

SUPPLEMENTARY DOCUMENT – DESCRIPTION OF THE MISSINGNESS PATTERN AND THE IMPUTATION MODEL USED FOR THE PAPER:

The relation between birthweight, childhood body mass index, and overweight and obesity in late adolescence: a longitudinal cohort study from Norway, The Tromsø Study, Fit Futures.

A total of 43% participants (411) were missing one or more of the height/length and weight variables from the birth registry (5% were missing birthweight) and/or childhood measurements (29% missing height/weight at 2-4 years of age, 23% missing height/weight at 5-7 years of age). In 31% (302) of participants data were missing from follow-up at Fit Futures 2, leaving 47.0% (452) of the core study population with five complete measurements.

Due to the relatively large number of missing explanatory variables, we performed multiple imputations to handle missing data. The missingness pattern was checked and found to be arbitrary, thus data was assumed to be missing at random (MAR). Multiple imputations were performed for nine variables from birth and childhood: birthweight, length, gestational age, exact age, and height and weight at 2-4 and 5-7 years of age. In addition, body mass index (BMI) for two girls at 15-17 years of age was imputed.

Multiple imputations were performed by chained equations, generating twenty duplicate datasets. The missing values were replaced by imputed values based on the observed information from all five ages. The imputation model included all variables from the final generalised estimating equations analysis, and it included BMI at 15-17 and 18-20 years of age as outcome variables. Calculated variables based on imputed values, such as weight class, were registered as passive variables. To account for a

significant interaction term between sex and BMI at 2-4 years of age, separate imputations were performed for boys and girls. To increase the predictive power of the imputation model, we included auxiliary variables: height, weight, waist circumference and exact age at Fit Futures 1, a dichotomous variable of missing or not at Fit Futures 2, and variables from the birth registry (maternal age at birth, caesarean section, multiple births, maternal disease or diabetes). Variables were imputed using linear regression for continuous variables and predictive mean matching (using 100 nearest neighbours) for variables that had a restricted range or were slightly skewed (the age variables, gestational age, weight, and BMI variables).

Participants with missing BMI at follow-up did not have a significantly higher BMI at Fit Futures 1 vs. those attending Fit Futures 2 (girls: 23.0 vs. 22.2 kg/m²; p= 0.06, boys: 22.7 vs. 22.2 kg/m²; p=0.20). Significantly more boys (39.8%) than girls (22.6%) were missing at Fit Futures 2. We chose not to use imputed missing BMI for those lost to follow-up in Fit Futures 2, since imputation of outcome variables are questionable and we cannot rule out the possibility of missingness that is not random (MNAR). Therefore all analyses performed on the imputed dataset include 961 subjects at 15-17 years of age and complete-case data for 659 subjects at 18-20 years of age.