RECENT CHILD AND ADOLESCENT OBESITY PATTERNS IN NYC: A QUANTILE/BINARY REGRESSION ANALYSIS

STUART SWEENEY[1 2], KEVIN KONTY[3], AND KATHRYN GRACE[4]

Abstract. U.S. obesity prevalence continues to trend upwards and policies have thus far failed to achieve widespread change. Within this broader context, child and adolescent obesity has also trended upwards and public schools have been recruited to monitor the trends and to serve as the test bed for nutritional outreach and fitness programs. New York City has been particularly aggressive in this regard with a large data collection effort since 2006 and extensive interventions targeting the schools. Our overriding goal in this paper is to describe recent patterns of child and adolescent obesity in NYC with particular focus on differences among race/ethnic groups, foreign-born/native-born, and association with socioeconomic status. We use quantile regression to assess patterns of association and to describe distributional changes among groups. We also assess patterns of prevalence difference using a recent methodology advance that links binary outcome models and quantile regression.

1. Introduction

Over the last several decades obesity has emerged as a global epidemic with the highest prevalence and rapid increases in affluent countries. U.S. obesity trends and patterns have been well documented and various policies designed to slow or reverse the trends, through nutrition and fitness, have thus far failed. A recent comprehensive review noted that as of 2003-04, fully two-thirds of US adults were overweight or obese (Wang and Beydoun, 2007). Within this broader context, child and adolescent obesity has also trended upwards and public schools have been recruited to monitor the trends and to serve as the test bed for nutritional outreach and fitness programs. New York City has been particularly aggressive in this regard with periodic large surveys prior to 2006, routine collection of fitnessgram data from all children in public schools since 2006, and an extensive and expanding program of exercise and nutrition interventions targeting the schools. A study using data from a 2003 survey found that approximately 25% of public elementary school children were obese and almost half were overweight or obese (Thorpe et al., 2004).

Our overriding goal in this paper is to describe recent patterns of child and adolescent obesity in NYC with particular focus on differences among race/ethnic groups, foreign-born/native-born, and association with socioeconomic status. We use quantile regression to assess patterns of association and to describe distributional changes among groups. We also assess patterns of prevalence difference using a recent methodology advance that links binary outcome models and quantile regression (Sweeney et al., 2012). As a precursor to that empirical analysis we have to confront some difficult measurement issues. There is an extensive literature on the problematic nature of indirect obesity measurement in children and adolescents and the care needed in selecting reference distributions and thresholds (for example, see Pelletier (2006) and references therein). Our use of quantile regression is partly in response to those problems. We use two data sources. The first is the National Health and Nutrition Examination Surveys (NHANES); a nationally representative

[1]Department of Geography, UC Santa Barbara
[2]Institute for Social, Behavioral, and Economic Research, UC Santa Barbara
[3]Methodology Unit, Department of Health and Mental Hygiene, New York City, NY
[4]Department of Geography, University of Utah
E-mail addresses: sweeney@isber.ucsb.edu,kkonty@health.nyc.gov,grace@geog.utah.edu.

sample that serves as the basis for the construction of US BMI reference distributions. The second is fitnessgram data from New York City. We use a subset of the records that contains complete four-year longitudinal histories for children between ages 5 and 18.

2. Measurement of Child/Adolescent Obesity

For research focused on population level assessment of adult obesity there is widespread agreement on the use of a BMI[1] thresholds of 25 for overweight and 30 for obese. For adults, even though BMI is an indirect measure of percent body fat there is strong evidence that BMI is both strongly correlated with percent fat and associated higher incidence several obesity-related chronic diseases. For example, in a life table study using the Framingham heart study panel non-smoking men and women with BMI above 25 at age 40, died approximately three years earlier than individuals with non-elevated BMI (Peeters et al., 2003).

The use of BMI for epidemiological studies of children and adolescents is far more problematic. The percent body fat in children follows a well established pattern that tracks developmental stages – adiposity rebound, sexual maturation – and those stages unfold differently for males and females (Rodríguez et al., 2011). A key problem is that changes in body composition unfold on a developmental clock and there is large variation among individuals in the chronological age when developmental stages occur. The natural variation in the timing of maturation presents as a peak in the coefficient of variation of age-specific BMI during mid-adolescence (Cole, 1989). Those developmental stages also unfold differently in different countries perhaps due to genetic differences in country-specific populations but also due to broader cultural and socio-economic influences. Clinical research suggest that there are race-specific differences in the timing of maturation and qualitative differences in adeposity, both of which complicate the interpretation of a single BMI threshold (Daniels et al., 1997). Another issue is that the connection between BMI and subsequent chronic health outcomes is more tenuous for children/adolescents compared to adults (cites). There is simply far greater uncertainty because any health outcomes are *chronic* rather than immediate, and thus thirty years of other causal factors may intervene between childhood obesity and subsequent health outcomes (Pelletier, 2006). All of these reason bring into question the use of BMI as both a screening tool and for population level obesity prevalence studies. On balance, the reality is that despite all these complications, it is the only available indirect measure that can be easily collected for large scale population studies.

2.1. **Reference distributions, LMS, and quantile regression.**

Given this complex setting, Pelletier (2006) provides a careful assessment of the issues involved in selecting cutoff values for child and adolescent BMI and recommends careful consideration in selecting appropriate reference distributions for the population under study. The dominant approach in constructing reference distributions for BMI is known as the *LMS* method; this is also the approach used in developing the CDC 2000 BMI growth centiles and in developing the IHOC international reference centiles. Given a reference population of healthy individuals, the LMS method proceeds by selecting – for each age×sex distribution – a suitable Box-Cox transformation (λ or L) to transform the raw BMI to normal, then uses the estimated mean (μ or M) and coefficient of variation (σ or S) to convert to standard normal distribution. Centiles from the transformed distributions can then be back-transformed to quantiles of BMI. Smoothly varying curves can be estimated simultaneously using penalized maximum likelihood (Cole and Green, 2006). In applied epidemiological studies, the resulting CDC reference distributions amount to transforming the BMI measures from a study population using the LMS parameters estimated from the reference population.

Some examples of how this looks in the New York City application are shown in Figure 1. In the top panel the LMS method is applied directly to the NYC study population for age 14 females

[1]Body mass index (BMI) is defined as an individuals weight in kilograms divided by their squared height in meters.

(left column) and age 10 males (right column). The Box-Cox transformation results in an approximately normal, thus symmetric, distribution. The results of using the CDC reference distribution parameters are on the second row. Because the positive-skewness of the study population is less than that of the reference population the transformed study population is now slightly (females) to strongly (males) left-skewed. While it is clearly not appropriate to study BMI using a direct LMS transformation of the study population, the two examples underscores that differences in the shape – not only the spread and center – between the study population and reference distribution are imbedded in prevalence studies based on the CDC, or other, standards. More problematic is that binary regression models that are frequently used to study relationships between covariates and BMI will yield inconsistent parameter estimates if the underlying index variable is not symmetric (Greene, 2003; Sweeney et al., 2012). Predictive prevalence differences based on those models will also be biased but to a lesser degree(Sweeney et al., 2012).

Quantile regression provides an alternative approach to fitting growth charts and for developing models with covariates (Wei et al., 2005). These alternative models are of the form:

$$Q_{Y_i|t_i,x_i}[\tau|t_i, x_i] = g_\tau(t_i) + x_i^T\gamma(\tau)$$

where τ is the quantile parameter such that τ proportion of the observation lie below the fitted curve or surface, Y is BMI, $g()$ is a smooth additive term on age, and x is a group indicator or some other individual-level covariate. Parameter estimates, γ are interpreted as a shift in the location of quantile, $Q_{Y_i|t_i}[\tau|t_i]$, given a unit change in covariate x. In contrast to the LMS method that imposes the assumption that BMI follows a skewed normal distribution, quantile regression is non-parametric and only relies on the order statistics of the empirical distribution. With the inclusion of smooth additive effects on age in the models it is possible to recover growth curves pertaining to the τth percentile.

2.2. **Growth curve comparisons based on quantile regression.** Graphical results from models including only the smooth age term are in Figure 2 – 6. The first growth curve comparison (Figure 2) is between historical NHANES data (pooling of waves I, II, and III spanning 1966 to 1990) and recent NHANES data (pooling of two waves spanning 2007-10). The dots in the background are individual data points and each line defines age-specific quantiles of BMI such that $100\tau\%$ of the data points lie below the line. Notice that historic growth curves trace a much narrower envelope than the recent data; thus capturing the dramatic shift in US child/adolescent obesity. The historic NHANES data is the basis for the current CDC 2000 growth charts.

The next comparison is between the current US population and current NYC public school population (see Figure 3). While the median and lower quantiles track fairly closely, the upper tails of the distribution differ at all ages but particularly after age 14. It is unclear at this point whether those differences reflect different compositions of US and NYC samples – the latter is poorer, more non-white, and contains more foreign-born – or if it reflects systematic measurement error. While measurement error is a possibility, it is noteworthy that since quantile regression only uses order statistics – and not actual values of BMI – the estimates are not impacted by outliers as is true of conditional mean models (such as ordinary least squares). Comparing the NYC pattern to historic NHANES data (figure 4), the two lowest quantiles (pertaining to $\tau = \{0.05, 0.15\}$) still track closely and perhaps represent biological lower bounds. The rest of the distribution has shifted towards higher BMI over time and the differences in inflection point suggest the overall shape of the distribution has changed in different ways and different ages. Again, this may reflect compositional differences in the two samples in addition to real shifts towards higher prevalence of adiposity.

A recent growth charts developed specifically for cross-national comparative studies is based on several national samples. It differs from past reference charts in attempting to link to adult obesity by selecting a obesity reference curve that passes through BMI=30 at age 18, and an overweight reference curve that passes through BMI=25 at age 18. As a final graphical exercise we can construct similar curves using the historic NHANES, recent NHANES, and NYC data. Focusing only on an obesity reference, the curves in Figures 5 and 6, each pass through BMI=30 at age 18. Notice that the NHANES curves from the two different periods follow essentially similar paths. The difference is that the historic path represents that 94th centile for males and 90th centile for women, whereas the recent path defines the roughly the 80th centiles for both men and women. The NYC curve defines the 87th and 85th centiles for men and women. Figure 6 provides a decomposition of the NYC population into race groups in an attempt to understand the source of the different composite distributional shape. For white males, the path has a similar shape to the national and the pronounced concavity of the overall curve is clearly due to the remaining racial groups. For females the tail of the differences in the upper tail for ages 14 and above appear to be present for all race groups.

3. Child/Adolescent Obesity in NYC: Differences by Race/Ethnic, FB, and SES

- This section to be completed. It will focus on using quantile regression and binary outcome models together to understand associations between BMI and race, ethnic, foreign-born, SES covariates. Quantile regression allows us to explore shifts in the shape of the distributions without imposing distributional assumptions. Binary outcome models will allow us to relate those changes to predictive differences in prevalence. The theory linking the quantile and binary models was published by two of the authors in an application to child stunting (Sweeney et al., 2012).
- Preliminary models have been fitted for the quantile regression component (see Tables 1–6). The data has cross-sectional and panel elements to it with ages 5 to 18 represented, and all individuals present for four observation periods (not necessarily consecutive). This allows us to construct lagged terms (BMI and height) and change in height per unit time to get a sense of presence of growth spurts. Height is particularly important to include because for children and adolescents, height is correlated with BMI.
- Binary outcome models will require choice of a one or more reference distributions. Following recommendations in Pelletier (2006) we plan to only examine prevalence of obesity using thresholds comparable to those in Figure 5.

References

T.J. Cole. The British, American NCHS, and Dutch weight standards compared using the lms method. *American Journal of Human Biology*, 1(4):397–408, 1989.

T.J. Cole and P.J. Green. Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in medicine*, 11(10):1305–1319, 2006.

S.R. Daniels, P.R. Khoury, and J.A. Morrison. The utility of body mass index as a measure of body fatness in children and adolescents: Differences by race and gender. *Pediatrics*, 99(6):804–807, 1997.

William Greene. *Econometric Analysis*. 5th edition. Prentice Hall, 2003.

A. Peeters, JJM Barendregt, F. Willekens, J.P. Mackenbach, A. Al Mamun, and LGA Bonneux. Obesity in adulthood and its consequences for life expectancy: A life-table analysis. *Annals of Internal Medicine*, 138(1):24–33, 2003.

D. Pelletier. Theoretical considerations related to cutoff points. *Food & Nutrition Bulletin*, 27 (Supplement 5):224–236, 2006.

G. Rodríguez, A. Pietrobelli, Y. Wang, and L.A. Moreno. Methodological aspects for childhood and adolescence obesity epidemiology. *Epidemiology of Obesity in Children and Adolescents*, pages 21–40, 2011.

S. Sweeney, F. Davenport, and K. Grace. Combining insights from quantile and ordinal regression: Child malnutrition in guatemala. *Economics & Human Biology*, 2012. URL `http://www.sciencedirect.com/science/article/pii/S1570677X12000767`.

L.E. Thorpe, D.G. List, T. Marx, L. May, S.D. Helgerson, and T.R. Frieden. Childhood obesity in New York City elementary school students. *American Journal of Public Health*, 94(9):1496–1500, 2004.

Y. Wang and M.A. Beydoun. The obesity epidemic in the United States – gender, age, socioeconomic, racial/ethnic, and geographic characteristics: A systematic review and meta-regression analysis. *Epidemiologic Reviews*, 29(1):6–28, 2007.

Y. Wei, A. Pere, R. Koenker, and X. He. Quantile regression methods for reference growth charts. *Statistics in Medicine*, 25(8):1369–1382, 2005.

		Race (ref=W)			Lunch (ref=full price)		POB
τ	Int	B	H	A	reduced	free	foreign
0.10	13.902	0.164	0.416	-0.482	0.021	0.082	-0.137
	(0.044)	(0.022)	(0.021)	(0.023)	(0.029)	(0.016)	(0.023)
0.20	14.560	0.197	0.575	-0.415	0.034	0.107	-0.124
	(0.041)	(0.021)	(0.020)	(0.022)	(0.028)	(0.016)	(0.022)
0.50	15.712	0.076	1.105	-0.345	0.079	0.195	-0.312
	(0.050)	(0.029)	(0.030)	(0.030)	(0.039)	(0.022)	(0.029)
0.80	17.542	0.378	1.606	-0.745	0.403	0.432	-0.669
	(0.085)	(0.052)	(0.047)	(0.047)	(0.065)	(0.035)	(0.046)
0.90	19.241	0.971	1.598	-1.092	0.472	0.559	-1.006
	(0.118)	(0.070)	(0.061)	(0.063)	(0.085)	(0.046)	(0.057)
0.95	21.194	1.337	1.515	-1.538	0.482	0.599	-1.331
	(0.135)	(0.089)	(0.078)	(0.080)	(0.101)	(0.058)	(0.075)

Table 1. NYC males

		Race (ref=W)			Lunch (ref=full price)		POB
τ	Int	B	H	A	reduced	free	foreign
0.10	13.394	0.367	0.436	-0.638	0.027	0.093	-0.109
	(0.048)	(0.023)	(0.023)	(0.024)	(0.031)	(0.017)	(0.023)
0.20	14.064	0.418	0.575	-0.723	0.058	0.131	-0.163
	(0.043)	(0.022)	(0.021)	(0.022)	(0.028)	(0.016)	(0.021)
0.50	15.391	0.870	1.083	-0.875	0.104	0.227	-0.408
	(0.051)	(0.030)	(0.028)	(0.027)	(0.038)	(0.021)	(0.026)
0.80	17.428	1.731	1.544	-1.403	0.202	0.483	-0.966
	(0.076)	(0.053)	(0.047)	(0.047)	(0.063)	(0.035)	(0.043)
0.90	18.898	2.198	1.493	-1.833	0.315	0.604	-1.262
	(0.105)	(0.070)	(0.063)	(0.064)	(0.085)	(0.046)	(0.059)
0.95	20.551	2.629	1.589	-2.122	0.447	0.722	-1.565
	(0.125)	(0.089)	(0.081)	(0.084)	(0.125)	(0.061)	(0.081)

Table 2. NYC females

τ	Int	$\mathrm{BMI}_{t-\Delta}$	$\Delta\mathrm{BMI}_{t-\Delta}$	Race (ref=W) B	H	A	Lunch (ref=full price) reduced	free
0.10	13.837	-0.013	0.016	0.152	0.406	-0.501	0.036	0.096
	(0.048)	(0.003)	(0.004)	(0.022)	(0.022)	(0.023)	(0.029)	(0.017)
0.20	6.219	0.523	0.016	0.069	0.250	-0.218	0.022	0.058
	(1.222)	(0.078)	(0.002)	(0.022)	(0.051)	(0.037)	(0.019)	(0.013)
0.50	2.133	0.829	0.031	0.048	0.180	-0.118	0.046	0.091
	(0.574)	(0.036)	(0.001)	(0.013)	(0.037)	(0.022)	(0.021)	(0.013)
0.80	2.146	0.885	0.051	0.203	0.276	-0.142	0.053	0.148
	(0.609)	(0.037)	(0.002)	(0.022)	(0.042)	(0.035)	(0.035)	(0.020)
0.90	3.232	0.861	0.063	0.390	0.392	-0.225	0.102	0.197
	(0.773)	(0.046)	(0.003)	(0.034)	(0.056)	(0.060)	(0.040)	(0.022)
0.95	4.936	0.815	0.067	0.628	0.590	-0.316	0.172	0.268
	(1.035)	(0.059)	(0.005)	(0.067)	(0.052)	(0.097)	(0.087)	(0.034)

TABLE 3. NYC males

τ	Int	$\mathrm{BMI}_{t-\Delta}$	$\Delta\mathrm{BMI}_{t-\Delta}$	Race (ref=W) B	H	A	Lunch (ref=full price) reduced	free
0.10	5.870	0.487	0.020	0.173	0.179	-0.340	0.052	0.088
	(0.606)	(0.041)	(0.002)	(0.024)	(0.027)	(0.028)	(0.024)	(0.014)
0.20	3.059	0.710	0.019	0.164	0.133	-0.230	0.062	0.089
	(0.411)	(0.027)	(0.001)	(0.017)	(0.022)	(0.023)	(0.018)	(0.011)
0.50	1.304	0.880	0.035	0.216	0.147	-0.166	0.048	0.116
	(0.242)	(0.016)	(0.001)	(0.016)	(0.018)	(0.019)	(0.017)	(0.010)
0.80	1.495	0.921	0.056	0.398	0.237	-0.191	0.054	0.168
	(0.267)	(0.017)	(0.001)	(0.027)	(0.025)	(0.026)	(0.025)	(0.014)
0.90	2.573	0.907	0.066	0.627	0.368	-0.270	0.083	0.224
	(0.343)	(0.022)	(0.001)	(0.042)	(0.031)	(0.042)	(0.035)	(0.022)
0.95	4.261	0.861	0.072	1.047	0.629	-0.375	0.070	0.292
	(0.456)	(0.029)	(0.002)	(0.052)	(0.042)	(0.072)	(0.059)	(0.036)

TABLE 4. NYC females

τ	Int	$BMI_{t-\Delta}$	$\Delta BMI_{t-\Delta}$	Race (ref=W) B	H	A	Poverty (ref=low) medium	high	extreme
0.10	13.895	-0.013	0.015	0.187	0.443	-0.471	-0.072	0.029	-0.011
	(0.048)	(0.003)	(0.004)	(0.023)	(0.022)	(0.023)	(0.021)	(0.024)	(0.025)
0.20	6.220	0.523	0.015	0.089	0.271	-0.210	0.020	0.042	0.007
	(1.228)	(0.078)	(0.001)	(0.026)	(0.055)	(0.033)	(0.019)	(0.015)	(0.018)
0.50	2.133	0.829	0.031	0.053	0.185	-0.115	0.059	0.086	0.090
	(0.580)	(0.036)	(0.001)	(0.014)	(0.042)	(0.020)	(0.012)	(0.015)	(0.015)
0.80	2.100	0.885	0.051	0.193	0.271	-0.123	0.062	0.091	0.221
	(0.607)	(0.037)	(0.002)	(0.025)	(0.045)	(0.033)	(0.019)	(0.020)	(0.023)
0.90	3.268	0.862	0.062	0.360	0.360	-0.213	0.058	0.124	0.334
	(0.774)	(0.046)	(0.003)	(0.034)	(0.052)	(0.062)	(0.027)	(0.030)	(0.035)
0.95	5.074	0.815	0.067	0.579	0.536	-0.283	0.027	0.105	0.469
	(1.039)	(0.060)	(0.005)	(0.076)	(0.061)	(0.093)	(0.045)	(0.051)	(0.057)

Table 5. NYC males

τ	Int	$BMI_{t-\Delta}$	$\Delta BMI_{t-\Delta}$	Race (ref=W) B	H	A	Poverty (ref=low) medium	high	extreme
0.10	5.887	0.487	0.019	0.203	0.214	-0.322	0.035	0.052	0.003
	(0.614)	(0.041)	(0.002)	(0.025)	(0.029)	(0.027)	(0.018)	(0.018)	(0.020)
0.20	3.071	0.711	0.018	0.186	0.158	-0.215	0.055	0.039	0.037
	(0.410)	(0.027)	(0.001)	(0.017)	(0.022)	(0.022)	(0.013)	(0.014)	(0.016)
0.50	1.339	0.881	0.035	0.231	0.168	-0.149	0.043	0.044	0.096
	(0.243)	(0.016)	(0.001)	(0.017)	(0.019)	(0.019)	(0.012)	(0.013)	(0.014)
0.80	1.551	0.922	0.056	0.394	0.235	-0.174	0.044	0.086	0.227
	(0.270)	(0.017)	(0.001)	(0.028)	(0.026)	(0.024)	(0.016)	(0.019)	(0.022)
0.90	2.610	0.908	0.065	0.612	0.355	-0.249	0.052	0.131	0.358
	(0.349)	(0.022)	(0.001)	(0.044)	(0.034)	(0.041)	(0.025)	(0.029)	(0.033)
0.95	4.421	0.863	0.071	0.976	0.571	-0.357	-0.001	0.134	0.505
	(0.459)	(0.030)	(0.002)	(0.058)	(0.046)	(0.072)	(0.045)	(0.052)	(0.060)

Table 6. NYC females

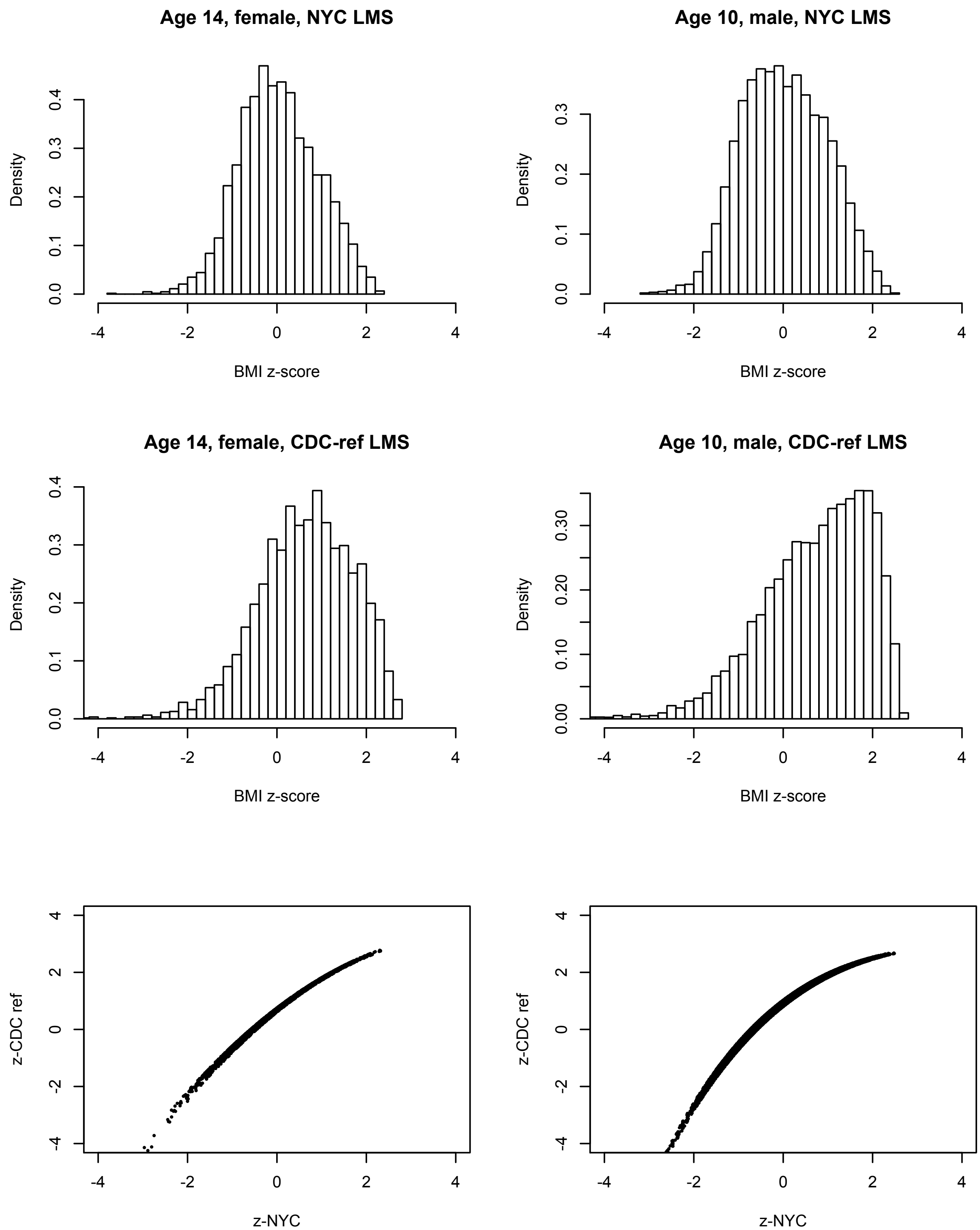

FIGURE 1. BMI transformations. Top panel: Direct application of LMS transformation from raw to normal; Middle panel: LMS transformation using parameters from CDC national reference distribution; Bottom panel: QQ-plot of transformations from panels 1 and 2.

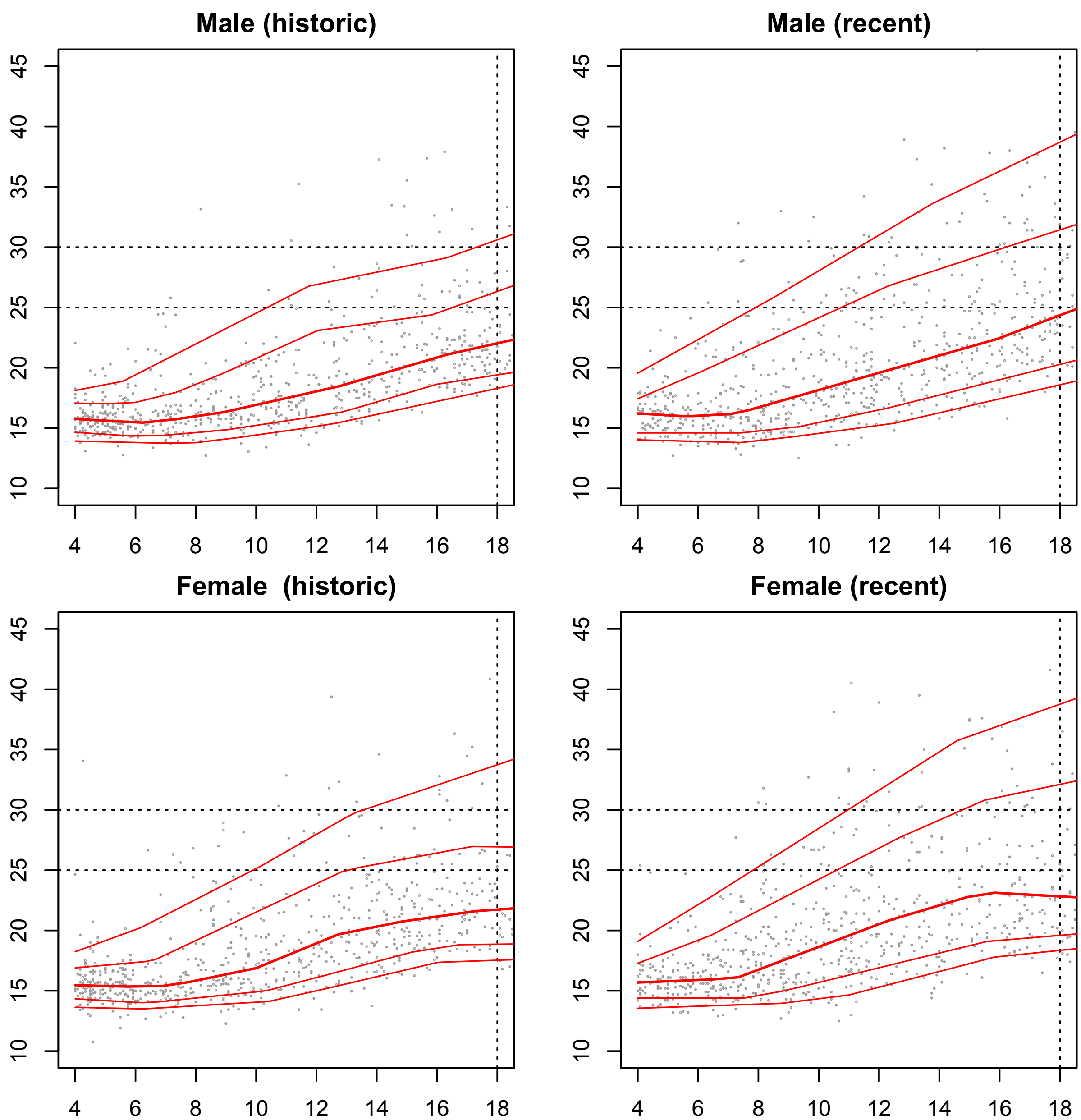

FIGURE 2. Growth charts from quantile regression. Plotted lines are from a smooth additive age term in quantile regression of raw BMI on age. Left panel is based on nationally representative NHANES data waves I, II, and III. Right panel is from 2007-10 NHANES data. Quantiles shown from lowest to highest pertain to $\tau = \{0.05, 0.15, 0.5, 0.85, 0.95\}$

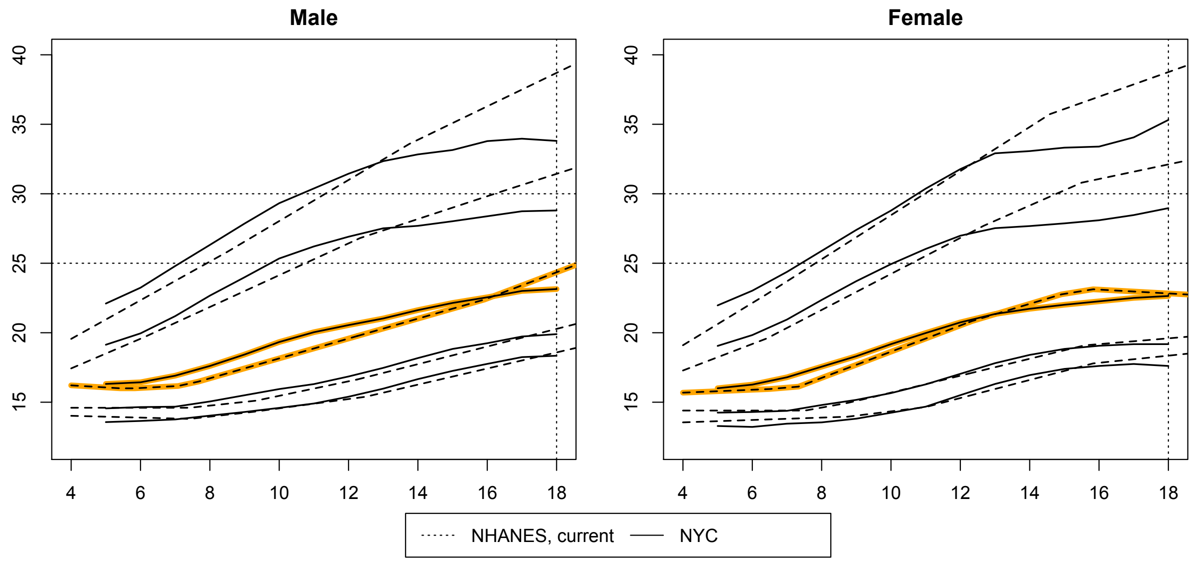

FIGURE 3. Growth charts from quantile regression. Quantiles shown pertain to $\tau = \{0.05, 0.15, 0.5, 0.85, 0.95\}$ and orange shading highlights the median growth path.

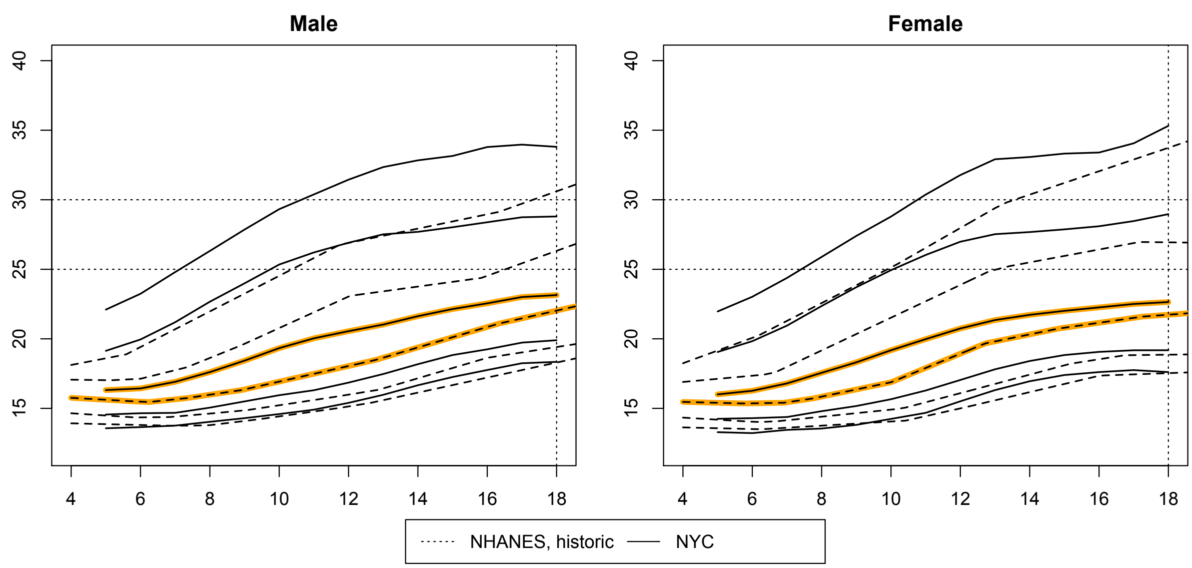

FIGURE 4. Growth charts from quantile regression. Quantiles shown pertain to $\tau = \{0.05, 0.15, 0.5, 0.85, 0.95\}$ and orange shading highlights the median growth path.

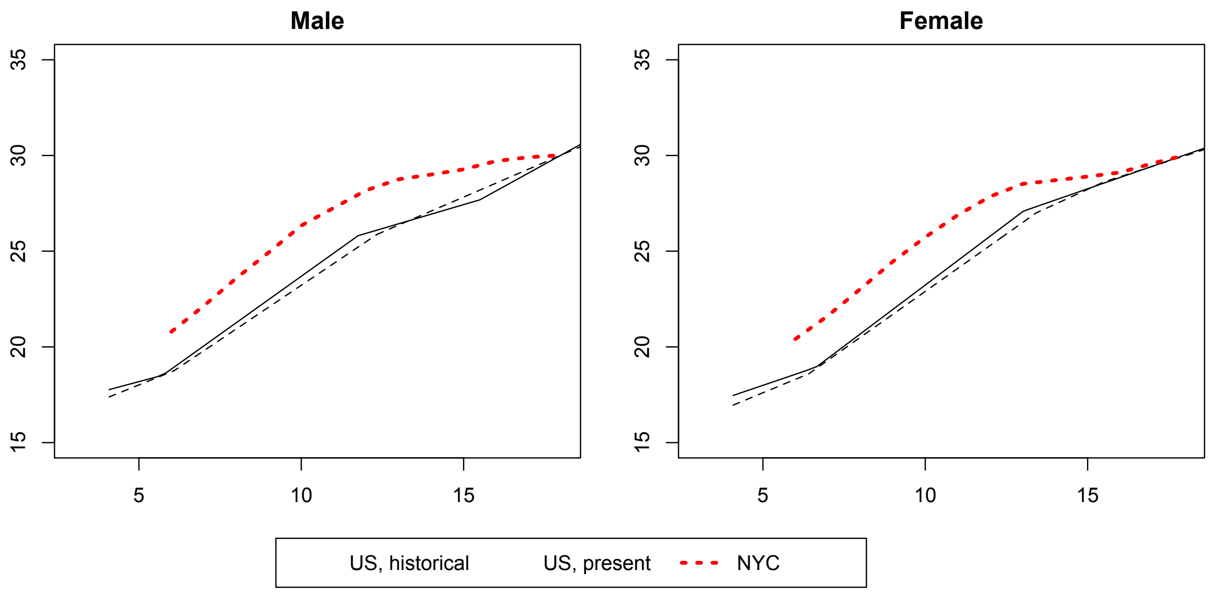

FIGURE 5. IOTF-type obesity cutpoints defined using historic NHANES, recent NHANES, and NYC fitnessgram. Each line passes through BMI 30 at age 18.

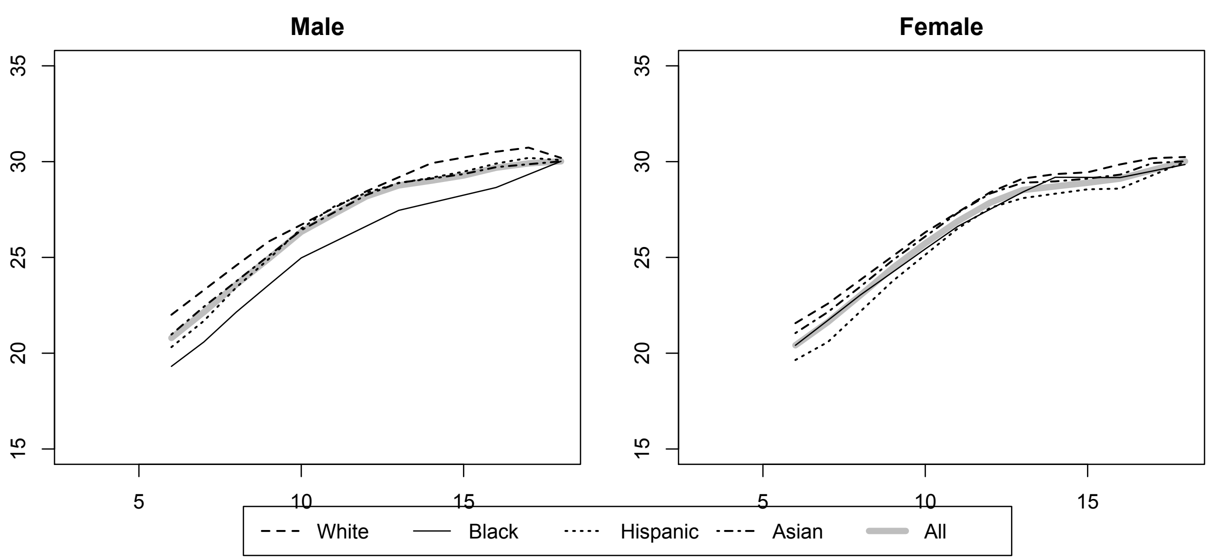

FIGURE 6. IOTF-type race group obesity cutpoints defined using NYC fitnessgram. Each line passes through BMI 30 at age 18.