1   Title: The current obesity "epidemic": segregation of familial genetic risk in NHANES

2   cohort supports a major role for large genetic effects

3

4   Running title: Segregation of familial genetic obesity risk

5

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

7   [1] School of Medicine, University of Wollongong, Wollongong NSW 2522 Australia

8   [2] Statistical Consulting Centre, School of Mathematics and Applied Statistics, University of

9   Wollongong NSW 2522 Australia

10   [3] Diabetes Service, St Vincent's Hospital, Sydney, NSW 2010 Australia

11   [4] Diabetes and Metabolism Division, Garvan Institute of Medical Research, Darlinghurst, NSW 2010

12   Australia

13   * Corresponding author

14   E-mail: ajenkins@uow.edu.au

15

16   The authors declare no competing financial or non-financial interests.

17   Abstract

18   **Background/Objectives:** The continuing increase in many countries in adult body mass index (BMI

19   kg/m^2) and its dispersion is contributed to by interactions between genetic susceptibilities and an

20   increasingly obesogenic environment. Whether population susceptibility to obesogenic environments

21   is mainly determined by a subgroup with high genetic susceptibility or susceptibility is more evenly

22   distributed throughout the population is unresolved, due to uncertainties around relevant genetic and

23   environmental architecture. We aimed to test the predictions of a Mendelian genetic architecture

24   based on collectively common but individually rare large-effect variants and its ability to account for

25   current trends in a large population-based sample.

26   **Subjects/Methods:** We studied publicly available adult BMI data (n = 9102) from 3 cycles of

27   NHANES (1999, 2005, 2013) adjusted for age, gender and race/ethnicity. A first degree family

28   history of diabetes (FH) served as a binary marker (FH_0/FH_1) of genetic obesity susceptibility. We

29   tested for multi-modality in BMI non-parametrically using a runs tests in conditional quantile

30   regression (CQR) models of FH effects, obtained parametric fits to a Mendelian model in FH_1, and

31   estimated FH-environment interactions in CQR models and in ANCOVA models incorporating

32   secular time.

33   **Results:** A unimodal FH effect on BMI was excluded ($p \leq 0.0001$) in CQR models and parametric fits

34   to a Mendelian model in FH_1 identified 3 modes at 25.8±1.0 (SEM), 32.1±1.8 and 40.6±2.5 kg/m^2.

35   Mode separation accounted for ~40% of BMI variance in FH_1 providing a lower bound for the

36   contribution of large effects. CQR analysis identified strong interactions between FH and

37   environmental factors ($p \leq 0.01$) and FH_1 accounted for ~60% of the secular trends in BMI and its SD

38   in ANCOVA models.

39   **Conclusions:** Multimodality in the FH effect is inconsistent with a predominantly polygenic small

40   effect architecture. We conclude that large genetic effects interacting with obesogenic environment

41   provide a better, quantitative explanation for current trends in BMI.

2

Introduction

42

43   The recent and continuing increase in the global mean adult BMI, first seen in high income countries,

44   is now apparent in most countries across a wide range of ethnic composition and socio-economic

45   conditions (1) and is accompanied by increases in measures of dispersion (2, 3). Although BMI is

46   known from family-based studies to be under strong genetic influences (4) population genetic

47   backgrounds have been effectively constant over this time, implying that BMI trends are driven by

48   change in environmental factors (obesogenic environment, OE). Evidence from twin studies, which

49   demonstrate increased genetic variance over time, supports an important role for interactions between

50   OE and genetic susceptibility (G x OE) on both mean and dispersion of BMI (3, 5), but how large a

51   role is not yet clear. Defining the role of G x OE in "epidemic" obesity, and hence of genetic

52   susceptibility itself, is hindered by problems of measurement and modeling of interactions (6) and by

53   uncertainty around both the genetic architecture (4) and the exact nature of the environmental drivers

54   (7). Whether a population's susceptibility to OE is predominantly determined by a subgroup with high

55   genetic susceptibility or is more evenly spread within populations is unresolved despite important

56   implications for the management of obesity and related disorders at population and individual levels

57   (2, 8–10).

58
59   The genetic variants responsible for obesity susceptibility remain largely unknown. Genome-wide

60   association studies (GWAS) have identified significant associations with >200 markers with small

61   effects on BMI (polygenes), together explaining only approximately 3-4 % of total variance

62   compared to family-based heritabilities (h^2) of 50-75% (11). Few causative mechanisms responsible

63   for these phenotypically weak associations are known (4). The sources of the h^2 unaccounted for by

64   GWAS are uncertain; proposals include overestimation of h^2, large numbers of common genetic

65   variants with small, statistically insignificant effects on phenotypes (12, 13) and importantly,

66   candidates not tested in most GWAS. Among the latter are rare genetic variants with large phenotypic

3

67   effects and G x OE interactions (4). Recently significant G x OE interactions have been detected in

68   individual GWAS loci and in composite genetic risk scores, which explain little of the missing

69   component of h^2 (14, 15).

70
71   A family history of diabetes (FH) is a potent, predominantly genetic (16, 17) risk factor for diabetes

72   diagnosis (DM) and for obesity-related phenotypes (18–21) consistent with the strong association

73   between type 2 DM and overweight/obesity. Familial effects on obesity-related phenotypes in adults

74   are also predominantly genetic (3, 22), so to the extent that the DM generating FH is of type 2

75   (approximately 94% of DM in the US population (23)), FH is a prevalent and readily obtained marker

76   of genetic susceptibility both to diabetes and to the obesity commonly preceding it. We have reported

77   evidence from a small sample of a multi-modal effect of FH on a composite adiposity index

78   consistent with segregation in families of discrete obesity risk (21). Polygenic risk scores (PRS)

79   based on large numbers of small effects are expected to be, and appear to be, normally-distributed

80   (24, 25) and cannot account for familial segregation of risk. The present work is based on the

81   alternative hypothesis that individually rare, but collectively common, genetic variants with large

82   phenotypic effects are the source of most of the missing h^2 and of most of G x OE, and that their

83   effects can be detected through analyses of phenotypic segregation in high-risk families (26).

84
85   The Continuous National Health and Nutrition Examination Survey (NHANES) is a continuing

86   (1999-) large-scale population-based survey incorporating an index of adiposity (Body Mass Index,

87   BMI) and first-degree FH (FH_0/FH_1) together with potential covariates and confounders. Although

88   BMI has recognized limitations as an adiposity phenotype (26, 27) it is the basis for most large-scale

89   genetic studies and like other authors, we assume that a large enough scale and appropriate modeling

90   of covariates will reduce effects of imprecision and bias (11). We aimed to test in a large multi-cycle

91   NHANES sample for the presence of familial segregation of genetic risk and to estimate the

92   contribution of FH, and by extension all discrete genetic risk, to recent secular trends in adult BMI.

93   Our results support a predominant role for large genetic effects interacting with OE in the obesity

94   "epidemic".

95

Data and Methods

Data

98   We used data from the 1999-2000, 2005-2006 and 2013-2014 cycles of NHANES

99   (https://www.cdc.gov/nchs/nhanes/index.htm accessed 25 Aug 2017). We extracted records for

100   participants age 20-65 years with non-missing gender, race/ethnicity, BMI, diabetes family history

101   (FH) and diabetes diagnosis data, and current smoking status if available. The definitions of two

102   fields changed over the sampling period: 1) Diabetes family history was defined in terms of 1° and 2°

103   relatives in 1999-2000 but by 1° relatives only in subsequent cycles. We recoded the 1999-2000

104   diabetes family history data to conform to the later definition using the separately collected data for

105   affected parents and siblings. 2) The self-identified race/ethnicity field (RIDRETH1) code was used

106   excluding other races (OR) to maintain consistency across cycles (Supplementary Methods). We

107   excluded from the primary analyses subjects diagnosed with diabetes because of possible

108   confounding by effects of either diabetes or diabetes therapies on BMI. The resulting data set is

109   summarized in Table 1.

Statistical analyses

111   *Approach*

112   We treat the data as a convenience sample and took no account of the sampling weights provided by

113   NHANES to permit nationally representative estimates. Our results are not intended to be

114   representative of the US population.

115   Our primary analyses are based on non-parametric visualization (kernel-smoothing) and analyses

116   (conditional quantile regression, CQR) of distributions requiring no prior distributional assumptions.

117   Parametric fits to multimodal distributions were then used to quantify the contributions of the

118   predicted large genetic effects model. FH(0/1) is treated as a binary genetic risk marker and calendar

119   time as a continuous surrogate of OE. Effects of OE interacting with FH were assessed in CQR

120   models, and also in least-squares ANCOVA models using bootstrap resampling to minimize

121   distributional assumptions in the calculation of effect size estimates and errors. All analyses were

122   performed using R 3.3.1 (28).

123

124   *Summary statistics*

125   Heterogeneity of the samples across cycles was assessed by Chi square test for categorical variables

126   and by one-way ANOVA for age. Effects on BMI were assessed by ANCOVA against continuous

127   time (yr = calendar start year - 1999). Effects on phenotype SD's were assessed by Bartlett's test in

128   one-way ANOVA models.

129

130   *Adjustment*

131   Prior to analysis BMI was adjusted for effects of age, gender and race/ethnicity in a linear model (age

132    $+ age^2 +$ race x gender). The adjustment model accounted for 4 % of the total variance in BMI

133   (Supplementary Table S1).

134

135   *Distributions*

136   Visualization

137   The effect of FH on the distribution of adjusted BMI was visualized using kernel-smoothed density

138   estimates by FH status (R base function density). The degree of smoothing is controlled by the

139   bandwidth parameter (bw) which was obtained in the full non-diabetic data set (bw = 0.99) from a

140   measure of the dispersion of the data (29). This produces a continuous distribution function and is

141   widely used to visualise features of potential interest which may be obscured in histograms. The

142   credibility of apparent effects of FH on the shape of the distribution was assessed by post-hoc

143   analysis of density differences ($FH_1 - FH_0$) by quantiles (n=20) of the full sample. Mean density

144   differences with SE were obtained by quantile by bootstrap resampling with replacement (1000

145   draws, stratified by FH status with resample sizes = strata sizes) and tested against zero using a Z-

146   test.

147

148   Conditional Quantile Regression (CQR)

149   Conditional quantile regression is a powerful tool for analyzing the effect of covariates on

150   distributions without assumptions of distributional shape. In contrast to ordinary least-squares (OLS)

151   regression which characterizes effects on global features of a distribution, CQR analyses local effects

152   of covariates independently at any specified quantiles and can detect and test variations in covariate

153   effects across quantiles. Originating in econometrics (30) it is now used in other areas including

154   genetics (31). In the CQR framework developed by Abadi et al (14) for analysis of genomic markers,

155   trends in effect sizes across quantiles represent interactions between genetic effects and unobserved

156   environmental and/or genetic factors. We treat FH as a binary genetic risk marker (FH_0/FH_1) and a

157   linear trend in quantile regression coefficients (β_{1i}) across quantiles (τ_i) represents summed linear

158   interactions of FH with unobserved factors. We analysed the effects of FH on adjusted BMI by CQR

7

159   using the R package quantreg. The effect of all interactions on the FH effect was tested in in a 2

160   parameter linear model:

161   for each quantile τ_i in y,

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

163   where y($\tau_i \mid$ FH=fh) = the ith quantile of adjusted BMI conditional on the value of FH (0,1), the

164   intercept β_{0i} is the ith quantile value in FH_0 and β_{1i} is the FH effect size in quantile i.

165

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

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

168   where yr = cycle start year – 1999, β_{0i} is the ith quantile value in FH_0 at yr =0 and β_{2i} and β_{3i} are CQR

169   coefficients for time and time*FH interaction effects in quantile i.

170

171   Equality of CQR parameter effect sizes across quantiles was tested using a Khmaladze test (32) (R

172   package quantreg).

173

174   The shape of the interaction relationship on the BMI scale was assessed in an OLS relationship

175   between quantile coefficients β_1 and β_0, which represents a linear model of a transformed quantile-

176   quantile plot between FH_1 and FH_0 after adjustment for covariates, and models a unimodal effect.

177   Residuals from the relationship were tested against randomness using a Wald-Wolfowitz runs test (R

178   package randtests), a simple, general non-parametric test of randomness in ordered binary

179   observations. The number of runs is compared to expectations under a normal approximation to the

180   random sampling distribution. Randomness is rejected in a two-sided test if the number of observed

181   runs is too high indicating high frequency oscillations, or too low indicating longer structures. We

182   applied a left-sided runs test to residuals from regression relationships testing for large structures and

8

183 interpret significance as excluding a unimodal effect and therefore as evidence of multiple modes in

184 the effect.

185

186 Parametric fits

187 We obtained fits to a 3-component normal mixture distribution representing a simple Mendelian

188 model of fixed genetic effect using an expectation-maximization algorithm (normalmixEM function

189 in the R package mixtools) . The models are characterised by the fitted means (μ_i), standard

190 deviations (σ_i) and mixing proportions (λ_i) of the three component distributions. Full model fits

191 were obtained in FH_1 but were not obtainable in FH_0 or DM_1 groups and we constrained μ_i in those

192 groups to values estimated in FH_1 in order to obtain comparable estimates of σ_i and λ_i. Risk allele

193 frequencies (q) under an additive Mendelian model of large genetic effects were calculated from the

194 fitted λ_i:

195 $$q = 0.5*\lambda_2 + \lambda_3$$

196 where λ_2 and λ_3 represent the proportions of carriers of 1 and 2 risk alleles respectively.Within-

197 sample consistency of calculated q across the three groups analysed was assessed by comparing

198 fitted q_{FH_1} with the prediction from random mating of DM_1 into the full sample:

199 $$\text{predicted } q_{FH_1} = (q_{DM} + n\text{-weighted mean}(q_{DM}, q_{FH_1}, q_{FH_0}))/2$$

200 *Secular trends*

201   Effects of diabetes family history status (FH_0, FH_1) and continuous time (yr = calendar start year -

202 1999) on adjusted BMI and its standard deviation (SD) were assessed in ANCOVA models of the

203 form:

204 $$y = \beta_0 + \beta_1 * FH + \beta_2*yr + \beta_3*FH*yr$$

205   where y = adjusted BMI mean or SD by FH status (0/1) and cycle, and yr = cycle start year – 1999.

206   Each OLS fit estimated 4 parameters from the 6 data points. Mean parameter estimates with 95% CI

207   were obtained by bootstrap resampling with replacement (1000 draws stratified by FH status and

208   cycle with resample sizes = strata sizes).

209

210   *Comparison of cross-sectional and secular trend effects*

211   Effects of FH on BMI distribution and on secular trends in BMI were compared by calculating the

212   contribution (%) of FH_1 to the effect in the full non-diabetic sample for calculated risk allele

213   frequency (q%) and to the slope (ß%) of the relationships between time and BMI in ANCOVA model

214   described above. Mean (± SE where possible) q% and ß% were calculated in the relevant bootstrap

215   samples.

216

217   Results

218   **Participant characteristics**

219   Data from 9102 non-diabetic subjects met the inclusion criteria, approximately equally distributed

220   across the 3 cycles. Gender balance varied little but there was a cycle effect in race/ethnicity, most

221   obvious in the reduced representation of MA in the two later cycles. Average age varied across cycles

222   but not its SD, while adjusted BMI and its SD showed linear trends with cycle time. FH_1 prevalence

223   was higher in the two later cycles compared to 1999-2000 as was DM_1 prevalence. Current smoking

224   status was predominantly missing in the data (55%) and was not included in the BMI adjustment

225   model. However smoking status was not related to FH whether analysed in the full data ($X^2 = 2.80$, 2

226   df , p = 0.25) or in those with non-missing smoking status ($X^2 = 0.43$, 1 df , p = 0.51), hence is

10

227 unlikely to confound analyses of FH effects. The mean age at diagnosis of DM (43.6 yr) is consistent

228 with predominantly type 2 DM in the sample.

Distributions

230 Visualization

231 Adjusted BMI in the non-diabetic sample showed an apparently unimodal distribution, right-skewed

232 compared to a normal model and closer to a log-normal model (Fig 1A). When visualized by FH

233 status (Fig 1B) the predicted multimodality in FH_1 was indicated with modes in the normal weight,

234 overweight and obese regions of the BMI distribution. In contrast FH_0 showed an apparently

235 unimodal distribution and a difference in shape between the two groups was supported by the pattern

236 of formal significance in the post-hoc analysis of density differences between groups (Fig 1B). BMI

237 distribution in the diabetic sample appeared to be depleted in the lower mode and enriched in the

238 upper modes compared to FH_1 (Fig 1C).

239

240 CQR

241 Analysis of the effect of FH status on the shape of the BMI distribution using CQR demonstrated

242 increasing FH effect size at higher levels of BMI ($p \leq 0.01$, Fig 2A, main panel), indicating strong

243 interactions between FH_1 and other variables not included in the model. FH_1 effect size ranged from

244 $< 1 \text{ kg/m}^2$ in the lower quantiles to $\sim 3 \text{ kg/m}^2$ in the upper quantiles, substantially different in both

245 regions to the OLS estimate (1.7 kg/m^2). The approximate linearity of FH_1 effects by quantile seen in

246 main panel of Fig 2A supports simple 1st order interactions, with that linearity emphasised by an OLS

247 analysis on the BMI scale between quantile coefficients β_1 (FH_1) and β_0 (FH_0) (slope $= 0.148 \pm 0.003$

248 (SE), $R^2 = 0.98$). Adjustment for main and interaction effects of calendar time in the model

249 (Supplementary Fig S2) weakened the trend in β_1 across quantiles ($p > 0.1$, Fig 2B main panel) and

250 the OLS relationship between β_1 and β_0 (slope $= 0.103 \pm 0.006$, $R^2 = 0.86$), supporting the conclusion

11

251   that calendar time is a strong surrogate of obesogenic environmental influences interacting with

252   genetic factors as represented by FH status.

253   While the overall OLS relationship between ß_1 and ß_0 was linear there was strong evidence for

254   additional non-linear structure in an analysis of residuals from the relationship (Fig 2A inset) which

255   found significant non-randomness (runs test, $p = 7.4 \times 10^{-5}$) attributable to a small number of broad

256   excursions. Adjustment for time effects accentuated the pattern (Fig 2B inset, runs test $p = 2.2 \times 10^{-5}$).

257   While these patterns in the conditional quantiles do not map directly onto the unconditional quantile

258   plots in Fig 1B, by rejecting the unimodal effect hypothesis they add strong support to the conclusion

259   that FH_1 has multimodal discrete effects on BMI. The pattern of residuals after adjustment for time

260   effects (Fig 2B inset) exhibits 3 broad peaks, with suggestions of finer structure particularly in the

261   central peak. The lower peak in the low-normal weight range may derive from the presence of type 1

262   diabetes family history in the sample, while the upper two are consistent with the predicted discrete

263   effects of FH_1 derived from type 2 diabetes.

264   Parametric analysis

265   The distribution of BMI in FH_1 appears consistent with a simple Mendelian model and fitting a 3-

266   component normal distribution model to the FH_1 data resulted in robust estimates of component

267   means and separations (Fig 3A), as well as mixing coefficients and SDs (Table 2). Approximately

268   50% of FH_1 occupied the upper two modes and separation between modes accounted for

269   approximately 40% of the total variance in adjusted BMI with the remainder assigned to dispersion

270   within modes (Fig 2B). Under a Mendelian model the variance due to mode separation represents a

271   lower bound on the contribution of large effects as some of the dispersion within modes represents

272   variance in effect sizes of individual contributing causal loci (see Discussion) which will contribute to

273   the ~60% of variance assigned to within-modes. Estimates of component SDs and mixing proportions

274   with component means, constrained for FH_0 and DM_1 to those identified in the FH_1 data, support

275   enrichment in the two upper components in FH_1 compared to FH_0 (48% vs. 33%) and more strongly

12

276   in DM_1 (72%). Predicted risk allele frequencies in FH_1 (q - Table 2) express these distributional

277   properties in Mendelian terms and show within-sample consistency in that q_{FH1} predicted from

278   random mating of DM_1 (0.37) is not different to the fitted estimate (0.30 ± 10).

279

280   **Secular trends**

281   Adjusted BMI mean (Fig 2A) and SD (Fig 2B) increased over the sampling period significantly faster

282   in FH_1 compared to FH_0 in the bootstrapped ANCOVA model, and estimates of ß and Δß in the mean

283   data were indistinguishable from the OLS estimates provided by the CQR analysis (Supplementary

284   Fig S2). Similar results were obtained with log-transformed BMI (Supplementary Fig S3). FH_1

285   accounted for 62% of the BMI mean trend and 60% of the BMI SD trend in this sample over the

286   period 1999-2014, effects similar in magnitude to the estimated FH_1 contribution to the sample risk

287   allele frequency (50%, Supplementary Table S2).

288

289   Discussion

290   **Summary**

291   We tested the prediction of segregation of discrete effects of FH on adult BMI (21), modeled as

292   modes of distribution, and estimated the contribution of FH_1 to recent trends in BMI mean and

293   dispersion in a large population-based sample. The results support a predominant role in the recent

294   obesity "epidemic" for rare genetic variants with large effects interacting with OE.

295

296   **Segregation of genetic susceptibility**

297   The non-parametric analysis provided evidence for a multi-modal distribution in the FH_1 group

298   consistent with the prediction of segregation of large genetic effects in families (21). Multi-modality

299   was supported by the analysis of density differences between FH_1 and FH_0 by unconditional quantiles

13

300   (Fig 1B) and by evidence of shape in the relationship between CQR coefficients on the BMI scale

301   (Fig 2A&B insets). Polygenic risk scores (PRS) in population-based samples are expected to be

302   normally-distributed, and appear to be so (24, 25). Any elevated polygenic obesity risk in DM_1 will

303   dilute into the mating population resulting in a right-shifted distribution in FH_1 compared to FH_0, not

304   multi-modality. Alternative explanations for multi-modality might be discrete stratification of OE

305   which seems unlikely, or un-modeled interactions between FH and other covariates. Un-modeled

306   interactions between FH and stratified residual confounders may exist and contribute but we found no

307   evidence for this in plots of distributions by gender and race/ethnicity (Supplementary Fig S1) or in

308   an analysis of smoking status against FH. Discrete inheritance of genetic variants with large effects

309   remains the most likely explanation for multi-modality in the BMI distribution.

310

311   Approximately 40% of the adjusted BMI variance in FH_1 was accounted for by between-modes

312   variance (Fig 3B) but this represents a lower bound since the identified modes are likely to be

313   synthetic ie composed of a range of effect sizes due to rare variants at different loci. Indications of

314   fine structure within the broad peaks (Fig 2B inset) are suggestive. Examples of rare variants with

315   large effects on BMI in adults (ß) are known from studies of candidate genes and monogenic obesity

316   loci (26) while more recently a common variant in Samoans (EAF = 0.26, $ß \approx 1.4 \text{ kg/m}^2$), very rare

317   in other populations (33), and an African-specific rare variant (EAF = 0.008, $ß = 4.6 \text{ kg/m}^2$)

318   undetected in Europeans and Asians (34) have been identified by GWAS. Overall, ß in these nine

319   examples ranges from 1.4- 9 kg/m^2 and a similar range in effect sizes in the NHANES sample would

320   contribute substantially to the within-mode variance estimated here. A combination of within-subject

321   variance ($\sim 5\%$ (35)) with polygenic variance ($\sim 5\%$ (4)) sets a lower bound for within-modes

322   variance and hence the upper bound for between modes, implying that between 40% and 90% of total

323   variance in FH_1 may be attributed to large genetic effects.

324

G x OE

325

326 FH_1 is a prevalent (36%) and powerful determinant of the rate of change of mean BMI and its

327 dispersion over time, accounting for 62% of the BMI trend and 60% of the BMI SD trend in this

328 sample over 1999-2014. Under a polygenic model the familial risk would be distributed normally

329 over FH_1 which would then be a marker of a large fraction of the at-risk population. However under

330 the Mendelian model supported here genetic risk would segregate in families and only approximately

331 50% of FH_1 would acquire the excess familial risk and only ~18% of the sample would then account

332 for ~60% of the trends. Individuals with DM_1 must represent a fraction of individuals with elevated

333 genetic obesity risk and it is likely that the remainder, particularly those with a family history of

334 obesity without DM, would substantially increase the genetic component of the trends consequent to

335 the high heritability of BMI (22). This Mendelian model is internally consistent in estimates of risk

336 allele frequencies (q) in FH_1, FH_0 and DM_1 (Table 2) and in comparisons of FH_1 effect sizes in cross

337 section (q, ~50%) and over time (ß, ~60%) (Supplementary Table S2). Our results support the

338 proposition that the largest part and perhaps all of recent trends in mean and dispersion of BMI are

339 due to a minor subset of individuals with elevated genetic susceptibility to OE.

340

Limitations

341

342 The design and interpretation of fits to parametric mixture distribution models involves choices

343 concerning the number of components, parameter starting values and algorithms, and fit to a specific

344 model cannot be taken in isolation as support for its structural validity. We base our choice and

345 structural interpretation of 3-component normal mixture model fits and parameters on the *a priori*

346 hypothesis of Mendelian segregation of obesity risk in families (21) supported by the non-parametric

347 distributional plots (Fig 1B,C) and CQR analysis (Fig 2A,B insets). Like Abadi et al (14) we

348 interpret interactions in the CQR analysis as predominantly G x OE although a contribution from G x

349 G interactions cannot be excluded. Our interpretation is supported by the effects on the interaction of

15

350   including calendar time in the CQR model (Fig 2B). Other limitations discussed above include our

351   inability to exclude discreet stratification of OE and the possible influence of unmeasured/unknown

352   confounders of FH.

353

354   **Conclusions**

355   We conclude that a Mendelian model of individually rare but collectively common genetic risk

356   variants with large effects interacting with OE provides a plausible quantitative explanation for recent

357   trends in obesity and should be favored over a predominantly polygenic model which does not. The

358   evidence for a predominant role for polygenes (13) can appear to be strong (eg "Polygenic obesity is

359   the most common form of obesity in modern societies..." (36)) but recent interpretations seek to

360   explain the still missing heritability in obesity in terms of unidentified large genetic effects and G x

361   OE (4, 37) and recommend a renewed focus on family-based designs and on specific populations in

362   which large effect variants may be enriched (33, 34). Our results strengthen that view by showing that

363   a model based on unidentified segregating variants with large effects interacting with OE can account

364   for the largest part of the secular trend in obesity and its dispersion in a large population-based

365   sample. We can't predict what positive clinical or public health changes might follow from these

366   findings but they should help refocus research away from the under-performing but entrenched big

367   idea (38) that is polygenic obesity.

368

369   Supplementary information is available at The International Journal of Obesity's website.

370   **Supplementary Methods:** Race/Ethnicity coding

371   **Supplementary Table S1**: BMI-adjustment model parameters

372   **Supplementary Table S2:** Comparison of effects of diabetes family history on secular trends in

373   adjusted BMI and its SD and on calculated risk allele frequencies (q) in non-diabetic participants.

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

375 **Supplementary Fig S2:** time and time x FH interaction effects in a Conditional quantile regression

376 of adjusted BMI in an ANCOVA model.

377 **Supplementary Fig S3:** Secular trends in log-transformed adjusted BMI mean and SD

378

Competing Interests

379

380 The authors declare no competing financial or non-financial interests.

381

382

References

384   1.   Di Cesare M, Bentham J, Stevens GAA et al. Trends in adult body-mass index in 200 countries

385   from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2

386   million participants. Lancet. 2016;387:1377-1396.

387   2.   Krishna A, Razak F, Lebel A, Davey Smith G, Subramanian SV. Trends in group inequalities

388   and interindividual inequalities in BMI in the United States, 1993–2012. Am J Clin Nutr.

389   2015;101:598-605.

390   3.   Silventoinen K, Jelenkovic A, Sund R et al. Differences in genetic and environmental variation

391   in adult BMI by sex, age, time period, and region: an individual-based pooled analysis of 40

392   twin cohorts. Am J Clin Nutr. 2017;106:457-466.

393   4.   Loos RJ. The genetics of adiposity. Curr Opin Genet Dev. 2018;50:86-95.

394   5.   Rokholm B, Silventoinen K, Ängquist L, Skytthe A, Kyvik KO, Sørensen TI. Increased genetic

395   variance of BMI with a higher prevalence of obesity. PLoS One. 2011;6:e20816.

396   6.   Franks PW, McCarthy MI. Exposing the exposures responsible for type 2 diabetes and obesity.

397   Science. 2016;354:69-73.

398   7.   Hall KD. Did the Food Environment Cause the Obesity Epidemic? Obesity. 2018;26:11-13.

399   8.   Kivimaki M, Stenholm S, Kawachi I. The widening BMI distribution in the United

400   States[letter]. Am J Clin Nutr 2015;101(6):1307-1308.

401   9.   Jenkins AB, Campbell LV. Variation in Genetic Susceptibility Drives Increasing Dispersion of

402   Population BMI[letter]. Am J Clin Nutr 2015;101:1308.

403   10.   Razak F, Smith GD, Krishna A, Lebel A, Subramanian SV. Variation in genetic susceptibility

404   drives increasing dispersion of population BMI Reply[letter]. Am J Clin Nutr

405   2015;101(6):1308-1309.

406   11.   Speakman JR, Loos RJF, O'Rahilly S, Hirschhorn JN, Allison DB. GWAS for BMI: a treasure

407   trove of fundamental insights into the genetic basis of obesity. Int J Obes. 2018;42:1524-1531.

408   12.   Locke AE, Kahali B, Berndt SI et al. Genetic studies of body mass index yield new insights for

409   obesity biology. Nature. 2015;518:197-206.

410   13.   Khera AV, Chaffin M, Wade KH et al. Polygenic Prediction of Weight and Obesity

411   Trajectories from Birth to Adulthood. Cell. 2019;177:587-596.e9.

412   14.   Abadi A, Alyass A, Robiou du Pont S et al. Penetrance of Polygenic Obesity Susceptibility

413   Loci across the Body Mass Index Distribution. Am J Hum Gen. 2017;101:925-938.

414   15.   Nagpal S, Gibson G, Marigorta U. Pervasive Modulation of Obesity Risk by the Environment

415   and Genomic Background. Genes. 2018;9:doi: 10.3390/genes9080411.

416   16.   Hemminki K, Li X, Sundquist K, Sundquist J. Familial risks for type 2 diabetes in Sweden.

417   Diabetes Care. 2010;33:293-297.

418   17.   Willemsen G, Ward KJ, Bell CG et al. The Concordance and Heritability of Type 2 Diabetes in

419   34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN)

420   Consortium. Twin Res Hum Genet. 2015;18:762-771.

421   18.   Ghosh A, Liu T, Khoury MJ, Valdez R. Family History of Diabetes and Prevalence of the

422   Metabolic Syndrome in U.S. Adults without Diabetes: 6-Year Results from the National Health

423   and Nutrition Examination Survey (1999-2004). Public Health Genom. 2010;13:353-359.

424   19.   Tirosh A, Shai I, Afek A et al. Adolescent BMI trajectory and risk of diabetes versus coronary

425   disease. N Engl J Med. 2011;364:1315-1325.

426   20.   Scott RA, Consortium I. The link between family history and risk of type 2 diabetes is not

427   explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study.

428   Diabetologia. 2013;56:60-69.

429   21.   Jenkins AB, Batterham M, Samocha-Bonet D, Tonks K, Greenfield JR, Campbell LV.

430   Segregation of a latent high adiposity phenotype in families with a history of type 2 diabetes

431   mellitus implicates rare obesity-susceptibility genetic variants with large effects in diabetes-

432   related obesity. PLoS One. 2013;8:e70435.

433 22. Stryjecki C, Alyass A, Meyre D. Ethnic and population differences in the genetic predisposition

434   to human obesity. Obes Rev. 2018;19:62-80.

435 23. Xu G, Liu B, Sun Y et al. Prevalence of diagnosed type 1 and type 2 diabetes among US adults

436   in 2016 and 2017: population based study. BMJ. 2018;362:k1497.

437 24. Llewellyn CH, Trzaskowski M, Plomin R, Wardle J. From modeling to measurement:

438   developmental trends in genetic influence on adiposity in childhood. Obesity. 2014;22:1756-

439   1761.

440 25. Rask-Andersen M, Karlsson T, Ek WE, Johansson A. Gene-environment interaction study for

441   BMI reveals interactions between genetic factors and physical activity, alcohol consumption

442   and socioeconomic status. PLoS Genet. 2017;13:e1006977.

443 26. Jenkins AB, Campbell LV. Future management of human obesity: understanding the meaning

444   of genetic susceptibility. Adv Genomics Genet. 2014;4:219-232.

445 27. Müller MJ, Geisler C, Blundell J et al. The case of GWAS of obesity: does body weight control

446   play by the rules. Int J Obes. 2018

447 28. R Development Core Team. R: A Language and Environment for Statistical Computing. R

448   Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/; 2016.

449 29. Sheather SJ, Jones MC. A Reliable Data-Based Bandwidth Selection Method for Kernel

450   Density Estimation. J Roy Stat Soc B Met. 1991;53:683-690.

451 30. Koenker R. Quantile Regression: 40 Years On. Ann Rev of Econ. 2017;9:155-176.

452 31. Briollais L, Durrieu G. Application of quantile regression to recent genetic and -omic studies.

453   Hum Genet. 2014;133:951-966.

454 32. Koenker R, Xiao Z. Inference on the quantile regression process. Econometrica. 2002;70:1583-

455   1612.

20

456   33.   Minster RL, Hawley NL, Su CT et al. A thrifty variant in CREBRF strongly influences body

457   mass index in Samoans. Nat Genet. 2016;48:1049-1054.

458   34.   Chen G, Doumatey AP, Zhou J et al. Genome-wide analysis identifies an African-specific

459   variant in SEMA4D associated with body mass index. Obesity. 2017;25:794-800.

460   35.   Wormser D, White IR, Thompson SG, Wood AM. Within-person variability in calculated risk

461   factors: Comparing the aetiological association of adiposity ratios with risk of coronary heart

462   disease. Int J Epidemiol. 2013;42:849-859.

463   36.   Albuquerque D, Nóbrega C, Manco L, Padez C. The contribution of genetics and environment

464   to obesity. Br Med Bull. 2017;123:159-173.

465   37.   Saeed S, Arslan M, Froguel P. Genetics of Obesity in Consanguineous Populations: Toward

466   Precision Medicine and the Discovery of Novel Obesity Genes: Obesity and Consanguinity.

467   Obesity. 2018;26:474-484.

468   38.   Joyner MJ, Paneth N, Ioannidis JP. What Happens When Underperforming Big Ideas in

469   Research Become Entrenched? JAMA. 2016;316:1355-1356.

470   Figure legends

471   Fig 1: Distribution of adjusted BMI in non-diabetic (DM_0) and diabetic (DM_1) participants in

472   combined NHANES 1999-2000, 2005-2006 and 2013-2014 cycles. A) Full non-diabetic sample

473   (n=9102) binned by quantiles (n=20) with superimposed kernel-smoothed and fitted densities in

474   normal and log-normal models. B) Kernel-smoothed adjusted BMI density by FH status. * significant

475   (p<0.05) difference in empirical density ($FH_1 - FH_0$) by quantiles of the full sample shown in panel

476   A. C) Kernel -smoothed adjusted BMI density in non-diabetic FH_1 (n=3297) compared to diabetic

477   participants (n=793).

478   Fig 2: Conditional quantile regression effects of FH on adjusted BMI in non-diabetic participants in

479   models consisting of FH alone (A) and in interaction with calendar time (B). The main panels show

480   the FH main effect size (β_1) by quantile with 95% CI (grey-shaded area), the OLS estimate of the

481   average effect (solid black line) with 95% CI (dotted black lines) and p-values associated with

482   Khmaladze tests for equality of β_1 across quantiles. The insets show the patterns of residuals ($\Delta\beta_1$)

483   from linear regressions of β_1 against conditional quantiles in FH_0 (β_0) , with 95% CI on the fits

484   (dotted lines) around the lines $\Delta\beta_1 = 0$ representing perfect fits.

485   Fig 3: A) Adjusted BMI density in FH_1 by quantile (grey bars) and kernel-smoothed (black line) with

486   fits to a three component normal mixture distribution. B) Estimated contributions in FH_1 of the

487   components of the mixture distribution to the prevalence (mixture coefficients, λ) and variance of

488   adjusted BMI.

489   Fig 4: Effects of diabetes family history ($FH_{0/1}$) on linear secular trends in age-, gender- and

490   race/ethnicity-adjusted BMI mean ± se (A) and standard deviation ± se (B). Parameter estimates

491   with 95% CI were obtained in ANCOVA models by stratified bootstrap resampling of all non-

492   diabetic individuals (see Methods). Dotted lines enclose 95% CI on fitted values at each point; β =

493   regression slope vs. time (kg/m^2 per year); $\Delta\beta = \beta_{FH1} - \beta_{FH0}$ (95%CI).
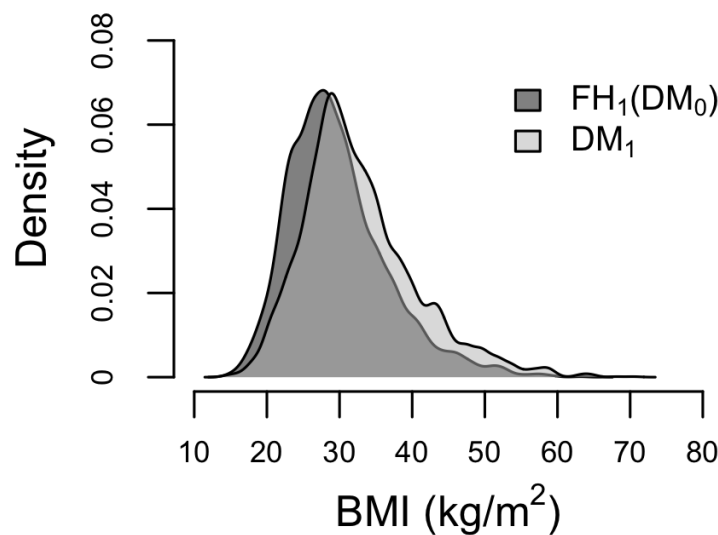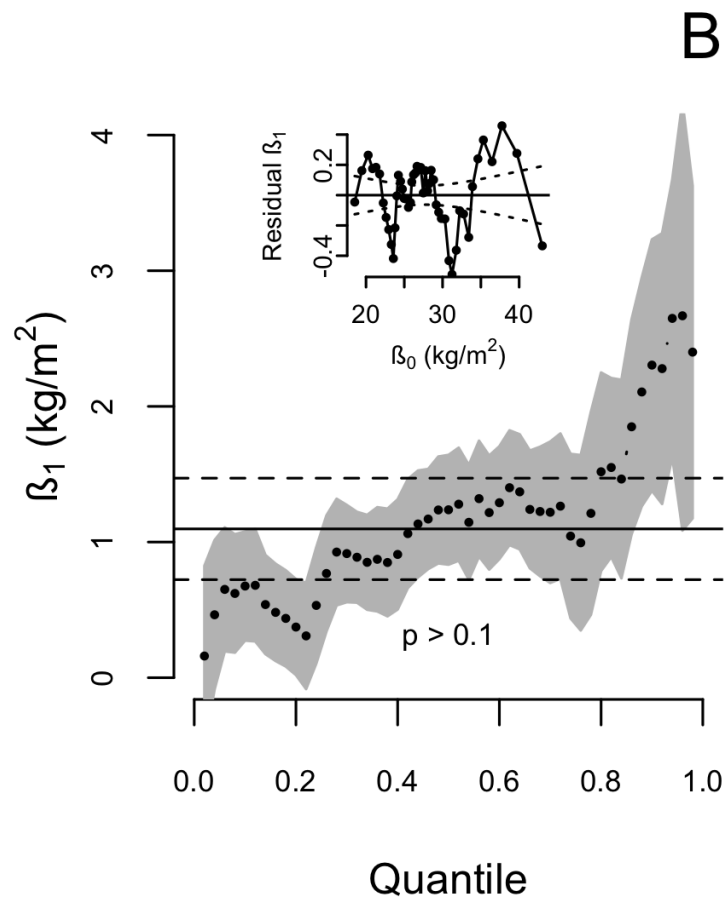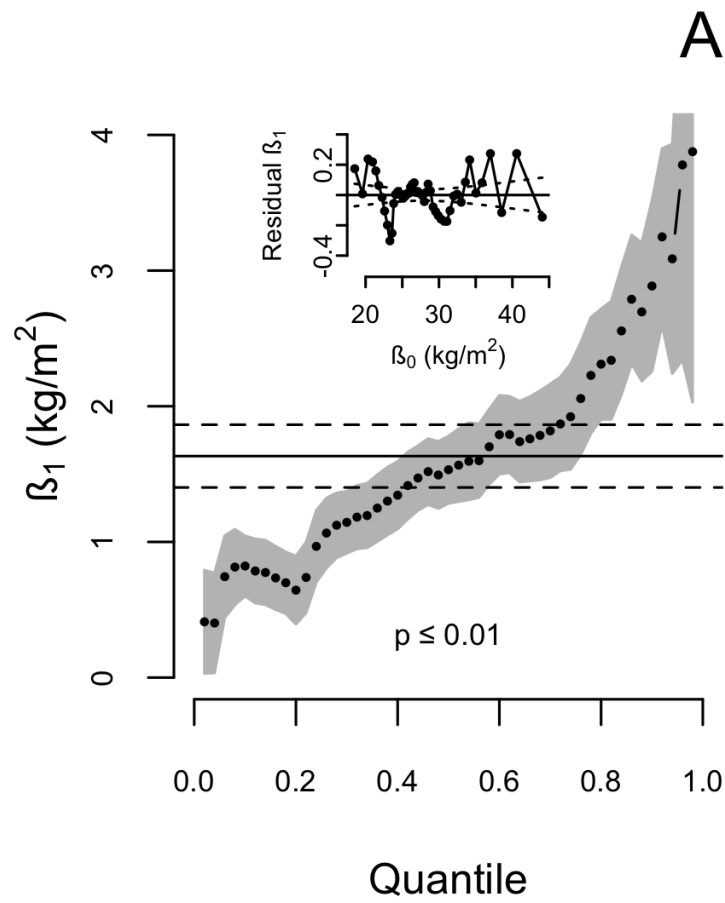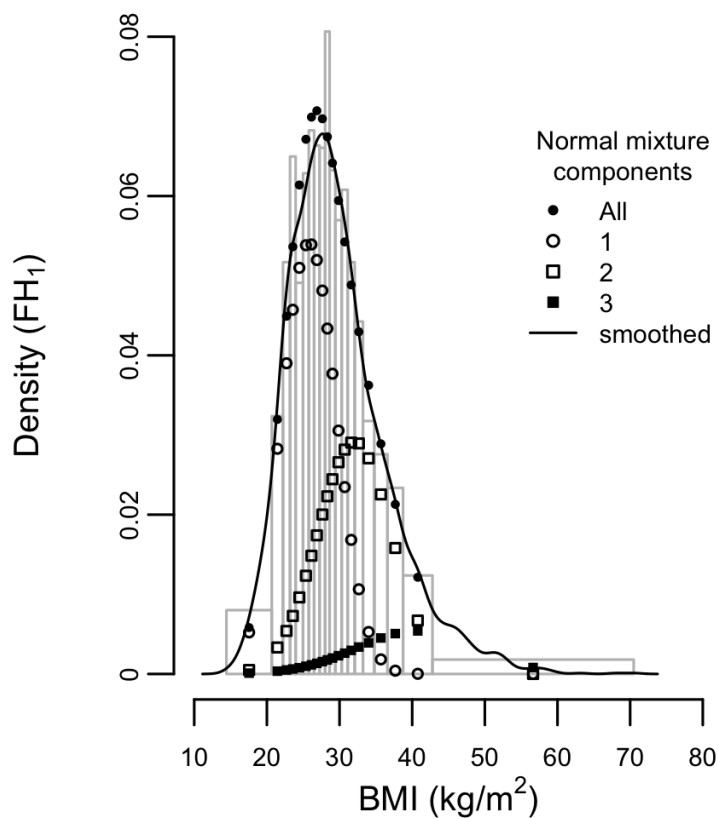
494
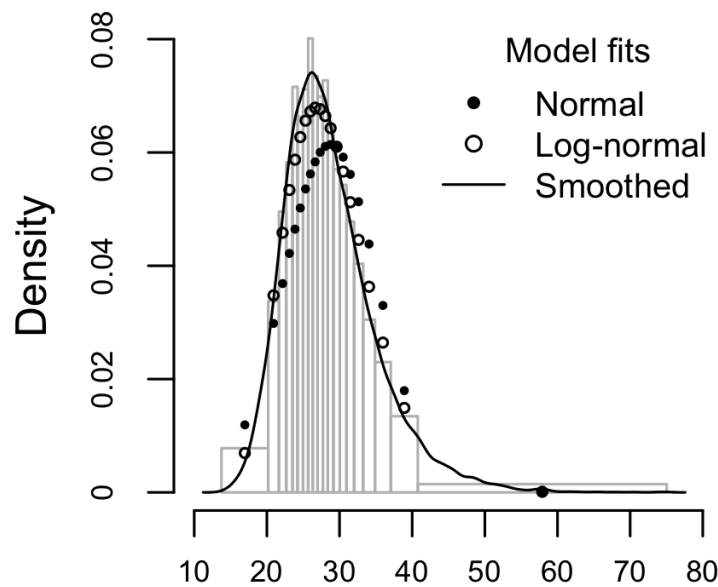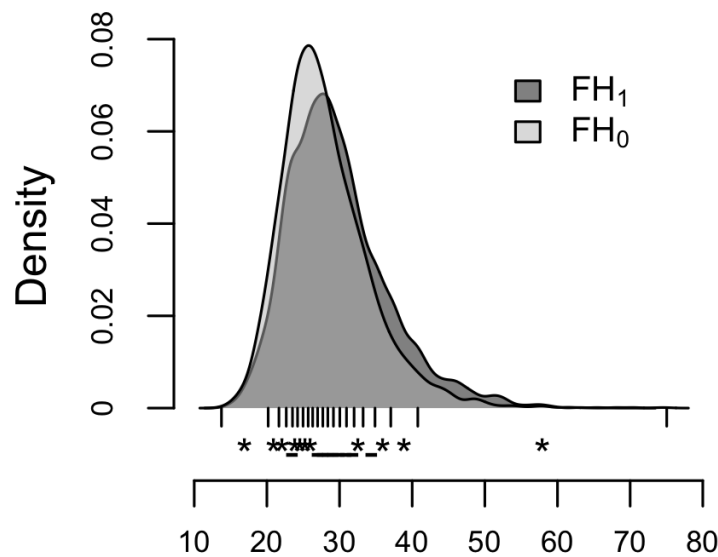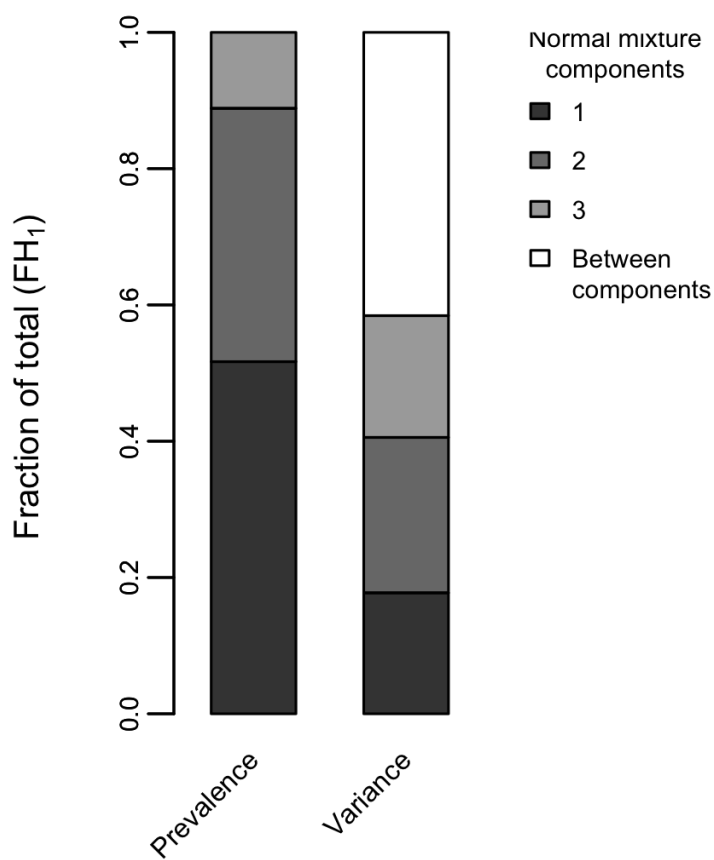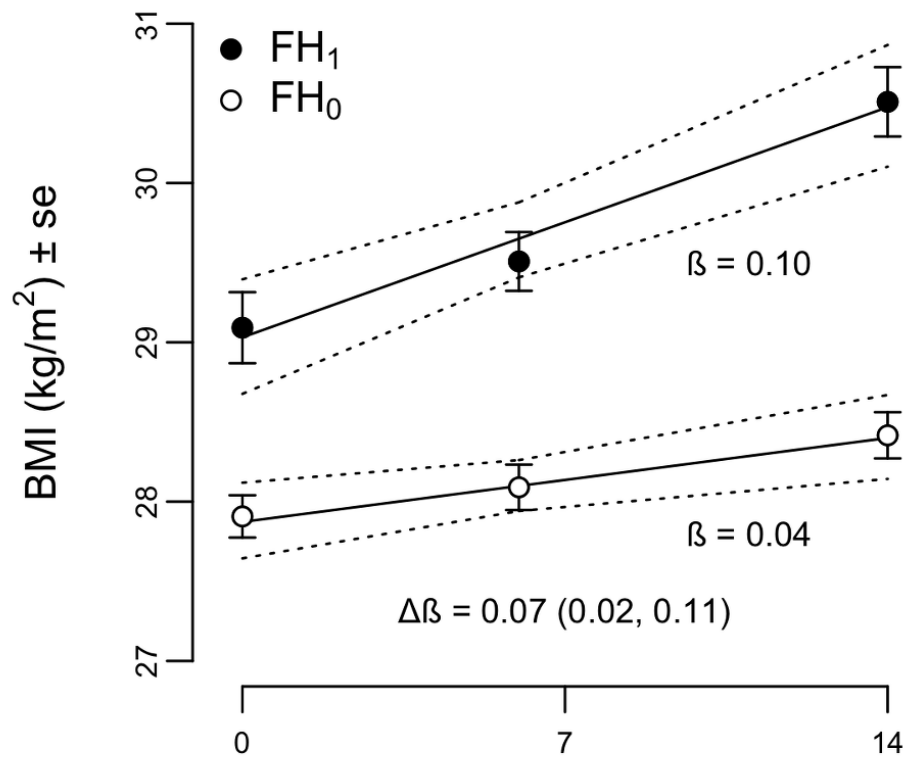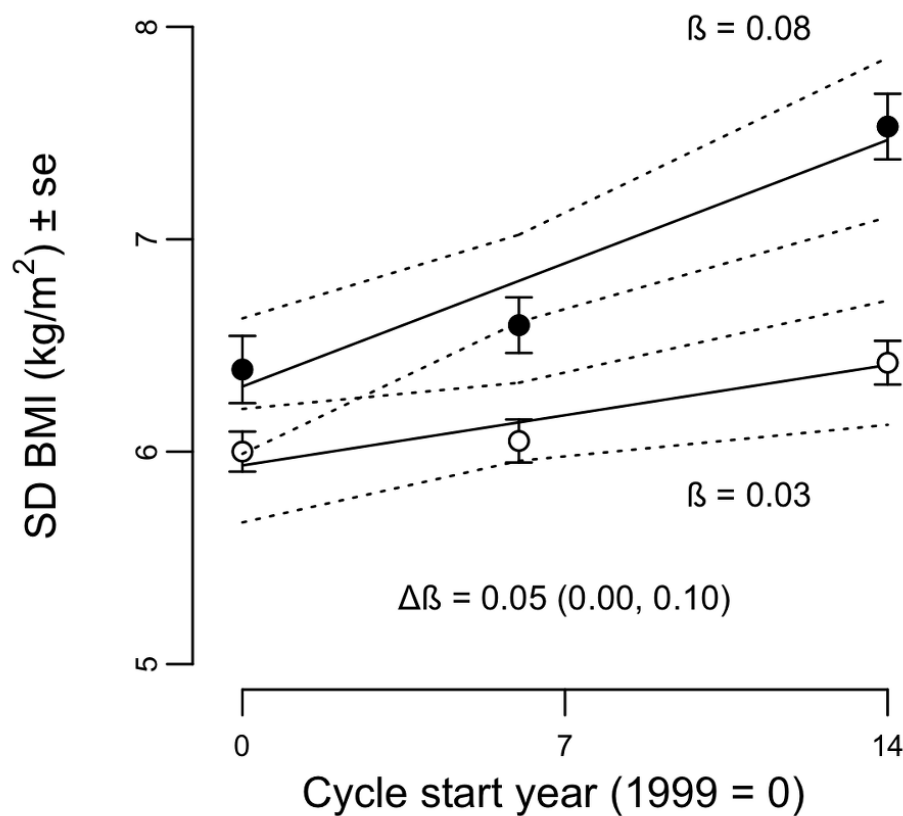
Table 1: Participants by diabetes status (DM_0/DM_1)

	All cycles	1999-2000	2005-2006	2013-2014	$p^\dagger$
		NHANES cycle			
		DM_0			
n	9102	2865	3076	3161	-
Gender (F%)	53	55	54	52	0.06
Race/Ethnicity (%) (MA/OH/NHW/NHB)$^\yen$	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

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

$^\yen$ 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[¥]

	Component	DM_0 FH$_0$	DM_0 FH$_1$	DM_1
	1	*	25.8±1.0	*
mean	2	*	32.1±1.8	*
	3	*	40.6±2.5	*
	1	3.8	3.8±0.9	3.7
sd	2	5.7	5.1±1.2	4.5
	3	10	8.2±0.8	8.4
	1	0.67	0.52±0.16	0.28
λ	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.