Polygenic Influences on Boys' & Girls' Pubertal Timing & Tempo

Gregor Horvath, Valerie Knopik, Kristine Marceau

Purdue University

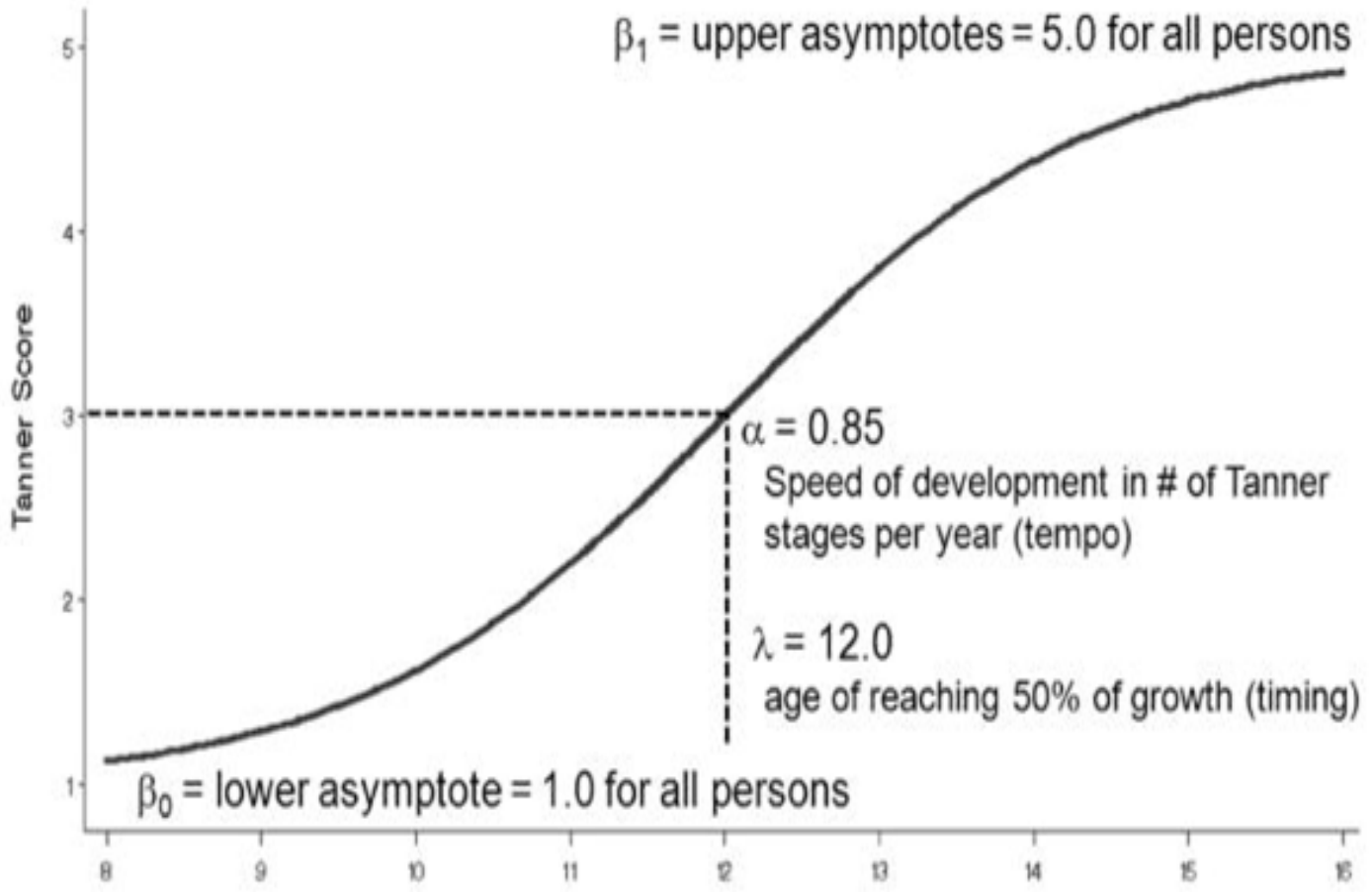
Biobehavioral Development Laboratory

PURDUE UNIVERSITY

Timing & Tempo of Puberty

- Varies by individual
  (Marceau et al., 2011)
- Risk factor for Cancer, depression, substance use, & obesity
  (Day et al., 2017;
  Lanza, Collins, & Linda, 2002;
  Nolen-Hoeksema & Girgus, 1994)

Figure from Marceau et al., 2011



Genetic Measurement

- Status
  Timing
  Tempo are heritable (Beunen et al., 2000, Ge et al., 2007)
- Candidate gene studies identify direct gene effects (Witchel et al., 2001)
- GWAS report many common loci in pubertal development (Day et al., 2015, 2017; Elks et al., 2010; Ong et al., 2009; Perry et al., 2014)

Why Polygenic Scoring Techniques?

- Many genes of small effect (Day et al., 2017)
- More reliable & powerful than candidate genes (Holz et al., in press)
- Could provide information on biological pathways

Research Questions

- How do different polygenic scoring techniques (k-fold & empirical scoring) differ in their ability to predict pubertal timing & tempo in boys and girls?

ALSPAC Cohort & Sample

Cohort

- 13,988 children from 14,541 mothers with expected delivery dates between April 1st, 1991 and December 31st, 1992
  - Sample bolstered with 713 additional children at 7 years

Analytic Sample

- 8801 youth from the bolstered ALSPAC sample with data who had data on pubertal development and genetics
- 96% White, 3% Other, <1% Asian, Black, or Other

Procedure

Measures

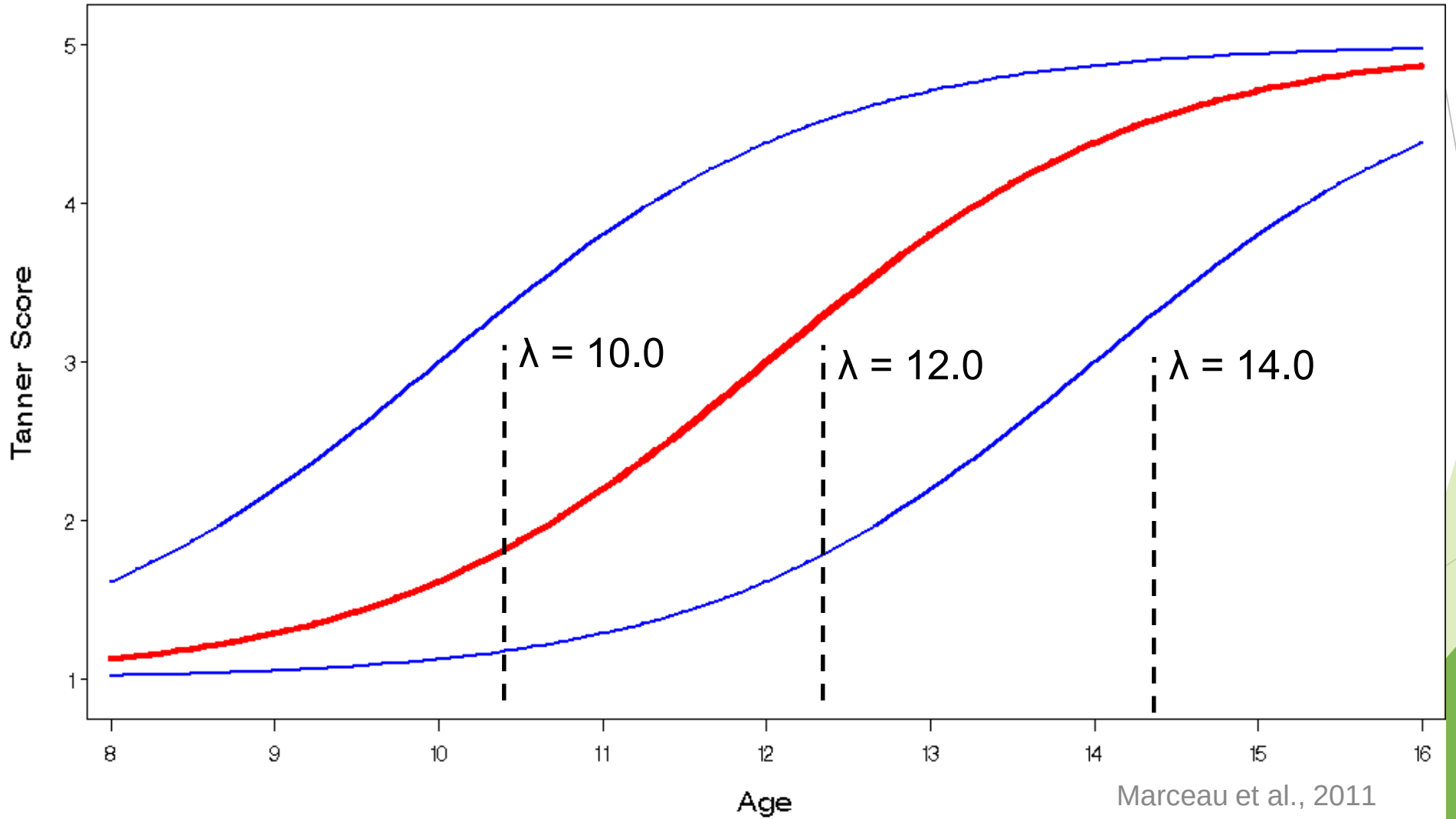
- 9 Assessments of Nurse Reported Tanner Stages (ages 8-17) (*n* = 13,994 with at least one measure)
  (Marshall & Tanner, 1969, 1970; Tanner, 1962)
- Timing & Tempo derived from logistic growth models
  (e.g., Marceau et al., 2011)

SNP Measurement

- 9,912 children genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform
- Quality control conducted in Plink (Purcell et al., 2007)
  - Data excluded for missingness per individual (>.1), marker (>.1), allele frequency (<.01), & Hardy-Weinberg Principle (<.001)
  - SNPs pruned based on linkage Disequilibrium (LD) threshold (.3)

Logistic Growth Model: Timing

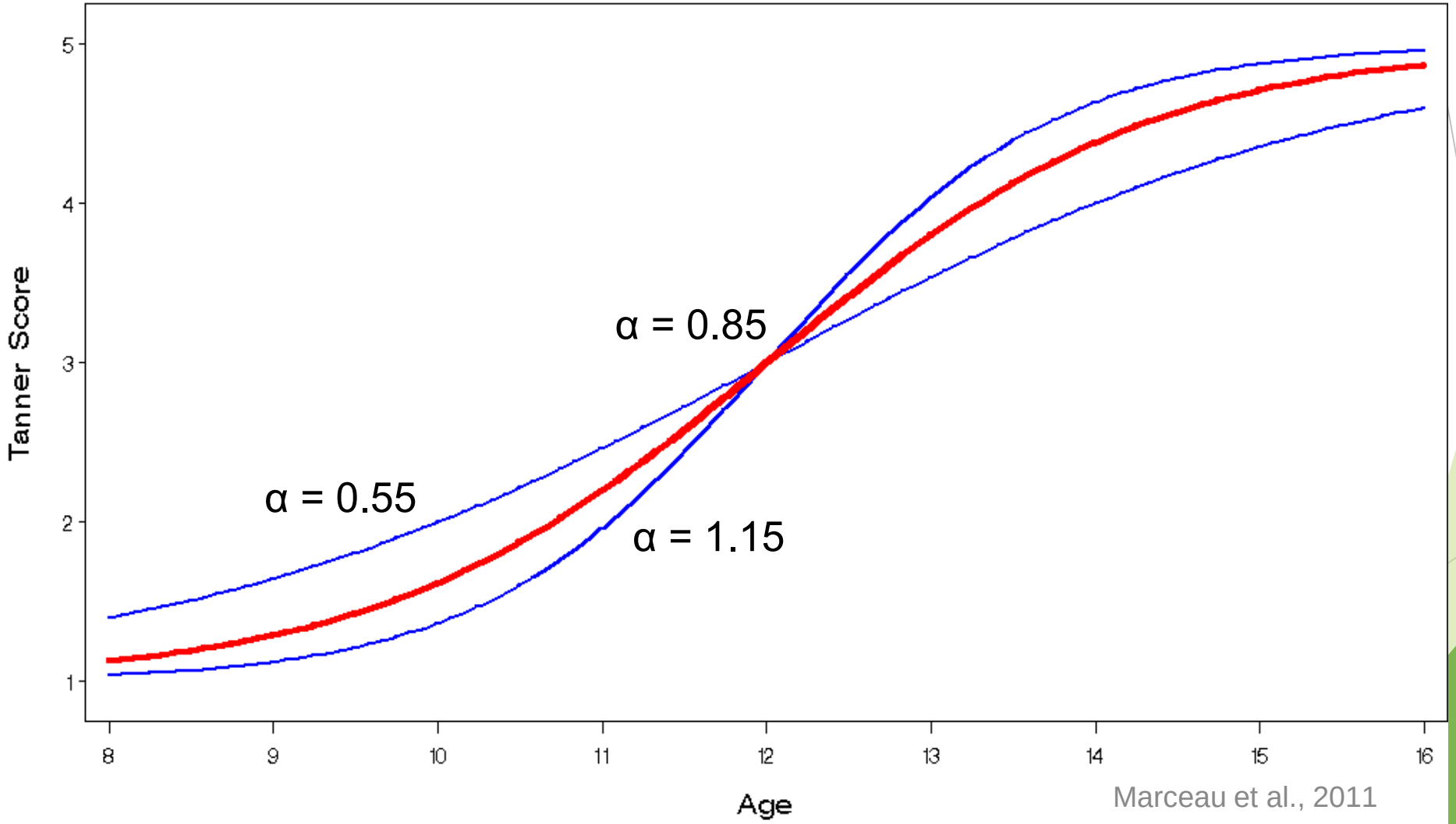
Age at which individuals reach relevant milestones



Marceau et al., 2011

Logistic Growth Model: Tempo

How fast an individual progresses through puberty



Marceau et al., 2011

Techniques

- Empirical
  - 96 SNPs that significantly predicted pubertal timing and/or tempo in Day et al. (2017)
- K-Fold - theoretical
  - 3410 SNPs, 313 genes
  - Variations in number of fold, P-threshold, and direction of SNP effect

K-fold strategy

Loop repeated for each SNP

Loop repeated K times

1. Select test (K=1) and discovery (K > 1) sets
Code & mean center each SNP
identify monomorphic SNPs
run baseline, individual SNP regressions
SNPs → Timing & Tempo (separately)
save coefficients

Training Set

Tested 2, 3, 4, & 5 folds

Tested p-thresholds of .05, .1, .25, .5, .75, & 1

Tested using only SNPs with + β and only SNPs with - β

2. Create matrix of genotypes in test sample (after centering)

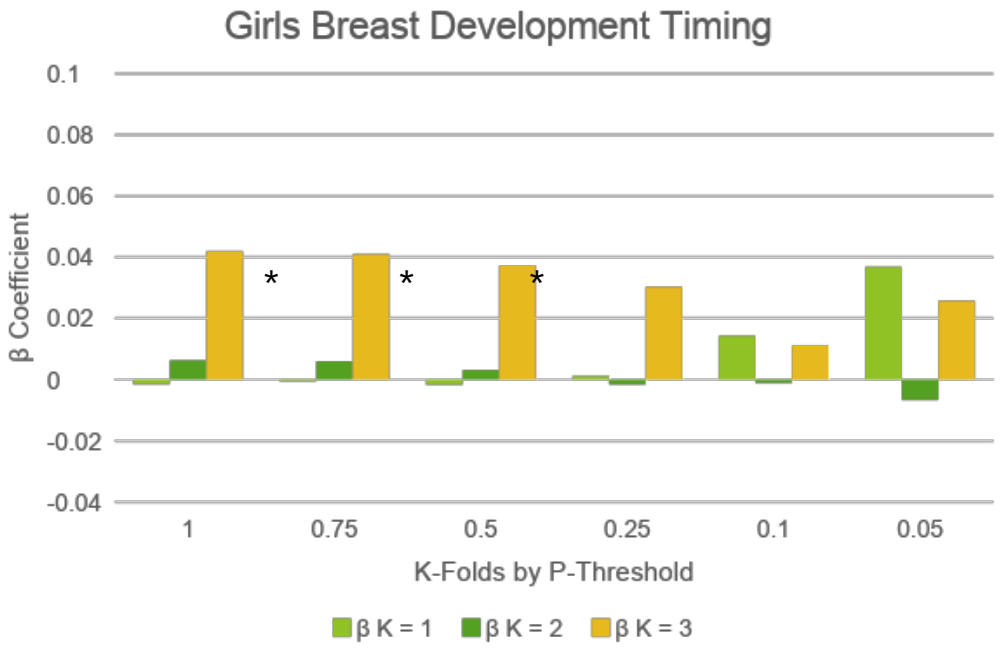
3. Multiply test matrix by SNP coefficients → polygenic scores
as in the Empirical scoring procedure

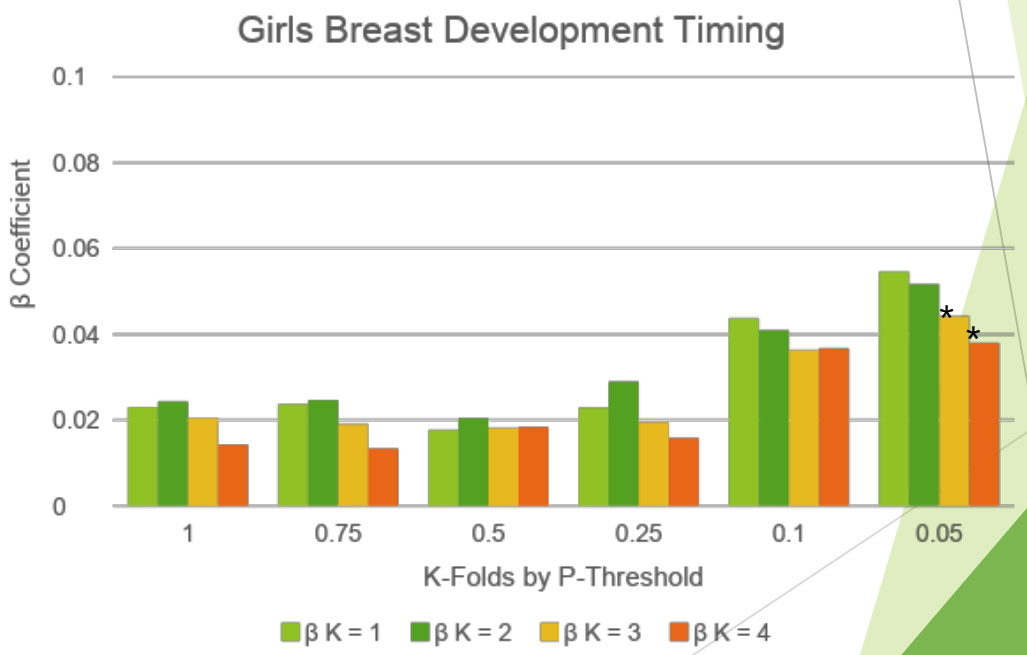
Test Set

4. Run hypothesis testing regressions (same as above)

Results: K-fold

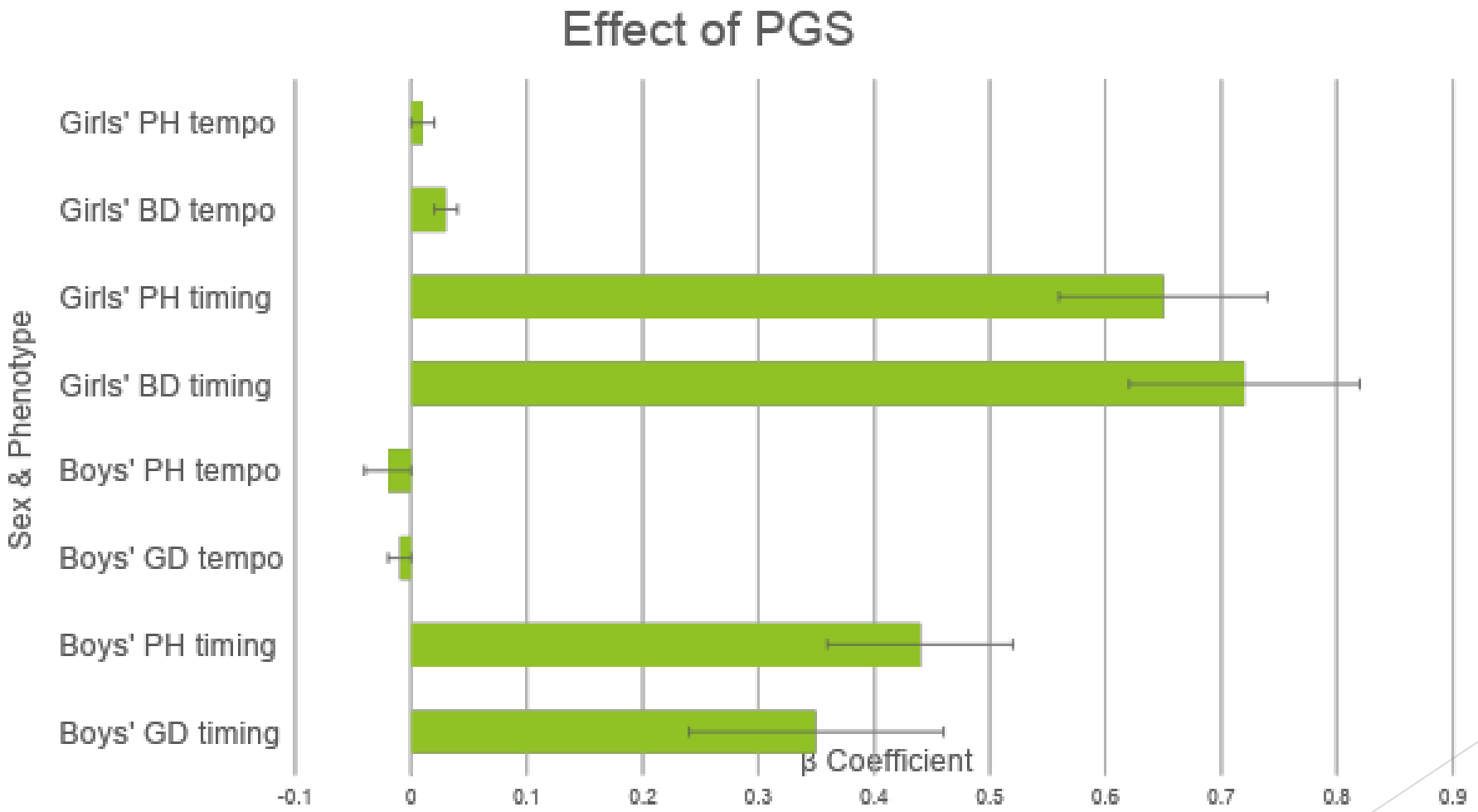
- K-fold produced no internally replicated effects of PGS on puberty phenotypes





* p< .05

Results: Empirical



Conclusions, Question 1: Empirical Scoring vs. K-Fold, theoretical

- Empirical version of PGS provided more reliably significant results in the prediction of pubertal timing
  - For boys and girls
    - More highly predictive for girls
- K-Fold likely lacking in power
- Neither method reliably predicted tempo

Future Measures & Implications

Implications

- The success of empirical PGS may allow us to look more closely at the biological underpinnings of puberty in smaller samples with genetic data
- Reliable predictor of puberty for girls *and* boys

Future Directions

- Replication
- Alternative PGS methods
  - Use derived weights from previous literature for theory-based score instead of k-fold
- Empirical scores for other puberty phenotypes

Acknowledgements

Biobehavioral Development Laboratory

Kristine Marceau
Valerie Knopik

Funding:

- 2017-18 Ross Fellowship (Horvath)
- K01 DA039288 (Marceau)

Thank you!

Empirical Strategy

- Scores created from SNPs identified in a recent meta-analysis (Day et al., 2017)
- SNPs were also measured or imputed in ALSPAC
- SNPs survived quality control & LD pruning (as K-fold)

	ß from Day et al., 2017	Person 1 # effect alleles	Person 1 SNP score	Person 2 # effect alleles	Person 2 SNP score
SNP1	.5	0	0	1	.5
SNP2	.1	2	.2	1	.1
Polygenic score			0.2		0.6