

Supplementary Material

A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a European cohort of children and adolescents

Ronja Foraita¹, Janine Witte^{1,2}, Claudia Börnhorst¹, Wencke Gwozdz^{3,4}, Valeria Pala⁵, Lauren Lissner⁶,
Fabio Lauria⁷, Lucia Reisch^{4,8}, Dénes Molnár⁹, Stefaan De Henauw¹⁰, Luis Moreno¹¹, Toomas
Veidebaum¹², Michael Tornaritis¹³, Iris Pigeot^{1,2}, Vanessa Didelez^{1,2}

Methods

Study Population

The IDEFICS/I.Family cohort ^{1,2} is a European cohort study initiated with the overall aims to identify and prevent dietary and lifestyle induced health effects in infants, children and adolescents. The baseline survey (B) was conducted from September 2007 to May 2008 in eight European countries (Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain and Sweden) with 16,229 participating children (2 to 9.9 years old). Two years (± 1 month) later, 13,596 children were included in the first follow-up examinations (FU1, from September 2009 to June 2010). The second follow-up examination (FU2) was conducted from January 2013 to June 2014, in which 7,105 children participated who already participated at B or FU1. The examinations covered a spectrum of parameters following a detailed and standardised study protocol. Parents filled in all questionnaires during B, FU1 and in FU2 if their child was less than 12 years old. Teens aged 12 years or more reported for themselves in FU2. Ethical approval was obtained from the responsible ethics committees in each country. All children and their parents provided oral and written informed consent, respectively, before examinations and/or the collection of samples, subsequent analysis and storage of personal data and collected samples. Teens older than 12 years were asked to provide their written consent using a simplified version of the consent form. Study subjects and their parents could opt out of each single study component. Information on early life factors was obtained from records of routine visits as well as from parental reports. Pregnancy-related questions were posed to biological mothers only.

Information on consumption frequencies was obtained from a food frequency questionnaire (FFQ) with 44 food items (B, FU1) and 59 food items (FU2) from 14 food groups ³. Daily family meals was further included as binary indicator for healthy food choices and family cohesion ⁴. The variables sleep duration and well-being were included as they can indicate a child's stress levels. Total sleep in hours was assessed by self-reports in FU1 and FU2. The average nocturnal sleep (hours/night) was calculated as the weighted average of reported usual weekday and weekend sleeping times. At baseline, total sleep was derived based on 24-h dietary recall data where the parents were asked 'What time did your child go to bed?' and 'What

time did your child get up??. The homeostatic model assessment (HOMA-IR, short HOMA) index ⁵ served as a marker for insulin resistance. Different stages of pubertal status of teens were estimated in FU2 by a self-administered questionnaire based on the development of voice (boys) and menarche (girls) ^{6,7}.

Unhealthy substance use at FU2 was measured by ever alcohol drinking and ever tobacco smoking. The family's socioeconomic position was assessed by educational level using the International Standard Classification of Education (ISCED) ⁸ and household income (net income after taxes and deductions), which were harmonised between countries ⁹. See Ahrens, Bammann ¹, Ahrens, Siani ² for more details.

Graph characteristics

Adjacent, neighbour, path: Two nodes are said to be adjacent, if they are connected by an edge. Such nodes are also referred to as neighbours. A path is an alternating sequence of distinct adjacent nodes and edges as for example $v_0 - v_1 \rightarrow \dots - v_l$. A directed path proceeds from v_0 to v_l along directed edges that point all into the same direction.

Distance, diameter and average path length: The distance between nodes in a graph is commonly defined as the length of the shortest path(s) between these nodes. The distance is infinity when no path exists. The diameter of a graph is defined as the longest path in the graph, which is the maximal distance of any pair of nodes. The average path length of a graph is defined as the average distance between all pairs of nodes. The path length is an indicator for the connectivity of the graph.

Hamming distance and structural Hamming distance: The Hamming distance ¹⁰ outputs the minimum number of edge insertions or deletions that are necessary to transform one graph into another graph, where edge directions are not taken into account. Whereas the structural Hamming distance ¹¹ considers edge directions and additionally counts the number of required edge flips that are necessary for a full transformation.

Root mean squared edge uncertainty (RMSEU): The root mean squared edge uncertainty (RMSEU) is a descriptive measure to assess the uncertainty of an undirected graphical model ¹². It reduces the multidimensionality of a graphical model by making use of the edge frequencies of graphs selected from

multiple datasets with the same variables. Calculation of this measure was based on 100 bootstrap replications. The bootstrap graph BOOT35 includes edges that were selected in more than 35% of the respective bootstrap graphs, i.e. edges with a frequency around 0.35% have the highest uncertainty to be selected into the graph or not. An RMSEU of 26.7% (cf. Table 3), for instance, can be obtained for an arbitrary graph with 51 nodes where each edge was either selected in 82.6% or 9.4% of the reruns.

Background on causal graphs and causal inference

Causal effects: A causal effect is defined as the effect of a hypothetical intervention, $\text{do}(X=x)$, on an exposure X , setting it to x versus x' , on the distribution of an outcome Y . In the present paper we use the difference in expectation of the outcome as the effect contrast, $E(Y | \text{do}(X=x)) - E(Y | \text{do}(X=x'))$. In a causal linear main effects model, x and x' are taken to be one unit apart and the average effect is assumed constant, and equals the slope (no effect modification by included covariates).

Causal response curve: When the exposure is not binary, or when linearity is not appropriate we may want to compare the expected value of a (continuous) outcome Y across different interventional values of X , i.e. we want to estimate $E(Y | \text{do}(X=x))$ as a function of the continuous or multi-valued X . We denote the function $E(Y | \text{do}(X=x))$ as causal response curve (aka 'expected outcome under hypothetical interventions (EOHI)').

Key structural assumptions: The estimation of causal effects or causal response curves typically relies on the following key assumptions: (1) causal consistency, meaning that an intervention on the exposure must be well defined such that what we actually observed would have been observed if the exposure value had been set to its value by the intervention; (2) there is no unobserved confounding so that the covariates included in the analysis are sufficient to adjust for any confounding; (3) positivity, meaning that each individual could have in principle been subject to any other exposure value (within the range being compared) by a corresponding intervention.

Causal DAG: A causal directed acyclic graph (causal DAG) consists of nodes representing the variables and directed edges representing direct causal relations; it has no cycles. More precisely, it is the absence of

edges that imply the absence of direct causal effects, and consequently the absence of any directed path from X to Y implies the absence of a total (overall) causal effect of X on Y . Probabilistically, a causal DAG implies conditional independencies between variables (the causal Markov properties) which can be read off using d-separation¹³.

CPDAG: A completed partially directed acyclic graph represents the equivalence class of DAGs, i.e. the set of DAGs that encode the same conditional independencies but not necessarily the same causal relations. For instance $X \rightarrow Y \rightarrow Z$ implies the same conditional independence as $X \leftarrow Y \leftarrow Z$, even though the causal meaning is very different; the corresponding CPDAG is $X - Y - Z$. An undirected edge in a CPDAG means that the equivalence class contains at least one DAGs where the edge is directed in one direction and at least one other DAGs where it is directed in the reverse direction. Causal discovery methods that solely rely on conditional independencies found in the data cannot distinguish between different DAGs contained in a CPDAG, i.e. without any external information (or randomization or parametric assumptions) we cannot distinguish $X \rightarrow Y \rightarrow Z$ from $X \leftarrow Y \leftarrow Z$; thus the only information that the CPDAG $X - Y - Z$ carries is that there is no direct causal relation between X and Z and that $X \rightarrow Y \leftarrow Z$ can be excluded. The DAGs contained in a CPDAG can be obtained by finding all possible edge orientations for the undirected edges such that no cycles and no new V-structures ($X \rightarrow Y \leftarrow Z$) are created. Asymptotically, the PC-algorithm outputs a CPDAG, but for finite-samples this cannot be guaranteed.

MPDAG: A maximally oriented partially directed acyclic graph is a subset of an equivalence class of DAGs, i.e. a subset of the DAGs contained by a CPDAG; the subset is obtained by adding background knowledge on absence or presence of edges to the conditional independencies. For example, if we knew that X is in time before Z (in addition to X and Z being conditionally independent given Y) then we obtain that either $X \leftarrow Y \rightarrow Z$ or $X \rightarrow Y \rightarrow Z$ must hold, which are summarized in the MPDAG $X - Y \rightarrow Z$.

Faithfulness: The assumption that every conditional independence in the data corresponds to the absence of some edge (and thus to some d-separation) in the underlying causal DAG is known as faithfulness. It can be

violated if, for instance, a positive and a negative effect along different pathways cancel out each other exactly.

Causal sufficiency: The assumption that the observed variables can be represented in a causal DAG without additional latent variables (nodes) being common causes of two or more observed nodes is known as causal sufficiency. This is a strong assumption, but while approaches exist to relax causal sufficiency, these are more time-consuming, much more difficult to interpret and have not yet been generalized for time-ordered data. The output of the PC-algorithm can still be interpreted in terms of conditional independencies even in the absence of causal sufficiency. The absence of edges can then still be interpreted as the absence of direct causal relations (under the assumption of faithfulness).

PC-algorithm: The PC-algorithm is named after Peter Spirtes and Clark Glymour; it proceeds by determining conditional independencies in the data and then finding a CPDAG that is compatible with the independencies. With perfect conditional independence information, the PC-algorithm is valid (sound and complete) under the assumptions of faithfulness and causal sufficiency. Under additional assumptions on the underlying data generating mechanism it is also consistent¹⁴⁻¹⁶

tPC: The tPC-algorithm^{17, 18} is a variant of the PC-algorithm and uses additional prior-knowledge on a partial temporal ordering of the variables (nodes) to exclude certain edge directions. Its output is an MPDAG (which cannot be guaranteed for finite samples). A recent tutorial¹⁹ describes how to apply tPC for causal discovery on cohort data with missing data.

MICD: Multiple imputation causal discovery is described in Foraita, Friemel²⁰ and Witte, Foraita¹⁸ and provided as R-package²¹. It proceeds by first creating multiply imputed data sets. Conditional independence tests are then performed separately on and then pooled across the multiple datasets. The resulting test decisions are entered into the (t)PC-algorithm resulting in an MPDAG.

Multiset of causal effects: An equivalence class of DAGs or MPDAG contains possibly many different causal DAGs. For example, $X - Z \rightarrow Y$ contains (i) $X \rightarrow Z \rightarrow Y$ and (ii) $X \leftarrow Z \rightarrow Y$; in (i) the effect of X on Y

could be non-zero, in (ii) it would be zero. Also, in (i) we would not adjust for Z when estimating the effect of X on Y as it is a mediator, while in (ii) we should adjust for Z as it is a confounder of the $X - Y$ relation (this would only be relevant if there are further paths from X to Y in the graph). This example shows that based on an MPDAG we may find different causal effects and different adjustment sets for the same exposure-outcome pair as we need to allow for the different causal DAGs that cannot be distinguished. Thus, instead of estimating a single causal effect, we estimate a multiset of causal effects, one value for each DAG contained in the MPDAG.

IDA: The **I**ntervention calculus when the **D**AG is **A**bsent algorithm provides a general principle to identify from an MPDAG the different adjustment sets for a given exposure-outcome pair. These can be characterized graphically^{22, 23}. We selected for the analysis local adjustment sets, which tend to be more robust to possible errors in the graph structure, and optimal adjustment sets, which optimise precision of effect estimates given that the graph was correctly recovered. The original IDA then proceeds by assuming a linear main effects causal model and uses a linear regression of the outcome on exposure and adjustment set to estimate the causal effect. However, this can be generalized to the non-linear case²⁴ as the graphical argument underlying the adjustment set does not require linearity.

Double robustness (DR): Doubly-robust estimation of causal effects combines a weighting approach using the estimated propensity score with adjustment for confounding in an outcome regression (augmented inverse probability weighting). The double robustness property means that the estimator is consistent if either the model for the propensity score or the outcome model is correctly specified.

Double machine learning: The principle of DR estimation can be extended so as to employ a machine learning approach to select the propensity score model and the outcome model nonparametrically. In the main paper, for continuous exposures, we apply the kernel smoothing approach of Kennedy et al.²⁵, where an ensemble learner combining the Lasso, random forests, generalized linear models, and weighted mean prediction is used to determine the weights.

References

1. Ahrens W, Bammann K, Siani A, et al. The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes* 2011; **35**: S3-S15.
2. Ahrens W, Siani A, Adan R, et al. Cohort Profile: The transition from childhood to adolescence in European children—how I.Family extends the IDEFICS cohort. *International Journal of Epidemiology* 2017; **46**: 1394-5j.
3. Hebestreit A, Börnhorst C, Barba G, et al. Associations between energy intake, daily food intake and energy density of foods and BMI z-score in 2–9-year-old European children. *Eur J Nutr* 2014; **53**: 673-81.
4. Fruh SM, Fulkerson JA, Mulekar MS, Kendrick LAJ, Clanton C. The surprising benefits of the family meal. *The Journal for Nurse Practitioners* 2011; **7**: 18-22.
5. Peplies J, Jiménez-Pavón D, Savva SC, et al. Percentiles of fasting serum insulin, glucose, HbA1c and HOMA-IR in pre-pubertal normal weight European children from the IDEFICS cohort. *Int J Obes* 2014; **38**: S39-S47.
6. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child* 1969; **44**: 291-303.
7. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child* 1970; **45**: 13-23.
8. UNESCO Institute for Statistics (UIS). *International Standard Classification of Education: ISCED 2011*. Montreal, Quebec: UIS; 2012.
9. Bammann K, Gwozdz W, Lanfer A, et al. Socioeconomic factors and childhood overweight in Europe: results from the multi-centre IDEFICS study. *Pediatr Obes* 2013; **8**: 1-12.
10. Hamming RW. Error Detecting and Error Correcting Codes. *Bell System Technical Journal* 1950; **29**: 147-60.
11. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning* 2006; **65**: 31-78.
12. Pigeot I, Sobotka F, Kreiner S, Foraita R. The uncertainty of a selected graphical model. *J Appl Stat* 2015; **42**: 2335-52.
13. Didelez V. Causal concepts and graphical models In: Maathuis MH, Drton M, Lauritzen S, Wainwright M, editors. *Handbook of Graphical Models*. Boca Raton: CRC Press; 2018.
14. Kalisch M, Bühlmann P. Estimating high-dimensional directed acyclic graphs with the PC-Algorithm. *Journal of Machine Learning Research* 2007; **8**: 613–36.
15. Robins JM, Scheines R, Spirtes P, Wasserman L. Uniform consistency in causal inference. *Biometrika* 2003; **90**: 491-515.
16. Zhang J, Spirtes PL. *Strong faithfulness and uniform consistency in causal inference*. In: Kjarulff U, Meek C, editors. Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence (UAI-03); 2012: Morgan Kaufmann Publishers Inc.; 2012. p. 632–9.
17. Witte J. *tPC - Causal discovery with temporal background knowledge*; Version: 1.0.0. R-package; 2021. <https://github.com/bips-hb/tpc>
18. Witte J, Foraita R, Didelez V. Multiple imputation and test-wise deletion for causal discovery with incomplete cohort data. 2021. <https://arxiv.org/abs/2108.13331>
19. Andrews RM, Foraita R, Didelez V, Witte J. A practical guide to causal discovery with cohort data. 2021. <https://arxiv.org/abs/2108.13395>
20. Foraita R, Friemel J, Günther K, et al. Causal discovery of gene regulation with incomplete data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2020; **183**: 1747-75.
21. Foraita R, Witte J. *Multiple imputation in causal graph discovery*; Version: 1.2.1. R-Package; 2021. <https://github.com/bips-hb/micd>
22. Maathuis MH, Kalisch M, Bühlmann P. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics* 2009; **37**: 3133-64.
23. Witte J, Henckel L, Maathuis MH, Didelez V. On efficient adjustment in causal graphs. *Journal of Machine Learning Research* 2020; **21**: 1-45.
24. Rotnitzky A, Snucler E. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *Journal of Machine Learning Research* 2020; **21**: 1-86.
25. Kennedy EH, Ma Z, McHugh MD, Small DS. Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2017; **79**: 1229-45.

Results

Table S1: Characteristics of children in the IDEFICS/I.Family cohort participating in all three surveys from 2007 to 2014 in 10 datasets imputed by chained equations using random forest

Time-invariant variables	N = 51,120 ^l		
Region			
<i>Central (Belgium, Germany, Hungary)</i>	27 %		
<i>North (Estonia, Sweden)</i>	29 %		
<i>South (Cyprus, Italy, Spain)</i>	44 %		
Female	49 %		
Migration background	6.5 %		
Completed weeks of pregnancy	39.24 (1.60)		
Tobacco smoking during pregnancy			
<i>never</i>	89 %		
<i>rarely</i>	3.4 %		
<i>several occasions a week</i>	3.0 %		
<i>daily</i>	4.5 %		
Mother's age at birth	29.8 (4.9)		
Birthweight [g]	3,344 (571)		
Total breastfeeding [months]	6.7 (6.3)		
Was fed with formula milk	52 %		
Fully integrated into household's diet [month]	14.4 (6.3)		
Time-varying variables	Baseline N = 51,120 ^l	FU1 N = 51,120 ^l	FU2 N = 51,120 ^l
Age	5.89 (1.78)	7.87 (1.79)	11.69 (1.81)
School			
<i>kindergarten</i>	52 %	23 %	-
<i>school</i>	47 %	77 %	-
<i>neither</i>	0.8 %	0.2 %	-
BMI z-score	0.32 (1.17)	0.43 (1.17)	0.55 (1.11)
Well-being [%]	84 (9)	82 (10)	82 (10)
Audiovisual media consumption [h/day]	1.57 (0.87)	1.89 (0.92)	2.93 (1.78)
Physical activity [h/week]	18 (10)	18 (10)	17 (9)
Total sleep [h/day]	10.20 (0.89)	10.00 (0.86)	9.28 (1.00)
Youth healthy eating index [%]	63 (11)	63 (11)	57 (10)
Daily family meals	75 %	78 %	69 %
Homa index z-score	0.05 (1.07)	0.42 (0.94)	0.09 (1.19)
Pubertal	-	-	41 %
Ever alcohol drinking	-	-	26 %
Ever tobacco smoking	-	-	6.2 %
Mother's BMI	23.7 (4.2)	24.0 (4.2)	24.9 (4.9)
Household's income			
<i>low</i>	37 %	31 %	28 %
<i>middle</i>	26 %	25 %	36 %
<i>high</i>	37 %	43 %	36 %
ISCED			

<i>low</i>	5.0 %	4.6 %	4.9 %
<i>middle</i>	42 %	42 %	42 %
<i>high</i>	53 %	54 %	53 %

¹ %; Mean (SD)

Table S2: Characteristics of the discovered graph without singletons

Characteristics	Main	MI-0.1	TWD	EM	BG35	BG75
Longest path	9 ¹	8 ²	7 ³	7 ⁴	10 ⁵	5 ⁶
Avg. path length	2.8	2.7	2.4	2.4	2.7	1.5
Number of selected edges	104	113	139	150	119	44
Number of undirected edges	12	13	14	0	9	0
Max. number of neighbors	8 ⁷	9 ⁸	13 ⁹	25 ¹⁰	11 ¹¹	5 ¹²
Avg. number of neighbors	4.3	4.4	5.5	5.9	4.7	1.7
Hamming distance ¹⁰	-	19	205	114	67	68
Structural Hamming distance ¹¹	-	34	214	125	82	87
Root mean squared edge uncertainty ¹²	-	-	-	-	19.4	4.5

¹Age (FU1) > School (FU1) > Daily family meals (FU1) > YHEI (FU1) > Well-being (FU1) > AVM (FU1) > HOMA (FU1) > HOMA (FU2) > zBMI (FU2)

²Age (FU1) > School (FU1) > Daily family meals (FU1) > YHEI (FU1) > AVM (FU1) > HOMA (FU1) > HOMA (FU2) > zBMI (FU2)

³Weeks of pregnancy > Daily family meals > Sleep > AVM > Well-being > HOMA > zBMI (FU2)

⁴Smoking during pregnancy > Weeks of pregnancy > Birthweight > Mother`s BMI > AVM > Well-being > Well-being (FU1)

⁵Weeks of pregnancy > Birthweight > Mother`s BMI > Daily family meals > YHEI > Well-being > Well-being (FU1) > Well-being (FU2) > Smoking (FU2) > Alcohol (FU2)

⁶Age > School > Well-being > Well-being (FU1) > Well-being (FU2)

⁷Region, AVM (FU1), Well-being ⁸AVM (FU1), Well-being ⁹Migrant ¹⁰Region ¹¹School ¹²School, Sex

BG35: Summarized bootstrap graph with edges that occurred at least 35 times in 100 bootstrap replications; EM: structural EM algorithm; Main: multiple imputation with nominal level of 0.05; MI-0.1: multiple imputation with nominal level of 0.1; TWD: test-wise deletion

Table S3: Selected edges in the main graph and their selection frequency in the bootstrap graphs.

Edges			Edges		
from	to	%	from	to	%
Region	Income	100.0	Sleep.FU1	Sleep.FU2	84.0
Region	ISCED	100.0	Sex	Birthweight	82.0
Migrant	Income	100.0	Birthweight	Mothers_BMI	78.0
HOMA	HOMA.FU1	100.0	School	Well-being	78.0
Age.FU1	School.FU1	100.0	PA	PA.FU2	73.0
Age.FU2	Puberty.FU2	100.0	Region	Breastfeeding	70.0
Age	School	99.0	AVM	AVM.FU2	68.0
Age.FU2	Alcohol.FU2	99.0	Income	Income.FU1	67.0
PA	PA.FU1	98.0	Weeks_of_pregnancy	Birthweight	66.0
Mothers_BMI	BMI	97.0	Age.FU1	Sleep.FU1	66.0
YHEI	YHEI.FU1	97.0	Age.FU2	AVM.FU2	66.0
HOMA.FU1	HOMA.FU2	97.0	BMI.FU2	Puberty.FU2	66.0
Sex	AVM.FU1	95.0	Formula_milk	Breastfeeding	65.5
Sex	PA.FU2	95.0	ISCED	Income	64.5
Familymeal	Familymeal.FU1	95.0	BMI.FU1	HOMA.FU1	64.0
Income.FU1	Income.FU2	95.0	HOMA.FU1	Puberty.FU2	63.0
ISCED.FU1	ISCED.FU2	95.0	Alcohol.FU2	Smoking.FU2	63.0
Well-being.FU1	Well-being.FU2	95.0	Age_at_birth	Mothers_BMI	60.0
YHEI.FU1	YHEI.FU2	95.0	Mothers_BMI.FU2	BMI.FU2	58.0
Sex	AVM.FU2	94.0	Region	Formula_milk	56.0
AVM.FU1	AVM.FU2	94.0	Age_at_birth	HH_Diet	56.0
YHEI	YHEI.FU2	93.0	BMI.FU2	HOMA.FU2	54.0
Age.FU2	Sleep.FU2	93.0	Age.FU2	YHEI.FU2	53.0
AVM	AVM.FU1	92.0	Region	HH_Diet	52.0
Well-being	Well-being.FU1	92.0	PA.FU1	Well-being.FU1	52.0
School.FU1	Familymeal.FU1	92.0	Region	Birthweight	50.0
Age.FU2	Smoking.FU2	91.0	ISCED	ISCED.FU1	49.0
Birthweight	BMI	90.0	YHEI.FU1	Familymeal.FU1	48.0
Age	AVM	88.0	Sleep	Sleep.FU2	47.0
School	AVM.FU2	88.0	YHEI	Familymeal	47.0
PA.FU1	PA.FU2	88.0	HH_Diet	Breastfeeding	46.5
School	Sleep	86.0	PA	YHEI	45.5
Sleep	Sleep.FU1	86.0	Familymeal.FU1	YHEI.FU1	45.0
Sex	AVM	85.0	HOMA.FU2	BMI.FU2	45.0
Weeks_of_pregnancy	Formula_milk	85.0	Well-being	YHEI	44.0
School	Sleep.FU2	84.0	Age	PA	42.0

Edges		
from	to	%
Age.FU2	Well-being.FU2	40.0
Income	PA	39.0
Well-being.FU1	PA.FU1	38.0
PA.FU1	YHEI.FU1	37.0
Smoking.FU2	Alcohol.FU2	37.0
YHEI.FU1	Well-being.FU1	36.5
Income	ISCED	35.5
Income.FU1	ISCED.FU1	35.0
AVM.FU2	YHEI.FU2	34.5
HOMA.FU1	BMI.FU1	34.0
Sex	Well-being	33.0
BMI	HOMA	33.0
YHEI	AVM.FU1	33.0
AVM.FU2	Sleep.FU2	33.0
Formula_milk	Sleep	32.0
Well-being.FU1	YHEI.FU1	31.5
Migrant	Well-being	31.0
AVM	YHEI	30.5
AVM.FU1	Well-being.FU1	30.5
Mothers_BMI	BMI.FU2	30.0
HOMA.FU1	AVM.FU1	30.0
Well-being	Well-being.FU2	28.0
Migrant	HH_Diet	26.0
AVM.FU1	YHEI.FU1	24.5
ISCED	Breastfeeding	24.0
Age	BMI	24.0
Well-being.FU1	Sleep.FU1	24.0
PA	Well-being	23.5
Region	BMI	23.0
ISCED.FU1	Income.FU1	23.0
AVM.FU1	Sleep.FU1	21.0
Region	AVM	18.0
Age_at_birth	Well-being	17.0
Sleep	Well-being.FU1	16.0
Sleep	HOMA.FU1	15.0
Well-being.FU1	AVM.FU1	14.5

Edges		
from	to	%
YHEI.FU1	PA.FU1	13.0
Well-being.FU2	YHEI.FU2	13.0
Age	Sleep.FU2	11.0
AVM.FU1	HOMA.FU1	10.0
ISCED	BMI	8.0
Age_at_birth	YHEI.FU2	5.0
Sleep.FU2	Well-being.FU2	4.5
YHEI.FU2	Well-being.FU2	2.0

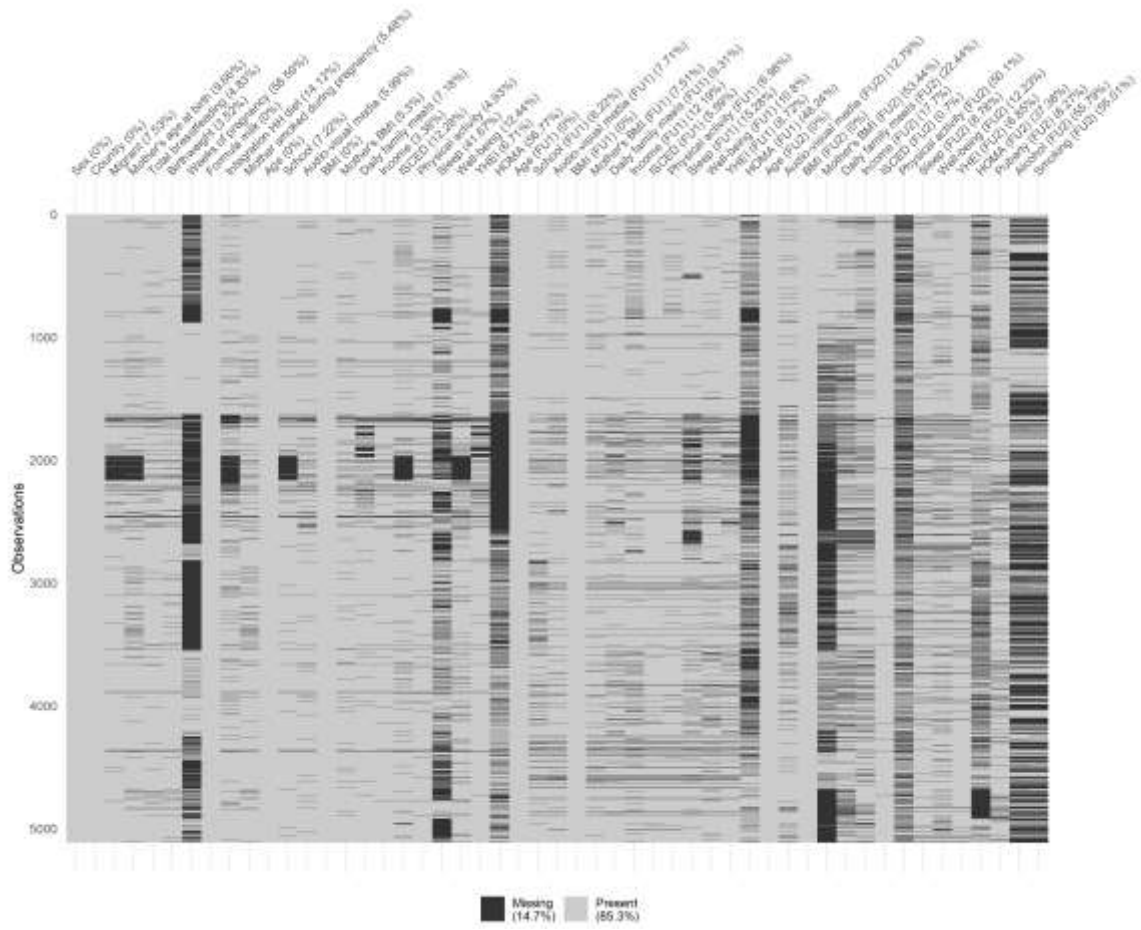
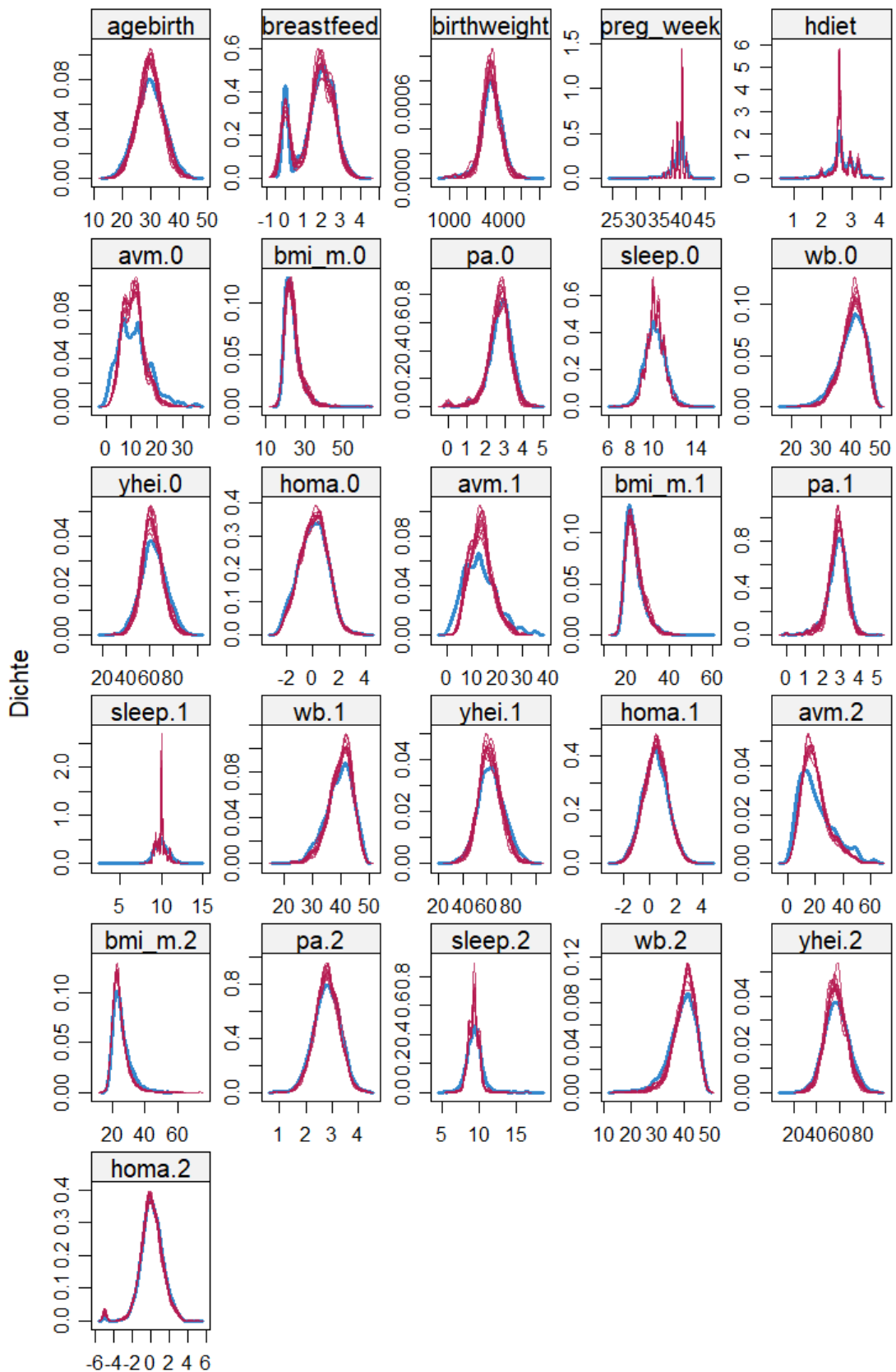
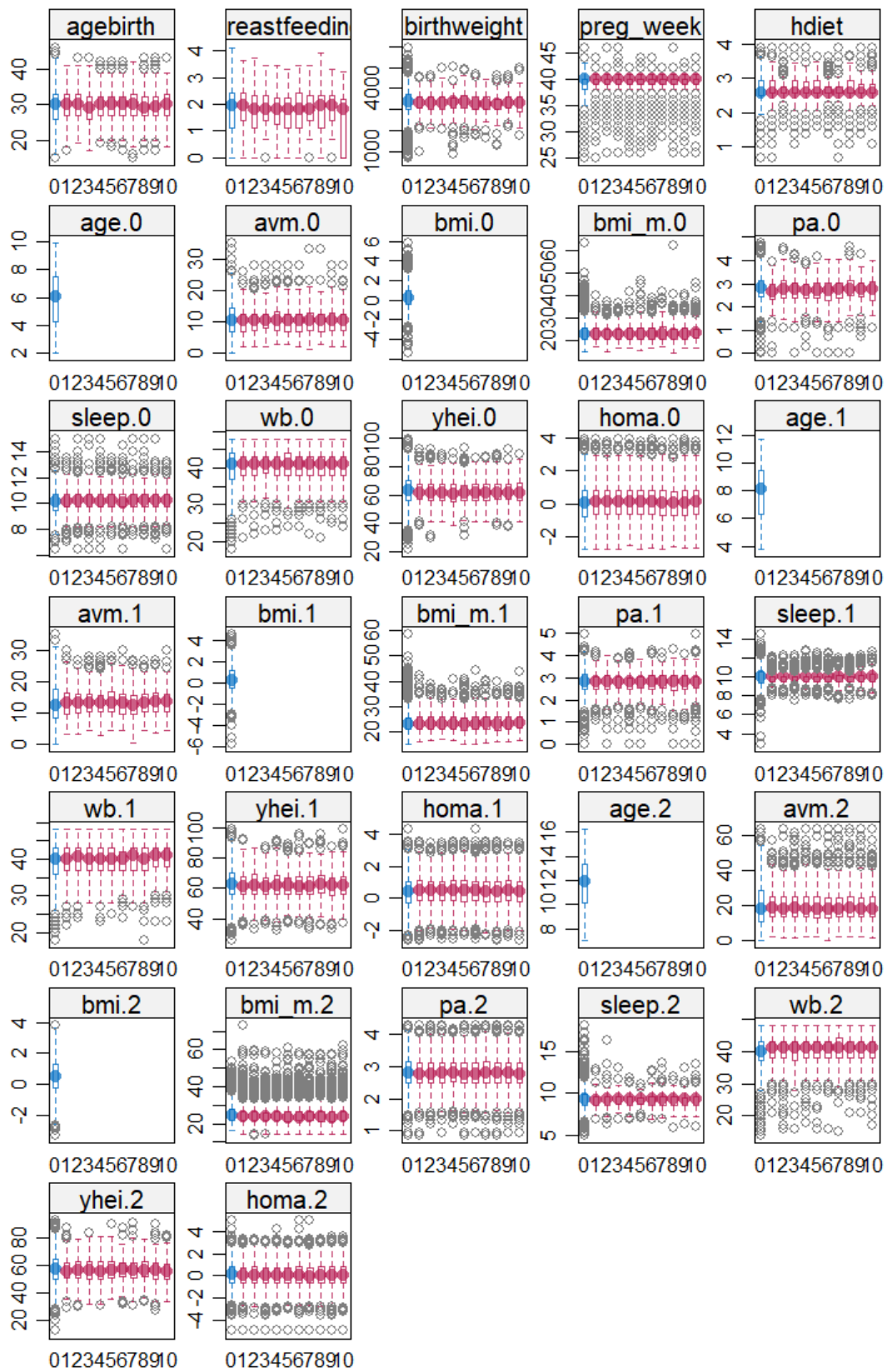


Figure S1: Missing values in the cohort dataset where black cells indicate a missing observation.





Imputation number



Figure S2: Diagnostic plots (kernel density estimates and boxplots for continuous and barplots for discrete variables) of the observed data (blue) and the multiply imputed data (red) with $m=10$.

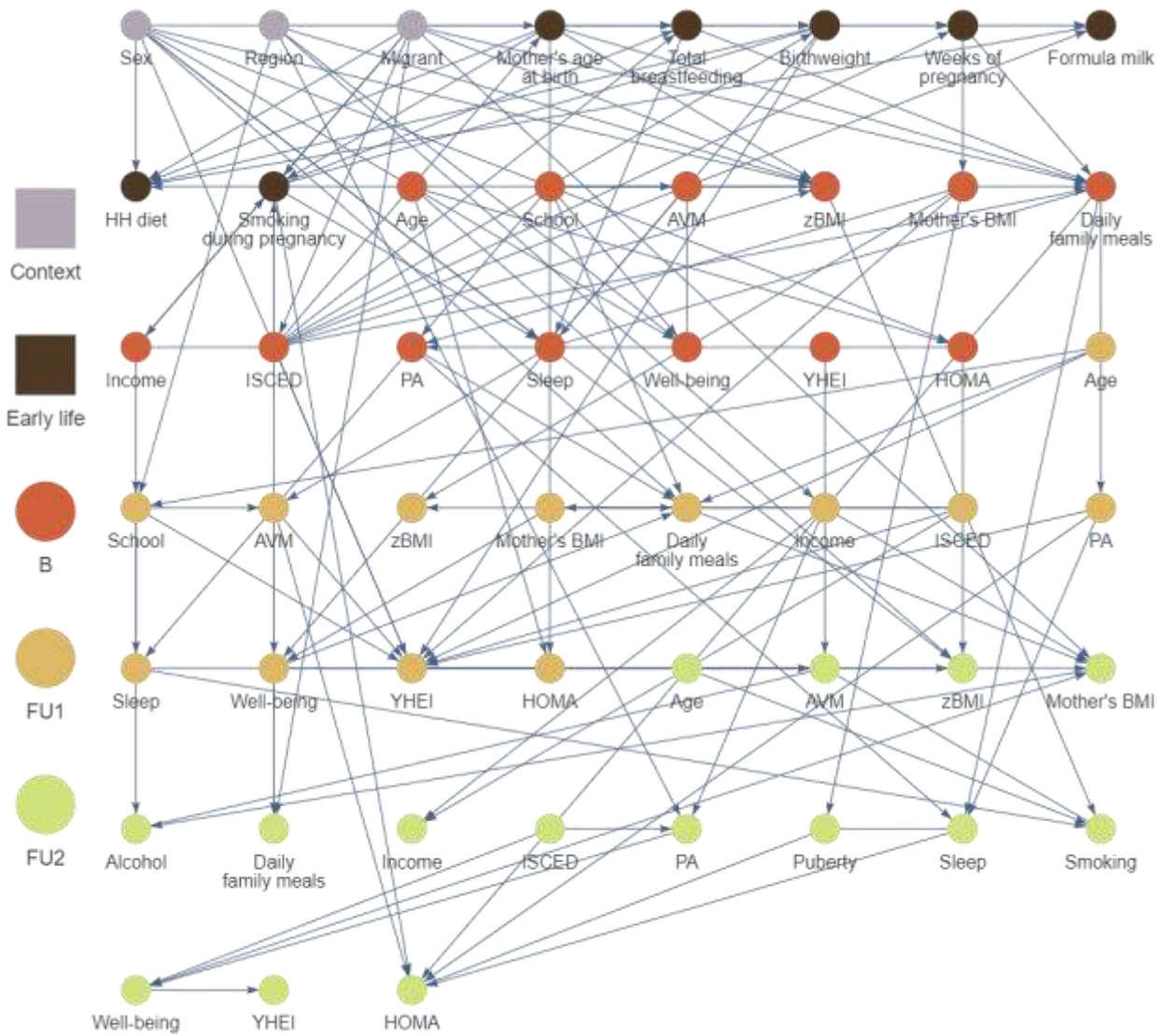


Figure S3: Causal graph of childhood obesity based on N = 5,112 European children and adolescents born between 1997 and 2006 estimated by the time-ordered pc-algorithm using **test-wise deletion**. Nodes are coloured with respect to their appearance in the life course. Edges without arrowheads could not be orientated by the algorithm.

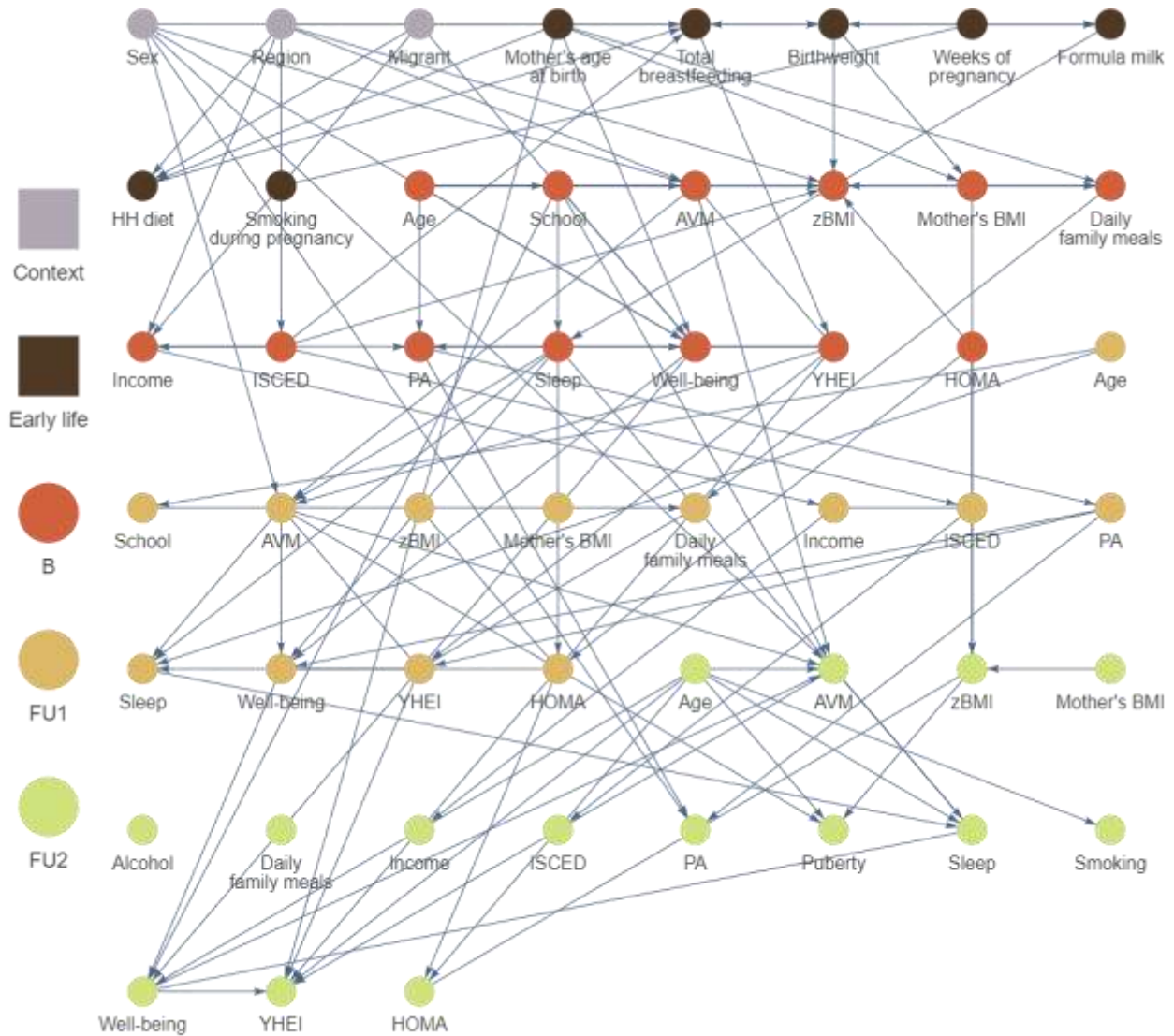


Figure S4: Causal graph of childhood obesity based on N = 5,112 European children and adolescents born between 1997 and 2006 estimated by the time-ordered pc-algorithm using **multiple imputation with $\alpha = 0.1$** . Nodes are coloured with respect to their appearance in the life course. Edges without arrowheads could not be orientated by the algorithm.

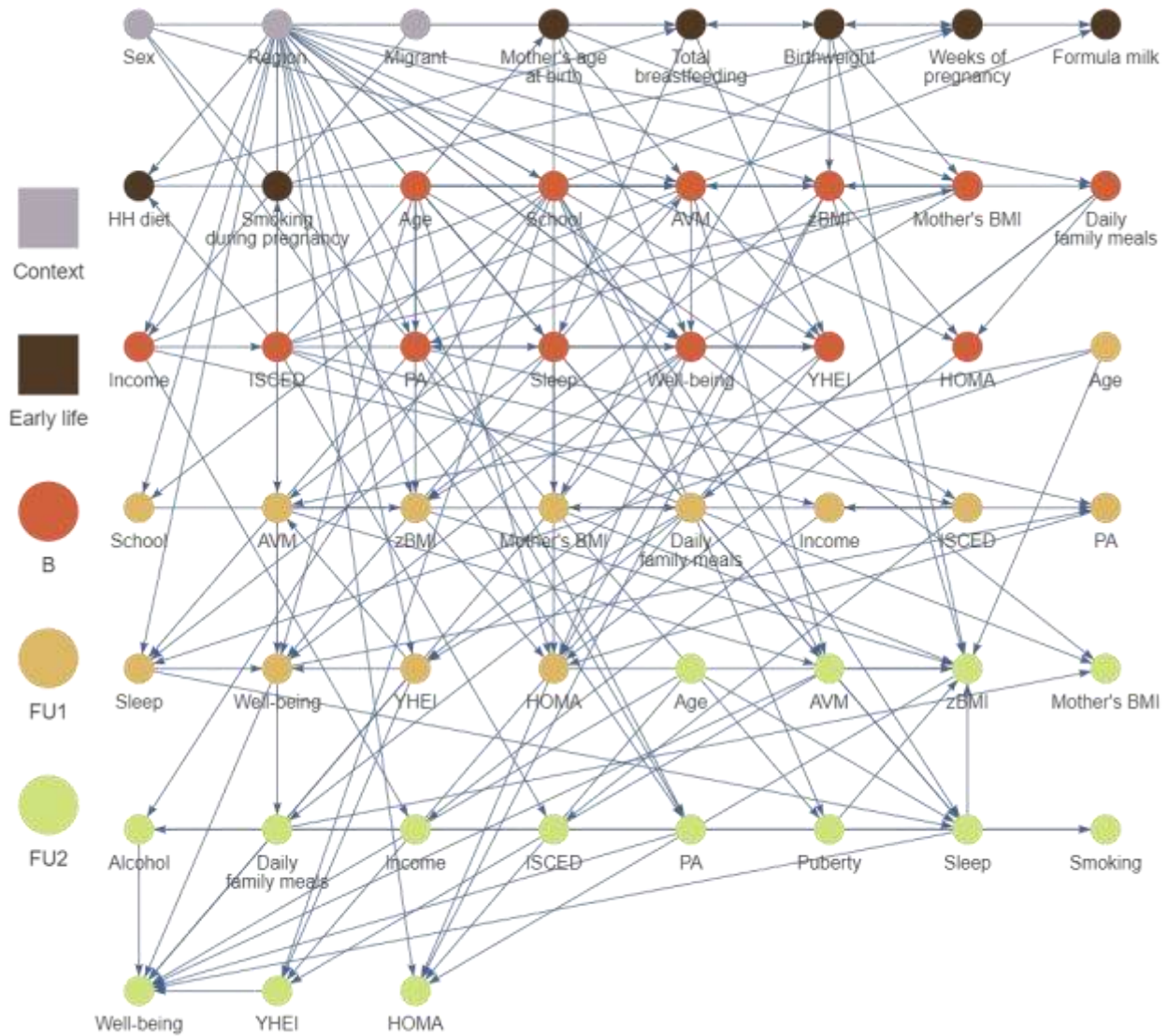


Figure S5: Causal graph of childhood obesity based on N = 5,112 European children and adolescents born between 1997 and 2006 estimated by the **Structural EM algorithm**. Nodes are coloured with respect to their appearance in the life course. Edges without arrowheads could not be orientated by the algorithm.

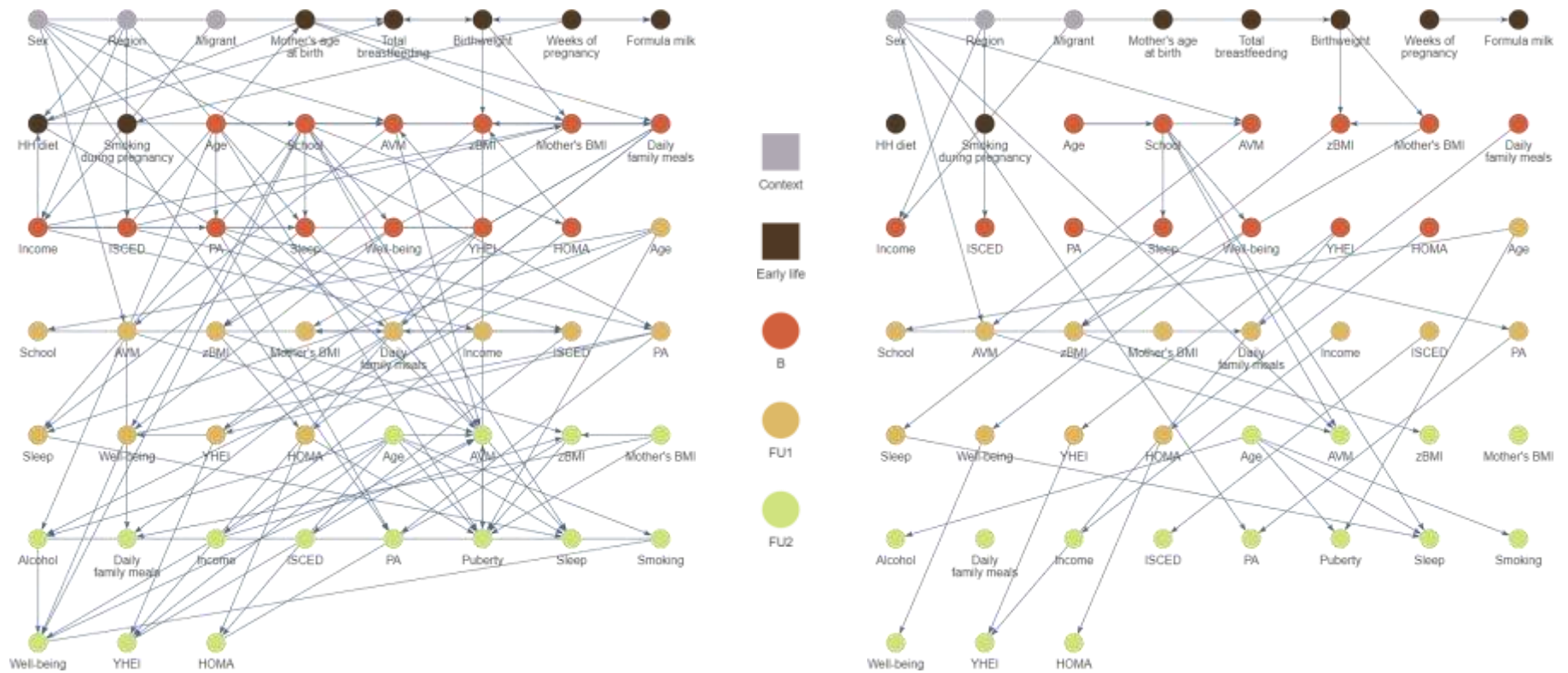


Figure S6: Cumulated causal graphs of childhood obesity based on $N = 5,112$ European children and adolescents born between 1997 and 2006 estimated by the **time-ordered pc-algorithm** based on 10 multiple imputed datasets for each of 100 independent **bootstrap** samples. The left and right graphs represent edges that appeared in more than 35 % and 75 %, respectively, of the “bootstrap graphs”. Nodes are coloured with respect to their appearance in the life course. Edges without arrowheads could not be orientated by the algorithm.

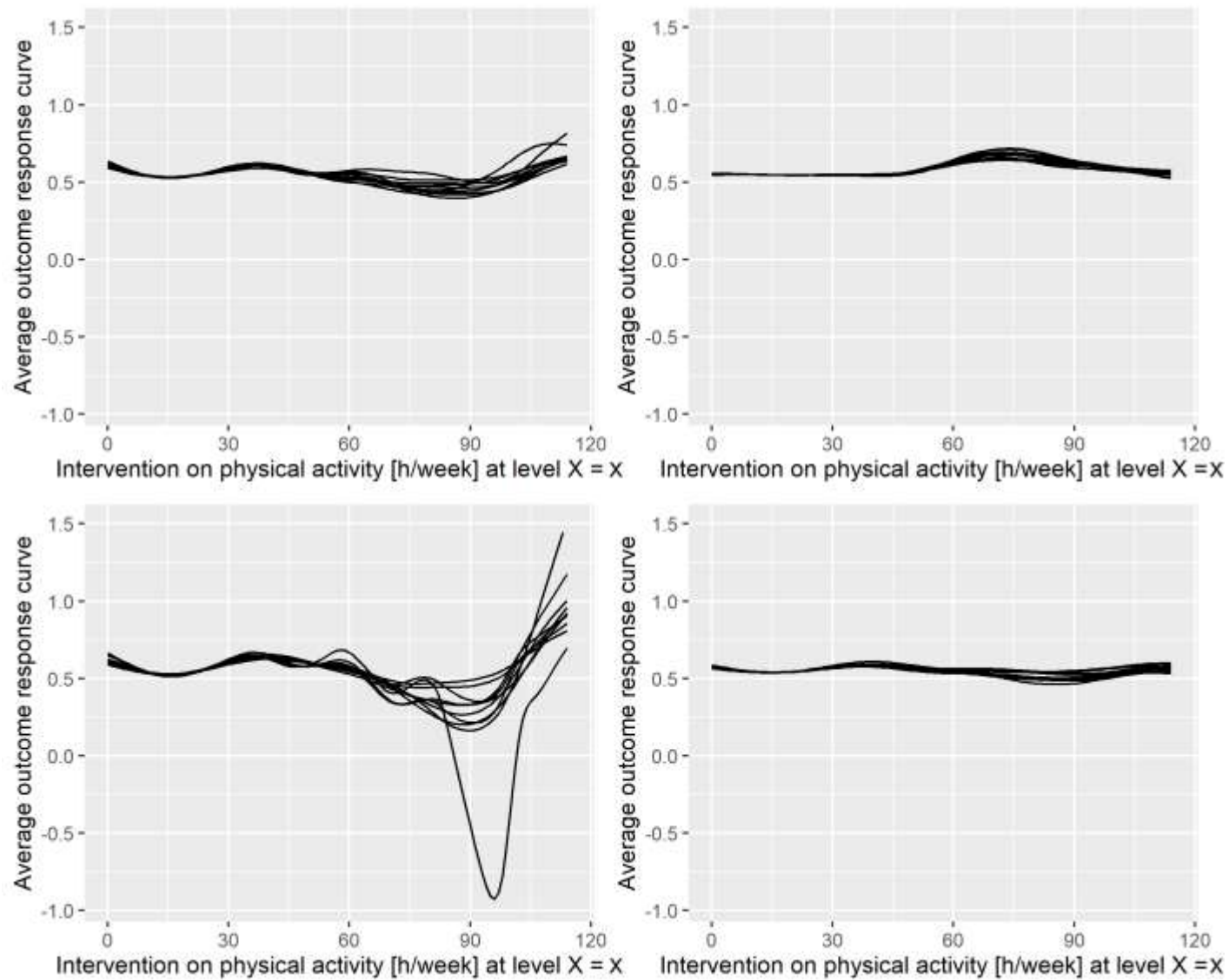


Figure S7: Estimated average response curves for physical activity at baseline on BMI z-scores (FU2)

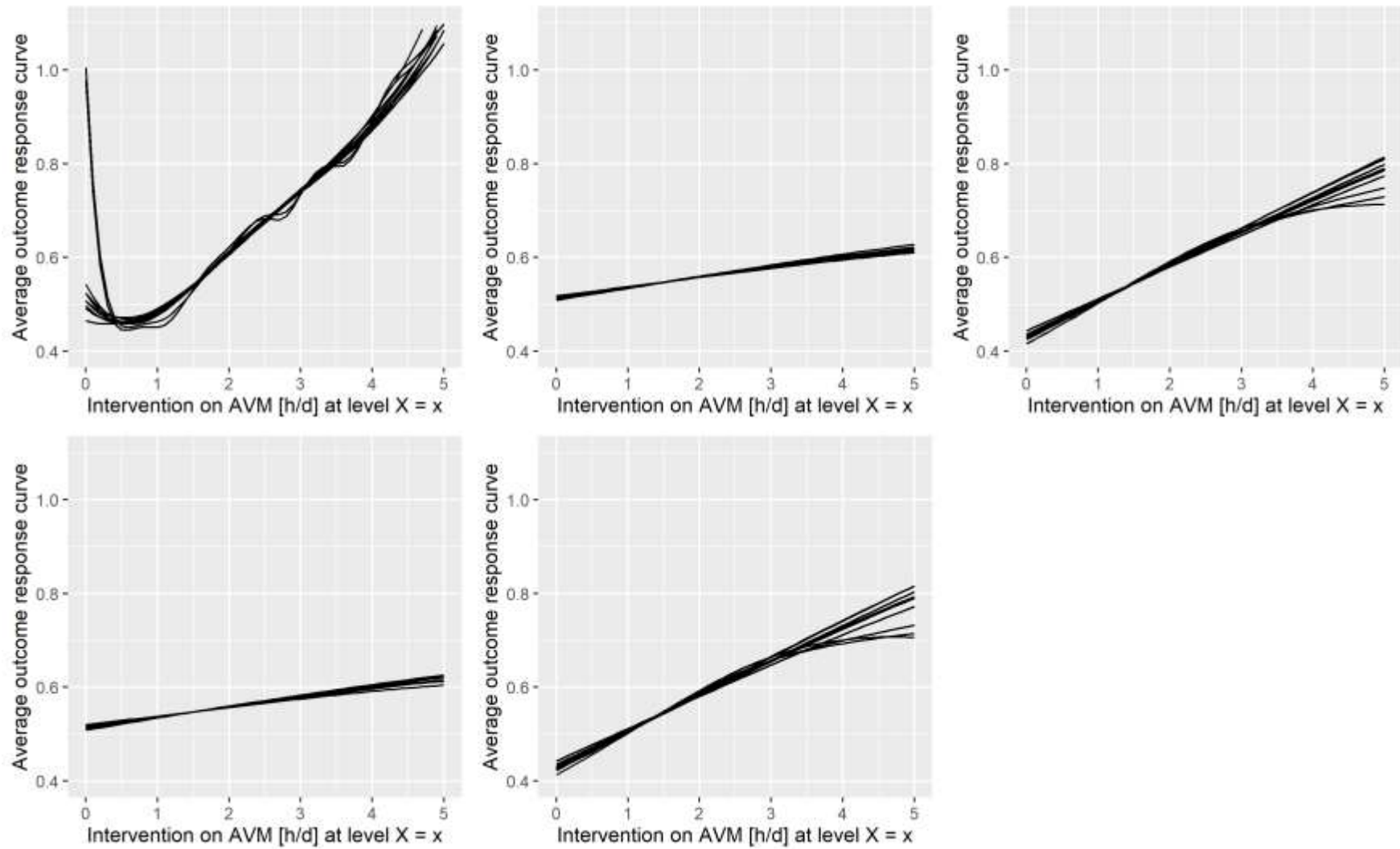


Figure S8: Estimated average response curves for audio-visual media consumption (AVM) at baseline on BMI z-scores (FU2)

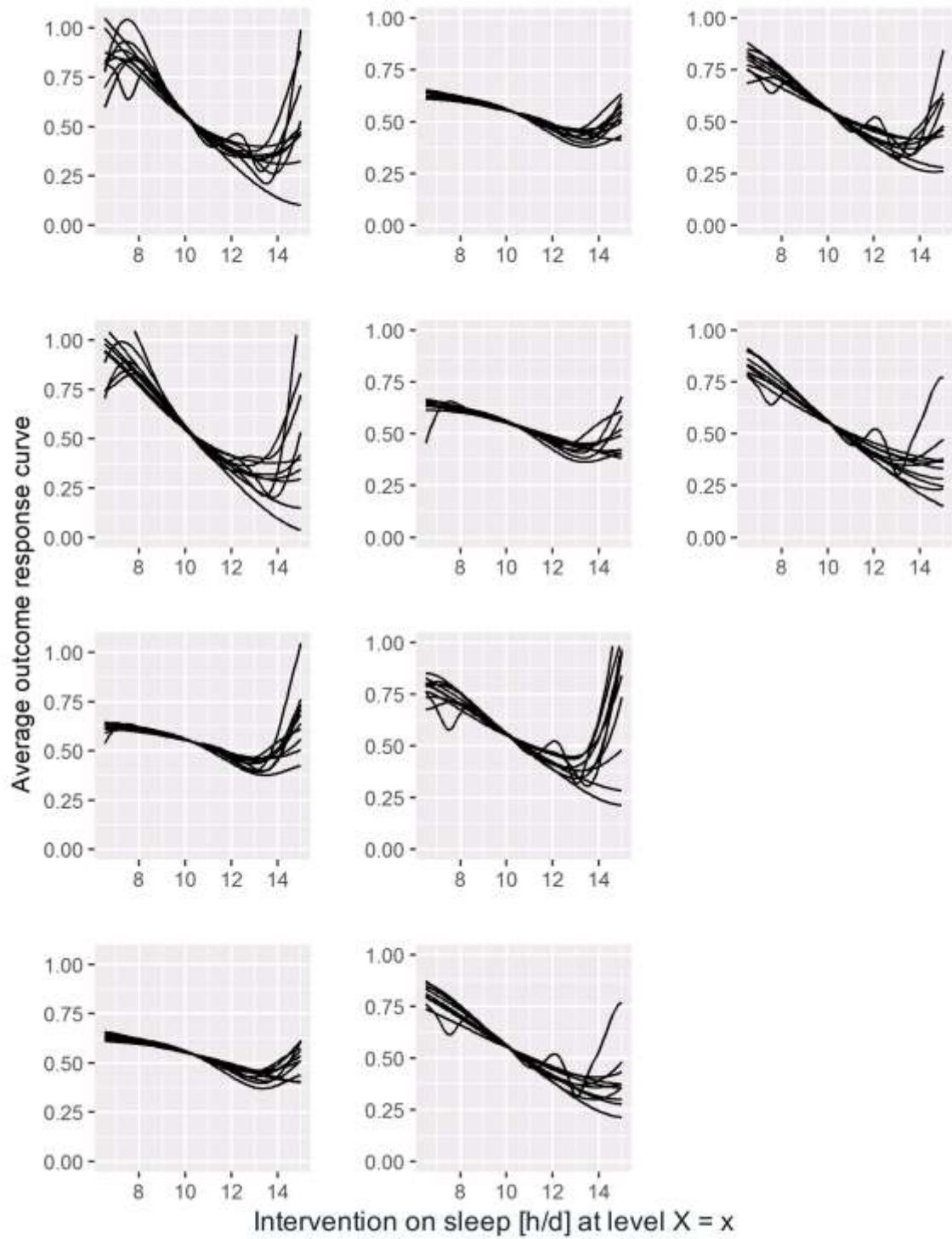


Figure S9: Estimated average response curves for total sleep at baseline on BMI z-scores (FU2)

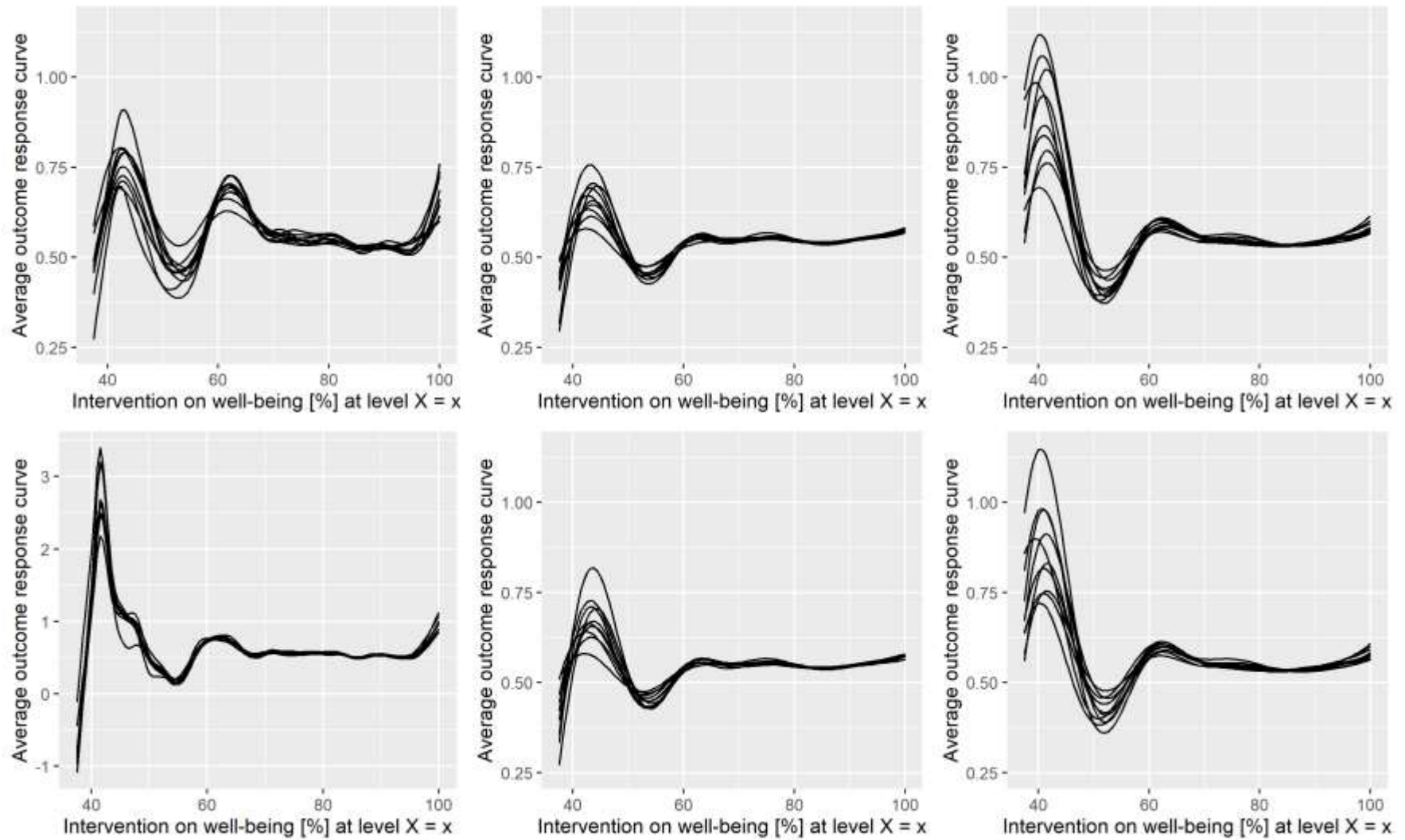


Figure S10: Estimated average response curves for well-being at baseline on BMI z-scores (FU2)