



Published in final edited form as:

Obesity (Silver Spring). 2014 May ; 22(5): E149–E156. doi:10.1002/oby.20618.

Use of Quantile Regression to Investigate the Longitudinal Association between Physical Activity and Body Mass Index

Matteo Bottai¹, Edward A. Frongillo¹, Xuemei Sui³, Jennifer R. O'Neill³, Robert E. McKeown^{5,6}, Trudy L. Burns⁴, Angela D. Liese⁵, Steven N. Blair^{3,5}, and Russell R. Pate³

¹Division of Biostatistics, Institute of Environmental Medicine, Karolinska Institutet

²Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina

³Department of Exercise Science, Arnold School of Public Health, University of South Carolina

⁴Department of Epidemiology, University of Iowa

⁵Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina

⁶Greenville Health System, Greenville, SC

Abstract

Objective—To examine associations among age, physical activity (PA), and birth cohort on body mass index (BMI) percentiles in men.

Design and Methods—Longitudinal analyses using quantile regression were conducted among men with two examinations between 1970 and 2006 from the Aerobics Center Longitudinal Study (n=17,759). Height and weight were measured; men reported their PA and were categorized as inactive, moderately or highly active at each visit. Analyses allowed for longitudinal changes in PA.

Results—BMI was greater in older than younger men and in those born in 1960 than those born in 1940. Inactive men gained weight significantly more rapidly than active men. At the 10th percentile, increases in BMI among inactive, moderately active, and highly active men were 0.092, 0.078, and 0.069 kg/m² per year of age, respectively. The 10th percentile increased by 0.081 kg/m² per birth year and by 0.180 kg/m² at the 90th percentile, controlling for age.

Conclusion—Although BMI increased with age, PA reduced the magnitude of the gradient among active compared to inactive men. Regular PA had an important, protective effect against weight gain. This study provides evidence of the utility of quantile regression to examine the specific causes of the obesity epidemic.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding author: Russell R. Pate, PhD, Department of Exercise Science, Arnold School of Public Health, 921 Assembly Street Suite 212, Columbia, SC 29208, Telephone: 803-777-2456, Fax: 803-777-2504, rpate@mailbox.sc.edu.

Competing Interests

The authors have no competing interests.

Keywords

obesity; physical activity; quantile regression; longitudinal; men

Introduction

The world-wide obesity epidemic (1-4) can be attributed to a widespread imbalance between energy intake and energy expenditure. The prevalence of obesity, defined as body mass index (BMI) ≥ 30.0 kg/m², has increased dramatically among men over the past 50 years from 10.4% in 1960-62 (5) to 35.5% in 2009-10 (6). The specific factors that cause the energy imbalance are still poorly understood. Some have suggested that physical activity has been essentially unchanged during the obesity epidemic, and conclude that the cause of the epidemic must be an increase in energy intake (7, 8). However, a major factor to consider is the rapid change in occupational energy expenditure over the past 50 years, with a large decline in manufacturing, mining, and farming; and a consistent increase in service jobs with substantially lower energy requirements (9). Importantly, mean daily energy expenditure from occupational physical activity declined by more than 100 calories over the past five decades, and that decrease accounted for a significant portion of the mean weight gain during that time period (9).

Multiple individual and environmental factors may affect an individual's ability to achieve energy balance and maintain a stable weight over time, and a number of observational and interventional studies have examined the potential effects of these factors (10-13).

Assessments of the potential influences on obesity tend to focus on the upper percentiles of the frequency distribution of BMI in categorical logistic regression analyses or on the mean as in linear regression analyses. Both approaches are limited because they sacrifice what can be learned about the *entire* distribution. For instance, the influence of age, physical activity, and birth cohort on BMI may affect subgroups of the population differently; thus, the effect on mean BMI may not adequately convey the potential varying impact on the entire distribution. Quantile regression is an analytical method that is compatible with assessing associations throughout the distribution of BMI (14-19). To date, no study has used quantile regression to examine the influences of age, physical activity, and birth cohort prospectively on obesity among adult men.

Therefore, the primary purpose of this paper was to determine the associations among age, physical activity, and birth cohort on the BMI percentiles of the distribution in a large sample of men. We hypothesized that BMI values would be centered on higher values in 60-year-old men than in 20-year-old men, and that BMI would be higher in 40-year-old men born in 1960 than in 40-year-old men born in 1940. We also expected that the BMI distribution would be shifted towards larger values with age to a greater degree in inactive men than in active men, but in a way that would not be uniform across the BMI distribution. The secondary purpose of this paper was to describe the application of an underutilized statistical method, quantile regression (14, 15), to study factors influencing BMI, an application for which the method seems particularly well suited.

Methods and Procedures

Sample selection

The Aerobics Center Longitudinal Study (ACLS) is a prospective observational study (20). Participants came to the Cooper Clinic in Dallas, TX for periodic preventive health examinations and counseling regarding diet, exercise, and other lifestyle factors associated with increased risk of chronic disease. Between 1970 and 2006, participants received at least one comprehensive medical examination and maximal graded treadmill exercise test at the clinic, and were enrolled in the ACLS. Most study participants were non-Hispanic whites from middle-to-upper socioeconomic strata, and were either referred by their employers or physicians or were self-referred. The study was reviewed and approved annually by the Cooper Institute Institutional Review Board, and all participants gave written informed consent. From the initial sample of 120,649 observations from 50,787 men, we included men without any history of heart attack, stroke or cancer (observations = 103,379, participants = 46,132), 25 to 75 years old (observations = 102,229, participants = 45,515), and men with at least two visits (observations = 74,473, participants = 17,759). In the final sample, 7,334 men had two visits, 3,566 men had three visits, 1,989 men had four visits, and 4,870 men had five or more visits.

Measures

The comprehensive health evaluation is described in detail elsewhere (20, 21). The outcome of interest in this study was BMI (kg/m^2). Height and weight were measured on a physician's scale and stadiometer. The exposures of interest were self-reported physical activity, diet, and smoking behavior. Physical activity was categorized based on participants' responses to questions about their regular physical activity habits over the past three months (1 = no activity, 2 = some sports or activity or walk/jog/run up to 10 miles per week, 3 = walk/jog/run more than 10 miles per week) (21-23). Categories of physical activity were defined at each visit as "inactive" if physical activity = 1, "moderate" if physical activity = 2, and "high" if physical activity = 3. The analysis allowed for changes in physical activity level over time. Smoking habits were obtained from a standardized questionnaire. Participants were classified as a nonsmoker or current smoker at the time of each examination. Eating habits were self-reported as eating: 1) much less, 2) somewhat less, 3) just what, 4) somewhat more, or 5) much more than I want. Birth cohort was defined as each participant's year of birth.

Statistical Analyses

We employed quantile regression to assess associations of predictor variables at the 10th, 25th, 50th, 75th, and 90th percentiles of BMI. Quantile regression has the advantages of allowing examination at multiple points in the distribution of BMI rather than only at the mean. Quantile regression does not require any assumption about the distribution of the regression residuals and, unlike ordinary linear regression, is not influenced by outliers or skewness in the distribution of the dependent variable, providing greater statistical efficiency when outliers are present. In addition, inference on quantiles can accommodate transformation of the dependent variable without the problems encountered in ordinary linear regression (24).

Quantile regression parameters are interpreted similarly to normal linear regression parameters except that the parameter indicates the change in the value at the modeled percentile, not the mean, of the dependent variable for each unit change in the independent variable. For example, a parameter estimate of 0.133 for age in the 75th percentile model would indicate that the 75th percentile of BMI increased by 0.133 kg/m² for each one year increase in age.

The densities shown were smoothed by applying the Epanechnikov kernel function, $K(x) = 0.75(1 - x^2)I(|x| < 1)$, with bandwidth 3 to a dense set of estimated quantiles (2nd, 4th, ..., 98th percentile). A kernel function gives the weights of the nearby data points in making an estimate while ensuring that the result is a probability density function and that the average of the corresponding distribution is equal to that of the sample used. The repeated observations of BMI taken on the same men may be dependent. The quantile regression estimator is consistent when the data are dependent (25). Because we were interested in population-level, not individual-level, estimates, we estimated the standard errors and confidence intervals with 1,000 cluster bootstrap samples to account for the dependence (26-29).

For completeness, we complemented our inference on quantiles with that from two other, more traditional approaches: linear regression and multinomial logistic regression. The former permits inference about the mean of BMI whereas the latter allows estimation of the conditional probability of being in any given BMI class (18-25, 25-30, and ≥ 30 kg/m²). To take the potential intra-individual dependence into account, the standard errors were estimated by applying generalized estimating equations (GEE) with an exchangeable working covariance matrix for the linear regression on the mean (30) and the robust cluster sandwich estimator for multinomial logistic regression (31).

Results

The shape of the BMI distribution differed across the levels of physical activity with respect to location, spread, and skewness (Figure 1). Across the three levels of physical activity, there were no statistically significant differences in the mean or quartiles of age or height of the men (Table 1). There were, however, gradients in weight, waist circumference, BMI, and body fat mass in the expected direction, with inactive men having the greatest relative weight and fat, and highly active men having the least.

BMI was higher at older ages for all three physical activity levels, but with a smaller gradient in the physically active as compared to the inactive men (Table 2). The difference in the magnitude of increase was significant at the 10th and 25th percentiles, as indicated by the statistically significant cross-product interaction terms for age and physical activity level (high vs. inactive) in those models. At the 10th percentile, below which was the leanest 10% of the population, the gradients in BMI with age in the inactive, moderately active, and highly active were 0.092, 0.078, and 0.069 kg/m² per year of age, respectively. The 10th percentile of BMI increased with year of birth by 0.081 kg/m² per birth cohort year and by 0.180 kg/m² at the 90th percentile, adjusting for age, and the magnitude of increase associated with year of birth was larger at each successive percentile. Eating habits were

significantly associated with BMI at all percentiles, and the category “eat just what I want” showed the largest reduction of BMI values at all percentiles. Smoking and drinking habits were not significant predictors or confounders, and were omitted from all models.

Figure 2 illustrates the results of the quantile regression analysis and compares the estimates for the distribution of BMI in the inactive and highly active populations at age 30, 50, and 70 (chart rows), in the cohorts born in 1940 and 1960 (chart columns) in those who reported that they eat “just what I want.” The distributions for moderate physical activity were similar to those for high physical activity and are not shown. The distribution in the active population in the 1940 cohort at age 30 is included as a shaded area in all graphs for reference. Moving down each column, the distribution of BMI shifted toward larger values with older age. Skewness abated with age in the inactive men because of the comparatively larger increase in the lower percentiles than in the higher ones. Conversely, skewness increased in the active population. Comparing the two columns, there was a conspicuous, significant cohort effect on both location and spread of the BMI distribution. The later generation shifted toward higher values and showed an accentuated elongation.

Table 3 reports the estimated coefficients and associated confidence intervals from GEE linear regression models for mean BMI, which increased with age at all levels of physical activity. All main effects were statistically significant ($P < 0.05$). The difference in the slopes of the increase in BMI over age across levels of physical activity is borderline significant ($P = 0.060$). These estimates however, were obtained after removing outliers (BMI < 14 or BMI > 50). Inference was dependent on which values were identified as outliers and removed. If all data were utilized, the estimated coefficients were substantially different, the standard errors inflated, the confidence intervals wider, and the difference in slopes far from statistically significant (data not shown).

Table 4 shows the estimated probability of being in one of three BMI categories at age 30, 50 and 70, for the 1940 cohort that reported “eat what I want.” All main effects were statistically significant ($P < 0.05$). The probability of having normal BMI was lower with older age. At age 70 the probability of having normal BMI was more than twice as great for the highly active men than for the inactive men. Further, the probability for highly active obese men was less than half that of the inactive men.

DISCUSSION

Quantile regression permitted us to describe and quantify that, as expected, BMI was centered on larger values in older compared to younger men and in those born in 1960 compared to those born in 1940. With the use of quantile regression, we also found that the distribution of BMI was shifted towards larger values in older ages to a greater degree in inactive men than in those who were physically active. It also allowed us to show that these relationships were not uniform across the distribution of BMI. For example, the association of physical activity and BMI was greater at the larger percentiles of BMI than at the smaller percentiles, as shown by the greater magnitude of the regression coefficients at the larger percentiles. Therefore, men in the normal BMI range who led inactive lives tended to have higher weight gradients with age than men who maintained an active lifestyle. For example,

at age 70, nearly half of the active men had a BMI that was below the 10th percentile of the inactive men. Quantile regression also permitted estimation of the entire distribution of BMI by age and year of birth adjusted for eating habits. Furthermore, quantile regression was statistically efficient and insensitive to extreme values of BMI.

Quantile regression has several advantages that apply directly to the analysis of our data set. First, research interest lies not in the mean of BMI but in its quantiles. Our study interest was in the complete distribution of BMI: the underweight, normal, overweight, and obese men. Inference on mean BMI alone would not be as informative as inference on multiple quantiles throughout the distribution. Quantile regression permits inference on multiple percentiles of BMI given a set of covariate values. Second, quantile regression has robustness to outliers and statistical efficiency. Large, outlying values have a major impact on the mean and therefore on linear regression estimates. Conversely, quantile regression is robust to them. Robustness to outliers makes the quantile estimator more efficient than the mean estimator when the population being sampled contains outliers. Third, quantile regression does not require transformations. When the relationship between dependent and independent variables is non-linear or the distribution of the dependent variable is skew, transforming the outcome may simplify modeling when using linear regression. Transformations such as the logarithm and the square root are frequently applied in linear regression, but are often challenging to implement in practice because of inconsistent back transformation and challenges in interpretation (32). In contrast, quantile regression accommodates skewed distributions seamlessly. This property of quantile regression has been exploited to great advantage in other settings, for example, power-transformations (33) and censored data (34), and is applicable to the analysis of BMI. Other measures of obesity may be bounded from above or below, such as percentage of body fat mass bounded between 0 and 100%, and quantile regression is also applicable for these measures (24).

These complex relationships could not be described by linear or multinomial logistic regression analyses. Inference about mean BMI from linear regression was unsatisfactory because it did not permit understanding the differential effect of physical activity across the distribution of BMI, and it was highly affected by the extreme values of BMI. Unless outliers were identified and removed through an ad hoc and somewhat arbitrary process, the amount of change in mean BMI with age did not appear to differ significantly across levels of physical activity, whereas in the quantile regression results, the differences in slope with age for differing activity levels were apparent in some quantiles. Multinomial logistic regression performed better than linear regression because it allowed inference about the upper tail of the distribution of BMI (i.e., overweight and obese), but it was inferior to quantile regression because it categorized the outcome BMI into a small number of groups that did not permit examination of the entire BMI distribution.

Quantile regression could be extended to the analysis of the effects of other risk factors on BMI or the analysis of other obesity measures. This analytic approach has the potential to contribute greatly to forming a fuller picture of the extent and causes of the obesity epidemic, and understanding the impact of large-scale obesity interventions, given that interventions may primarily affect one portion of the distribution of the outcome (35). Quantile regression is readily implemented by available statistical software (e.g., Stata, SAS,

S-plus). It should be used instead of linear regression in any study for which the effects of explanatory variables may differ across the range of the outcome variable and affect the shape of the distribution.

Therefore, analyses should not focus merely on the mean BMI, for these would provide diluted, tangential measures of the effects of interest. For instance, influences on BMI may differ across the population: stronger associations at the upper end of the distribution ($> 70^{\text{th}}$ percentile), moderate associations in the middle of the distribution (30^{th} - 70^{th} percentile), and low or no association at the lower end of the distribution ($< 30^{\text{th}}$ percentile). Thus, the effect on mean BMI may not adequately convey the impact on the entire distribution. Further, if, as Rose (36) argued there are potentially greater population benefits from approaches that encompass the large portion of the population at moderate risk, then we must employ methods that can provide information about effects *throughout* the distribution.

This study has strengths and limitations that deserve mention. A major strength is the large sample with multiple measurements over the period of thirty-six years. Another strength is the use of quantile regression which allowed for a comprehensive examination of the relationship between physical activity and BMI across the entire distribution of BMI. A limitation is that the ACLS cohort is predominately white, well-educated, and of middle-to-upper socio-economic status, and is not representative of the general population (20). However, we contend that one of the advantages of the method proposed here is that it allows for more appropriate comparisons of findings from similar studies in other populations where distributions may be shifted or associations may vary across the distributions in different ways.

In summary, quantile regression was an effective statistical method that allowed us to examine how physical activity affected BMI across the entire distribution of BMI. Our findings demonstrated that the distribution of BMI ranged over higher values in older men than in younger men. However, the shift in the BMI distributions between younger and older men was smaller among regularly active men than among inactive men. The beneficial effect of regular physical activity in attenuating the BMI increase with ageing was most evident among the lower percentiles, which in younger men were within the range of normal BMI values. For example, the 25^{th} percentile of BMI increased with age at a rate that was 24% smaller in active men (76 g/m^2 per year) than inactive men (100 g/m^2 per year). This study provides compelling evidence of the utility of quantile regression to examine the specific causes of the obesity epidemic.

Acknowledgements

This work was supported by the U.S. Department of Defense (W81XWH-08-1-0082) and the National Institutes of Health grants AG06945, HL62508, and R21 DK088195 (to X.S. from the National Institute of Diabetes and Digestive and Kidney Diseases).

The authors thank the Cooper Clinic physicians and technicians for collecting the data and the staff at the Cooper Institute for data entry and data management. In addition, this work was undertaken by the collaborative effort of the TRIM Research Group, and the TRIM authors express their appreciation to other members: **Clemson University:** Susan Barefoot, Margaret Condrasky, Ellen Granberg; **Cooper Institute:** Susan Campbell; **Medical University of South Carolina:** Patrick M. O'Neil; **Pennington Biomedical Research Center:** David W. Harsha; **South Carolina Research Authority:** Kate Beaver, Robert Davis, Stephen L. Jones; **South Carolina State**

University: Bonita Manson; **University of Iowa:** Kathleen F. Janz; **University of South Carolina:** Robert R. Moran; **Winthrop University:** Patricia G. Wolman.

References

1. James WP. WHO recognition of the global obesity epidemic. *Int J Obes (Lond)*. 2008; 32:S120–S126. [PubMed: 19136980]
2. World Health Organization. Obesity and overweight. 2011. [WWW document]. URL <http://www.who.int/mediacentre/factsheets/fs311/en/>
3. Yach D, Stuckler D, Brownell KD. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nat Med*. 2006; 12:62–66. [PubMed: 16397571]
4. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999–2008. *JAMA*. 2010; 303:235–241. [PubMed: 20071471]
5. Flegal KM, Carroll MD, Kuczmarski RJ, Johnson CL. Overweight and obesity in the United States: Prevalence and trends. *Int J Obes*. 1998; 22:39–47.
6. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *JAMA*. 2012; 307:491–497. [PubMed: 22253363]
7. Swinburn B, Sacks G, Ravussin E. Increased food energy supply is more than sufficient to explain the US epidemic of obesity. *Am J Clin Nutr*. 2009; 90:1453–1456. [PubMed: 19828708]
8. Westerterp KR, Plasqui G. Physically active lifestyle does not decrease the risk of fattening. *PLoS One*. 2009; 4:e4745. [PubMed: 19270740]
9. Church TS, Thomas DM, Tudor-Locke C, Katzmarzyk PT, Earnest CP, Rodarte RQ, et al. Trends over 5 decades in U.S. occupation-related physical activity and their associations with obesity. *PLoS One*. 2011; 6:e19657. [PubMed: 21647427]
10. French SA, Story M, Jeffery RW. Environmental influences on eating and physical activity. *Ann Rev Public Health*. 2001; 22:309–335. [PubMed: 11274524]
11. Loos RJJ, Bouchard C. Obesity - is it a genetic disorder? *J Int Med*. 2003; 254:401–425.
12. Dietz WH, Gortmaker SL. Preventing obesity in children and adolescents. *Ann Rev Public Health*. 2001; 22:337–353. [PubMed: 11274525]
13. Jakicic JM. The role of physical activity in prevention and treatment of body weight gain in adults. *J Nutr*. 2002; 132:3826S–3829S. [PubMed: 12468633]
14. Koenker R, Hallock KF. Quantile regression. *J Econ Perspect*. 2001; 15:143–156.
15. Koenker, R. Quantile regression. Cambridge University Press; New York: 2005.
16. Mitchell JA, Pate RR, Espana-Romero V, O’Neill JR, Dowda M, Nader PR. Moderate-to-vigorous physical activity is associated with decreases in body mass index from ages 9 to 15 years. *Obesity (Silver Spring)*. 2013; 21:E280–E293. [PubMed: 23592682]
17. Mitchell JA, Rodriguez D, Schmitz KH, Audrain-McGovern J. Greater screen time is associated with adolescent obesity: A longitudinal study of the BMI distribution from Ages 14 to 18. *Obesity (Silver Spring)*. 2013; 21:572–575. [PubMed: 23592665]
18. Mitchell JA, Hakonarson H, Rebbeck TR, Grant SF. Obesity-susceptibility loci and the tails of the pediatric BMI distribution. *Obesity (Silver Spring)*. 2013; 21:1256–1260. [PubMed: 23408508]
19. Espana-Romero V, Mitchell JA, Dowda M, O’Neill JR, Pate RR. Objectively measured sedentary time, physical activity and markers of body fat in preschool children. *Pediatr Exerc Sci*. 2013; 25:154–163. [PubMed: 23406703]
20. Blair SN, Kohl HW III, Paffenbarger RS Jr, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality: A prospective study of healthy men and women. *JAMA*. 1989; 262:2395–2401. [PubMed: 2795824]
21. Kampert JB, Blair SN, Barlow CE, Kohl HW III. Physical activity, physical fitness, and all-cause and cancer mortality: a prospective study of men and women. *Ann Epidemiol*. 1996; 6:452–457. [PubMed: 8915477]
22. Kohl HW III, Blair SN, Paffenbarger RS Jr, Macera CA, Kronenfeld JJ. A mail survey of physical activity habits as related to measured physical fitness. *Am J Epidemiol*. 1988; 127:1228–1239. [PubMed: 3369421]

23. Banda JA, Hutto B, Feeney A, Pfeiffer KA, McIver K, LaMonte MJ, et al. Comparing physical activity measures in a diverse group of midlife and older adults. *Med Sci Sports Exerc.* 2010; 42:2251–2257. [PubMed: 20421836]
24. Bottai M, Cai B, McKeown RE. Logistic quantile regression for bounded outcomes. *Stat Med.* 2010; 29:309–317. [PubMed: 19941281]
25. Jung SH. Quasi-likelihood for median regression models. *J Am Stat Assoc.* 1996; 91:251–257.
26. Lipsitz SR, Fitzmaurice GM, Molenberghs G, Zhao LP. Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus. *J R Stat Soc Ser C Appl Stat.* 1997; 46:463–476.
27. Geraci M, Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics.* 2007; 8:140–154. [PubMed: 16636139]
28. Liu Y, Bottai M. Mixed-effects models for conditional quantiles with longitudinal data. *Int J Biostat.* 2009; 5:28.
29. Geraci M, Bottai M. Linear quantile mixed models. *Stat Comput.* 2013 doi 10.1007/s11222-013-9381-9.
30. Liang KY, Zeger SL. Longitudinal data-analysis using generalized linear-models. *Biometrika.* 1986; 73:13–22.
31. Huber, PJ. The behavior of maximum likelihood estimates under nonstandard conditions; *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press; Berkeley. 1967; p. 221-233.
32. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ.* 2001; 20:461–494. [PubMed: 11469231]
33. Mu Y, He X. Power transformation toward a linear regression quantile. *J Am Stat Assoc.* 2007; 102:269–279.
34. Powell JL. Censored regression quantiles. *J Econom.* 1986; 32:143–155.
35. Frith AL, Naved RT, Ekstrom EC, Rasmussen KM, Frongillo EA. Micronutrient supplementation affects maternal-infant feeding interactions and maternal distress in Bangladesh. *Am J Clin Nutr.* 2009; 90:141–148. [PubMed: 19439457]
36. Rose G. Sick individuals and sick populations. *Int J Epidemiol.* 1985; 14:32–38. [PubMed: 3872850]

What Is Already Known About This Subject

- The specific factors that cause energy imbalance are still poorly understood.
- Studies of the potential influences on obesity tend to focus on the upper percentiles of the frequency distribution of BMI in logistic regression or on the mean in linear regression.
- Both approaches are limited because they sacrifice what can be learned about the *entire* distribution.

What This Study Adds

- More complete understanding of the distribution of BMI that showed higher values and more elongated shapes among the older individuals and more recent generations of our study population.
- Insights into the important, protective effect of regular physical activity against age and birth year that shifted and reshaped the distribution of BMI values toward a desirable range.
- An illustration of an application of quantile regression as an effective statistical method to examine the effects of possible specific causes of the obesity epidemic.

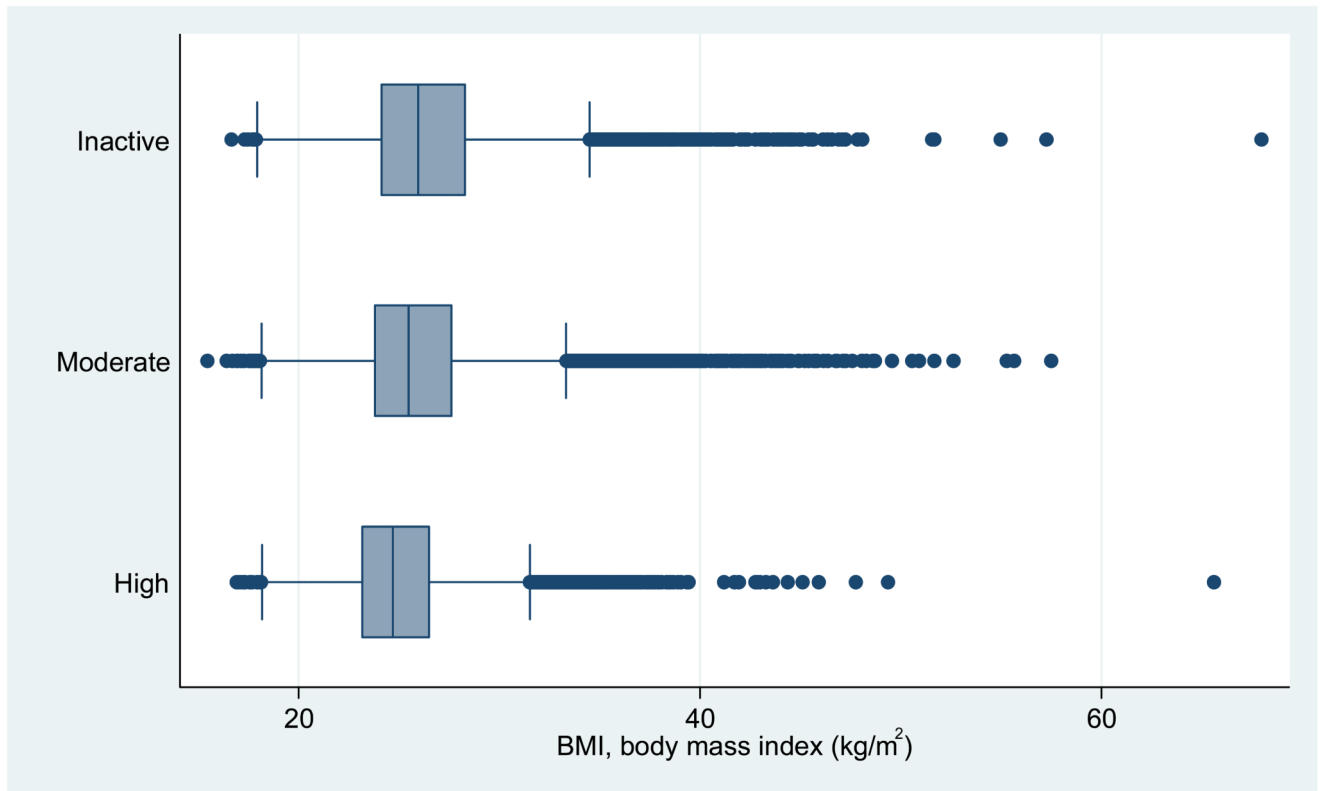


Figure 1. Box plots of BMI by levels of physical activity of men, Dallas, Texas, 1970-2006. Values with BMI > 70 kg/m² were excluded.

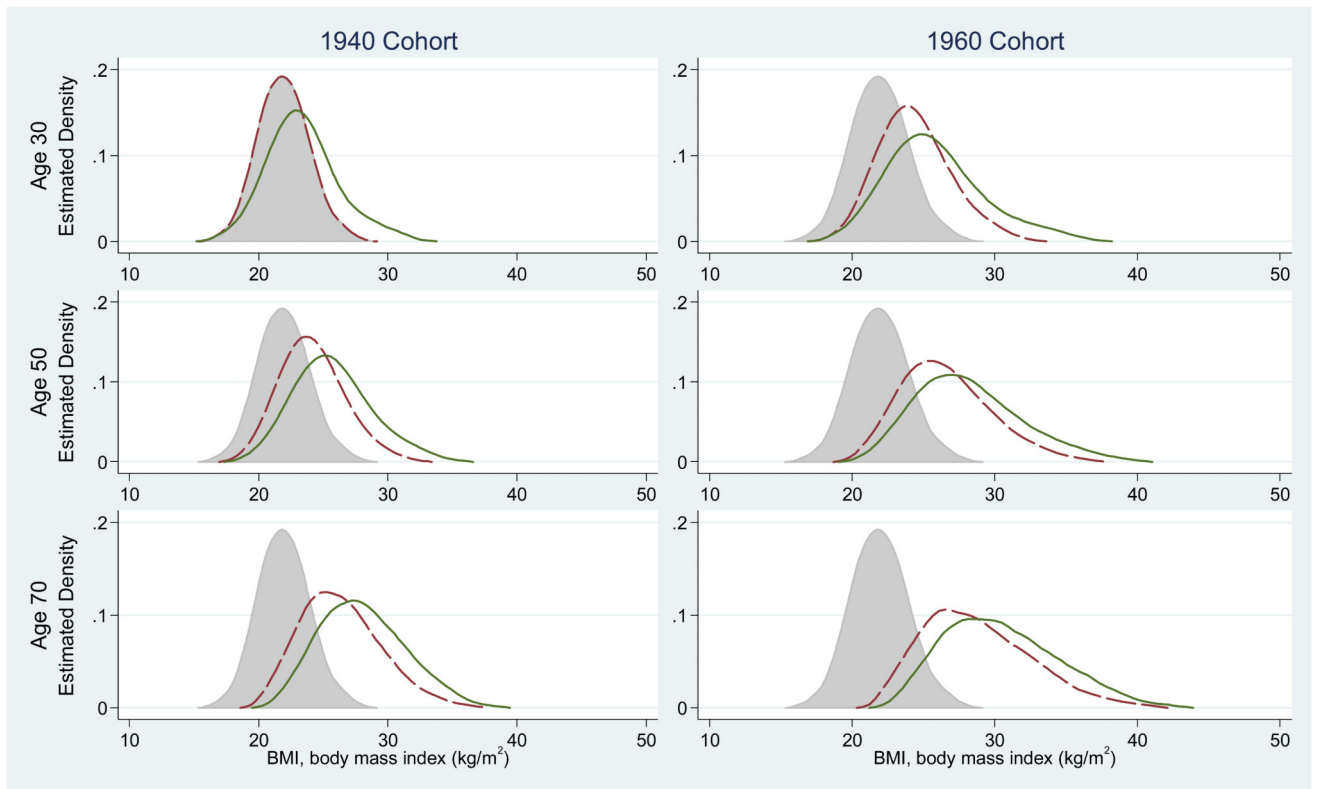


Figure 2.

Quantile regression estimates of the BMI distributions from the model shown in Table 2 in inactive (solid, green curve) and highly active (dashed, red curve) men at ages 30, 50, and 70 in the 1940 cohort and 1960 cohort for men who “eat just what I want,” Dallas, Texas, 1970-2006 (men = 8,885; observations = 17,304). Shaded area in all panels is for reference and represents physically active 30-year-old men in the 1940 cohort.

Table 1

Descriptive statistics of men by physical activity level (inactive, moderate PA, high PA) across all visits, Dallas, Texas, 1970–2006 (men = 17,759, observations = 74,473)^a

	Inactive (obs. = 18,552)				Moderate PA (obs. = 37,058)				High PA (obs. = 18,863)			
	25 th	50 th	75 th	Mean or %	25 th	50 th	75 th	Mean or %	25 th	50 th	75 th	Mean or %
Age, years	40	47	54	47.2	41	48	54	48.0	42	48	55	48.6
Height, cm	174.6	179.1	182.9	178.9	175.3	179.1	183.5	179.2	174.6	179.1	182.9	178.9
Weight, kg	76.4	83.4	92.0	85.2	75.5	82.0	90.3	83.7	73.3	79.3	86.3	80.4
Waist circumference, cm	88.0	94.0	101.0	94.3	86.0	92.0	98.0	91.0	84.0	89.0	94.0	88.0
BMI, kg/m ²	24.2	26.0	28.3	26.5	23.8	25.5	27.6	26.1	23.2	24.7	26.5	25.1
Body fat, %	18.1	22.1	26.1	22.2	17.0	20.8	24.4	20.8	15.0	18.8	22.5	18.8
Current smoker				14%				12%				8%
Alcohol consumption (0/wk)				27%				26%				25%
Alcohol consumption (1-7/wk)				51%				50%				49%
Alcohol consumption (8/wk)				22%				23%				26%
Eat much less than I want				7%				5%				6%
Eat somewhat less than I want				35%				43%				43%
Eat just what I want				44%				41%				41%
Eat somewhat more than I want				12%				10%				9%
Eat much more than I want				2%				1%				1%

PA, physical activity; obs., observations; BMI, body mass index; GEE, generalized estimating equations.

^a Differences across physical activity levels in the percentiles (tested with quantile regression with cluster bootstrapped standard errors), the means (GEE), and the proportions (multinomial logistic regression with robust cluster sandwich estimator) are significant ($P < 0.05$) for all variables.

Table 2

Effects of predictors at five percentiles (10th, 25th, 50th, 75th, and 90th) of the distribution of body mass index (kg/m²) estimated by quantile regression in men, Dallas, Texas, 1970-2006 (men = 8,885; observations = 17,304)

	10 th percentile	25 th percentile	50 th percentile	75 th percentile	90 th percentile
PA (Moderate: vs. Inactive)					
Coefficient ^a	-0.358	-0.576	-0.850	-1.000	-1.368
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI ^b	-0.553, -0.164	-0.732, -0.420	-1.026, -0.674	-1.226, -0.774	-1.756, -0.979
PA (High vs. Inactive)					
Coefficient ^a	-0.834	-1.154	-1.563	-1.899	-2.512
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI ^b	-1.039, -0.629	-1.330, -0.978	-1.744, -1.383	-2.156, -1.641	-2.939, -2.085
Age (years, centered at 50)					
Coefficient ^a	0.092	0.100	0.120	0.133	0.143
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI ^b	0.069, 0.115	0.078, 0.122	0.098, 0.142	0.104, 0.162	0.100, 0.186
Interaction (Age × Moderate PA)					
Coefficient ^a	-0.014	-0.011	-0.016	-0.003	0.015
<i>P</i>	0.184	0.248	0.132	0.828	0.425
95% CI ^b	-0.035, 0.007	-0.030, 0.008	-0.037, 0.005	-0.030, 0.024	-0.022, 0.053
Interaction (Age × High PA)					
Coefficient ^b	-0.023	-0.024	-0.017	0.005	0.009
<i>P</i>	0.026	0.017	0.128	0.702	0.674
95% CI ^a	-0.043, -0.003	-0.043, -0.004	-0.038, 0.005	-0.022, 0.033	-0.033, 0.052
Eating Habit (2 vs. 1) ^c					
Coefficient ^a	-1.329	-1.893	-2.661	-3.499	-3.930
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI ^b	-1.844, -0.815	-2.235, -1.551	-3.093, -2.229	-3.960, -3.038	-4.625, -3.235
Eating Habit (3 vs. 1) ^c					
Coefficient ^a	-2.353	-2.861	-3.685	-4.592	-4.809

	10 th percentile	25 th percentile	50 th percentile	75 th percentile	90 th percentile
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI ^b	-2.880, -1.827	-3.198, -2.523	-4.139, -3.232	-5.089, -4.095	-5.520, -4.099
Eating Habit (4 vs. 1)^c					
Coefficient ^a	-0.692	-0.967	-1.575	-2.363	-2.330
<i>P</i>	0.029	<0.001	<0.001	<0.001	<0.001
95% CI ^b	-1.312, -0.071	-1.354, -0.580	-2.055, -1.095	-2.888, -1.837	-3.222, -1.438
Eating Habit (5 vs. 1)^c					
Coefficient ^a	-0.102	0.392	0.664	0.480	1.205
<i>P</i>	0.821	0.388	0.097	0.372	0.086
95% CI ^b	-0.984, 0.781	-0.498, 1.282	-0.121, 1.448	-0.572, 1.532	-0.172, 2.582
Cohort^e					
Coefficient ^a	0.081	0.090	0.110	0.144	0.180
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI ^b	0.067, 0.096	0.077, 0.102	0.095, 0.125	0.129, 0.159	0.153, 0.207
Intercept^d					
Coefficient ^a	24.516	26.530	29.144	32.005	34.734
<i>P</i>	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI ^b	23.991, 25.040	26.173, 26.886	28.704, 29.583	31.507, 32.503	33.985, 35.484

BMI, body mass index; PA, physical activity.

^aThe coefficient represents the change in the value at the nth percentile of BMI for each unit change in the independent variable. For interactions, the coefficient is the difference in the change in the value of BMI at the nth percentile compared to the main relative to the change when the interacting variable is at its reference level, so, for example, at the 10th percentile, BMI increases by 0.096 for each year of age for those who are inactive, but by 0.023 less than that per year of age for those with high physical activity.

^bConfidence intervals (CI) are based on 1,000 cluster bootstrap samples. Test for interaction terms: 10th (*P* = 0.082), 25th (*P* = 0.048), 50th (*P* = 0.255), 75th (*P* = 0.695), 90th (*P* = 0.722).

^c1=Eat much less than I want; 2=Eat somewhat less than I want; 3=Eat just what I want; 4=Eat somewhat more than I want; 5=Eat much more than I want.

^dThe intercept is the value of the nth percentile of BMI when all other variables are zero.

^eBirth year centered at 1940.

Table 3

Effects of predictors at the mean of the distribution of body mass index estimated by generalized estimating equations after removing outliers^a in men, Dallas, Texas, 1970-2006 (men = 8,882; observations = 17,295)

	Coefficient	P value	95% CI
PA (Moderate vs. Inactive)	-0.335	<0.001	-0.413, -0.257
PA (High vs. Inactive)	-0.787	<0.001	-0.881, -0.693
Age (years, centered at 50)	0.106	<0.001	0.096, 0.115
Interaction (Age × Moderate PA) ^b	-0.003	0.433	-0.011, 0.005
Interaction (Age × High PA) ^b	0.006	0.226	-0.004, 0.016
Eating Habit (2 vs. 1) ^c	-0.868	<0.001	-1.000, -0.736
Eating Habit (3 vs. 1) ^c	-1.010	<0.001	-1.149, -0.870
Eating Habit (4 vs. 1) ^c	-0.363	<0.001	-0.520, -0.206
Eating Habit (5 vs. 1) ^c	0.895	<0.001	0.594, 1.196
Cohort (birth year, centered at 1940)	0.114	<0.001	0.105, 0.123
Intercept	27.184	<0.001	27.029, 27.339

BMI, body mass index; CI, PA, physical activity.

^aDefined as BMI < 14 kg/m² or BMI > 50 kg/m².

^bTest for interaction terms: $P = 0.060$.

^c1=Eat much less than I want; 2=Eat somewhat less than I want; 3=Eat just what I want; 4=Eat somewhat more than I want; 5=Eat much more than I want.

Table 4

Predicted probabilities for being in a defined BMI category (normal weight, overweight, obese) based on multinomial regression in the 1940 cohort of men who “Eat Just What I Want,” Dallas, Texas, 1970-2006 (men = 8,885; observations = 17,304)

PA Level	Age	Normal Weight		Overweight		Obese	
		Pred. Prob.	95% CI ^a	Pred. Prob.	95% CI ^a	Pred. Prob.	95% CI ^a
Inactive	30	0.77	0.81, 0.71	0.22	0.18, 0.27	0.02	0.01, 0.02
	50	0.45	0.48, 0.42	0.46	0.43, 0.48	0.09	0.08, 0.10
	70	0.15	0.20, 0.11	0.54	0.53, 0.54	0.31	0.27, 0.35
Moderate	30	0.82	0.85, 0.77	0.17	0.14, 0.21	0.01	0.01, 0.01
	50	0.56	0.59, 0.54	0.39	0.37, 0.40	0.05	0.05, 0.06
	70	0.25	0.30, 0.21	0.55	0.53, 0.57	0.19	0.17, 0.22
High	30	0.87	0.89, 0.83	0.13	0.10, 0.16	0.00	0.00, 0.01
	50	0.66	0.68, 0.64	0.31	0.29, 0.33	0.03	0.02, 0.03
	70	0.35	0.41, 0.29	0.52	0.48, 0.54	0.14	0.11, 0.17

BMI, body mass index, PA, physical activity; Pred. Prob., predictive probability; CI, confidence interval.

^aThe 95% confidence intervals are based on robust, cluster, sandwich estimator for the standard error.