Title: The current obesity "epidemic": segregation of familial genetic risk in NHANES cohort supports a major role for large genetic effects

Running title: Segregation of familial genetic obesity risk

Authors: Arthur B. Jenkins[1,4]*, Marijka Batterham[2], Lesley V. Campbell[3,4]

[1] School of Medicine, University of Wollongong, Wollongong NSW 2522 Australia

[2] Statistical Consulting Centre, School of Mathematics and Applied Statistics, University of Wollongong NSW 2522 Australia

[3] Diabetes Service, St Vincent's Hospital, Sydney, NSW 2010 Australia

[4] Diabetes and Metabolism Division, Garvan Institute of Medical Research, Darlinghurst, NSW 2010 Australia

* Corresponding author

E-mail: ajenkins@uow.edu.au

Abstract

The continuing increase in many countries in adult body mass index (BMI kg/m^2) and its dispersion is contributed to by interaction between genetic susceptibilities and an increasingly obesogenic environment (OE). The determinants of OE-susceptibility are unresolved, due to uncertainty around relevant genetic and environmental architecture. We aimed to test the multi-modal distributional predictions of a Mendelian genetic architecture based on collectively common, but individually rare, large-effect variants and their ability to account for current trends in a large population-based sample. We studied publicly available adult BMI data (n = 9102) from 3 cycles of NHANES (1999, 2005, 2013). A first degree family history of diabetes served as a binary marker (FH_0/FH_1) of genetic obesity susceptibility. We tested for multi-modal BMI distributions non-parametrically using kernel-smoothing and conditional quantile regression (CQR), obtained parametric fits to a Mendelian model in FH_1, and estimated FH x OE interactions in CQR models and ANCOVA models incorporating secular time. Non-parametric distributional analyses were consistent with multi-modality and fits to a Mendelian model in FH_1 reliably identified 3 modes. Mode separation accounted for ~40% of BMI variance in FH_1 providing a lower bound for the contribution of large effects. CQR identified strong FH x OE interactions and FH_1 accounted for ~60% of the secular trends in BMI and its SD in ANCOVA models. Multimodality in the FH effect is inconsistent with a predominantly polygenic, small effect architecture. and we conclude that large genetic effects interacting with OE provide a better quantitative explanation for current trends in BMI.

Introduction

The recent and continuing increase in the global mean adult BMI, first seen in high income countries, is now seen in most countries across a wide range of ethnic composition and socio-economic conditions[1] and is accompanied by increases in measures of dispersion[2, 3]. Although BMI is known from family-based studies to be under strong genetic influences[4] population genetic backgrounds

have been effectively constant over this time, implying that BMI trends are driven by change in environmental factors (obesogenic environment, OE). Evidence from twin studies, which demonstrate increased genetic variance over time, supports an important role for interactions between OE and genetic susceptibility (G x OE) on both mean and dispersion of BMI[3, 5], but how large a role is not yet known. Defining the role of G x OE in "epidemic" obesity, and hence of genetic susceptibility itself, is hindered by problems of measurement and modeling of interactions[6] and by uncertainty around both the genetic architecture[4] and the exact nature of the environmental drivers[7]. Whether population susceptibility to OE is predominantly determined by a subgroup with high genetic susceptibility or is more evenly spread within populations is unresolved despite important implications for the management of obesity and related disorders at population and individual levels[2, 8–10].

The genetic variants responsible for obesity susceptibility remain largely unknown. Genome-wide association studies (GWAS) have identified significant associations with >200 markers with small effects on BMI (polygenes), together explaining only approximately 3-4 % of total variance compared to family-based heritabilities (h^2) of 50-75%[11]. Few causative mechanisms responsible for these phenotypically weak associations are known[4]. The sources of the h^2 unaccounted for by GWAS are uncertain; suggestions include overestimation of h^2, large numbers of common genetic variants with small, statistically insignificant effects on phenotypes[12, 13] and importantly, candidates not tested in most GWAS. Among the latter are rare genetic variants with large phenotypic effects and G x OE interactions[4]. Recently significant G x OE interactions have been detected in individual GWAS loci and in composite genetic risk scores, which however explain little of the missing component of h^2[14, 15].

A family history of diabetes (FH) is a potent, predominantly genetic[16, 17] risk factor for diabetes diagnosis (DM) and for obesity-related phenotypes[18–21] consistent with the strong association between type 2 DM and overweight/obesity. Familial effects on obesity-related phenotypes in adults are also predominantly genetic[3, 22], so to the extent that the DM generating FH is of type 2 (approximately 94% of DM in the US population[23]), FH is a prevalent and readily obtained marker of genetic susceptibility both to diabetes and to the obesity commonly preceding it. We have previously reported evidence from a small sample of a multi-modal effect of FH on a composite adiposity index consistent with segregation in families of discrete obesity risk[21]. Polygenic risk scores (PRS) based on large numbers of small effects are expected to be, and appear to be, unimodally-distributed[24, 25] and thus cannot account for familial segregation of discrete risk. The present work is based on the alternative hypothesis that individually rare, but collectively common, genetic variants with large phenotypic effects are the source of most of the missing h^2 and of most of G x OE, and that their effects can be detected through analyses of phenotypic segregation in high-risk families[26].

The Continuous National Health and Nutrition Examination Survey (NHANES) is a continuing (1999-) large-scale population-based survey incorporating an index of adiposity (Body Mass Index, BMI) and first-degree FH (FH_0/FH_1) together with potential covariates and confounders. Although BMI has recognized limitations as an adiposity phenotype[26, 27] it is the basis for most large-scale genetic studies and like other authors, we assume that a large enough scale and appropriate modeling of covariates will reduce effects of imprecision and bias[11]. We aimed to test in a large multi-cycle NHANES sample for the presence of familial segregation of genetic risk and to estimate the contribution of FH, and by extension all discrete genetic risk, to recent secular trends in adult BMI. The results support a predominant role for large genetic effects interacting with OE in the obesity "epidemic".

Subjects and Methods

Subjects

We used data from the 1999-2000, 2005-2006 and 2013-2014 cycles of NHANES (see Section Web Resources). We extracted records for participants age 20-65 years with non-missing gender, race/ethnicity, BMI, diabetes family history (FH) and diabetes diagnosis data, and current smoking status if available. The definitions of two fields changed over the sampling period: 1) Diabetes family history was defined in terms of 1° and 2° relatives in 1999-2000 but by 1° relatives only in subsequent cycles. We recoded the 1999-2000 diabetes family history data to conform to the later definition using the separately collected data for affected parents and siblings. 2) The self-identified race/ethnicity field (RIDRETH1) code was used excluding other races (OR) to maintain consistency of categories across cycles (Supplementary Methods). We excluded from the primary analyses subjects diagnosed with diabetes because of possible confounding by effects of either diabetes or diabetes therapies on BMI. The resulting data set is summarized in Table 1.

Statistical analyses

Approach

We treated the data as a convenience sample and took no account of the sampling weights provided by NHANES to permit nationally representative estimates. Our results are not intended to be representative of the US population.

Our primary analyses are based on non-parametric visualization (kernel-smoothing) and analyses (conditional quantile regression, CQR) of distributions requiring no prior distributional assumptions. Parametric fits to multimodal distributions were then used to quantify the contributions of the predicted large genetic effects model. $FH(_{0/1})$ is treated analytically as a binary genetic risk marker but the distribution of its effect across quantiles is interpreted in a Mendelian model in which FH_1 represents enrichment of a mixture of single and double carriers of risk variants. Calendar time is

treated as a continuous surrogate of OE. Effects of OE interacting with FH were assessed in CQR models, and also in least-squares ANCOVA models using bootstrap resampling to minimize distributional assumptions in the calculation of effect size estimates and errors. All analyses were performed using R 3.6.1[28].

Summary statistics

Heterogeneity of the samples across cycles was assessed by Chi square test for categorical variables and by one-way ANOVA for age. Effects on BMI were assessed by ANCOVA against continuous time (yr = calendar start year - 1999). Effects on phenotype SD's were assessed by Bartlett's test in one-way ANOVA models.

Adjustment

Prior to analysis BMI was adjusted for effects of age, gender and race/ethnicity in a linear model (age + age^2 + race x gender). The adjustment model accounted for 4 % of the total variance in BMI (Supplementary Table S1).

Distributions

Visualization

The effect of FH on the distribution of adjusted BMI was visualized using kernel-smoothed density estimates by FH status (R base function density). The degree of smoothing is controlled by the bandwidth parameter (bw) which was obtained in the full non-diabetic data set (bw = 0.99) from a measure of the dispersion of the data[29]. This produces a continuous distribution function and is widely used to visualise features of potential interest which may be obscured in histograms. The

credibility of the apparent effect of FH on the shape of the distribution was assessed by post-hoc analysis of density ratios (FH_1 / FH_0) by quantiles (20) of the full sample. Mean density ratios with SEM were obtained by quantile by bootstrap resampling with replacement (1000 draws, stratified by FH status with resample sizes = strata sizes)., and compared to the predictions of normal and log-normal mixture distributions characterised by the proportions, means and SD's in the sample stratified by FH status and cycle. Lack of fit to the mixture distributions was assessed by X^2 tests in 1/variance-weighted linear regressions.

Conditional Quantile Regression (CQR)

Conditional quantile regression is a powerful tool for analyzing the effect of covariates on distributions without assumptions of distributional shape. In contrast to ordinary least-squares (OLS) regression which characterizes effects on global features of a distribution, CQR analyses local effects of covariates independently at any specified quantiles and can detect variations in covariate effects across quantiles. Originating in econometrics[30] it is now used in other areas including genetics[31]. In the CQR framework developed by Abadi et al[14] for analysis of genomic markers, trends in effect sizes across quantiles represent interactions between genetic effects and unobserved environmental and/or genetic factors. We treat FH as a binary genetic risk marker (FH_0/FH_1) and a linear trend in quantile regression coefficients ($ß_{1i}$) across quantiles (τ_i) represents summed linear interactions of FH with unobserved factors. We analysed the effects of FH on adjusted BMI by CQR using the R package quantreg. The effect of all interactions on the FH effect was tested in in a 2 parameter linear model:

for each quantile τ_i in y,

$$y(\tau_i \mid FH=fh) = ß_{0i} + ß_{1i} * FH$$

where $y(\tau_i \mid FH=fh)$ = the ith quantile of adjusted BMI conditional on the value of FH (0,1), the intercept $ß_{0i}$ is the ith quantile value in FH_0 and $ß_{1i}$ is the FH effect size in quantile i.

The interaction between FH and continuous calendar time was estimated in the ANCOVA model:

$$y(\tau_i \mid FH=fh) = ß_{0i} + ß_{1i} * FH + ß_{2i}*yr + ß_{3i}*FH*yr$$

where yr = cycle start year – 1999, $ß_{0i}$ is the ith quantile value in FH_0 at yr =0 and $ß_{2i}$ and $ß_{3i}$ are CQR coefficients for time and time*FH interaction effects in quantile i. The coefficients for the time-related effects represent the effects in FH_0 ($ß_{2i}$) and the additional effects in FH_1 ($ß_{3i}$) so that the coefficients for total time effects in $FH_1 = ß_{2i} + ß_{3i}$.

Equality of CQR parameter effect sizes across quantiles was tested using the anova.rq function in the R package quantreg.

The strengths of the CQR effects across quantiles were also assessed in linear meta-regression analyses of relationships between quantile coefficients $ß_{1\text{-}3i}$ and quantiles using the R package metafor[14]. Regression coefficients ($ß_{MR}$) with SEM are reported in units of $kg.m^{-2}$ over the full quantile scale (0-1). The structure of the FH effects in relation to the multi-modal Mendelian hypothesis was assessed in an analysis of residuals from linear OLS fits of $ß_{1i}$ to $ß_{0i}$ with Durbin-Watson tests for residual 1st order autocorrelation (function durbinWatsonTest in the R package car).

Parametric fits

We obtained fits to a 3-component normal mixture distribution representing a simple Mendelian model of fixed genetic effect using an expectation-maximization algorithm (normalmixEM function in the R package mixtools). The models are characterised by the fitted means (μ_i), standard deviations (σ_i) and mixing proportions (λ_i) of the three component distributions. Full model fits were obtained in FH_1 but were not obtainable in FH_0 or DM_1 groups and we constrained μ_i in those groups to values

estimated in FH_1 in order to obtain comparable estimates of σ_i and λ_i. Risk allele frequencies (q) under an additive Mendelian model of large genetic effects were calculated from the fitted λ_i:

$$q = 0.5*\lambda_2 + \lambda_3$$

where λ_2 and λ_3 represent the proportions of carriers of 1 and 2 risk alleles respectively.Within-sample consistency of calculated q across the three groups analysed was assessed by comparing fitted q_{FH_1} with the prediction from random mating of DM_1 into the full sample:

$$\text{predicted } q_{FH_1} = (q_{DM} + \text{n-weighted mean}(q_{DM}, q_{FH_1}, q_{FH_0}))/2$$

Secular trends

Effects of diabetes family history status (FH_0, FH_1) and continuous time (yr = calendar start year - 1999) on adjusted BMI and its standard deviation (SD) were assessed in ANCOVA models of the form:

$$y = ß_0 + ß_1 * FH + ß_2*yr + ß_3*FH*yr$$

where y = adjusted BMI mean or SD by FH status (0/1) and cycle, and yr = cycle start year – 1999. Each OLS fit estimated 4 parameters from the 6 data points. Mean parameter estimates with 95% CI were obtained by bootstrap resampling with replacement (1000 draws stratified by FH status and cycle with resample sizes = strata sizes).

Comparison of cross-sectional and secular trend effects

Effects of FH on BMI distribution and on secular trends in BMI were compared by calculating the contribution (%) of FH_1 to the effect in the full non-diabetic sample for calculated risk allele frequency (q%) and to the slope (ß%) of the relationships between time and BMI in ANCOVA model described above. Mean (± SEM where possible) q% and ß% were calculated in the relevant bootstrap samples.

Results

Participant characteristics

Data from 9102 non-diabetic subjects met the inclusion criteria, approximately equally distributed across the 3 cycles. Gender balance varied little but there was a cycle effect in race/ethnicity, most obvious in the reduced representation of MA in the two later cycles. Average age varied across cycles but not its SD, while adjusted BMI and its SD showed linear trends with cycle time. FH_1 prevalence was higher in the two later cycles compared to 1999-2000 as was DM_1 prevalence. Current smoking status was predominantly missing in the data (55%) and was not included in the BMI adjustment model. However smoking status was not related to FH status whether analysed in the full data (X^2 = 2.80, 2 df , p = 0.25) or in those with non-missing smoking status (X^2 = 0.43, 1 df , p = 0.51), hence is unlikely to confound analyses of FH effects. The mean age at diagnosis of DM (43.6 yr) is consistent with predominant type 2 DM in the sample.

Distributions

Visualization

Adjusted BMI in the non-diabetic sample showed an apparently unimodal distribution, right-skewed compared to a normal model and closer to a log-normal model (Fig 1A). When visualized by FH status (Fig 1B) the predicted multimodality in FH_1 was indicated with modes in the normal weight, overweight and obese regions of the BMI distribution. In contrast FH_0 showed an apparently unimodal distribution. A difference in shape between the two groups was supported by the analysis of density ratios (Fig 1B inset) in which models based on mixtures (FH x cycle) of either normal or log-normal distributions did not provide adequate fits to the data ($p \leq 0.001$). BMI distribution in the

diabetic sample appeared to be depleted in the lower mode and enriched in the upper modes compared to FH_1 (Fig 1C).

CQR

Analysis of the effect of FH status alone on the shape of the BMI distribution using CQR demonstrated increasing FH effect size at higher levels of BMI ($ß_{1MR} = 2.2 \pm 0.2$ (SEM) $kg.m^{-2}$ Fig 2A, main panel), indicating strong interactions between FH_1 and other variables not included in the model. FH_1 effect size ranged from < 1 kg/m^2 in the lower quantiles to ~3 kg/m^2 in the upper quantiles, substantially different in both regions to the OLS estimate (1.7 kg/m^2). Inclusion of calendar time in the two-way model weakened the trend in $ß_1$ across quantiles ($ß_{1MR} = 1.5 \pm 0.3$, Fig 2B main panel) and exposed significant OLS effects and trends across quantiles for main ($ß_{2MR} = 0.07 \pm 0.02$) and interaction ($ß_{3MR} = 0.11 \pm 0.03$) effects of time (Fig 2C-D, main panels). The total OLS and MR interaction effect sizes in FH_1 ($ß_{2OLS} + ß_{3OLS} = 0.11 \pm 0.03$ (SE), $ß_{2MR} + ß_{3MR} = 0.18 \pm 0.04$) were approximately double those in FH_0 ($ß_{2OLS}$, $ß_{2MR}$, Fig 2C). The analysis supports the conclusion that calendar time is a strong surrogate of OE interacting with genetic factors as represented by FH status, and that susceptibility to OE increases with increasing BMI in both groups but more strongly in FH_1.

While the overall MR relationships between $ß_1$ and quantiles in both models were approximately linear (Fig 2A,B) there was strong evidence for additional non-linear structure in the OLS analysis of $ß_{1i}$ against $ß_{0i}$ (insets Fig 2A,B). The linear models provided good fits (one-way: slope = 0.15 ± 0.003 (SE), $R^2 = 0.98$; two-way: slope = 0.10 ± 0.01, $R^2 = 0.86$) but residual sequential structure was apparent in both models, confirmed by tests of autocorrelation in residuals ($p_{DW} \leq 0.002$) . The pattern of residuals in the two-way model (Fig 2B inset), adjusted for time-related effects, shows clear signs of discrete effects of FH_1 on the distribution of adjusted BMI with distinct peaks in the lower, middle and upper regions of the distribution. This pattern in the conditional quantile

coefficients does not map directly onto the unconditional quantile plots in Fig 1B, but does highlight similar regions in the distribution, and permits the conclusion that FH_1 has discrete, not continuous effects on BMI.

Parametric analysis

The distribution of BMI in FH_1 (Fig 1B) and the discrete pattern in the FH_1 effect by CQR (Fig 2B inset) appear consistent with a simple Mendelian model and fitting a 3-component normal distribution model to the FH_1 data resulted in robust estimates of component means (Fig 3A) and mixing coefficients and SDs (Table 2). Approximately 50% of FH_1 occupied the upper two modes and separation between modes accounted for approximately 40% of the total variance in adjusted BMI with the remainder assigned to dispersion within modes (Fig 2B). Under a Mendelian model the variance due to mode separation represents a lower bound on the contribution of large effects as some of the dispersion within modes represents variance in effect sizes of individual contributing causal loci (see Discussion) which will contribute to the ~60% of variance assigned to within-modes. Estimates of component SDs and mixing proportions with component means, constrained for FH_0 and DM_1 to those identified in the FH_1 data (Table 2), support enrichment in the two upper components in FH_1 compared to FH_0 (48% vs. 33%) and more strongly in DM_1 (72%). Predicted risk allele frequencies in FH_1 (q - Table 2) express these distributional properties in Mendelian terms and show within-sample consistency in that q_{FH1} predicted from random mating of DM_1 (0.37) is not different to the fitted estimate (0.30 ± 10).

Secular trends

Adjusted BMI mean (Fig 2A) and SD (Fig 2B) increased over the sampling period significantly faster in FH_1 compared to FH_0 in the bootstrapped ANCOVA model, and estimates of ß and Δß in the mean data were indistinguishable from the OLS estimates provided by the CQR analysis (Fig 2C,D). Similar results were obtained with log-transformed BMI (Supplementary Fig S2). FH_1 accounted for

62% of the BMI mean trend and 60% of the BMI SD trend in this sample over the period 1999-2014, effects similar in magnitude to the estimated FH_1 contribution to the sample risk allele frequency (50%, Supplementary Table S2).

Discussion

Summary

We tested the prediction of segregation of discrete effects of FH on adult BMI[21], modeled as modes of distribution, and estimated the contribution of FH_1 to recent trends in BMI mean and dispersion in a large population-based sample. The results support a predominant role in the recent obesity "epidemic" for rare genetic variants with large effects interacting with OE.

Segregation of genetic susceptibility

The non-parametric analysis provided evidence for a multi-modal distribution in the FH_1 group consistent with the prediction of segregation of large genetic effects in families[21]. Multi-modality was supported by the analysis of density differences between FH_1 and FH_0 by unconditional quantiles (Fig 1B) and by evidence of discrete signals in the OLS analysis of CQR coefficients (Fig 2A&B insets). The two upper peaks in Fig 2B inset are consistent with the predicted Mendelian effects of large effect variants on BMI but the potential lower peak is unexpected and may reflect the presence of type 1 diabetes family history in FH_1 group. Individuals with type 1 diabetes often present with BMI in the underweight (<18.5) – normal range (<25)[32] but a genetic basis for this has not been established.

Polygenic risk scores (PRS) in population-based samples are expected to be unimodally-distributed, and appear to be so[24, 25]. Any elevated polygenic obesity risk in DM_1 will dilute into the mating population resulting in a right-shifted distribution in FH_1 compared to FH_0, not discrete effects.

Alternative explanations for multi-modality might be discrete stratification of OE components not captured by calendar time which seems unlikely, or un-modeled interactions between FH and other covariates. Un-modeled interactions between FH and stratified residual confounders may exist and contribute but we found no evidence for this in plots of distributions by gender and race/ethnicity (Supplementary Fig S1) or in an analysis of smoking status against FH. Discrete inheritance of genetic variants with large effects is the most likely explanation for multi-modality in the FH effects on the BMI distribution .

Approximately 40% of the adjusted BMI variance in FH_1 was accounted for by between-modes variance (Fig 3B) but this represents a lower bound since the identified modes are likely to be synthetic ie composed of a range of effect sizes due to rare variants at different loci. Indications of fine structure within the broad central peak (Fig 2B inset) are suggestive. Examples of rare variants with large effects on BMI in adults (ß) are known from studies of candidate genes and monogenic obesity loci[26] while more recently a common variant in Samoans (EAF = 0.26, ß ≈ 1.4 kg/m^2), very rare in other populations[33], and an African-specific rare variant (EAF = 0.008, ß =4.6 kg/m^2) undetected in Europeans and Asians[34] have been identified by GWAS. Overall, ß in these nine examples ranges from 1.4- 9 kg/m^2 and a similar range in effect sizes in the NHANES sample would contribute substantially to the within-mode variance estimated here. A combination of within-subject variance (~5%[35]) with polygenic variance (~ 5%[4]) sets a lower bound for within-modes variance and hence the upper bound for between modes, implying that between 40% and 90% of total variance in FH_1 may be attributed to large genetic effects.

G x OE

FH_1 is a prevalent (36%) and powerful determinant of the rate of change of mean BMI and its dispersion over time in the ANCOVA models, accounting for 62% of the BMI trend and 60% of the

BMI SD trend in this sample over 1999-2014. Consistent results were obtained in the CQR models with $ß_{OLS}$ and $ß_{MR}$ in FH_1 approximately double those in FH_0. Under a polygenic model the familial risk would be distributed normally over FH_1 which would then be a marker of a large fraction of the at-risk population. However under the Mendelian model supported here genetic risk would segregate in families and only approximately 50% of FH_1 would acquire the excess familial risk and only ~18% of the sample would then account for ~60% of the trends. Individuals with a family history DM_1 must represent a fraction of individuals with elevated genetic obesity risk and it is likely that the remainder, particularly those with a family history of obesity without DM, would substantially increase the genetic component of the trends consequent to the high heritability of BMI[22]. This Mendelian model is internally consistent in estimates of risk allele frequencies (q) in FH_1, FH_0 and DM_1 (Table 2) and in comparisons of FH_1 effect sizes in cross section (q, ~50%) and over time (ß, ~60%) (Supplementary Table S2). Our results support the proposition that the largest part, and perhaps all, recent trends in mean and dispersion of BMI are due to a minor subset of individuals with elevated genetic susceptibility to OE.

Limitations

The design and interpretation of fits to parametric mixture distribution models involves choices concerning the number of components, parameter starting values and algorithms, and fit to a specific model cannot be taken in isolation as support for its structural validity. We base our choice and structural interpretation of 3-component normal mixture model fits and parameters on the *a priori* hypothesis of Mendelian segregation of obesity risk in families[21] supported by the non-parametric distributional plots (Fig 1B,C) and CQR analysis (Fig 2A,B insets). Like Abadi et al[14], we interpret interactions in the CQR analysis as predominantly G x OE although a contribution from G x G interactions cannot be excluded. Our interpretation is supported by the effects on the interaction of including calendar time in the CQR model (Fig 2B). Other limitations discussed above include our

inability to exclude discreet stratification of OE and the possible influence of unmeasured/unknown confounders of FH effects.

Conclusions

We conclude that a Mendelian model of individually rare but collectively common genetic risk variants with large effects interacting with OE provides a plausible quantitative explanation for recent trends in obesity and should be favored over a predominantly polygenic model which does not. The evidence for a predominant role for polygenes[13] can appear to be strong (eg "Polygenic obesity is the most common form of obesity in modern societies..."[36]) but recent interpretations seek to explain the still missing heritability in obesity in terms of unidentified large genetic effects and G x OE[4, 37] and recommend a renewed focus on family-based designs and on specific populations in which large effect variants may be enriched[33, 34]. Our results strengthen that view by showing that a model based on unidentified segregating variants with large effects interacting with OE can account for the largest part of the secular trend in BMI and its dispersion in a large population-based sample. We can't predict the clinical or public health changes which might follow from these findings but the results should help to refocus research away from the under-performing but entrenched big idea[38] that is polygenic obesity.

Supplemental Data:

Supplementary Methods: Race/Ethnicity coding

Supplementary Table S1: BMI-adjustment model parameters

Supplementary Table S2: Comparison of effects of diabetes family history on secular trends in adjusted BMI and its SD and on calculated risk allele frequencies (q) in non-diabetic participants.

Supplementary Fig S1: Kernel-smoothed adjusted BMI densities by FH, gender and race/ethnicity

Supplementary Fig S2: Secular trends in log-transformed adjusted BMI mean and SD

Declaration of Interests

The authors declare no competing interests.

Web Resources

NHANES data: https://www.cdc.gov/nchs/nhanes/index.htm accessed 25 Aug 2017

References

1. Di Cesare, M., Bentham, J., Stevens, G.A.A., Zhou, B., Danaei, G., Lu, Y., Bixby, H., Cowan, J., M., Riley, L.M. et al. (2016) Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet, *387*, 1377–1396.
2. Krishna, A., Razak, F., Lebel, A., Davey Smith, G., and Subramanian, S.V. (2015) Trends in group inequalities and interindividual inequalities in BMI in the United States, 1993–2012. Am J Clin Nutr, *101*, 598–605.
3. Silventoinen, K., Jelenkovic, A., Sund, R., Yokoyama, Y., Hur, Y.-M., Cozen, W., Hwang, A.E., Mack, T.M., Honda, C., Inui, F. et al. (2017) Differences in genetic and environmental variation in adult BMI by sex, age, time period, and region: an individual-based pooled analysis of 40 twin cohorts. Am J Clin Nutr, *106*, 457–466.
4. Loos, R.J. (2018) The genetics of adiposity. Curr Opin Genet Dev, *50*, 86–95.
5. Rokholm, B., Silventoinen, K., Ängquist, L., Skytthe, A., Kyvik, K.O., and Sørensen, T.I. (2011) Increased genetic variance of BMI with a higher prevalence of obesity. PLoS One, *6*, e20816.
6. Franks, P.W., and McCarthy, M.I. (2016) Exposing the exposures responsible for type 2 diabetes and obesity. Science, *354*, 69–73.
7. Hall, K.D. (2018) Did the Food Environment Cause the Obesity Epidemic? Obesity, *26*, 11–13.
8. Kivimaki, M., Stenholm, S., and Kawachi, I. (2015) The widening BMI distribution in the United States. Am J Clin Nutr, *101*, 1307–1308.
9. Jenkins, A.B., and Campbell, L.V. (2015) Variation in Genetic Susceptibility Drives Increasing Dispersion of Population BMI. Am J Clin Nutr, *101*, 1308.

10. Razak, F., Smith, G.D., Krishna, A., Lebel, A., and Subramanian, S.V. (2015) Variation in genetic susceptibility drives increasing dispersion of population BMI Reply. Am J Clin Nutr, *101*, 1308–1309.

11. Speakman, J.R., Loos, R.J.F., O'Rahilly, S., Hirschhorn, J.N., and Allison, D.B. (2018) GWAS for BMI: a treasure trove of fundamental insights into the genetic basis of obesity. Int J Obes, *42*, 1524–1531.

12. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J. et al. (2015) Genetic studies of body mass index yield new insights for obesity biology. Nature, *518*, 197–206.

13. Khera, A.V., Chaffin, M., Wade, K.H., Zahid, S., Brancale, J., Xia, R., Distefano, M., Senol-Cosar, O., Haas, M.E., Bick, A. et al. (2019) Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell, *177*, 587–596.e9.

14. Abadi, A., Alyass, A., Robiou du Pont, S., Bolker, B., Singh, P., Mohan, V., Diaz, R., Engert, J.C., Gerstein, H.C., Anand, S.S. et al. (2017) Penetrance of Polygenic Obesity Susceptibility Loci across the Body Mass Index Distribution. Am J Hum Gen, *101*, 925–938.

15. Nagpal, S., Gibson, G., and Marigorta, U. (2018) Pervasive Modulation of Obesity Risk by the Environment and Genomic Background. Genes, *9*, doi: 10.3390/genes9080411.

16. Hemminki, K., Li, X., Sundquist, K., and Sundquist, J. (2010) Familial risks for type 2 diabetes in Sweden. Diabetes Care, *33*, 293–297.

17. Willemsen, G., Ward, K.J., Bell, C.G., Kaare Christensen, K., Bowden, J., Dalgard, C., Harris, J.R., Kaprio, J., Lyle, R., Magnusson, P.K.E. et al. (2015) The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet, *18*, 762–771.

18. Ghosh, A., Liu, T., Khoury, M.J., and Valdez, R. (2010) Family History of Diabetes and Prevalence of the Metabolic Syndrome in U.S. Adults without Diabetes: 6-Year Results from

the National Health and Nutrition Examination Survey (1999-2004). Public Health Genom, *13*, 353–359.

19. Tirosh, A., Shai, I., Afek, A., Dubnov-Raz, G., Ayalon, N., Gordon, B., Derazne, E., Tzur, D., Shamis, A., Vinker, S. et al. (2011) Adolescent BMI trajectory and risk of diabetes versus coronary disease. N Engl J Med, *364*, 1315–1325.

20. Scott, R.A., and Consortium, I. (2013) The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study. Diabetologia, *56*, 60–69.

21. Jenkins, A.B., Batterham, M., Samocha-Bonet, D., Tonks, K., Greenfield, J.R., and Campbell, L.V. (2013) Segregation of a latent high adiposity phenotype in families with a history of type 2 diabetes mellitus implicates rare obesity-susceptibility genetic variants with large effects in diabetes-related obesity. PLoS One, *8*, e70435.

22. Stryjecki, C., Alyass, A., and Meyre, D. (2018) Ethnic and population differences in the genetic predisposition to human obesity. Obes Rev, *19*, 62–80.

23. Xu, G., Liu, B., Sun, Y., Du, Y., Snetselaar, L.G., Hu, F.B., and Bao, W. (2018) Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study. BMJ, *362*, k1497.

24. Llewellyn, C.H., Trzaskowski, M., Plomin, R., and Wardle, J. (2014) From modeling to measurement: developmental trends in genetic influence on adiposity in childhood. Obesity, *22*, 1756–1761.

25. Rask-Andersen, M., Karlsson, T., Ek, W.E., and Johansson, A. (2017) Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. PLoS Genet, *13*, e1006977.

26. Jenkins, A.B., and Campbell, L.V. (2014) Future management of human obesity: understanding the meaning of genetic susceptibility. Adv Genomics Genet, *4*, 219–232.

27. Müller, M.J., Geisler, C., Blundell, J., Dulloo, A., Schutz, Y., Krawczak, M., Bosy-Westphal, A., Enderle, J., and Heymsfield, S.B. (2018) The case of GWAS of obesity: does body weight control play by the rules. Int J Obes,

28. R Development Core Team (2016) R: A Language and Environment for Statistical Computing. ed. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/),

29. Sheather, S.J., and Jones, M.C. (1991) A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. J Roy Stat Soc B Met, *53*, 683–690.

30. Koenker, R. (2017) Quantile Regression: 40 Years On. Ann Rev of Econ, *9*, 155–176.

31. Briollais, L., and Durrieu, G. (2014) Application of quantile regression to recent genetic and -omic studies. Hum Genet, *133*, 951–966.

32. Manyanga, T., Sellers, E.A.C., Wicklow, B.A., Doupe, M., and Fransoo, R. (2016) Is the change in body mass index among children newly diagnosed with type 1 diabetes mellitus associated with obesity at transition from pediatric to adult care. Pediatric Diabetes, *17*, 599–607.

33. Minster, R.L., Hawley, N.L., Su, C.T., Sun, G., Kershaw, E.E., Cheng, H., Buhule, O.D., Lin, J., Reupena, M.S., Viali, S. et al. (2016) A thrifty variant in CREBRF strongly influences body mass index in Samoans. Nat Genet, *48*, 1049–1054.

34. Chen, G., Doumatey, A.P., Zhou, J., Lei, L., Bentley, A.R., Tekola-Ayele, F., Adebamowo, S.N., Baker, J.L., Fasanmade, O., Okafor, G. et al. (2017) Genome-wide analysis identifies an African-specific variant in SEMA4D associated with body mass index. Obesity, *25*, 794–800.

35. Wormser, D., White, I.R., Thompson, S.G., and Wood, A.M. (2013) Within-person variability in calculated risk factors: Comparing the aetiological association of adiposity ratios with risk of coronary heart disease. Int J Epidemiol, *42*, 849–859.

36. Albuquerque, D., Nóbrega, C., Manco, L., and Padez, C. (2017) The contribution of genetics and environment to obesity. Br Med Bull, *123*, 159–173.

37. Saeed, S., Arslan, M., and Froguel, P. (2018) Genetics of Obesity in Consanguineous Populations: Toward Precision Medicine and the Discovery of Novel Obesity Genes: Obesity and Consanguinity. Obesity, *26*, 474–484.

38. Joyner, M.J., Paneth, N., and Ioannidis, J.P. (2016) What Happens When Underperforming Big Ideas in Research Become Entrenched? JAMA, *316*, 1355–1356.

Figure legends

Fig 1: Distribution of adjusted BMI in non-diabetic (DM_0) and diabetic (DM_1) participants in combined NHANES 1999-2000, 2005-2006 and 2013-2014 cycles. A) Full non-diabetic sample (n=9102) binned by quantiles (n=20) with superimposed kernel-smoothed and fitted densities in normal and log-normal models. B) Main: kernel-smoothed adjusted BMI density by FH status. Inset: density ratios (FH_1 / FH_0) ± SEM obtained by bootstrap resampling by quantiles of the full sample compared to predictions of normal (solid line) and log-normal (dotted line) mixture models with p associated with lack-of-fit testing (p_{lof}). C) Kernel -smoothed adjusted BMI density in non-diabetic FH_1 (n=3297) compared to diabetic participants (n=793).

Fig 2: Conditional quantile regression effects of diabetes family history (FH) on adjusted BMI in non-diabetic participants in models consisting of FH alone (A) and in interaction with continuous calendar time (B-D). The main panels show the effect sizes ($ß_{1-3}$) by quantile with 95% CI (grey-shaded area), the OLS estimate of the average effect (solid green line) with 95% CI (dotted green lines) with p-values (p_{ANOVA}) from anova tests for equality of $ß_i$ across quantiles and the meta-regression fits ± 95%CI (magenta lines) with $ß_{MR}$ ± SEM. The insets in panels A and B show the patterns of residuals ($\Delta ß_{1i}$) ± residual SEM from linear OLS fits of $ß_{1i}$ (FH_1) to $ß_{0i}$ (FH_0) with 95% CI on the fits (dotted lines) around the lines $\Delta ß_1 = 0$ representing perfect fits, and with the p value (p_{DW}) from a Durbin-Watson test of autocorrelation of residuals. CQR estimates of the SEM of $ß_{0i}$ are also depicted but are mostly obscured by the point symbols.

Fig 3: A) Adjusted BMI density in FH_1 by quantile (grey bars) and kernel-smoothed (black line) with fits to a three component normal mixture distribution. B) Estimated contributions in FH_1 of the components of the mixture distribution to the prevalence (mixture coefficients, λ) and variance of adjusted BMI.

Fig 4: Effects of diabetes family history ($FH_{0/1}$) on linear secular trends in age-, gender- and race/ethnicity-adjusted BMI mean ± sem (A) and standard deviation ± sem (B). Parameter estimates with 95% CI were obtained in ANCOVA models by stratified bootstrap resampling of all non-diabetic individuals (see Methods). Dotted lines enclose 95% CI on fitted values at each point; ß = regression slope vs. time (kg/m^2 per year); $\Delta ß = ß_{FH1} - ß_{FH0}$ (95%CI).

Tables

Table 1: Participants by diabetes status (DM_0/DM_1)

	NHANES cycle				
	All cycles	**1999-2000**	**2005-2006**	**2013-2014**	**p†**
DM_0					
n	9102	2865	3076	3161	-
Gender (F%)	53	55	54	52	0.06
Race/Ethnicity (%) (MA/OH/NHW/NHB)¥	23/8/46/23	30/7/44/20	24/4/49/24	16/11/48/25	7.8×10^{-56}
Age (yr): Mean	40.5	40.7	39.4	41.5	4.5×10^{-9}
SD	13.2	13.2	13	13.2	0.63
BMI (kg/m^2)*: Mean	28.7	28.3	28.6	29.2	4.6×10^{-9}
SD	6.6	6.3	6.4	7.1	1.2×10^{-11}
Diabetes Family History (Y%)	36	29	42	38	4.1×10^{-25}
Current smoking (%, Y/N/missing)	25/20/55				-
DM_1					
n (%)	793 (8.0)	211 (6.9)	252 (7.6)	330 (9.5)	0.0003
Age at Diagnosis (yr)	43.6	44.3	42.9	43.6	0.47

† Cycle effects (p) by ANOVA (age), ANCOVA (BMI), Bartlett's test (SD's) and Chi-squared test for categorical variables.

¥ MA = Mexican American, OH = Other Hispanic, NHW = Non-Hispanic White, NHB = Non-Hispanic Black

* Adjusted for age, gender and race/ethnicity in a linear model (see Table S1).

Table 2: Three-component normal mixture distribution fits to adjusted BMI by FH and DM status[¥]

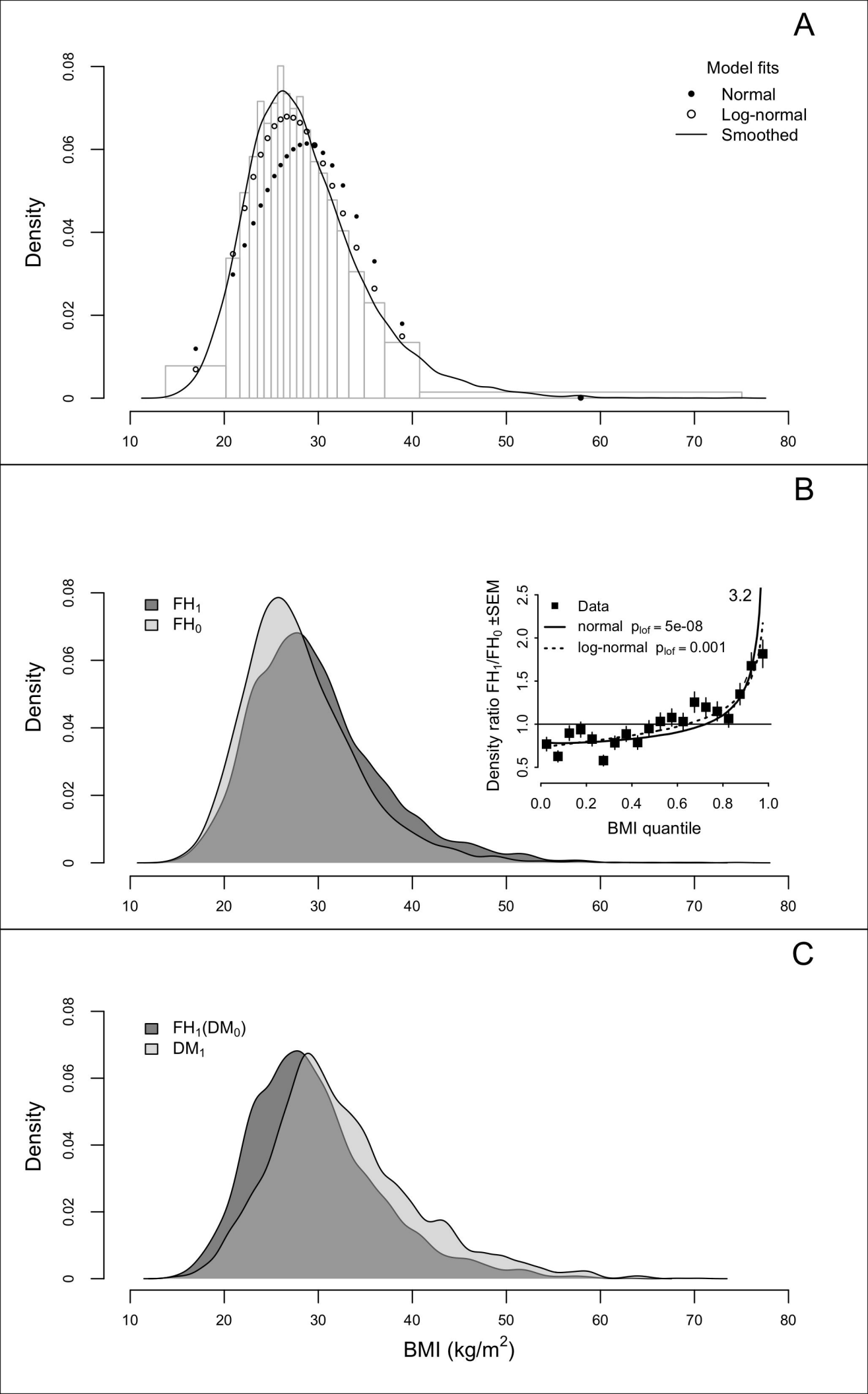
		DM_0		DM_1
	Component	FH_0	FH_1	
mean	1	*	25.8±1.0	*
	2	*	32.1±1.8	*
	3	*	40.6±2.5	*
sd	1	3.8	3.8±0.9	3.7
	2	5.7	5.1±1.2	4.5
	3	10	8.2±0.8	8.4
λ	1	0.67	0.52±0.16	0.28
	2	0.29	0.37±0.15	0.47
	3	0.04	0.11±0.06	0.25
q[†]	-	0.18	0.30±0.10 (0.37)[§]	0.49

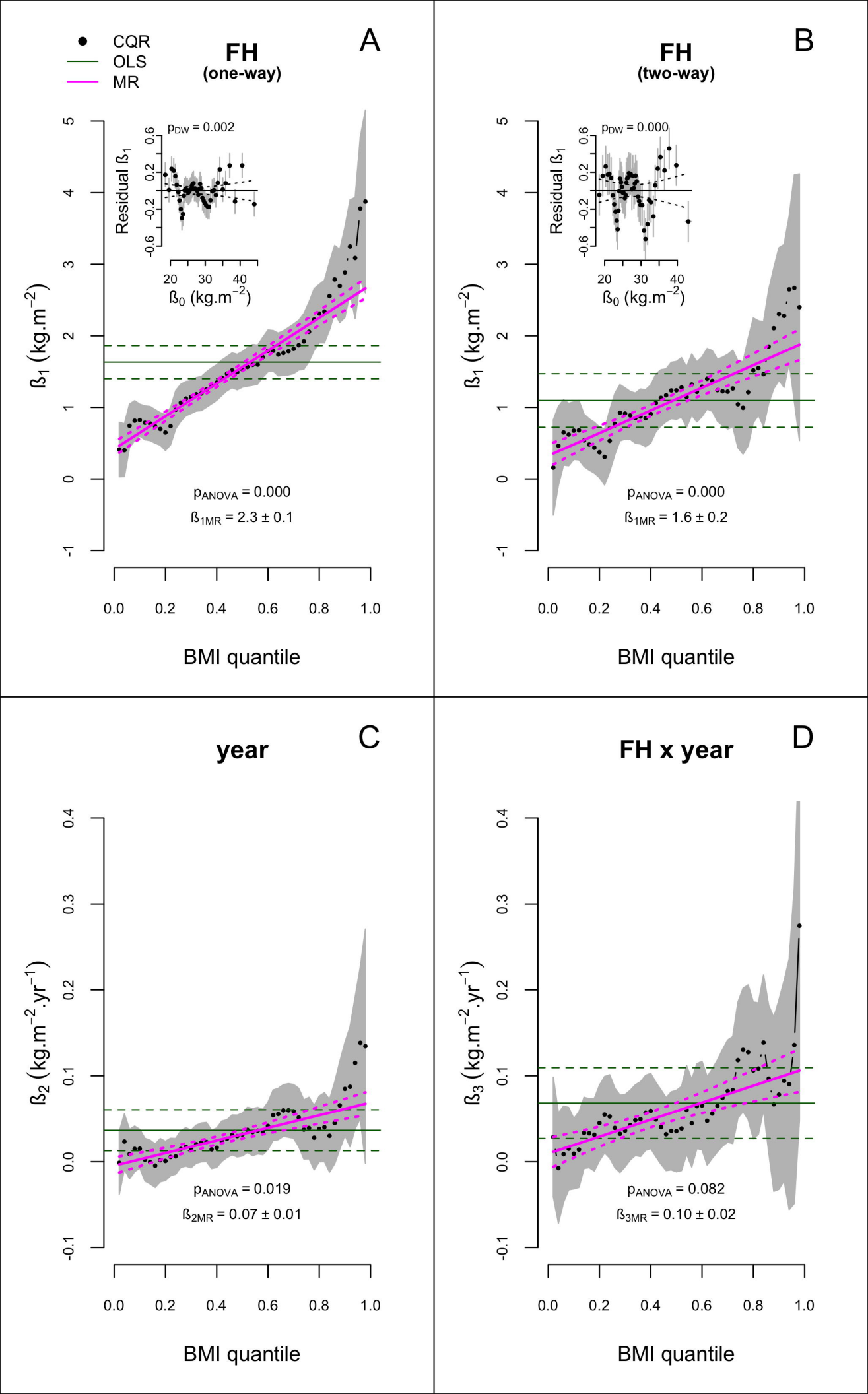
[¥] ± bootstrap standard error for FH_1 only

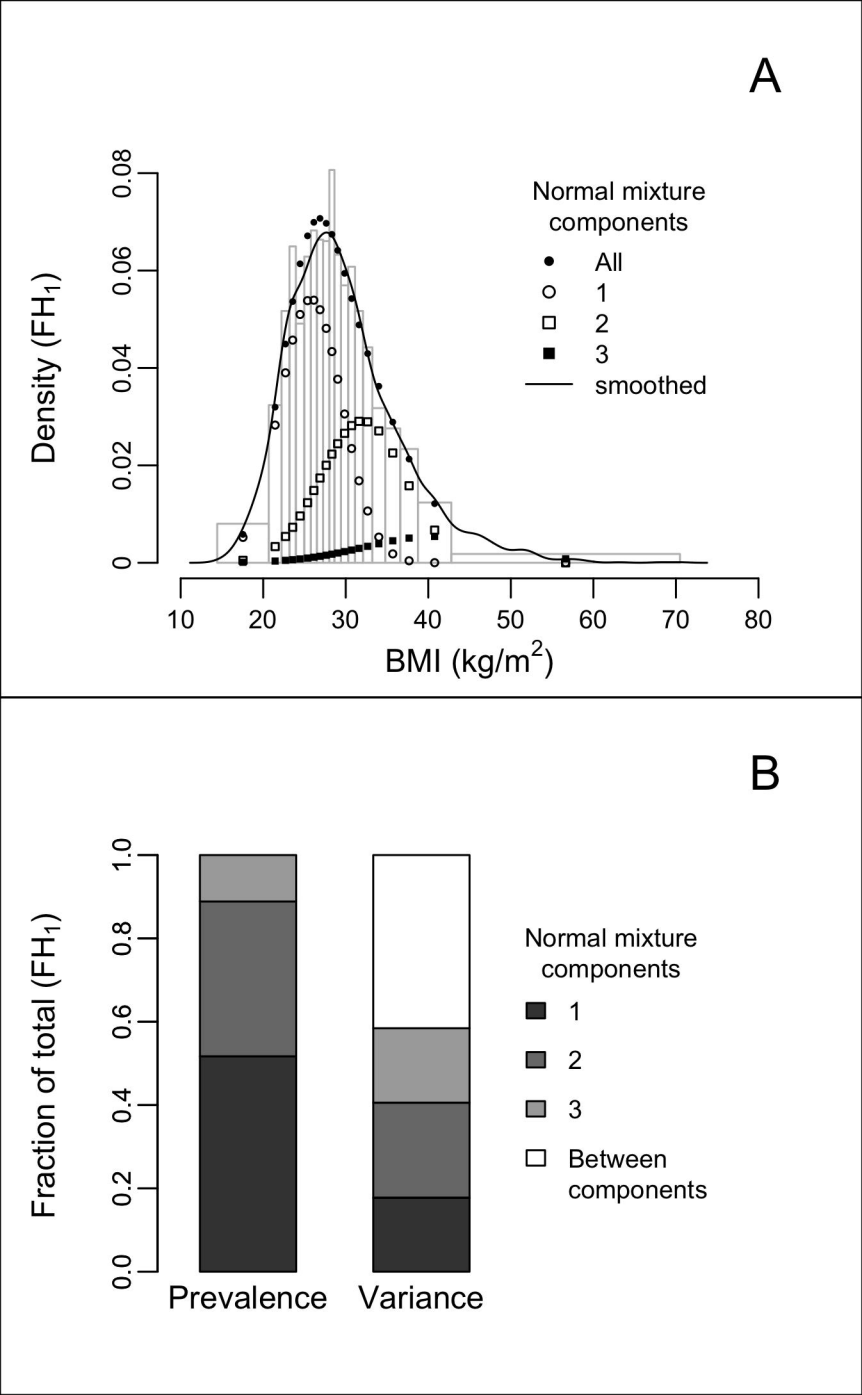
* Means in FH_0 and DM_1 constrained to fitted values in FH_1

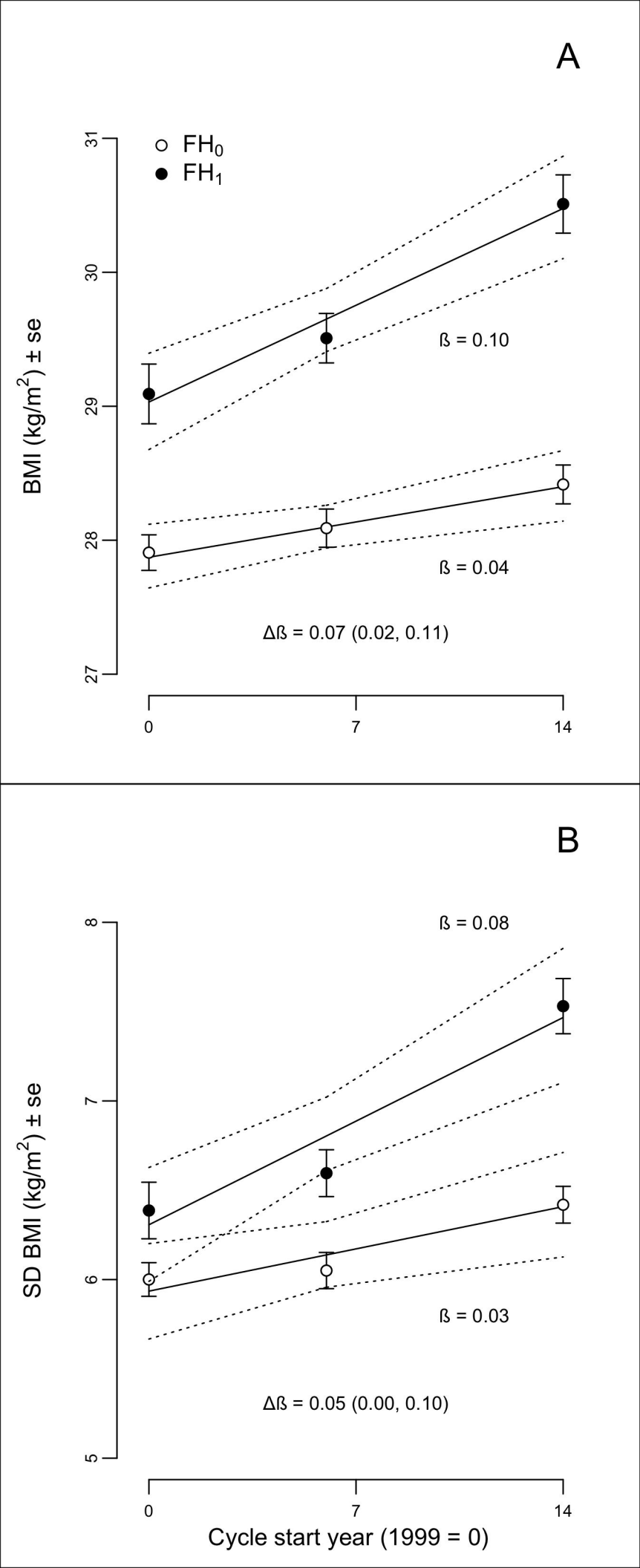
[†] Calculated risk allele frequency in an additive Mendelian model of large effects

[§] Predicted from DM_1 mating randomly into the full sample.


A
Model fits
Normal
Log-normal
Smoothed
Density
0
0.02
0.04
0.06
0.08
10
20
30
40
50
60
70
80
B
FH1
FH0
Density
Density ratio FH1/FH0 ±SEM
Data
normal plof = 5e-08
log-normal plof = 0.001
3.2
0.5
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0
BMI quantile
C
FH1(DM0)
DM1
BMI (kg/m2)


CQR
OLS
MR
FH
(one-way)
A
pDW = 0.002
Residual ß1
ß0 (kg.m−2)
ß1 (kg.m−2)
pANOVA = 0.000
ß1MR = 2.3 ± 0.1
BMI quantile
FH
(two-way)
B
pDW = 0.000
Residual ß1
ß0 (kg.m−2)
ß1 (kg.m−2)
pANOVA = 0.000
ß1MR = 1.6 ± 0.2
BMI quantile
year
C
ß2 (kg.m−2.yr−1)
pANOVA = 0.019
ß2MR = 0.07 ± 0.01
BMI quantile
FH x year
D
ß3 (kg.m−2.yr−1)
pANOVA = 0.082
ß3MR = 0.10 ± 0.02
BMI quantile


A
Density (FH1)
0
0.02
0.04
0.06
0.08
10
20
30
40
50
60
70
80
BMI (kg/m2)
Normal mixture components
All
1
2
3
smoothed
B
Fraction of total (FH1)
0.0
0.2
0.4
0.6
0.8
1.0
Prevalence
Variance
Normal mixture components
1
2
3
Between components


A
FH0
FH1
BMI (kg/m2) ± se
31
30
29
28
27
ß = 0.10
ß = 0.04
Δß = 0.07 (0.02, 0.11)
0
7
14
B
SD BMI (kg/m2) ± se
8
7
6
5
ß = 0.08
ß = 0.03
Δß = 0.05 (0.00, 0.10)
0
7
14
Cycle start year (1999 = 0)