Leveraging the use of PRS in longitudinal studies: Example based on the ALSPAC Cohort

Laurent Briollais
Lunenfeld-Tanenbaum Research Institute, Sinai Health System
Dalla Lana School of Public Health, University of Toronto
Toronto, CANADA

Joint work with:

Yanyan Wu
Office of Public Health Studies
U. of Hawai'i at Manoa, USA

Oct 12, 2021

BMI-related PRS

- PRS approaches have been successful in predicting overweight and obesity risks
  - Locke et al. (Nature, 2015) identifies 97 BMI-associated loci ($p < 5 \times 10^{-8}$) from $\sim$ 340K individuals from the GIANT consortium, which accounts for $\sim$ 2.7% of BMI variation in adults
  - Yengo et al. (Hum. Mol. Genet., 2018) identifies 941 height- and BMI-associated loci ($p < 1 \times 10^{-8}$) from $\sim$ 700K individuals from the (GIANT, UK Biobank), which accounts for $\sim$ 6.0% of BMI variation in adults
  - Khera et al. (Cell, 2019) using a PRS based on 2.1 million genetic variants from the GIANT consortium to predict the risk of obesity and other metabolic traits in the UK Biobank and other studies

Clinical importance of PRS

- Khera et al. (Cell, 2019) stressed the greatly increased risk of severe obesity among individuals in the top decile of the PRS
- For instance, 15.6% of individuals in the top decile of PRS went on to develop severe obesity compared with 5.6% of those in deciles 2-9 and 1.3% in the lowest decile
- The PRS in Khera et al. explains 23.4% of BMI heritability vs. 6.0% in Yengo et al. in adults

Questions of interest in longitudinal studies

1. How to estimate and test the effect of PRS on child BMI at different ages?
2. Can we mitigate the PRS-risk with environmental factors, e.g. breastfeeding?
3. Can we dynamically predict BMI given PRS and history of environmental factor(s)? $\Rightarrow$ Personalized medicine/prevention

Background papers

For the 2 first questions:

- YY. Wu, S. Lye, C-L. Dennis, L. Briollais. Exclusive breastfeeding can attenuate body-mass-index increase among genetically susceptible children: A longitudinal study from the ALSPAC cohort. PLoS Genet. 2020;16(6):e1008790.

For the dynamic predictions:

- D. Rizopoulos. Joint models for longitudinal and time-to-event data with applications in R. Chapman & Hall/CRC Biostatistics Series. 2012.

Child cohort

- ALSPAC = the Avon Longitudinal Study of Parents and Children
- Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study
- The core ALSPAC sample consists of 15,083 pregnancies
- 2,690 boys and 2,576 girls for our analyses

Construction of our BMI-related PRS

- 69 SNPs associated with BMI at genome-wide significance in GIANT consortium + 25 independent non-overlapping SNPs that we previously studied in relation to child BMI
- The sex-specific PRSs were created using the imputed dosages for the 94 SNPs
- Weights obtained from the GIANT consortium and UK Biobank meta-analysis
- PRS $= \sum_{i=j}^{94} \hat{\beta}_j G_j$
- Rescaled from 0 to 10 $\Rightarrow$ 1-unit = 1-decile

Modelling of BMI growth curves

- Separate models for boys and girls fitted from birth to 18 years of age
- Linear mixed effects model with cubic splines (R package *spida*)
- Let $Y_i(t)$ denote the longitudinal BMI measurement for the ith child at age t:

$$Y_i(t) = X_i^T\beta + sp(t).[W_i^T\beta_{w.sp(t)} + PRS_i^T\beta_{PRS.sp(t)}] + Z_i^T(t)b_i + \epsilon_i(t)$$

with $sp(t)$ is a cubic spline function of age t, X_i and W_i are confounding variables assumed time-invariant and time-varying, respectively.

Questions of interest

1. How to estimate and test the effect of PRS on child BMI at different ages?
2. Can we mitigate the PRS-risk with environmental factors, e.g. breastfeeding?
3. Can we dynamically predict BMI given PRS and history of environmental factor(s)?

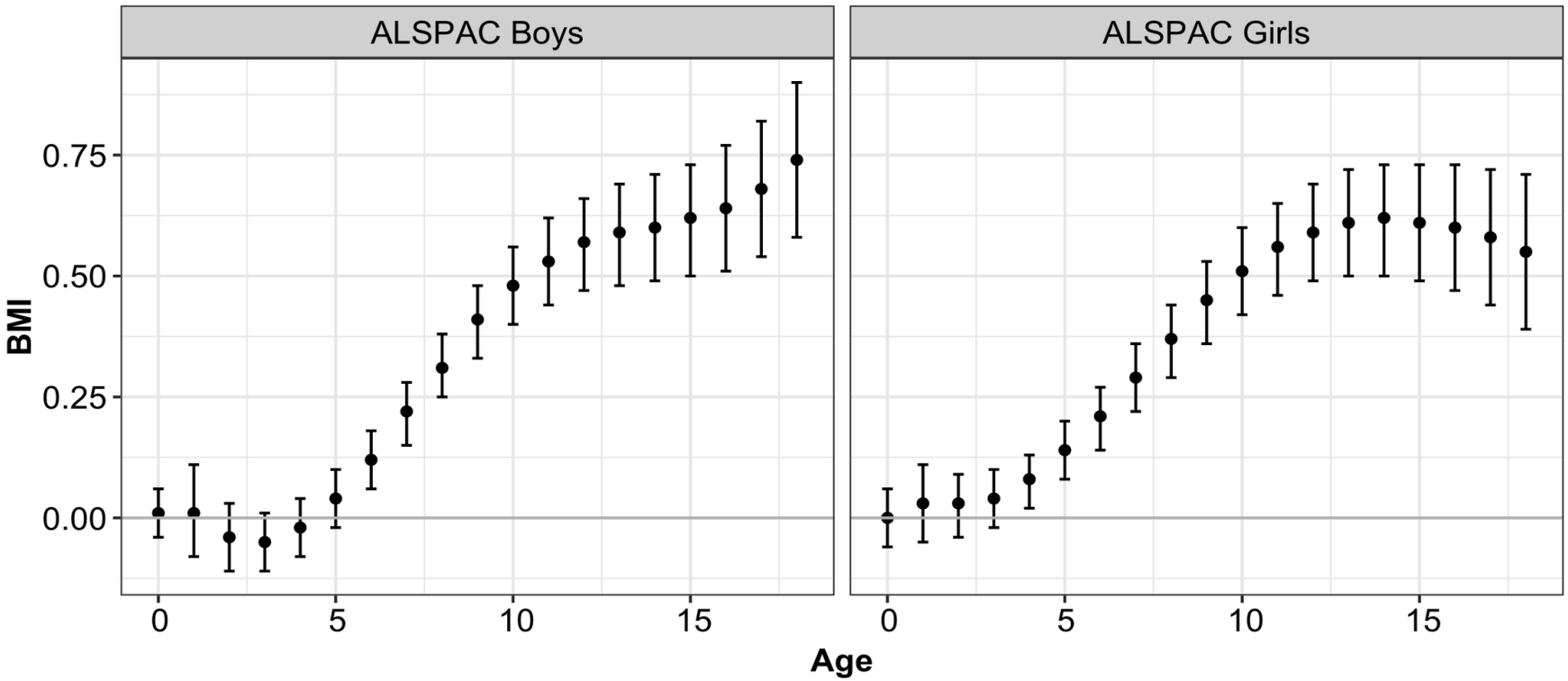
Estimate and test of PRS effect on longitudinal child BMI

- Use of the General Linear Hypothesis (GLH) approach (McDonald, 1975) for hypothesis testing

- $H_0 : L\beta = m$ and $L\hat{\beta} \sim N(L\beta, Lcov(\hat{\beta})L^T)$
  $\Rightarrow$ p-value and the 95% confidence interval can be obtained.

- Example: We want to estimate and test an 2.5-unit increase in PRS at age 10 in boys.
- $L\hat{\beta} = L_2^T\hat{\beta} - L_1^T\hat{\beta}$
  - $L_2^T\hat{\beta}$ is the estimated mean BMI assuming sex=boys, age=10 and PRS=7.5
  - $L_1^T\hat{\beta}$ is obtained assuming sex=boys, age=10 and PRS=5.0
  - Holding all other covariates constant.

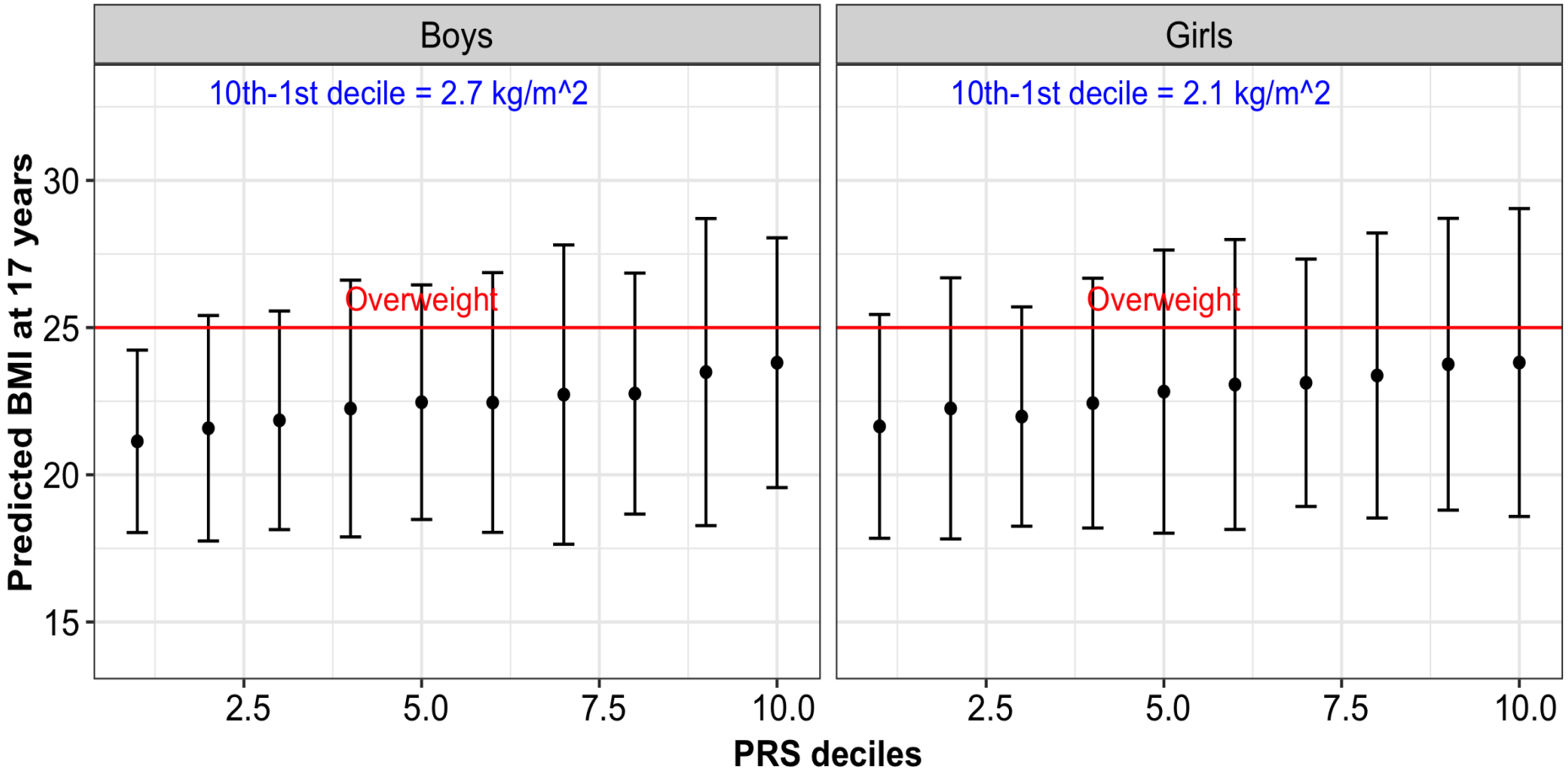
Effect size of PRS on longitudinal child BMI: Example of an 2.5-unit (=quartile) increase



Test of a 2.5 unit increase in PRS on longitudinal BMI

	Boys		Girls	
Age	Effect	*p*-value	Effect	*p*-value
0	0.01	0.79	0.00	0.91
1	0.01	0.78	0.03	0.48
2	-0.04	0.29	0.03	0.41
3	-0.05	0.11	0.04	0.17
4	-0.02	0.48	0.08	0.008
5	0.04	0.22	0.14	< 0.0001
6	0.12	0.0001	0.21	< 0.0001
7	0.22	< 0.0001	0.29	< 0.0001
8	0.31	< 0.0001	0.37	< 0.0001
9	0.41	< 0.0001	0.45	< 0.0001
10	0.48	< 0.0001	0.51	< 0.0001
11	0.53	< 0.0001	0.56	< 0.0001
12	0.57	< 0.0001	0.59	< 0.0001
13	0.59	< 0.0001	0.61	< 0.0001
14	0.60	< 0.0001	0.62	< 0.0001
15	0.62	< 0.0001	0.61	< 0.0001
16	0.64	< 0.0001	0.60	< 0.0001
17	0.68	< 0.0001	0.58	< 0.0001
18	0.74	< 0.0001	0.55	< 0.0001

Predicted BMI vs. PRS deciles



Questions of interest

1. How to estimate and test the effect of PRS on child BMI at different ages?
2. Can we mitigate the PRS-risk with environmental factors, e.g. breastfeeding?
3. Can we dynamically predict BMI given PRS and history of environmental factor(s)?

Modelling of BMI growth curves

- Model with breastfeeding: duration of exclusive breastfeeding (EBF) or duration of non-exclusive breastfeeding (BF)

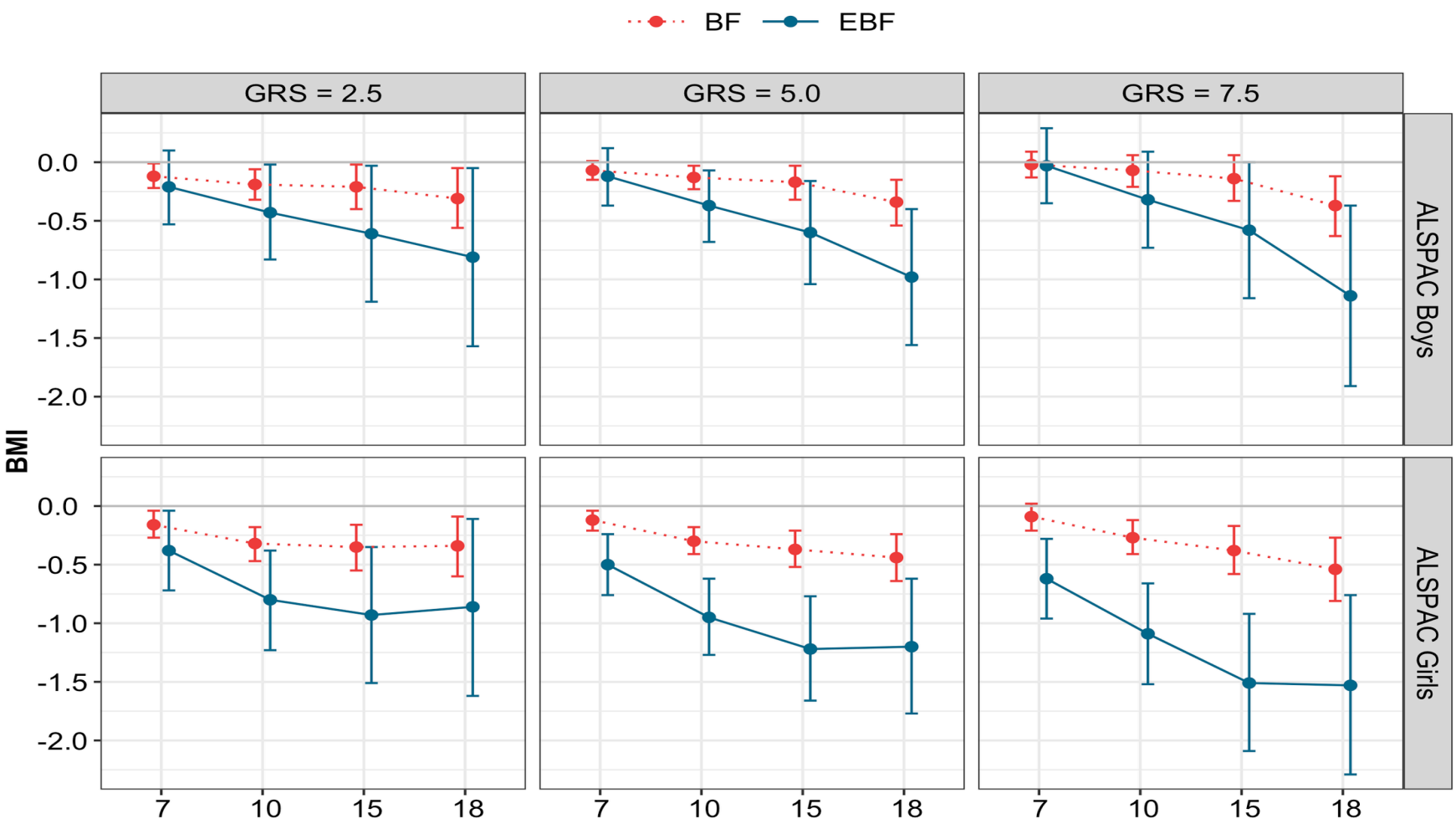
- Let $Y_i(t)$ denote the longitudinal BMI measurement for the ith child at age t:

$$Y_i(t) = X_i^T \beta + sp(t).[W_i^T \beta_{w.sp(t)} + PRS_i^T \beta_{PRS.sp(t)} +$$

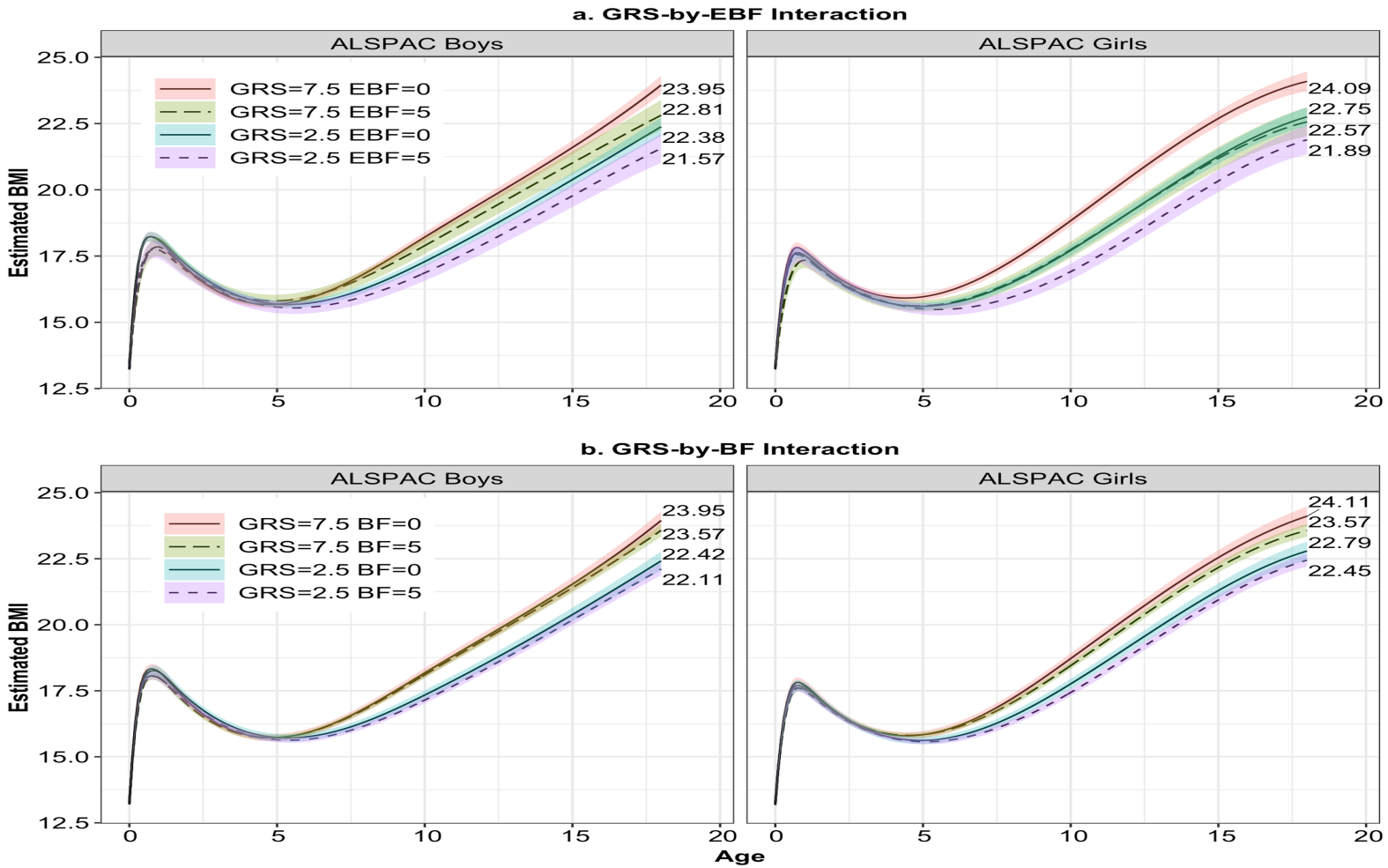
$$+ PRS_i : EBF_i^T \beta_{PRS:EBF.sp(t)}] + Z_i^T(t) b_i + \epsilon_i(t)$$

with EBF has time-varying effect through the interaction with the spline BUT no effect at baseline.

Risk-reduction effect of 5 months of EBF and non-exclusive BF on BMI



Predicted BMI growth trajectories for ALSPAC boys and girls given PRS and EBF



Other results

- Both the duration and exclusivity of breastfeeding are important for reducing PRS-related risk of BMI/obesity

Questions of interest

1. How to estimate and test the effect of PRS on child BMI at different ages?
2. Can we mitigate the PRS-risk with environmental factors, e.g. breastfeeding?
3. Can we dynamically predict BMI given PRS and history of environmental factor(s)?

Dynamic predictions of a longitudinal outcome

- The predicted BMI longitudinal profile can help determine the children most at risk given his/her PRS

- We are interested in the expected value of a child longitudinal outcome at time $u > t$ given his/her observed responses up to that time point $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$, i.e.,

$$w_i(u|t) = E(y_i(u)|T_i^* > t, \mathcal{Y}_i(t), PRS_i; \theta*\}, u > t.$$

- These predictions are dynamically updated, e.g., using $w_i(u|t')$ that uses additional longitudinal information up to this latter time point

Dynamic predictions of a longitudinal outcome

- To compute these predictions, because we do no know the true parameter θ^*, a Bayesian formulation is applied to calculate the expectation of $w_i(u|t)$ with respect to the posterior distribution of the parameters $\{\theta\}$ as

$$E\{y_i(u)|T_i^* > t, \mathcal{Y}_i(t)\} = \int E\{y_i(u)|T_i^* > t, \mathcal{Y}_i(t); \theta\} p(\theta) d\theta.$$

The first part of the integrand is obtained from the LME

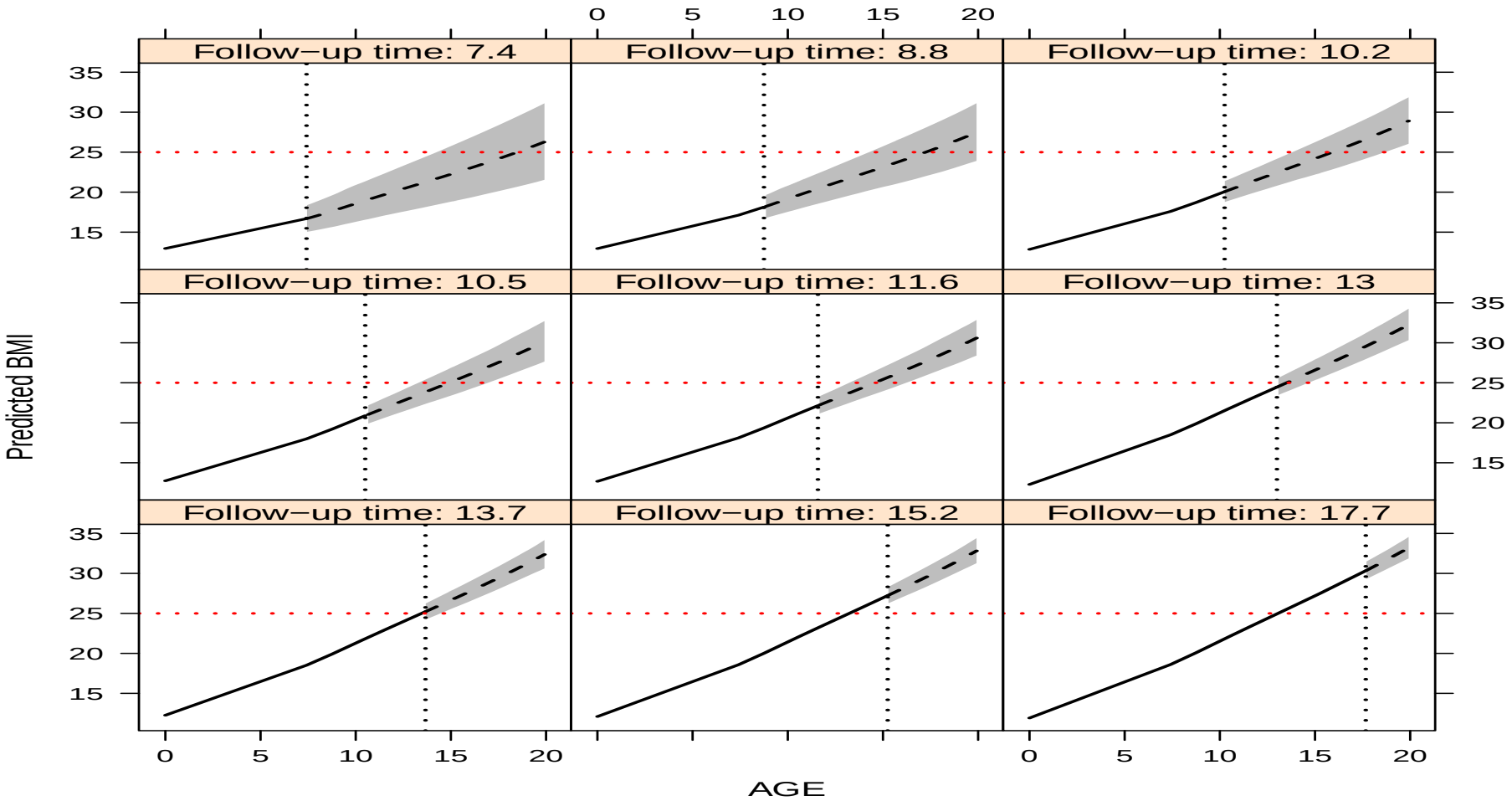
$$E\{y_i(u)|T_i^* > t, \mathcal{Y}_i(t); \theta\} = x_i^T(u)\beta + z_i^T(u)\bar{b}_i^{(t)},$$

But replacing the random effects by

$$\bar{b}_i^{(t)} = \int b_i p(b_i|T_i^* > t, \mathcal{Y}_i(t); \theta) db_i.$$

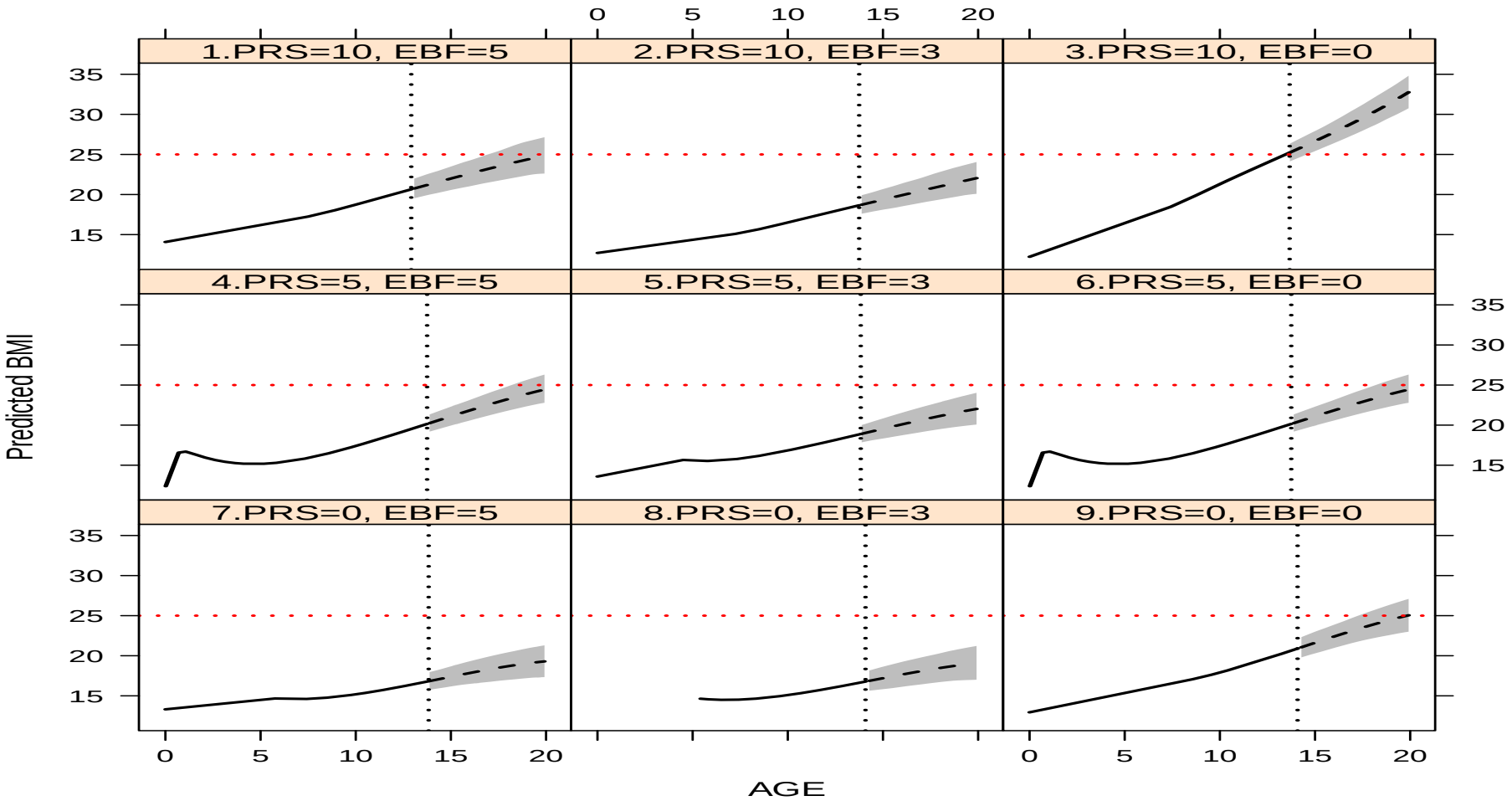
Examples of dynamic predictions of child BMI

Figure: ALSPAC boy in the the 10th decile of PRS



Examples of dynamic predictions of child BMI

Figure: Prediction profiles of 9 ALSPAC boys conditional on PRS and EBF



Conclusion/Discussion

- Longitudinal studies allow to better understand the time-dependent effect of PRS
- However, SNPs included in the PRS come from cross-sectional studies
- Additional efforts are needed to discover genetic variants that have time-dependent effects through large international cohort studies that include various age groups

Thanks!