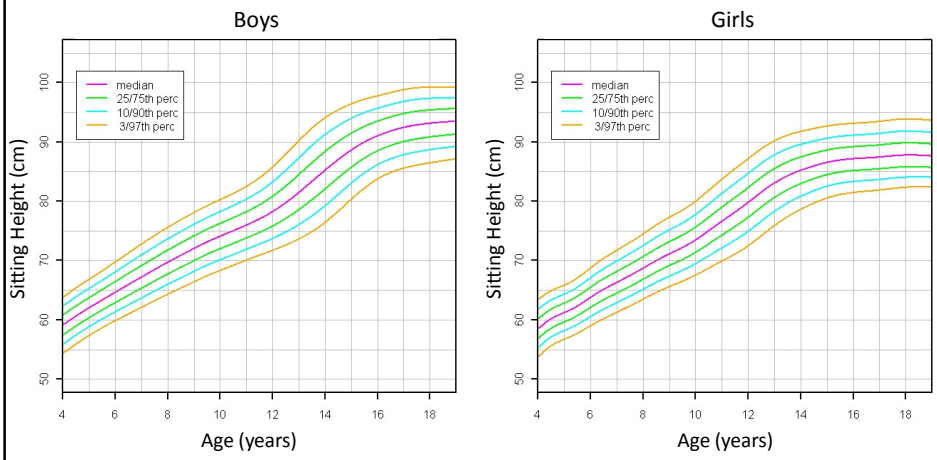
Height



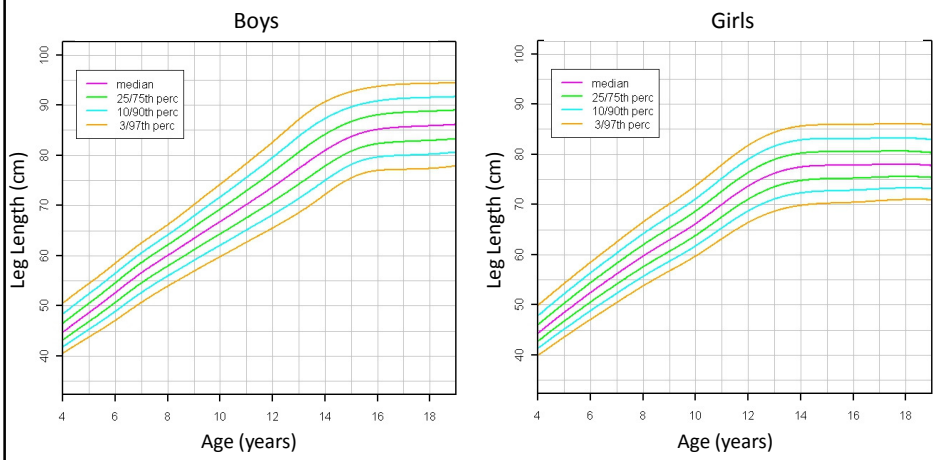
Height



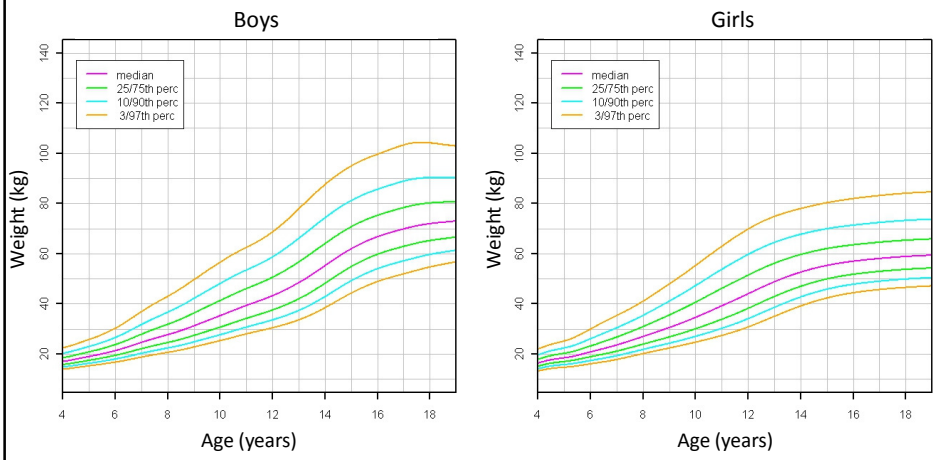
Sitting Height



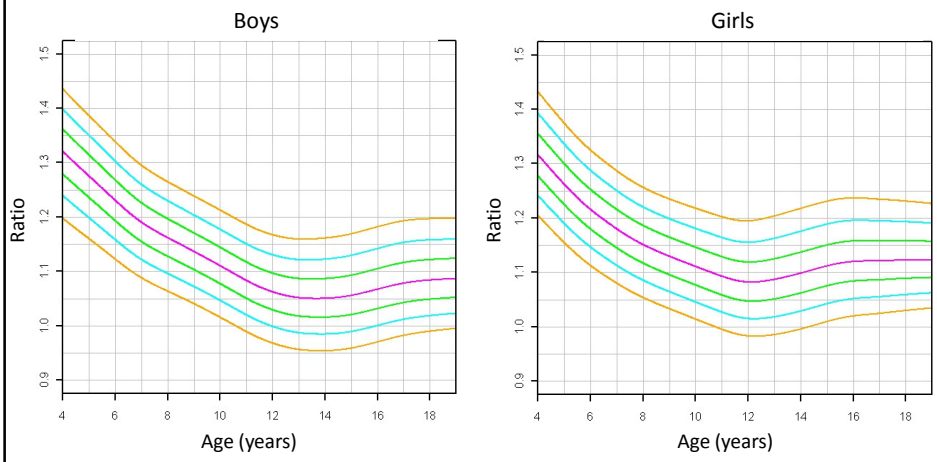
Subischial Leg Length



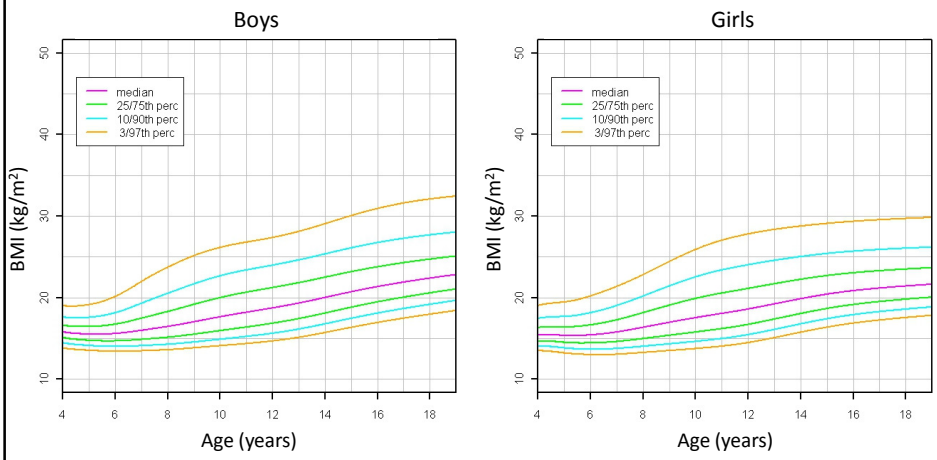
Weight



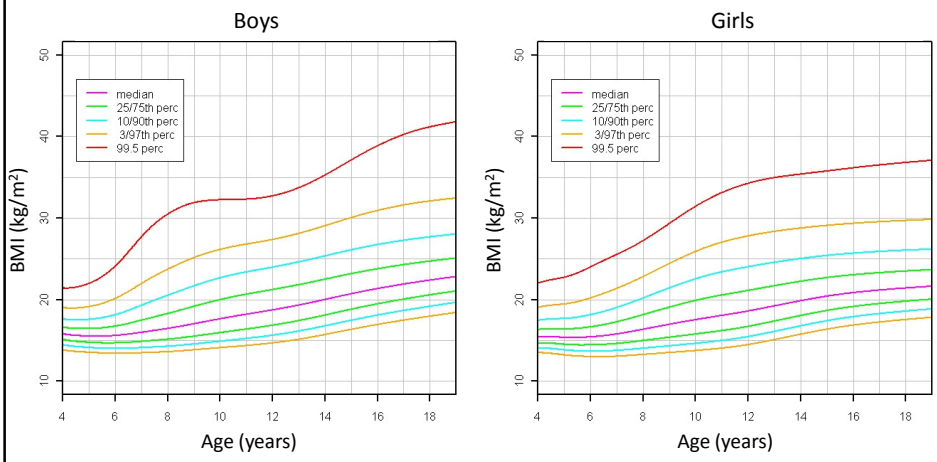
Ratio Sitting Height / Leg Length



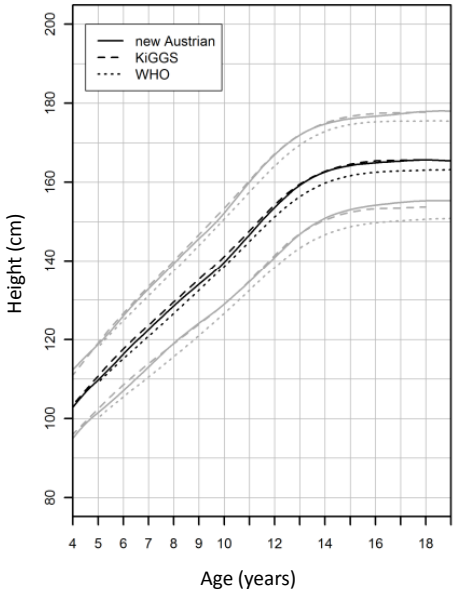
Body-Mass-Index



Body-Mass-Index



Comparisons (1/3)



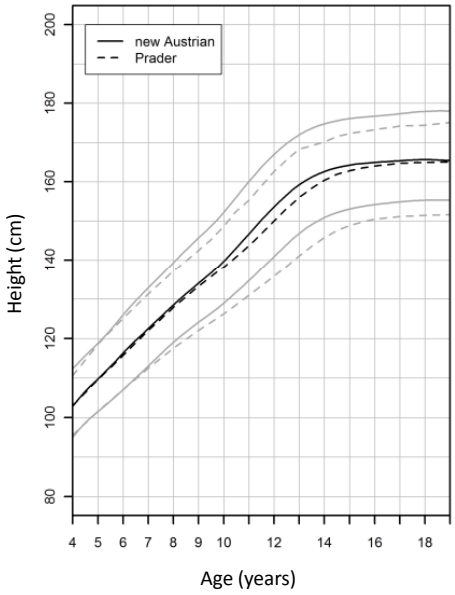
new Austrian vs. KiGGS vs. WHO

Height for Girls

black = median
grey = 3rd / 97th percentile

solid = new Austrian
dashed = KiGGS (Germany)
dotted = WHO

Comparisons (2/3)



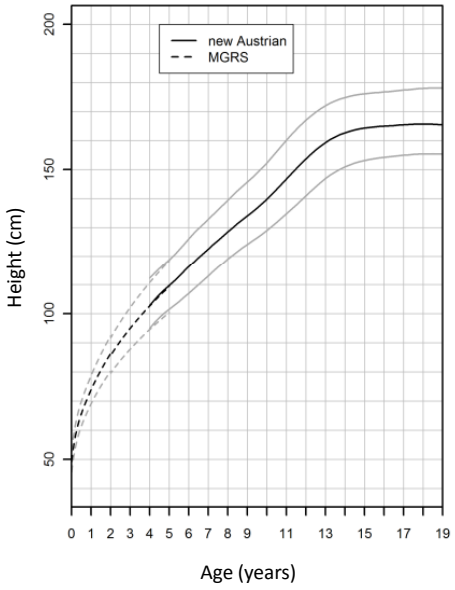
new Austrian vs. Prader

Height for Girls

black = median
grey = 3rd / 97th percentile

solid = new Austrian
dashed Prader

Comparisons (3/3)



new Austrian + MGRS

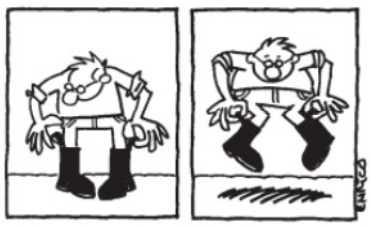
Height for Girls

black = median
grey = 3rd / 97th percentile

solid = new Austrian (4 to <19 years)
dashed = MGRS (0 to 5 years)

Sample Size Issues (1/3)

- 150 bootstrap samples
  - original sample size
  - higher / lower sample sizes
- on each bootstrap sample:
  - estimate GAMLSS model
  - calculate low percentile curve (e.g. -2.5 SDS)
- across all bootstrap samples:
  - estimate standard error of percentiles at different ages
- derive recommendation for adequate sample size



Sample Size Issues (2/3)

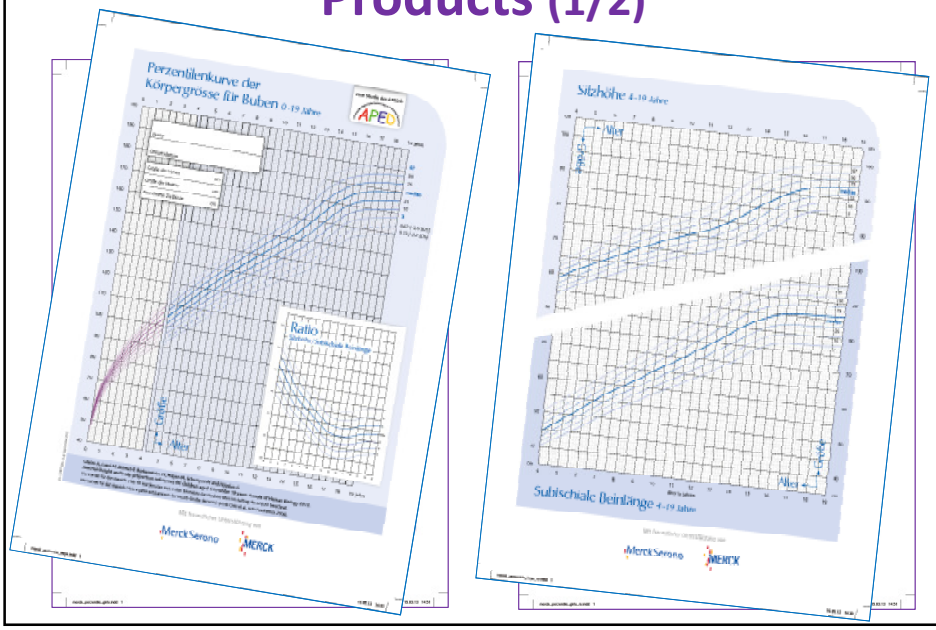
- Estimation of GAMLSS model on bootstrap samples:
  - Step 1: data preparation (before bootstrap)
  - Step 2: starting model for outlier removal
    → automatic
  - Step 3: interim model by d.f. optimization
    → automatic
  - Step 4: final model by "inspection":
    - residuals outside +/-3.5 or +/-3.0
    - or increase penalty *k* in GAIC

Sample Si

- Standard errors (cm)
  - for -2.5 SDS
  - height (boys)
- recommend sample size of ≥ 5000 per sex

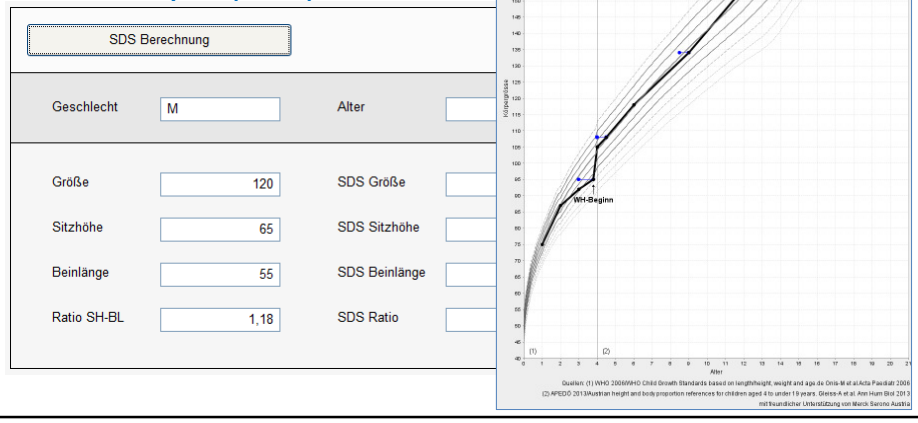
Age	n=2500	n=5000	n= 7663	n=10000
4	1.21	0.94	0.67	0.71
5	0.78	0.73	0.46	0.40
6	0.93	0.95	0.53	0.42
7	0.82	0.75	0.51	0.38
8	0.80	0.64	0.53	0.43
9	0.86	0.71	0.51	0.47
10	1.05	0.87	0.55	0.52
11	0.93	0.74	0.54	0.45
12	1.21	0.93	0.61	0.56
13	1.43	1.09	0.83	0.61
14	1.40	1.08	0.72	0.60
15	1.21	1.02	0.57	0.54
16	1.21	1.00	0.56	0.54
17	1.08	0.80	0.61	0.51

Products (1/2)



Products (2/2)

- Electronic:
  Research Documentation
  & Analysis (RDA)



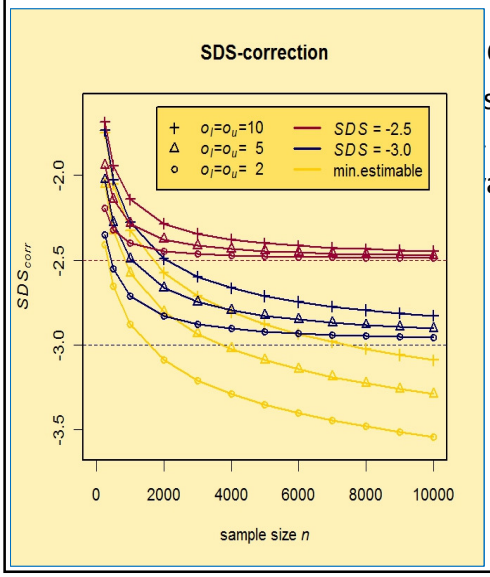
Conclusions

- 14.544 individuals
- balanced w.r.t. age, sex and province
- state-of-the-art modelling
- curves suitable for the next decades
- results for height and body proportions published (Gleiss et al., Ann.Hum.Biol., 2013)
- results for weight and BMI (over-weight and obesity) submitted
- secondary use of data

References

- **Borghi**, E. et al. (Statistics in Medicine, 2006)
- **Cole**, T.J. & **Green**, P.J. (Statistics in Medicine, 1992)
- **de Onis**, M. et al (Bulletin of the WHO, 2007)
- **Gleiss**, A. et al (Annals of Human Biology, 2013)
- **Kromeyer-Hauschild**, K. et al (Monatsschrift Kinderheilkunde, 2001)
- **Mayer**, M., Gleiss, A. et al (submitted)
- **MGRS** (Acta Paediatrica, 2006)
- **Prader**, A. et al (Helvetica Paediatrica Acta, 1989)
- **Rigby**, R.A. & **Stasinopoulos**, D.M. (Statistics in Medicine, 2004)
- **Rigby** , R.A. & **Stasinopoulos**, D.M. (Applied Statistics, 2005)
- **Rosario**, A.S. et al (Annals of Human Biology, 2011)
- **Stasinopoulos**, D.M. & **Rigby** , R.A. (J. of Statistical Software, 2007)
- **van Buuren**, S. & **Fredriks**, M. (Statistics in Medicine, 2001)

GAMLSS Method (8/9)



d:

stribution function
Stasinopoulos, 2004)
ations removed

$$SDS_{corr} = \Phi^{-1}\left(\left(\Phi(SDS)\cdot(n-o_l-o_u)+o_l\right)/n\right)$$

n: original # observations
o_l: # omitted at lower end
o_u: # omitted at upper end

Modelling (4/4)

Girls	outliers removed	penalty k in GAIC	d.f. μ	d.f. σ	d.f. ν	d.f. τ	Age Expon.
Height	2+0	2	10.66	6.89	4.09	1.67	0.27
Weight	1+0	2	9.32	6.59	4.34	3.90	0.28
BMI	2+0	3	6.32	3.90	0.16	2.77	0.23
Sit. height	3+2	2	13.46	7.44	3.43	0.10	0.50
Leg length	9+1	2	8.99	0.16	3.54	0.18	0.93
Ratio SH/LL	8+8	2	8.26	2.88	0.10	0.10	1.00