

# Effect of Environmental Factors on Obesity: A Quantile Regression Approach

Anthony J Payne<sup>1</sup>, Julia A Knight<sup>2</sup> and Taraneh Abarin<sup>1\*</sup>

<sup>1</sup>Department of Mathematics and Statistics, Memorial University, Canada

<sup>2</sup>Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada

## Abstract

**Objectives:** This study explored associations of environmental factors with percent trunk fat (PTF) and body mass index (BMI), using quantile regression to explain variability in these traits at percentiles of the distributions.

**Methods:** Using a sample of 1695 adults from Newfoundland and Labrador, multiple and quantile regression models were used to analyse the significance of environmental factors on the average population and upper percentiles of the BMI and PTF distributions.

**Results:** Higher physical activity was associated with significantly lower PTF and BMI in the average population and upper percentiles, regardless of age. Both genders in percentiles closer to the median of PTF had more benefit with increased physical activity compared to higher percentiles. Interestingly, adults in higher percentiles of BMI distribution seem to benefit more with increased physical activity compared to percentiles closer to the median.

**Conclusion:** Using quantile regression as a robust approach toward violation of normality assumptions and outliers, variations in PTF and BMI for individuals across upper percentiles of the distributions based on some lifestyle factors were described. This method may be used to estimate the impact of certain lifestyle on different percentiles of BMI and PTF, rather than average population.

**Keywords:** Trunk fat; Body mass index; Newfoundland population; Quantile regression; Environmental factors; Obesity

**Abbreviations:** NL: Newfoundland and Labrador; PTF: Percent Trunk Fat; DXA: Dual-energy X-ray Absorptiometry; OLS: Ordinary Least Squares; FFQ: Food Frequency Questionnaire; PCA: Principal Component Analysis

## Introduction

According to the World Health Organization (WHO), an adult with body mass index (BMI) of 30 or more is defined as obese. An adult with a BMI equal to or more than 25 is defined as overweight. Overweight and obesity is a major epidemic in Newfoundland and Labrador (NL), and throughout Canada. The results from several surveys conducted by Statistics Canada show that NL has had the highest percentage of overweight/obese residents in Canada since 2007, and this percentage had risen to 63.2% in 2012. BMI is commonly used when examining obesity due to its ease and low cost of measurement. In this study, we also examined percent trunk fat (PTF), which is not commonly explored in obesity studies. BMI estimates the total fat mass using anthropometric measurements, while PTF is the measure of fat mass in the trunk region using dual-energy X-ray absorptiometry (DXA). DXA can produce an accurate measurement of adipose tissue within the body with a low margin of error [1,2]. Research studies show that measures of abdominal fat are closely associated with obesity-related detrimental effects on health [3-5].

Although remedies such as log-transformation may improve the modelling process, an alternative and more robust method (compared to OLS) against non-normality and outliers is quantile regression, which has been widely used in modelling obesity-related traits [6-8]. Quantile regression models the relationship between covariates and the conditional quantiles of the response given the covariates, so it is especially useful in applications where extremes are important, such as analyzing BMI where upper (overweight and obese) quantiles of population levels are critical from a public health perspective. It

provides a more complete picture of the conditional distribution of the response given the covariates and it is relatively easy to incorporate different covariates into the analysis. In this study, we applied quantile regression methodology (as well as OLS), in order to more accurately assess the relationships between the covariates and the response.

Our goal is to quantify the effect of basic environmental factors on PTF and BMI in a way that is simple to calculate and more crafted for population quantiles, rather than the population average. This may lead to new methods for estimating one's likelihood of abdominal obesity based on PTF (in addition to BMI) using environmental factors that may be accessible to general physicians and potentially the general public.

## Materials and Methods

### Coding study

All subject data used was taken from the CODING (Complex Diseases in the Newfoundland Population: Environment and Genetics) study. The study was conducted by the Faculty of Medicine at Memorial University of Newfoundland. To be eligible to take part in this study, participants had to be at least 19 years of age, be a third generation or higher Newfoundlander, and not have any serious metabolic, cardiovascular, or endocrine diseases. Posters and handouts were

\*Corresponding author: Taraneh Abarin, Department of Mathematics and Statistics, Memorial University, 230 Elizabeth Ave, St. John's, NL A1B 3X9, Canada, Tel: +1(709) 864-8733; Fax: +1(709) 864-3010; E-mail: [tabarin@mun.ca](mailto:tabarin@mun.ca)

Received March 31, 2016; Accepted April 14, 2016; Published April 22, 2016

**Citation:** Payne AJ, Knight JA, Abarin T (2016) Effect of Environmental Factors on Obesity: A Quantile Regression Approach. J Biom Biostat 7: 293. doi:10.4172/2155-6180.1000293

**Copyright:** © 2016 Payne AJ, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

distributed in public facilities in the city of St. John's, NL to recruit individuals to take part. Individuals were required to complete a series of questionnaires in order to obtain information regarding diet and physical activity. All patients consented to participation in all aspects of the CODING study.

### Anthropometrics and body composition

Subjects were weighed to the nearest 0.1 kg in standardized hospital gowns (Health O Meter, Bridgeview, IL). Height was measured using a fixed stadiometer (nearest 0.1 cm). BMI was calculated as weight in kilograms divided by height in meters squared ( $\text{kg}/\text{m}^2$ ). PTF was measured using Lunar Prodigy (DXA) (GE Medical Systems, Madison, WI). DXA measurements were performed on subjects following the removal of all metal accessories, while lying in a supine position. Quality assurance was performed on the DXA scanner daily and the typical coefficient of variation was 1.3% during the study period. All anthropometric and body composition measurements followed a 12-hour fast.

### Physical activity

Physical activity patterns were measured using the Atherosclerosis Risk in Communities Questionnaire [9], which consists of a Work Index, Sports Index, and Leisure Time Activity Index. A variable exclusive of workplace activity was used in the data analysis. This variable is continuous; however three categories were developed to classify individuals into low (index less than 5), moderate (index 5-8), and high (index greater than 8) physical activity in order to generalize results and make them more accessible.

### Dietary intake, smoking status, and alcohol intake

Dietary intake information was collected using a 124-item semi-quantitative Willett Food Frequency Questionnaire (FFQ) [10]. The number of weekly servings of various food items consumed was collected using the FFQ, and then NutriBase Clinical Nutrition Manager (version 8.2.0; Cybersoft Inc, Phoenix, AZ) was used to calculate the daily intake (mg/day) for all dietary information, for all subjects in the study. One alcohol consumption variable was collected through the FFQ and analysed with NutriBase Clinical Nutrition Manager, which provided a continuous quantitative variable (g/day/kg of body weight). Individuals also responded to a separate question on their alcohol consumption habits and this information was categorized as no alcohol, low, moderate, and high consumption.

### Statistical analysis

Statistical analyses were conducted using R version 3.0.0 GUI. Out of over 3000 individuals that participated in the CODING study, 1695 unrelated individuals were used for the analysis, in order to ensure that results were not biased due to the shared environment or genetics of related individuals (Table 1). Only individuals with complete data for the variables we explored were included. For all models, gender was first used as a covariate. However, due to significant differences (t-tests) in both BMI and PTF for males and females, all individuals were stratified by gender for further analysis. Our results also confirm that environmental factors seem to have different impacts on females than males.

Although the OLS method was only used as a benchmark for the quantile regression approach, a complete series of required standard assumptions (normality, homogeneity, goodness-of-fit) were assessed.

In the CODING study, there are over 200 variables measured for individuals. There are no variables representing socio-economic status available in the dataset. In order to examine the independent impact of the nutritional variables as well as other lifestyle variables (such as physical activity, drinking habit, smoking status), two separate sets of analyses were performed; one with and one without the nutritional variables.

### Models with nutritional variables

Principal Component Analysis (PCA) was applied to reduce the dimension of the total 85 nutritional variables. Due to high variability in the data, the calculation was first done using the eigenvalues of the correlation matrix. The computation was then repeated using a singular value decomposition of the data matrix with all variables scaled. Graphical tools were utilized in the selection process. Four components explaining 68.8% of the variability in the data were selected. PCA was conducted using the `princomp` and `prcomp` built-in functions in package `stats` in R. PTF and BMI were then separately modelled using multiple regression analysis, using four components (in addition to other variables), with and without natural log-transformation. The same approach was repeated for quantile regression analysis.

### Models without the nutritional variables

The variables included in the models were age, alcohol consumption, smoking status, and physical activity. The interaction terms between age and alcohol consumption and age and smoking status were also examined. All models were compared using analysis of variance, Akaike information criterion values, and  $R^2$  values. Results from testing hypotheses were considered significant for p-values less than 0.05.

The normality assumption for the OLS models was assessed by standard methods and not satisfied. Applying the Box-Cox transformation method resulted in natural log-transformation for these models.

In order to provide robust estimates for the upper percentiles of the distributions of PTF and BMI, quantile regression analysis was performed. A quantile regression model for a certain quantile  $q$  ( $0 < q < 1$ ) produces the expected response value of the quantile, based on the covariates.

Similar to the idea of OLS for the case of a single covariate, the  $q^{\text{th}}$  sample quantile regression estimator ( $\beta_q$ ) can be obtained by minimizing  $\sum p_q(y_i - \beta_q x_i)$  with respect to  $\beta_q$ , where  $y_i$  and  $x_i$  are the response and covariate for the  $i^{\text{th}}$  observation respectively, and  $p_q$  is the simple piecewise linear function, balancing between the number of observations lying above and below the fitted model [11]. The extension to the multivariate case is straightforward. To compare those with high levels of PTF/BMI with the median, we used the 50<sup>th</sup>, 80<sup>th</sup>, 90<sup>th</sup>, and 95<sup>th</sup> percentiles as cut-off values. The quantile regression analysis was conducted using the `quantreg` package (version 5.05) in R.

## Results

### Models with the nutritional variables

The first 4 principal components contained 6 nutritional variables across the components. As one may expect, calorie consumption played an important role in explaining the variability in the nutritional variables. Other variables in the four components were alpha carotene, lycopene, vitamin A, vitamin D, and calcium. Overall, adding the loading values for the nutritional principal components had negligible impact on any of the models (data not shown).

Variable	Category	Male (n=402)			Female (n=1293)		
		Mean (SD)	%	Median	Mean (SD)	%	Median
Age (years)	-	39.4 (14.3)	-	40.3	42.9 (12.1)	-	44.0
BMI	Overall	27.4 (4.5)	-	26.8	26.2 (5.1)	-	25.2
	Normal/underweight	22.9 (1.6)	32.8	-	22.4 (1.8)	48.8	-
	Overweight/obese	29.6 (3.8)	67.2	-	29.9 (4.6)	51.2	-
PTF	-	29.5 (9.8)	-	30.7	38.4 (8.9)	-	39.6
Smoking status	Non-smoking	-	87.8	-	-	89.4	-
	Smoking	-	12.2	-	-	10.6	-
Alcohol consumption	No alcohol	-	15.9	-	-	20.9	-
	Low	-	60.2	-	-	67.1	-
	Moderate	-	20.2	-	-	10.7	-
	High	-	3.7	-	-	1.3	-
Physical activity category	Low	-	20.9	-	-	22.1	-
	Moderate	-	61.2	-	-	67.7	-
	High	-	17.9	-	-	10.2	-
Physical activity score	-	6.4 (1.4)	-	6.5	6.2 (1.3)	-	6.0

Table 1: Summary data of unrelated individuals from the CODING study used in this analysis.

### Models without the nutritional variables

The natural log-transformation on PTF/BMI did not improve the models (results not shown). The final OLS models explained 23.8% and 15.4% of the variation in PTF for males and females, respectively (full model coefficients are in Table 2. The numbers were 7.7% and 6.5% for BMI models, for males and females respectively.

The age variable was centred, so corresponding coefficients can be directly interpreted as the respective PTF/BMI at the mean age. For every year increase in age, on average, PTF increased by 0.24 and 0.17 units, respectively, for males and females. For both genders, the impact of age on the PTF percentiles declined as the percentile increased (Figure 1). This implies that aging for the individuals in the higher tail of PTF is a less important factor in explaining PTF. This figure also presents the estimated impact of other factors on PTF, and their corresponding 95% confidence intervals, for the selected quantiles. The red line indicates the OLS estimate of average PTF. The impact of age on the average BMI for both genders was not as strong as for PTF. Figure 2 summarizes these results. This figure also shows that unlike PTF, for both genders, the impact of age on BMI stayed more or less unchanged in the upper percentiles.

For both males and females, increased physical activity level was significantly associated with a decrease in both average PTF and BMI. Figure 3 shows the impact of high physical activity (as compared to low physical activity) on the means and medians of both PTF and BMI for both genders. More importantly, this impact (except for median BMI for males) remained significant with increase in age. More specifically, for females with moderate or high physical activity, there was an average 3.3 and 9.4 unit reduction in PTF respectively. As Table 2 shows, females in all selected percentiles benefit from moderate to high physical activity. Interestingly, the impact of moderate to high physical activity on the average PTF declined as percentiles above median increased, but increased on the average BMI (Figures 1 and 2). This implies that females closer to the median of PTF have more benefit with increased physical activity compared to higher percentiles, but the benefit with respect to BMI is greater at higher percentiles.

For males with moderate and high physical activity, PTF was almost an average of 4.3 and 9 units lower, respectively, compared to those with low activity. Similar to females, as Table 2 shows, the impact of high physical activity for males declined with increase in percentiles.

Furthermore, males also exhibit more reduction in average BMI for individuals at higher BMI percentiles compared to closer to the median (Figures 1 and 2). It should also be mentioned that the results of the tests for the interaction of age with smoking and alcohol, for both genders, were not significant. It is noted from Table 2 that alcohol consumption may significantly affect the average PTF for females (based on the OLS method only) (Tables 1, 2 and Figures 1-3).

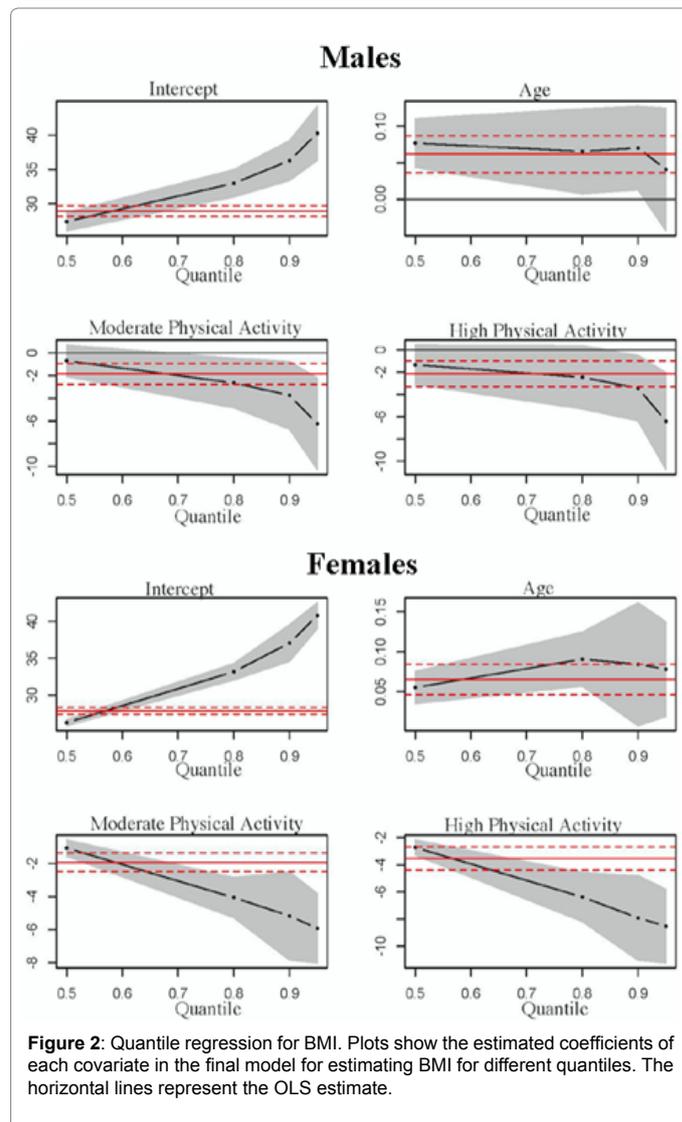
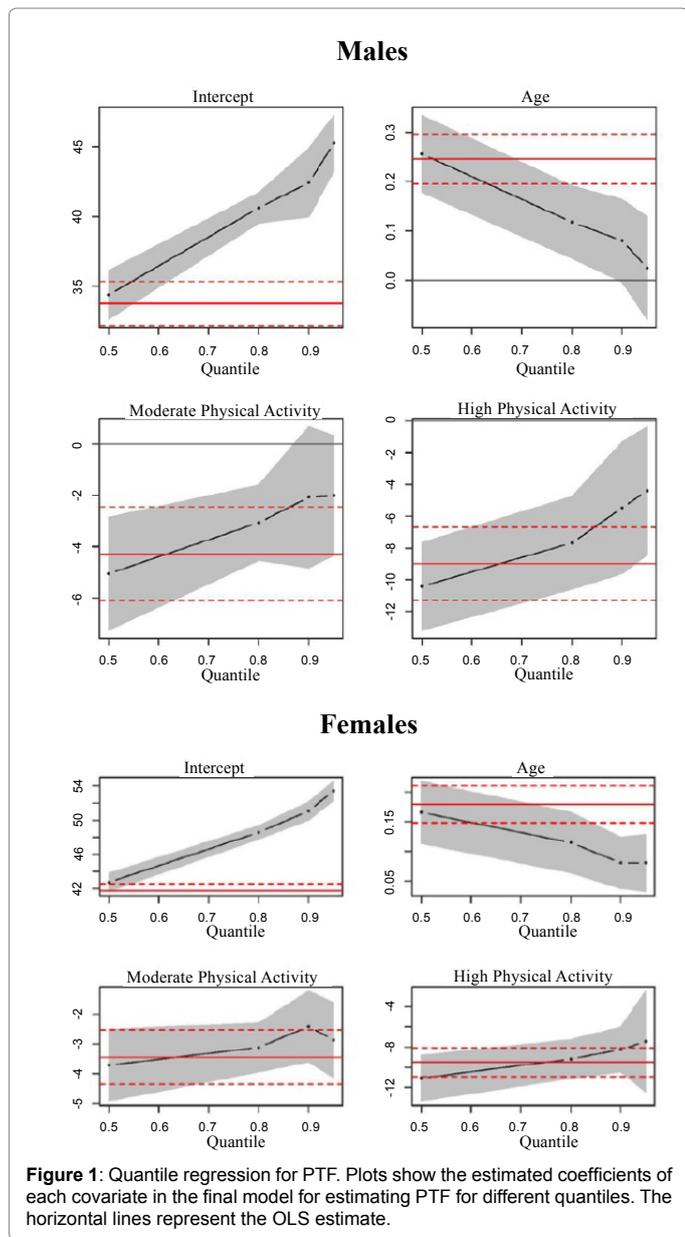
### Discussion

In this study we applied several models to test whether or not the magnitude of PTF and BMI in different percentiles may be modified by a series of environmental factors. Our results regarding the impact of physical activity on average PTF are consistent with the literature [12,13]. Our models, based on DXA measurements of PTF have not been explored before to our knowledge. Most studies of this kind have worked with easily-measured traits such as BMI and waist circumference [14-16]. The novelty of this study is to estimate the effect of different lifestyles on different percentiles of PTF compared to BMI.

With respect to the relationship between alcohol and PTF in women, the negative sign of the coefficients should be interpreted with caution as it is not a significant factor at any of the percentiles or in any of the BMI models. The effect of increased alcohol consumption on obesity is still debatable [17].

This study benefits from a large sample of the NL adult population. NLs generally isolated and homogeneous people provide a suitable population to model obesity-related traits based on environmental factors only, while controlling for confounding genetic factors.

The authors acknowledge the limitations of this study. A study with a larger sample size would provide higher power in detecting smaller effects. In addition, there are some limitations regarding the ability to generalize since it consists of volunteers rather than random individuals. There may be possible error or bias in self-reported physical activity level as well as dietary intake. It is also possible that other dietary or environmental information not considered in this study may potentially affect one's obesity level. We also acknowledge the limitations of PCA for looking at the nutritional variables. Additional analyses may reveal more information about diet and obesity. Furthermore, our study was based on cross-sectional models. Longitudinal data analysis is required to determine whether high physical activity is effective in explaining PTF over time.



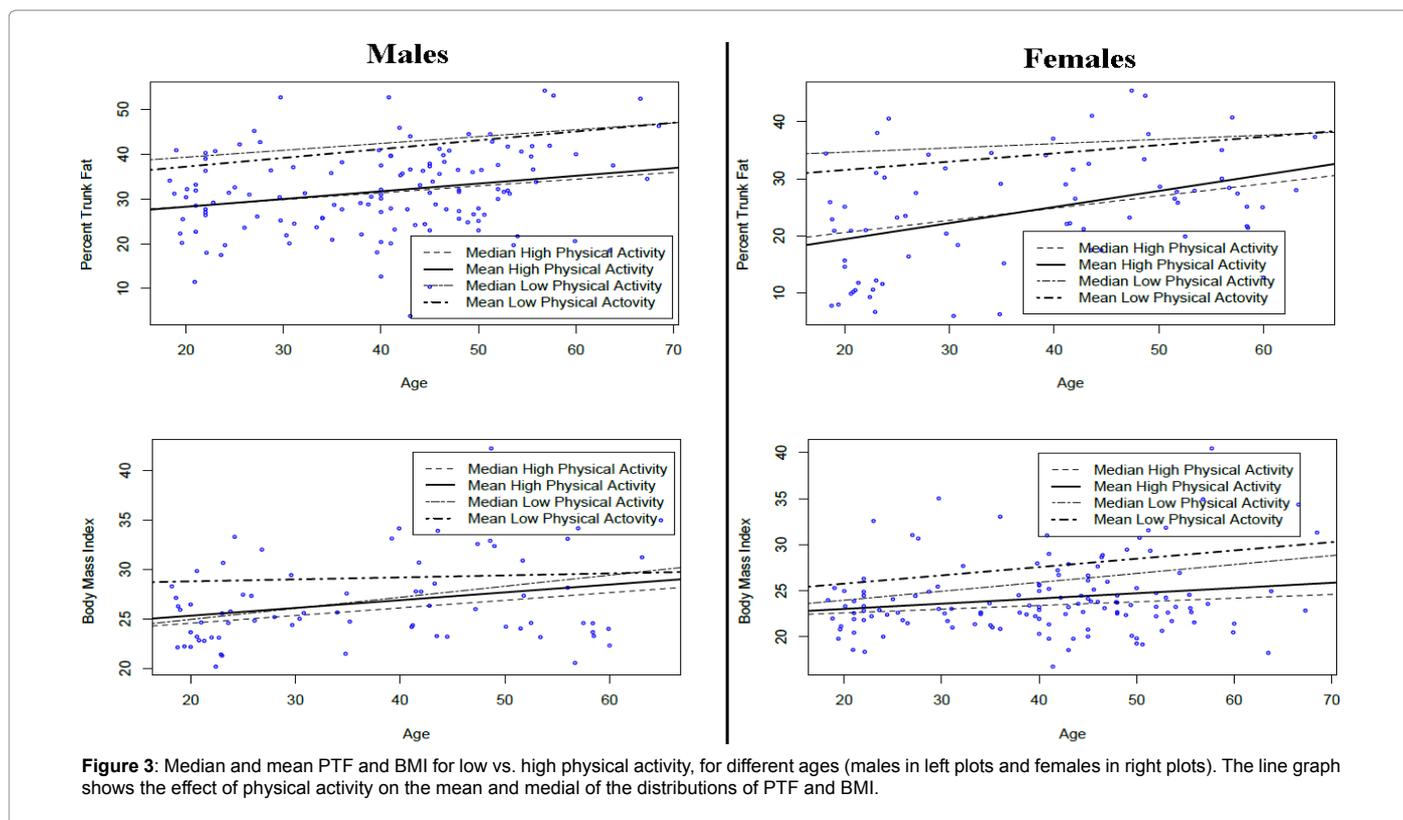
Variable		OLS	50 <sup>th</sup> percentile	80 <sup>th</sup> percentile	90 <sup>th</sup> percentile	95 <sup>th</sup> percentile
Male	Intercept	33.77 <sup>*</sup>	34.39 <sup>*</sup>	40.62 <sup>*</sup>	42.42 <sup>*</sup>	44.25 <sup>*</sup>
	Age	0.24 <sup>*</sup>	0.26 <sup>*</sup>	0.12 <sup>*</sup>	0.08	0.02
	Moderate activity	-4.28 <sup>*</sup>	-5.05 <sup>*</sup>	-3.07 <sup>*</sup>	-2.08 <sup>*</sup>	-2.01 <sup>*</sup>
	High activity	-8.98	-10.42 <sup>*</sup>	-7.67 <sup>*</sup>	-5.46 <sup>*</sup>	-4.38
Female	Intercept	42.70 <sup>*</sup>	42.73 <sup>*</sup>	48.58 <sup>*</sup>	51.07 <sup>*</sup>	53.43 <sup>*</sup>
	Age	0.17 <sup>*</sup>	0.16 <sup>*</sup>	0.11 <sup>*</sup>	0.08 <sup>*</sup>	0.08 <sup>*</sup>
	Low alcohol	-1.23 <sup>*</sup>	-	-	-	-
	Moderate and high alcohol	-1.73 <sup>*</sup>	-	-	-	-
	Moderate activity	-3.34 <sup>*</sup>	-3.70 <sup>*</sup>	-3.11 <sup>*</sup>	-2.40 <sup>*</sup>	-2.86 <sup>*</sup>
	High activity	-9.36 <sup>*</sup>	-11.08 <sup>*</sup>	-9.17 <sup>*</sup>	-8.24 <sup>*</sup>	-7.44 <sup>*</sup>

<sup>\*</sup>Indicates a significant coefficient at significance level 0.05.

**Table 2:** Final model coefficients for PTF.

As a by-product of our study, we created cut-off points for overweight and obesity based on PTF separately for males and females. We determined approximately which quantile of our observed sample

fell on the WHO defined BMI values for overweight and obese. In our sample, for males, the 33<sup>rd</sup> percentile of BMI falls approximately on 25 (cut-off point for normal weight as defined by WHO) and the



74<sup>th</sup> percentile falls on a BMI of 30 (the cut-off point for overweight). The corresponding PTF value for the standardized value of 25 for BMI is 24.2, which is approximately the 30<sup>th</sup> percentile for PTF. The corresponding PTF value for the standardized value of 30 for BMI is 35.2, which is approximately the 65<sup>th</sup> percentile for PTF. This means that if we use 35.2 as a potential obesity cut-off point for the adult male population, in our sample we have approximately 35% obese based on PTF and only 26% obese based on BMI. Similarly, for females, the corresponding WHO BMI percentiles observed were the 49<sup>th</sup> and 82<sup>nd</sup>. The corresponding PTF value for the standardized value of 25 is 24.2, which is approximately the 36<sup>th</sup> percentile for PTF. The corresponding PTF value for the standardized value of 30 is 44.99, which is approximately the 75<sup>th</sup> percentile for PTF. This means that if we use 45 as a potential cut-off point for the female adult population, in our sample we have 25% obese based on PTF and only 18% obese based on BMI. The estimates from this research may, in turn, be used with the obesity cut-off points obtained for PTF to determine whether an individual is at risk for obesity.

Our results show that both moderate and high physical activity levels are significantly associated with a reduction in both BMI and PTF in both males and females. Furthermore, we found that both genders in percentiles closer to the median of PTF have more benefit from increased physical activity compared to higher percentiles. Interestingly, in contrast, both genders in higher percentiles of the BMI distribution seem to benefit more with increased physical activity compared to percentiles closer to the median.

The novel goal of this study was to estimate the effect of certain lifestyle changes on different percentiles of BMI and PTF for the adult population. These estimates can determine whether individuals in different percentiles have different lifestyle contributions. Our results

may be useful for estimating the likelihood of obesity using a trait other than BMI, and also using environmental factors that are potentially accessible to general physicians and with advancing technology, hopefully, to the general public.

#### Acknowledgements

The authors thank Dr Guang Sun for providing medical expertise and Farrell Cahill for his help and expertise regarding the dataset. We thank the Natural Sciences and Engineering Research Council of Canada and Research & Development Corporation Newfoundland and Labrador for their generous support in this project.

#### References

- Chen Z, Wang Z, Lohman T, Heymsfield SB, Outwater E, et al. (2007) Dual-energy X-ray absorptiometry is a valid tool for assessing skeletal muscle mass in older women. *J Nutr* 137: 2775-2780.
- Kennedy AP, Shea JL, Sun G (2009) Comparison of the classification of obesity by BMI vs. dual-energy X-ray absorptiometry in the Newfoundland population. *Obesity (Silver Spring)* 17: 2094-2099.
- Bjorntorp P (1997) Body fat distribution, insulin resistance, and metabolic diseases. *Nutrition* 13: 795-803.
- Larsson B, Svärdsudd K, Welin L, Wilhelmsen L, Björntorp P, et al. (1984) Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: 13 year follow up of participants in the study of men born in 1913. *Br Med J (Clin Res Ed)* 288: 1401-1404.
- Donahue RP, Abbott RD, Bloom E, Reed DM, Yano K (1987) Central obesity and coronary heart disease in men. *Lancet* 1: 821-824.
- Zimmermann MB, Gubeli C, Puntener C, Molinari L (2004) Detection of overweight and obesity in a national sample of 6-12-y-old Swiss children: accuracy and validity of reference values for body mass index from the US Centers for Disease Control and Prevention and the International Obesity Task Force. *The American Journal of Clinical Nutrition* 79: 838-843.
- Wei Y, Pere A, Koenker R, He X (2006) Quantile regression methods for reference growth charts. *Stat Med* 25: 1369-1382.

8. Briollais L, Durrieu G (2014) Application of quantile regression to recent genetic and -omic studies. *Hum Genet* 133: 951-966.
9. Baecke JA, Burema J, Frijters JE (1982) A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *The American Journal of Clinical Nutrition* 36: 936-42.
10. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, et al. (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 122: 51-65.
11. Koenker R (2005) *Quantile Regression*. Cambridge University Press.
12. Kay SJ, Fiatarone Singh MA (2006) The influence of physical activity on abdominal fat: a systematic review of the literature. *Obes Rev* 7: 183-200.
13. Troisi RJ, Heinold JW, Vokonas PS, Weiss ST (1991) Cigarette smoking, dietary intake, and physical activity: effects on body fat distribution--the Normative Aging Study. *Am J Clin Nutr* 53: 1104-1111.
14. Thompson WG, Rostad Holdman N, Janzow DJ, Slezak JM, Morris KL, et al. (2005) Effect of energy-reduced diets high in dairy products and fiber on weight loss in obese adults. *Obes Res* 13: 1344-1353.
15. Chau JY, van der Ploeg HP, Merom D, Chey T, Bauman AE (2012) Cross-sectional associations between occupational and leisure-time sitting, physical activity and obesity in working adults. *Prev Med* 54: 195-200.
16. Hamer M, Brunner EJ, Bell J, Batty GD, Shipley M, et al. (2013) Physical activity patterns over 10 years in relation to body mass index and waist circumference: the Whitehall II cohort study. *Obesity (Silver Spring)* 21: E755-761.
17. Sayon-Orea C, Martinez-Gonzalez MA, Bes-Rastrollo M (2011) Alcohol consumption and body weight: a systematic review. *Nutr Rev* 69: 419-431.