**Introduction**

Increasing trends in obesity prevalence has led to 29% of adults in England being classified as obese (NHS, 2019). The shift towards higher body weights has not occurred evenly across the population. Median body mass index (BMI) rose sharply during the 1990s, only for the rate of increase to slow throughout the 2000s. This contrasts to both the left tail of the BMI distribution which has remained virtually unchanged since 1992, as well as the right tail of the distribution which has increased consistently resulting in a widening of the BMI distribution (Green, Subramanian, & Razak, 2016; Johnson, Li, Kuh, & Hardy, 2015). These trends are consistent by sex and socioeconomic status (Green et al., 2016). Similar trends have also been observed internationally (Hayes et al., 2017; Krishna et al., 2015; Ouyang et al., 2015; Wagner et al., 2019; Yamada et al., 2020).

The widening distribution of BMI has important implications. The j-shaped curve in the risk of multiple health conditions across BMI results in an exponential increase in risk with increasing BMI (Aune et al., 2016). If the distribution becomes increasingly skewed towards higher BMI values, it may increase the proportion of the population at greatest risk of developing obesity-related health conditions.

To our knowledge, there has been no attempt to understand *why* the distribution has shifted towards higher BMI values at a population level. Prior studies documenting the widening distribution have been descriptive in nature (Green et al., 2016). There has been some examination of how demographic, social, behavioural and health-related characteristics vary across the entire BMI distribution (Aizawa, 2019; Bann et al., 2020; Basu et al., 2015; Bottai et al., 2014; Davillas & Jones, 2020; Dutton & Mclaren, 2016; Norris et al., 2020; Ouyang et al., 2015; Siddiqi et al., 2018). This has been important in progressing thinking away from dichotomous measures of risk, towards conceptualising risk as operating throughout an entire BMI distribution and providing more detailed investigation of non-linear relationships (Basu et al., 2015; Siddiqi et al., 2018). However, most studies focus on single points in time. A critical gap remains in applying methods that can identify the relative contributions of characteristics in explaining the changing distribution of BMI over time (Aizawa, 2019; Kranjac & Wagmiller, 2020). Understanding how populations vary across the BMI distribution has the potential to better guide health policy interventions particularly through targeting differing levels of risk across the distribution (Bann et al., 2020; Dutton & Mclaren, 2016).

We utilise advances in unconditional quantile decompositions to explore why the BMI distribution has widened in England over a ten-year period. We seek to determine the relative importance of: (i) changes in the association of socio-demographic factors to BMI (i.e. certain characteristics are having less or more influence on BMI values at each point of the BMI distribution), and (ii) changes in population composition (i.e. changes in the characteristics of individuals located across the distribution).

**Materials and Methods**

*Data*

We used data from the Health Survey for England (HSE) (Mindell et al., 2012). The HSE is a representative annual cross-sectional survey which collects information on health behaviours and is used by the Department of Health for policy decision making. There were no alternative datasets available that contained sufficient sample sizes, was representative of England, contained all relevant variables and covered the time period of our study (2002-2004 to 2002-2014).

We selected two time periods of the survey which (a) maximised the time period of analysis, (b) had similar variables to aid comparisons, and (c) had similar median BMI values but were different at the right tail of the distribution (i.e. demonstrated a widening distribution). We pooled individual survey years to increase our sample size and reliability of our quantile regression analyses. This was important for minimising the common support problem whereby estimating relationships at specific quantiles can be unreliable where counts for each combination of our explanatory variables may be small (i.e. cross-group differences) (Basu et al., 2015). We used HSE data from 2002-2004 to 2002-2014. We included only adults who were aged at least 20 years old. Analytical sample size was 29,018 for 2002-2004 and 23,992 for 2012-2014.

BMI was calculated by dividing weight (kg) by height squared (m2). Anthropometric data were objectively collected by a nurse during data collection. While BMI is a limited measure (Rothman, 2008), it provides a useful measure of relative weight and other formal measures of body fat were not available.

Explanatory variables were selected based on a simple theoretical framework to capture major determinants of BMI. While there are numerous factors that influence body weight, we opted for a parsimonious model that included factors consistently demonstrated throughout the literature to explain body weight. Increasing model complexity can introduce statistical issues (e.g. over-fitting) and small number issues at specific centiles which is problematic when using quantile regression techniques. The trade-off between maximising detail and minimising complexity meant that we selected six factors: age, sex, race, physical activity, occupation and education.

Age, sex and race were included as personal characteristics. Age is positively associated with body weight (Rothman, 2008). Age was categorised into roughly 20-year age bands: ‘20-39’, ‘40-59’ and ‘60+’. While males have higher median BMI, they also have a narrower BMI distribution compared to females (Green et al., 2016). The association of BMI to racial groups is complex, but important to account for (El-Sayed, Scarborough, & Galea, 2011). We grouped individuals as ‘White’ and ‘non-White’ due to the small numbers in our sample not allowing a further disaggregation of ‘non-White’ individuals.

Physical activity was also included as it crudely reflects an individual’s energy balance (sufficient dietary data were not available). Physical activity was measured as whether individuals achieved 30 minutes of moderate or vigorous physical activity five days per week. The variable was included as it represented the UK government guideline for recommended level of physical activity in adults across the study period (Department of Health, 2004, 2011).

We examined the influence of an individual’s socioeconomic context on BMI. Individuals in poorer social circumstances are associated with higher BMI (Norris et al., 2020), due to a complex myriad of material and social factors including fewer resources to purchase healthy foods or greater access to obesogenic environments that facilitate unhealthy diets. Measuring socioeconomic context is complex and unlikely to be covered by a single variable. We elected to use two commonly used measures: educational attainment and occupation. While education and occupation are correlated together, their effects on body weight also operate on different causal pathways. Education represents both an individual’s ability to access higher paid jobs, as well as their cognitive ability to understand health-related information. We selected an individual’s highest educational attainment and grouped them as ‘no qualifications’, ‘secondary level or equivalent’, ‘A-level of equivalent’ and ‘degree level, equivalent or higher’. Occupation, representing an individual’s socioeconomic position as well as their access to material resources, was measured using the National Statistics Socio-economic Classification (NS-SeC) which categorised occupations into ‘low’, ‘medium’ and ‘high’ occupations.

*Statistical Analysis*

Our analytical strategy had two main stages. Firstly, we estimate unconditional quantile regression models to examine the association between BMI and our explanatory variables at each time period. The main advantage of quantile regression is the ability to examine how each explanatory variable varies in strength and association across the entire BMI distribution, rather than only at the mean (Aizawa, 2019; Bottai et al., 2014). In addition, with the population distribution of BMI positively skewed (Green et al., 2016), mean centric approaches may be less appropriate as measures of central tendency. We analysed the 10th, 50th and 90th percentiles. While our focus is to understand changes at the right tail (90th) of the distribution, we also examined changes at the left (10th) and middle (50th) of the distribution as points of reference to compare, interpret and situate the context of changes at the right tail (i.e. were findings observed across the entire distribution or just the right tail).

Secondly, we used decomposition analysis to identify the factors that explain differences in the BMI distribution between 2002-4 and 2012-4. Differences at the 10th, 50th and 90th percentiles were decomposed into two components: (1) differences in population composition (i.e. characteristics of individuals across the BMI distribution) and (2) differences in the relationships (i.e. coefficients) in our explanatory factors. We used two extensions of the Oaxaca-Blinder decomposition approach to quantile regression (Blinder, 1973; Oaxaca, 1973), as proposed by Machado and Mata (Machado & Mata, 2005), and by Firpo and colleagues based on recentered influence functions (RIFs) (Firpo, Fortin, & Lemieux, 2009). We used both of these approaches to check the robustness of our results. The basic idea is that the BMI distribution can be assumed to be a function of a set of explanatory variables ($X\_{s}$) and coefficients ($β$), and thus differences in the BMI distribution between 2002-04 and 2012-14 can be decomposed as follows:

$$\hat{BMI}\_{2012-14}-\hat{BMI}\_{2002-04}=\left(\overbar{X}\_{2012-14}\hat{β}\_{2012-14}\right)-\left(\overbar{X}\_{2002-04}\hat{β}\_{2002-04}\right) =\left(\overbar{X}\_{2012-14}\hat{β}\_{2012-14}- \overbar{X}\_{2002-04}\hat{β}\_{2012-04}\right)+\left(\overbar{X}\_{2002-14}\hat{β}\_{2012-14}- \overbar{X}\_{2002-04}\hat{β}\_{2002-04}\right)= ∆X+ ∆β $$

$∆X$ accounts for BMI differences due to the compositional characteristics between the population in 2002-04 and 2012-14. It estimates the expected BMI differential by assuming that the population in 2012-14 and 2002-04 display the same behavioural responses and only their composition differs. $∆β$ accounts for BMI differences in coefficients in 2002-04 and 2012-14. It estimates the expected BMI differences by assuming that the population composition in both years is the same but the coefficients capturing the relationship between individual behavioural responses and contextual factors differ. Sample weights were applied to all analyses. Missing values were assumed missing at random. We conducted all the analysis in StataMP v14 and all reproducible code has been made openly accessible via GitHub [https://github.com/fcorowe/bmi]. Further details of the methodology and its implementation are reported in Appendix A.

**Results**

Table 1 presents descriptive statistics of changes in percentiles of BMI between 2002-4 and 2012-4. Median BMI hardly changed from 26·6 in 2002-4 to 26·7 in 2012-4. There was almost no change at the left tail of the distribution with differences of 0·0 and 0·1 in BMI for the 10th and 25th percentiles respectively. In contrast, the 90th percentile grew over the period from 33·6 in 2002-4 to 34·4 in 2012-4, equating to a 2·5% increase. Table 2 presents sample characteristics at the left, middle and right tail of the BMI distribution for 2002-4 and 2012-4. It reveals that the population composition has experienced little change at the right tail of the distribution compared to the left tail and middle parts.

**Table 1.** Differences in BMI at selected centiles

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Percentile | 2002-4 | 2012-4 | Difference | % Change |
| 10th | 21.5 | 21.5 | 0.0 | 0.0 |
| 25th | 23.7 | 23.8 | 0.1 | 0.3 |
| 50th | 26.6 | 26.7 | 0.1 | 0.5 |
| 75th | 29.9 | 30.3 | 0.4 | 1.3\* |
| 90th | 33.6 | 34.4 | 0.8 | 2.5\* |
| Significance level \* p<0.05. Sample size was 29,018 for 2002/4 and 23,992 for 2012/4. |

**Table 2.** Variable weighted sample means at selected centiles of the BMI distribution.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Distribution point | Left (< 25th percentile) | Middle (25th - 75th percentile) | Right (> 75th percentile) | Complete cases, %a |
| Variable | Category | 2002-4 | 2012-4 | Difference | 2002-4 | 2012-4 | Difference | 2002-4 | 2012-4 | Difference | 2002-4 | 2012-4 |
| Sex |   |   |   |   |   |   |   |   |   |   | 100.0 | 100.0 |
|   | Male | 38.8 | 39.4 | 0.6 | 53.9 | 54.5 | 0.6 | 46.4 | 47.2 | 0.8 |  |  |
|   | Female | 61.2 | 60.6 | -0.6 | 46.1 | 45.5 | -0.6 | 53.6 | 52.8 | -0.8 |  |  |
| Age |   |   |   |   |   |   |   |   |   |   | 100.0 | 100.0 |
|   | 20-39 | 50.4 | 51.6 | 1.2 | 32.9 | 31.1 | -1.8 | 30.4 | 29.4 | -0.9 |  |  |
|   | 40-59 | 30.2 | 28.0 | -2.3 | 37.9 | 38.0 | 0.1 | 37.5 | 38.0 | 0.5 |  |  |
|   | 60+ | 19.3 | 20.4 | 1.1 | 29.1 | 30.9 | 1.8 | 32.1 | 32.6 | 0.5 |  |  |
| Race |   |   |   |   |   |   |   |   |   |   | 99.8 | 99.7 |
|   | White | 92.7 | 87.7 | -4.9 | 94.4 | 90.4 | -4.0 | 94.5 | 91.4 | -3.1 |  |  |
|   | Non-white | 7.3 | 12.3 | 4.9 | 5.6 | 9.6 | 4.0 | 5.5 | 8.6 | 3.1 |  |  |
| Physical activity |   |   |   |   |   |   |   |   |   | 76.7 | 88.6 |
|   | Yes | 35.2 | 46.5 | 11.3 | 33.4 | 45.5 | 12.1 | 24.2 | 37.1 | 12.9 |  |  |
|   | No | 64.8 | 53.5 | -11.3 | 66.6 | 54.5 | -12.1 | 75.8 | 62.9 | -12.9 |  |  |
| Educational attainment |   |   |   |   |   |   |   |   |   | 99.7 | 99.6 |
|   | None | 19.5 | 15.0 | -4.5 | 26.0 | 20.3 | -5.7 | 34.7 | 26.8 | -7.9 |  |  |
|   | Secondary | 29.2 | 21.6 | -7.7 | 20.3 | 24.8 | 4.5 | 28.6 | 25.4 | -3.2 |  |  |
|   | A level  | 26.8 | 28.2 | 1.5 | 24.1 | 26.3 | 2.2 | 21.5 | 25.1 | 3.6 |  |  |
|   | Degree | 24.4 | 35.1 | 10.7 | 19.5 | 28.5 | 9.0 | 15.2 | 22.7 | 7.5 |  |  |
| Occupation |   |   |   |   |   |   |   |   |   | 96.3 | 95.8 |
|   | Low  | 39.6 | 35.7 | -4.0 | 40.0 | 38.1 | -1.9 | 45.3 | 42.0 | -3.3 |  |  |
|   | Medium | 23.2 | 26.2 | 3.0 | 21.9 | 25.1 | 3.2 | 22.4 | 24.7 | 2.4 |  |  |
|   | High | 37.2 | 38.1 | 0.9 | 38.1 | 38.8 | 0.7 | 32.4 | 33.2 | 0.9 |  |  |
| N of obs. | 6,750 | 4,755 |   | 12,667 | 9,968 |   | 9,601 | 9,269 |   |  |  |

a The percentages reported for complete cases used a denominator 29,018 for 2002/4 and 23,992 for 2012/4 which are the maximum potential number of observations for each of the segments of the distribution considered in Table 2.

**Table 3.** Unconditional Quantile Regressions exploring Body Mass Index, 2002-4 and 2012-4.

|  |  |  |
| --- | --- | --- |
| Period | 2002-4 | 2012-4 |
| Centile | 10th  | 50th  | 90th | 10th  | 50th  | 90th |
| *Sex* |   |   |   |   |   |   |
| Male | 0.889\*\*\* | 1.046\*\*\* | -1.871\*\*\* | 1.112\*\*\* | 0.626\*\*\* | -2.460\*\*\* |
|  | (0.121) | (0.131) | (0.365) | (0.089) | (0.092) | (0.240) |
| *Age* |   |   |   |   |   |   |
| Age 40-59 | 1.298\*\*\* | 1.477\*\*\* | 0.361 | 1.517\*\*\* | 1.553\*\*\* | 0.192 |
|  | (0.154) | (0.163) | (0.437) | (0.118) | (0.116) | (0.295) |
| Age 60+ | 1.388\*\*\* | 1.440\*\*\* | 0.266 | 1.648\*\*\* | 1.588\*\*\* | -0.519 |
|  | (0.162) | (0.184) | (0.505) | (0.124) | (0.125) | (0.321) |
| *Race* |   |   |   |   |   |   |
| White | 0.293 | 0.353 | 0.204 | 0.292 | 0.665\*\*\* | 0.966\*\* |
|  | (0.320) | (0.317) | (0.855) | (0.205) | (0.186) | (0.470) |
| *Physical activity activity* |   |   |   |   |   |   |
| Moderate exercise | -0.365\*\*\* | -0.894\*\*\* | -2.548\*\*\* | -0.129 | -0.798\*\*\* | -2.815\*\*\* |
|  | (0.141) | (0.151) | (0.391) | (0.093) | (0.094) | (0.240) |
| *Educational attainment*  |   |   |   |   |   |
| Secondary | 0.035 | -0.511\*\*\* | -3.056\*\*\* | 0.210\* | -0.569\*\*\* | -1.557\*\*\* |
|  | (0.150) | (0.182) | (0.522) | (0.123) | (0.138) | (0.387) |
| A level | -0.451\*\* | -0.795\*\*\* | -2.521\*\*\* | -0.063 | -0.799\*\*\* | -1.808\*\*\* |
|  | (0.188) | (0.209) | (0.601) | (0.136) | (0.147) | (0.403) |
| Degree level | -0.457\*\* | -1.469\*\*\* | -3.686\*\*\* | -0.427\*\*\* | -1.431\*\*\* | -2.605\*\*\* |
|  | (0.214) | (0.242) | (0.688) | (0.158) | (0.164) | (0.434) |
| *Occupation* |   |   |   |   |   |   |
| Medium | 0.084 | -0.088 | 0.086 | 0.093 | -0.318\*\*\* | -0.818\*\*\* |
|  | (0.162) | (0.175) | (0.493) | (0.118) | (0.119) | (0.316) |
| High | 0.441\*\*\* | 0.020 | -0.352 | 0.313\*\* | -0.121 | -0.577\* |
|  | (0.165) | (0.175) | (0.490) | (0.124) | (0.126) | (0.323) |
| Constant | 20.274\*\*\* | 26.464\*\*\* | 41.797\*\*\* | 19.905\*\*\* | 26.692\*\*\* | 42.184\*\*\* |
|  | (0.362) | (0.351) | (0.980) | (0.235) | (0.220) | (0.584) |
| Observations | 21,380 | 21,380 | 21,380 | 20,484 | 20,484 | 20,484 |
| Standard errors in parentheses |
| Reference category: Age 20-39, Non-white, No physical activity, No qualification, Low-skilled occupation |
| \* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01QR = Quantile Regression |

Table 3 presents the unconditional quantile regression estimates showing the association of our explanatory variables to BMI at the 10th, 50th and 90th percentiles for 2002-4 and 2012-14. At the 10th and 50th percentiles, being male is positively associated to higher BMI but shows a negative association at the 90th percentile. Older age groups were positively associated to BMI at the 10th and 50th percentiles, however we detect no association at the 90th percentile. Race displayed no associations in 2002-4, but the White group was positively associated to BMI at the 50th and 90th percentiles in 2012-4. Physical activity was consistent across years, with negative associations at the 50th and 90th percentiles (and a larger effect size at the 90th percentile). Education also showed a consistent negative relationship across the BMI distribution, with larger effect sizes observed with increasing educational attainment. Occupation was mostly non-significant, but in 2012-14 medium and higher occupations were negatively associated to BMI at the 90th percentile. We stratified these analyses by sex (Appendix B), however the results are largely consistent with the patterns described above.

Next we performed the decomposition analysis (see Table 4; we also stratified these analyses by sex in Appendix C). In the analysis we accounted for the effects of education and occupation together, since they both comprise proxies for socio-economic context. The Machado-Mata and Melly (Table 4 – Panel B) and RIF (Table 4 – Panel C) decomposition estimates suggest that the increase at the 90th percentile is mostly attributable to changes in the coefficients of our explanatory variables, with changes in the composition of the population acting against these trends. According to our RIF estimates, changes in coefficients and population composition accounted for 140% (=100\*[2.269/1.615]) and -40% (=100\*[-0.645/1.615]) of the difference in BMI at the 90th percentile, respectively. The estimates suggests that in the absence of changes in population composition, the BMI at the 90th percentile would have increased by an additional 40%.

**Table 4.** Quantile decomposition of changes in Body Mass Index (BMI) between 2002-4 and 2012-4.

|  |  |  |  |
| --- | --- | --- | --- |
|   | 10th Percentile | 50th Percentile | 90th Percentile |
| *A: Raw BMI gap* |  |  |  |
| Qt(2012-4) - Qt(2002-4) | 0.001 | 0.134 | 0.845 |
|   | (0.080) | (0.069) | (0.136) |
| *B: Decomposition method: Machado-Mata–Melly* |   |   |   |
| Qt(2012-4) - Qt(2002-4) | 0.005 | 0.144\*\*\* | 0.880\*\*\* |
|  | (0.043) | (0.043) | (0.095) |
| Attributable to characteristics | -0.091 | -0.223\*\*\* | -0.425\*\*\* |
|  | (0.073) | (0.066) | (0.132) |
| Attributable to coefficients | 0.097 | 0.368\*\*\* | 1.306\*\*\* |
|   | (0.067) | (0.058) | (0.115) |
| *C: Decomposition method: RIF regression* |  |  |  |
| Qt(2012-4) - Qt(2002-4) | 0.348\*\*\* | 0.541\*\*\* | 1.615\*\*\* |
|  | (0.054) | (0.059) | (0.154) |
| Attributable to characteristics | 0.072\*\*\* | -0.123\*\*\* | -0.645\*\*\* |
|  | (0.020) | (0.023) | (0.058) |
| Attributable to coefficients | 0.276\*\*\* | 0.665\*\*\* | 2.269\*\*\* |
|  | (0.056) | (0.061) | (0.161) |
| *D: Attributable to characteristics* |   |   |   |
| Sex | 0.001 | 0.000 | -0.002 |
|  | (0.005) | (0.003) | (0.013) |
| Age | 0.128\*\*\* | 0.117\*\*\* | -0.044\* |
|  | (0.011) | (0.011) | (0.025) |
| Race | -0.004\* | -0.011\*\*\* | -0.020\*\*\* |
|  | (0.002) | (0.003) | (0.008) |
| Physical activity | -0.009 | -0.094\*\*\* | -0.352\*\*\* |
|  | (0.010) | (0.011) | (0.032) |
| Socio-economic | -0.044\*\*\* | -0.136\*\*\* | -0.227\*\*\* |
|  | (0.011) | (0.013) | (0.034) |
| *E: Attributable to coefficients* |   |   |   |
| Sex | 0.087\* | -0.101\* | -0.214 |
|  | (0.046) | (0.052) | (0.139) |
| Age | 0.005 | 0.011 | 0.039 |
|  | (0.006) | (0.007) | (0.019) |
| Race | -0.238 | 0.314 | 1.071\* |
|  | (0.218) | (0.236) | (0.626) |
| Physical activity | 0.083\*\* | 0.053 | -0.054 |
|  | (0.034) | (0.037) | (0.098) |
| Socio-economic | 0.005 | 0.051\*\*\* | 0.075\*\* |
|   | (0.018) | (0.020) | (0.053) |
| Constant | 0.335 | 0.336 | 0.942 |
|   | (0.231) | (0.250) | (0.662) |
| Observations | 41,864 | 41,864 | 41,864 |
| Standard errors in parentheses. Bootstrapped standard errors based on 100 replicates are reported for the Machado-Mata–Melly decomposition. |
| \* *p* < 0·10, \*\* *p* < 0·05, \*\*\* *p* < 0·01 |
| Qt = Quantile-tau – BMI value for a specific quantile |  |  |
| Sample size was 21,380 for 2002/4 and 20,484 for 2012/4. |  |  |

Table 4 – Panel D identifies the extent that changes in the composition of each factor helped explain differences in BMI values between 2002-4 and 2012-4 at each point of the distribution. Positive values represent factors that contributed to an increase in BMI at a specific percentile over the time period, with negative associations resulting in a decrease over time. A decline in the share of less educated individuals and White individuals (see Table 2) contributed to reduce the gap in BMI at the 90th percentile between 2002-4 and 2012-4 as observed in Table 4D. Similarly for race, BMI at the 90th percentile was 0.020 points lower in 2012-4. Changes in the composition of physically active population at the 90th percentile (i.e. a decline in the share of physically active people at the percentile) contributed to decrease the BMI at the 90th percentile. The effect size was larger with physical activity compositional changes contributing to BMI at the 90th percentile being 0.352 points lower in 2012-4. Similar patterns were observed for these factors at the other percentiles, other than age which was positively associated at the 10th and 50th percentiles.

Table 4 – Panel E disaggregates the effects attributable to changes in the coefficients. Positive values represented factors that were associated with higher BMI values for a percentile in 2012-4 than in 2002-4, with negative values contributing to a decline over time. The results reveal that race and socio-economic context were key factors contributing to a higher BMI value at the 90th percentile in 2012-4. Their positive coefficients suggest that they have seen increasing relative importance in explaining differences in BMI of the 90th percentile over time. Changes in the coefficients for race at the 90th percentile resulted in BMI at the 90th percentile being 1.071 units higher in 2012-4 compared to 2002-4. A lower effect size was observed for changes in the socio-economic coefficients (0.075).

**Conclusions**

*Key Results*

Our results comprise the first empirical evidence to identify the relative significance of the factors contributing to widening the BMI distribution in England. Changes in the strength of the relationships between contextual explanatory variables and BMI were important in our model for understanding increasing right tail of the BMI distribution. Race and socio-economic status are more strongly associated with BMI at the 90th percentile of the distribution in 2012-14 than in 2002-04 due to the changing nature of their relationships. Compositional changes with respect to age, race, socio-economic factors, and physical activity also made significant contributions to explaining changes in BMI at the 90th percentile of the distribution.

*Interpretation*

Changes in the association between our explanatory variables and BMI appear to reflect systematic changes in the determinants of BMI. The main underpinning forces seem to be associated with changes in the influences of race, socio-economic context and physical activity. Sex was not important in explaining the widening of the distribution (although was relevant for explaining differences in the distribution cross-sectionally) and this follows other studies (Green et al., 2016; Wagner et al., 2019).

The contextual effect of physical activity remained consistent over time, suggesting that it remains an important determinant of body weight. The negative association of physical activity to BMI grew stronger as we moved towards the right tail of the distribution supporting findings elsewhere (Bann et al., 2020; Bottai et al., 2014; Ouyang et al., 2015). Individuals at the 90th percentile who met government physical activity guidelines were associated with an average BMI of 2.815 units lower in 2012-14 than compared to those who did not. Compositional changes, observed as increases in physical activity levels over time, led to a narrowing of the distribution through reducing the value of BMI at the 90th percentile. These trends support the assertion that increasing physical activity levels represents an important policy strategy for tackling excess body weight (NHS, 2019). While our findings should be interpreted carefully given the simplicity of our physical activity variable, they do follow similar findings elsewhere (Bann et al., 2020).

The negative compositional associations for socio-economic context indicate that the decline in the share of socially disadvantaged populations at the 90th percentile resulted in a reduction of BMI at the 90th percentile. This implies that the widening in the BMI distribution would have been greater in absence of changes in population composition in socio-economic context, offsetting the widening effect of differences in coefficients. It implies a complex picture whereby inequalities may have narrowed due to fewer individuals of lower socio-economic status at the 90th percentile of the BMI distribution, yet socio-economic status remains important for understanding differences in BMI among the population at the 90th percentile given the distinct social gradient observed for coefficients in 2012-14. Greater investigation for the reasons behind these social inequalities is therefore paramount, especially given the notable and persistent inequalities in body weight and their health-related outcomes (Green et al., 2016; Norris et al., 2020).

Our results also demonstrate that demographic, social and behavioural predictors of BMI are not stable in their associations across the BMI distribution. While exploring how predictors of BMI vary across its distribution is nothing new (Aizawa, 2019; Basu et al., 2015; Bottai et al., 2014; Dutton & Mclaren, 2016; Kranjac & Wagmiller, 2020; Siddiqi et al., 2018), it demonstrates the importance of thinking beyond a mean-centric approach or using specific cut points (e.g. modelling obesity as a binary outcome). Understanding the heterogeneous nature of how correlates of BMI varies and utilising this knowledge to design interventions to target the different part of BMI distribution is paramount.

*Limitations*

There are several limitations to our study. We use repeated cross-sectional data which is limited in its ability to draw out causal inferences. The purpose was to study changes within populations and therefore our analyses are appropriate. Our results also follow descriptive analyses of the widening distribution of BMI using longitudinal data (Bann et al., 2020; Johnson et al., 2015). Current longitudinal applications focus on specific cohorts that are less generalisable to the wider population.

The choice of our explanatory variables was limited by our ability to use comparable variables between our time periods. Missing data may have introduced bias into our model estimates if records were not missing at random or differed by time period. The determinants of body weight are complex and future research should seek to examine the contribution of other factors. Some of our measures were determined by data availability, hence simplistic and/or self-reported by individuals (e.g. physical activity), and these issues may have contributed to introducing measurement bias into our results. Evidence, for example, suggests systematic low to moderate disagreement between self-reported and objective measured physical activity measures (Prince et al., 2008). Future research should consider additional determinants of obesity such as dietary intake or environmental features (e.g. access to healthy foods; Davillas & Jones, 2020) that we were unable to include in our study. Additional explanatory variables may help both to explain why and how the distribution of BMI has widened, as well as the pathways through which important characteristics such as race are associated to obesity. The inclusion of additional variables may result in issues of low statistical power and require larger datasets than the HSE for deploying our methods. An opportunity for future research may be afforded by the growing availability of ‘big data’ that can allow for greater detail in studying the heterogeneity in determinants of body weight.

We focus only on specific percentiles (10th, 50th and 90th). These choices were selected to capture differences in the BMI distribution at the left, middle and right tail. We acknowledge that the specific percentiles selected to represent these positions may be arbitrary and selection of other percentiles may determine our results (Aizawa, 2019). We tested additional percentiles (e.g. 25th and 75th), however our results did not alter significantly.

**References**

Aizawa, T. (2019). Transition of the BMI distribution in India: evidence from a distributional decomposition analysis. *Journal of Bioeconomics*, *21*(1), 3–36.

Aune, D., Sen, A., Prasad, M., Norat, T., Janszky, I., Tonstad, S., … Vatten, L. J. (2016). BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *BMJ* *353*, i2156.

Bann, D., Fitzsimons, E., & Johnson, W. (2020). Determinants of the population health distribution: an illustration examining body mass index. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyz245

Basu, S., Hong, A., & Siddiqi, A. (2015). Using Decomposition Analysis to Identify Modifiable Racial Disparities in the Distribution of Blood Pressure in the United States. *American Journal of Epidemiology*, *182*(8), 345–353.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, *8*(4), 436–455.

Bottai, M., Frongillo, E. A., Sui, X., Neill, J. R. O., Mckeown, R. E., Burns, T. L., … Pate, R. R. (2014). Use of Quantile Regression to Investigate the Longitudinal Association between Physical Activity and Body Mass Index. *Obesity*, *22*(5), 149–156.

Davillas, A., & Jones, A. (2020). Regional inequalities in adiposity in England: distributional analysis of the contribution of individual-level characteristics and the small area obesogenic environment. *Economics and Human Biology*, 100887.

Department of Health. (2004). At least five a week: Evidence on the impact of physical activity and its relationship to health. Retrieved from https://webarchive.nationalarchives.gov.uk/20130105001829/http://www.dh.gov.uk/prod\_consum\_dh/groups/dh\_digitalassets/@dh/@en/documents/digitalasset/dh\_4080981.pdf

Department of Health. (2011). Start Active, Stay Active: A report on physical activity from the four home countries’ Chief Medical Officers.

Dutton, D. J., & Mclaren, L. (2016). How important are determinants of obesity measured at the individual level for explaining geographic variation in body mass index distributions? Observational evidence from Canada using Quantile Regression and Blinder-Oaxaca Decomposition. *Journal of Epidemiology & Community Health*, *70*, 367–373.

El-Sayed, A. M., Scarborough, P., & Galea, S. (2011). Ethnic inequalities in obesity among children and adults in the UK: a systematic review of the literature. *Obesity Reviews*, *12*(5), e516-34.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973.

Green, M. A., Subramanian, S. V, & Razak, F. (2016). Population-level trends in the distribution of body mass index in England, 1992 – 2013. *Journal of Epidemiology & Community Health*, *70*, 832–835.

Hayes, A. J., Lung, T. W. C., Bauman, A., & Howard, K. (2017). Modelling obesity trends in Australia: unravelling the past and predicting the future. *International Journal of Obesity*, *41*, 178–185.

Johnson, W., Li, L., Kuh, D., & Hardy, R. (2015). How Has the Age-Related Process of Overweight or Obesity Development Changed over Time? Co-ordinated Analyses of Individual Participant Data from Five United Kingdom Birth Cohorts. *PLoS Medicine*, *12*(5), e1001828.

Kranjac, A. W., & Wagmiller, R. L. (2020). Decomposing Trends in Child Obesity. *Population Research and Policy Review*, *39*(2), 375–388.

Krishna, A., Razak, F., Lebel, A., Smith, G. D., & Subramanian, S. V. (2015). Trends in group inequalities and interindividual inequalities in BMI in the United States, 1993-2012. *American Journal of Clinical Nutrition*, *101*(3), 598–605.

Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distribution using quantile regression. *Journal of Applied Econometrics*, *20*, 445–465.

Mindell, J., Biddulph, J. P., Hirani, V., Stamatakis, E., Craig, R., Nunn, S., & Shelton, N. (2012). Cohort Profile: The Health Survey for England. *International Journal of Epidemiology*, *41*, 1585–1593.

NHS. (2019). Statistics on Obesity, Physical Activity and Diet, England, 2019. Retrieved from https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-obesity-physical-activity-and-diet/statistics-on-obesity-physical-activity-and-diet-england-2019

Norris, T., Bann, D., Hardy, R., & Johnson, W. (2020). Socioeconomic inequalities in childhood-to-adulthood BMI tracking in three British birth cohorts. *International Journal of Obesity*, *44*, 388–398.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*(3), 693–709.

Ouyang, Y., Wang, H., Su, C., Wang, Z., Song, Y., Xiao, Y., … Zhang, B. (2015). Use of quantile regression to investigate changes in the body mass index distribution of Chinese adults aged 18 – 60 years: a longitudinal study. *BMC Public Health*, *15*, 278.

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *5*, 56.

Rothman, K. J. (2008). BMI-related errors in the measurement of obesity. *International Journal of Health Geographics*, *32*, S56–S59.

Siddiqi, A., Shahidi, F. V., Hildebrand, V., Hong, A., & Basu, S. (2018). Illustrating a “consequential” shift in the study of health inequalities: a decomposition of racial differences in the distribution of body mass. *Annals of Epidemiology*, *28*(4), 236–241.

Wagner, K. J. P., Boing, A. F., Cembranel, F., Boing, A. C. da S., & Subramanian, S. V. (2019). Change in the distribution of body mass index in Brazil: analysing the interindividual inequality between 1974 and 2013. *Journal of Epidemiology and Community Health*, *73*(6), 544 LP – 548.

Yamada, G., Jones-Smith, J. C., Castillo-Salgado, C., & Moulton, L. H. (2020). Differences in magnitude and rates of change in BMI distributions by socioeconomic and geographic factors in Mexico, Colombia, and Peru, 2005–2010. *European Journal of Clinical Nutrition*, *74*(3), 472–480.