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

Daniel J Dutton, Lindsay McLaren

INTRODUCTION
Research from various disciplines has focused on quantifying the association between ‘determinants of obesity’ and body mass index (BMI). Important determinants include: income, education and occupation; diet (eg, fruit and vegetable consumption); physical activity levels; sex and ethnicity; smoking status; mental health conditions such as depression; age; social support and others. Establishing which individual characteristics are associated with obesity informs public health by identifying subgroups that might be at increased risk of obesity. However, the increase in obesity prevalence over time is most likely driven by changes in population-level determinants of obesity rather than changes in the distribution of individual-level determinants.

Population-level determinants of obesity are factors that influence obesity on a wide scale and may not register as individual-level risks due to their homogeneity. Examples include policies resulting in lower food prices, decreases in work and leisure time physical activity and urban design promoting driving. These act on all individuals in a population and play a role in the increase of obesity prevalence, but traditional data analysis using individuals from the same population may not identify these effects. While there is clearly an interplay between population-level and individual-level determinants of obesity in determining the population level of obesity, there is little evidence on the relative impact of these factors.

Individual-level determinants of BMI are sometimes suggested as potential targets for policy. However, those determinants might have different associations with BMI across different groups due to different population-level determinants. One important example of such groups in Canada is geographic regions: there is an East to West gradient whereby those in the Atlantic provinces (eastern part of Canada) have a higher prevalence of obesity than the rest of Canada, and British Columbia (BC) (westernmost province) has a low prevalence of obesity. If income is lower and obesity prevalence is higher in the Atlantic provinces than elsewhere, it seems reasonable to suggest that policy to raise income in those provinces could decrease obesity prevalence. However, whether this has the desired impact depends on the association between income and obesity prevalence in different regions.

Determining the relative importance of individual-level and population-level determinants of obesity is possible using decomposition techniques that separate the difference between two regions into two shares, one share attributable to the levels of the determinants and another attributable to the differential impact (or ‘associations’) of those determinants on BMI between regions. If the difference in BMI between two regions is mostly explained by levels of the correlates of BMI
included in the model, then it would be reasonable to suggest those correlates as targets for intervention. If the share of the difference attributable to the associations (the ‘unexplained’ share) is large, it would indicate that the determinants included do a poor job of explaining the difference between regions and are not a relevant policy target for reducing between-region differences in BMI.

When studying population-level obesity, there is additional information to be gained beyond mean BMI. Population-level studies of obesity show that there is an increasing right skew to the BMI distribution in Canada, so the distribution of BMI gives researchers a more nuanced picture of the BMI distribution than modelling the average or ranges. Quantile regression allows for the study of the association between individual-level determinants of BMI and the entire BMI distribution, and has been applied in the public health literature on obesity. For example, the association between local fast food prices and BMI at the county level in the USA had the largest magnitude association for the heaviest individuals: a $1 decrease in the price of fast food was associated with a 0.1 kg/m² increase in average BMI for males, but a 1.0 kg/m² increase in BMI for those at the 75th and 90th percentiles of BMI. Other American studies have found associations that change across the BMI distribution for food stamp programme participation among females (an increase of 1.6 kg/m² on average, but an increase of 2.9 kg/m² at the 90th percentile of BMI) and education (one additional year of education was associated with a 0.165 kg/m² decrease in BMI for females at the median BMI level, and a 0.3 kg/m² decrease in BMI for females at the 90th percentile of BMI). Analysis of Canadian BMI data showed a negative association between BMI and education for males and females on average, and that association was stronger for females at the 60th centile and up. Research modelling the distribution is consistent in its message: average estimates sometimes suppress interesting associations that are unique to certain sections of the BMI distribution (eg, the higher percentiles).

In Canada, some research decomposing geographic differences in BMI has been performed. One study found that obesity is concentrated among poorer individuals in the Atlantic provinces; that concentration is not present to the same extent in other regions, and is reversed in some provinces like Alberta. In another paper, these authors show that obesity at the health region level is partially attributable to health-region specific, time invariant factors (‘fixed effects’). Our earlier study found that the importance of individual-level characteristics in determining differences in average BMI between regions in Canada differed for men and women. Specifically, when statistically significant overall differences between regions were observed for females, the difference was unexplained by the individual-level variables. In males, the individual-level variables were able to explain most of the statistically significant differences; however, the overall difference available to explain was very low. This difference in trends for males and females warrants further investigation. Furthermore, the earlier paper’s focus on average BMI prompts expansion to the full BMI distribution.

Our objective was to examine whether equalising the identified individual-level determinants of BMI across geographic regions of Canada could be reasonably expected to reduce differences in BMI distributions between those regions. By ‘equalizing’, we mean using the average level of each covariate in one region and applying it to the other (see equations below). We use quantile regression and decomposition techniques to estimate the difference in BMI distributions between regions that would remain if we were to hypothetically equalise the determinants.

METHODS

Data

We used merged data from three cycles (those available at the time the data contract was initiated, namely 2001, 2003 and 2007) of the Canadian Community Health Survey (CCHS), which is a repeated, nationally representative cross-sectional survey conducted by Statistics Canada. The target population is individuals over age 12 living in Canada’s 10 provinces. Those living in institutions, remote areas, the northern territories, on Crown lands and military bases are excluded. The survey represents 98% of the target Canadian population. The CCHS uses a multistage stratified cluster sampling procedure. Weights are provided to account for this complex sampling technique; we adjusted the weight for each observation by a constant based on the number of cycles included.

The variables used are informed by past literature. Specifically, we used: age; income decile (household income adjusted for household and community size); smoking status (non, occasional, daily); alcohol use status (non, occasional, former, regular); physical activity level (active, moderate or inactive based on an index constructed by Statistics Canada using daily energy expenditure); fruit and vegetable consumption (number per day); whether the respondent had a family doctor (yes/no, included as a proxy for healthcare access / attitudes); marital status (single vs other); urban versus rural residence; education (postsecondary degree or higher vs not); and employment in the past 12 months (was employed vs not). Questions included in Statistics Canada surveys are extensively validated. The fruit and vegetable question is taken from the Behavioral Risk Factor Surveillance System (BRFSS) survey and has been validated. Pregnant women were excluded from the samples (as in other studies using survey data, eg, ref 31).

Separate models were run for males and females in each region (Atlantic provinces, Quebec, Ontario, Prairie provinces and BC). Our sample of interest was working age individuals (age 18–65). BMI was constructed from self-reported height and weight. We reran all analyses replacing self-reported BMI with two different corrected measures of BMI. Results were substantively identical in all cases, and those from self-reported data are presented.

Statistical analysis

We used quantile regression and a decomposition technique. First, we ran unconditional (or marginal) quantile regressions; we estimated the association of the variables with every second centile of BMI from the 2nd to the 98th. A simplified version of our regression model for each region was:

$$\text{BMI}_{q_t} = \left( \sum_{k=1}^{k-1} \beta_{q_t} X_k \right) + \varepsilon$$

In words, the BMI at each estimated quantile (Q, in our case every second centile) was estimated to be a function of all k-1 variables.

The 2005 cycle had only partial coverage for key variables in our analysis (eg., fruit and vegetable consumption) so was excluded. ‘Unconditional’ does not mean the model contained no regressors. It is a convention used in the quantile regression literature to describe the method used to estimate the model coefficients.
regressor variables identified above plus a constant (Xk) and an error term (ε). Each estimated quantile of BMI was related to the regressors by quantile-specific associations (βQ). These models are stratified by region.

Then we ran Blinder-Oaxaca-type decompositions across estimated quantiles. We use the Atlantic provinces as the comparison region because they have the most skewed BMI distribution and are thus the most informative counterfactual. For each comparison, at all estimated quantiles of BMI, we have variable values and quantile-specific associations for both regions, which we use to obtain an overall difference in BMI and explained and unexplained shares of the difference. We do this across all estimated quantiles to model the entire BMI distribution difference. So for any Qt we have the following expression:

\[ \Delta_{A,C}\text{BMI}_t = \beta_{A,Q_t} (X_C - X_A) + X_C (\beta_{C,Q_t} - \beta_{A,Q_t}) \]

The difference in BMI between the Atlantic provinces (A) and a comparator region (C) at Qt (\(\Delta_{A,C}\text{BMI}_t\)) is composed of a difference attributable to different levels of the individual-level determinants (X) evaluated at the Atlantic provinces’ coefficient estimates (\(\beta_{A,Q_t}(X_C - X_A)\)) and a remaining share unexplained by variables included in the model.14–16

Ethics approval for this project was obtained from the University of Calgary’s Conjoint Health Ethics Research Board (Ethics ID: E-23704). All analyses were conducted using Stata V12. The commands we used are adopted from the ‘rifreg’ command package, available from one of the authors of the technique’s website here: http://faculty.arts.ubc.ca/infortin/index.html.

### RESULTS

Table 1 shows summary statistics for males and females by region.

BMI is highest in the Atlantic provinces, and lowest in Quebec and BC. Education and income levels are similar across regions (with some sex differences). The Atlantic provinces have lower physical activity and fruit and vegetable consumption than other regions. Table 2 shows BMI deciles by sex and region.

At least 20% of the sample from the Atlantic provinces were obese. Males in the Prairie provinces were the only other group with a similar prevalence. All regions had at least 10% obesity prevalence. Figure 1 shows the distribution of BMI by region and sex.

Quantile regression decomposition results are presented graphically due to the very large volume of results (each point on the graph is a regression coefficient). The figures depict each region decomposed against the Atlantic provinces for males and females separately. The y-axis is BMI, so the ‘overall’ line shows the size of the overall BMI difference between the two regions at each BMI percentile (quantile) in kg/m². The ‘explained’ line shows the share of the overall difference at each percentile that is explained by the different levels of the individual-level variables between the two regions. The ‘unexplained’ line shows the remaining share.

The overall difference between any region and the Atlantic provinces was larger for females than for males at all quantiles. That means that the difference between the heaviest individuals in the Atlantic provinces and those in other regions was greater for females than for males. The difference between the regions increased more for females than for males as the BMI quantile increased.

For both males and females, the overall difference between regions (solid line) increased more than the explained difference.

### Table 1  Summary statistics for all study variables

<table>
<thead>
<tr>
<th></th>
<th>Atlantic Provinces</th>
<th>Quebec</th>
<th>Ontario</th>
<th>Prairies</th>
<th>British Columbia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>Self-reported BMI (mean, kg/m²)</td>
<td>27.32</td>
<td>26.56</td>
<td>26.12</td>
<td>26.76</td>
<td>25.49</td>
</tr>
<tr>
<td>Age (mean, years)</td>
<td>43.46</td>
<td>42.84</td>
<td>42.52</td>
<td>42.41</td>
<td>42.23</td>
</tr>
<tr>
<td>5th income decile or lower (%)</td>
<td>41</td>
<td>50</td>
<td>39</td>
<td>50</td>
<td>37</td>
</tr>
<tr>
<td>Education beyond high school (%)</td>
<td>59</td>
<td>61</td>
<td>63</td>
<td>62</td>
<td>64</td>
</tr>
<tr>
<td>Daily smoker (%)</td>
<td>29</td>
<td>26</td>
<td>30</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>Occasional smoker (%)</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Never smoker (%)</td>
<td>67</td>
<td>70</td>
<td>66</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>Occasional drinker (%)</td>
<td>15</td>
<td>13</td>
<td>10</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>Former drinker (%)</td>
<td>13</td>
<td>16</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Never drinker (%)</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Regular drinker (%)</td>
<td>69</td>
<td>48</td>
<td>79</td>
<td>65</td>
<td>76</td>
</tr>
<tr>
<td>Physically active (%)</td>
<td>22</td>
<td>18</td>
<td>23</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>Moderately physically active (%)</td>
<td>24</td>
<td>25</td>
<td>25</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>Physically inactive (%)</td>
<td>54</td>
<td>57</td>
<td>52</td>
<td>58</td>
<td>48</td>
</tr>
<tr>
<td>Fruits and vegetables per day (mean)</td>
<td>3.89</td>
<td>4.51</td>
<td>4.37</td>
<td>5.30</td>
<td>4.28</td>
</tr>
<tr>
<td>No family doctor (%)</td>
<td>15</td>
<td>8</td>
<td>36</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>Single (%)</td>
<td>32</td>
<td>36</td>
<td>41</td>
<td>43</td>
<td>36</td>
</tr>
<tr>
<td>Rural (%)</td>
<td>46</td>
<td>43</td>
<td>27</td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>Worked in the past 12 months (%)</td>
<td>84</td>
<td>73</td>
<td>85</td>
<td>73</td>
<td>89</td>
</tr>
</tbody>
</table>

Income deciles are computed by Statistics Canada in the CCHS at the provincial level from all respondents who provided a valid income value. The decile is based on the ratio of household income of the respondent to the low income cut-off appropriate for that respondent’s household and community size. The deciles are dummy variables, so the 10th income decile indicates whether an individual falls into the lowest 10% of such ratios in their province. The percentiles are computed for males and females separately. Our sample is not evenly split by decile since not all individuals with valid income measures are included in our study (ie, they did not all provide valid responses for all other variables used).

CCHS, Canadian Community Health Survey.
(dotted line) across the BMI distribution for most regions, so the ability of the explanatory variables to account for the difference between the regions became weaker for heavier individuals. That weakening of explanatory power was more pronounced for females.

Overall, variables measured at the individual level were not effective at explaining the between-region differences in BMI for females. For instance, when comparing the Atlantic provinces with Quebec (figure 2A) or BC (figure 2D), the unexplained share accounted for almost the entire difference between the regions across the whole BMI distribution. For example, the 80th BMI centile in the Atlantic provinces is approximately 3 BMI units higher than the same BMI percentile in Quebec (figure 2A), and approximately 2/3 of that difference is unexplained by observable differences in levels of variables between regions. The Atlantic provinces versus Ontario comparison (figure 2B) showed an increase in explanatory power of the variables as the BMI quantile increased. Still, this increase in explained difference made up less than half of the overall difference at any quantile. The Atlantic provinces versus Prairies comparison (figure 2C) showed a fairly even share of the overall difference split between the explained and unexplained shares.

Although males exhibited the same qualitative patterns as females, the magnitude is just much smaller (figure 2A–D). The Atlantic provinces versus Quebec comparison (figure 2A) showed that the overall difference between regions was almost entirely unexplained. The Atlantic provinces versus Ontario comparison (figure 2B) was similar in that the variables explained a small magnitude of the overall difference, but the overall difference was smaller, so the share explained is larger as a proportion of the total difference. The Atlantic provinces versus Prairies overall difference (figure 2C) was quite small, so the explained share was large compared to the overall difference despite being of tiny magnitude. The Atlantic provinces versus BC comparison (figure 2D) was fully unexplained until approximately the 60th centile, after which there was a slight increase in the ability of the variables to explain the overall difference between the two regions.

**DISCUSSION**

Our results show that, overall, the differences in BMI distributions between regions are not attributable to different levels of the variables included in the model. In other words, regional BMI distribution differences reflect factors beyond those captured by these individual-level variables. Therefore, equalisation of identified determinants of BMI cannot be reasonably expected to reduce differences in BMI distributions between regions.

**Table 2** Percentiles of BMI by sex and region

<table>
<thead>
<tr>
<th>Region</th>
<th>Atlantic Provinces</th>
<th>Quebec</th>
<th>Ontario</th>
<th>Prairies</th>
<th>British Columbia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>10th centile</td>
<td>22.2</td>
<td>20.5</td>
<td>21.4</td>
<td>19.4</td>
<td>22.0</td>
</tr>
<tr>
<td>20th centile</td>
<td>23.6</td>
<td>21.9</td>
<td>22.8</td>
<td>20.6</td>
<td>23.4</td>
</tr>
<tr>
<td>30th centile</td>
<td>24.9</td>
<td>23.0</td>
<td>23.8</td>
<td>21.6</td>
<td>24.2</td>
</tr>
<tr>
<td>40th centile</td>
<td>25.6</td>
<td>24.2</td>
<td>24.9</td>
<td>22.5</td>
<td>25.3</td>
</tr>
<tr>
<td>50th centile</td>
<td>26.7</td>
<td>25.5</td>
<td>25.6</td>
<td>23.6</td>
<td>26.3</td>
</tr>
<tr>
<td>60th centile</td>
<td>27.8</td>
<td>26.9</td>
<td>26.4</td>
<td>24.8</td>
<td>27.2</td>
</tr>
<tr>
<td>70th centile</td>
<td>29.0</td>
<td>28.4</td>
<td>27.7</td>
<td>26.4</td>
<td>28.4</td>
</tr>
<tr>
<td>80th centile</td>
<td>30.5</td>
<td>30.7</td>
<td>29.1</td>
<td>28.1</td>
<td>29.8</td>
</tr>
<tr>
<td>90th centile</td>
<td>32.9</td>
<td>33.9</td>
<td>31.3</td>
<td>31.4</td>
<td>32.0</td>
</tr>
</tbody>
</table>

BMI, body mass index.

**Figure 1** Kernel densities of BMI across all regions for males and females. The distributions for females are more skewed than for males across Canada. BMI, body mass index.
It has been proposed that the relative importance of determinants measured at the individual level in Canada has weakened over time as obesity prevalence has increased.\textsuperscript{20} Our findings further support theoretical models that claim that the population-level determinants of obesity are its main drivers. Previous work using this type of decomposition suggested that population-level determinants like low-cost, calorie-rich foods could overwhelm individual-level determinants.\textsuperscript{23} Specific region-level influences that are malleable by government action (eg, the level and distribution of healthcare resources) or not (eg, climate) are most likely working in tandem to create these region-level differences.

Outside of Canada, the decomposition of obesity literature has focused on other groupings such as ethnicity. US researchers have decomposed BMI across African–American and Caucasian males and females,\textsuperscript{37 38} concluding that unexplained differences between African–American and Caucasian were associated with one or a combination of potential factors including culture, genetics or unobserved (unmeasured) confounders. Other examples internationally have decomposed obesity between ethnic groups (eg, Chinese and Malays),\textsuperscript{39} between countries (eg, Italy and Spain),\textsuperscript{40} across sexes\textsuperscript{41} or over time\textsuperscript{42} and indicate that there can be large unexplained differences over the groups under consideration. Hypothesised unexplained factors include cultural differences;\textsuperscript{39 40} increased availability of non-traditional food which is differentially taken up by groups;\textsuperscript{41} and unmeasured behavioural differences.

Collectively, the literature suggests that interventions targeting individual-level behaviours in the face of powerful population-level determinants would be expected to have a minimal impact on population BMI prevalence and distributions. Other plausible population-level factors include the transportation and food marketing environments, and the policy environment they exist within.\textsuperscript{11} Modelling these population-level factors is an important avenue for future research.

One might argue that the between-region differences in BMI in Canada are small relative to those in other countries (eg, the USA\textsuperscript{35}). However, regional differences in obesity prevalence can reach a non-negligible 10%, and the point remains that the distribution of BMI is right-skewed, making the average a poor summary of the distribution. The public health benefits of addressing obesity are not realised by targeting differences in average BMI.

Our study has some limitations. First, the cross-sectional nature of our data means we cannot attribute changes in BMI outcomes to direct causal changes in the regressor variables. Second, we use behavioural variables which may be a source of endogeneity in our model.\textsuperscript{43} However, these variables have been cited as important determinants of BMI in the past and excluding them would lead to unmeasured confounding.\textsuperscript{44} Third, we used self-reported BMI as our outcome measure, which is considered biased compared to measured BMI when measuring obesity prevalence;\textsuperscript{45} however, we ascertained that results change very little on correction.\textsuperscript{32 33} A fourth limitation is that anything unmeasured in our decomposition analysis would appear to be part of the unexplained share of the difference. Better measures of diet and physical activity would give us more confidence that we captured unbiased variation in those individual-level variables. An implication is that the perceived effect of population-level determinants is attributed to omitted individual-level determinants. However, we did ensure inclusion of a broad cross-section of ultimately statistically significant variables (in OLS, not shown) in our within-region models in an effort to reduce this unmeasured variable problem. Finally, by combining multiple cycles of data, we introduced a time dimension. However, including dummy variables for the cycles included did not change our results.

When considered as a whole, these results provide quantitative support for the view that efforts aimed at encouraging individual-level behaviour change are insufficient to address

---

\textsuperscript{iv}Maps are available online at http://www.cdc.gov/obesity/data/prevalence-maps.html
obesity on a population scale, and indicate that equalising the modelled variables between regions would not be expected to reduce differences in BMI distributions between regions.

What is already known on this subject?

Obesity prevalence varies across geographic regions in Canada, as do individual-level correlates of obesity like income, education, and behavioural variables. Population-level determinants of obesity influence the impact these individual-level variables have on regional obesity prevalence. Public policy often targets individual-level variables; however, the relative importance of population-level and individual-level variables has not been well quantified.

What this study adds?

We quantified the relative importance of individual-level and population-level variables, using quantile regression and Blinder Oaxaca decomposition, and conclude that population-level variables are the main influences of cross-regional variation in obesity prevalence. The implication for public policy is that aiming to equalize individual-level variables across regions will not have as large an impact as targeting population-level variables.

Acknowledgements The authors acknowledge J C Herbert Emery for helpful comments on earlier versions of this manuscript. The analysis presented in this paper was conducted at the Prairie Regional Research Data Centre (PRRDC), which is part of the Canadian Research Data Centre Network (CRDCN). The services and activities provided by the PRRDC are made possible by the financial or in-kind support of the SSHRC, the CIHR, the CFI, Statistics Canada and the University of Calgary.

Contributors DJD and LM shared the responsibility of drafting the manuscript and interpreting the results, while DJD was responsible for the statistical analysis. Both DJD and LM approved this version of the manuscript. DJD is the guarantor.

Funding DJD was funded by a doctoral student award from the Population Health Intervention Research Network while conducting this research. LM acknowledges a Population Health Investigator award from Alberta Innovates—Health Solutions, and an Applied Public Health Chair award funded by the Canadian Institutes of Health Research, the Public Health Agency of Canada and Alberta Innovates—Health Solutions.

Competing interests None declared.

Ethics approval University of Calgary Conjoint Health Ethics Research Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Statistics Canada controls access to the master files of their population-level surveys through the Research Data Centres. These confidential computer labs are the only access point for the data used in this study. Thus, the raw data are not available. However, OLS and quantile regression results underlying the decomposition results presented here are available on request from DJD. The high volume of these results precludes them from being included in the published work.

REFERENCES

19. Williams PT. Evidence that obesity risk factor potencies are weight dependent, a phenomenon that may explain accelerated weight gain in western societies. PLoS ONE 2011;6:e27657.
33. Dutton DJ, McLaren L. The usefulness of “corrected” body mass index vs. self-reported body mass index: comparing the population distributions, sensitivity, specificity, and predictive utility of three correction equations using Canadian population-based data. BMC Public Health 2014;14:430.


