CMAJOpen-2018-0174	
Temporal trends of severe obesity prevalence in children and youth from primary care electronic medical records in Ontario, Canada: an observational study	
Authors: Sarah Carsley PhD, Eliza Pope BSc, Laura N. Anderson PhD, Mark S. Tremblay PhD, Karen Tu MD MSc, Catherine S. Birken MD MSc; on behalf of Team ABC	
Reviewer comments	**Author's response**
Reviewer 1 **Sina Gallo -** McGill University, School of Dietetics and Human Nutrition, Montréal, Que.	
1) The authors have categorized ethnicity as general population, Chinese or South Asian based on a validated surname method. In the US, ethnic minority children (i.e. American Indian, Hispanic and Black) have higher obesity rates compared to white counterparts. In addition, Chinese and South Asians accounted for <5% of the sample in this study. Please justify use of these categories and why racial and cultural background as per CHMS (i.e. includes Black, Filipino, Latin American, Southeast Asian, Arab, West Asian, Japanese, Korean, Aboriginal, and Other) was not used instead.	Other than national surveys such as the CHMS, there is unfortunately no data on reported ethnicity at a population level. The use of these categories is a reflection of the validity of the surname algorithm. In Canada, the recent census data showed the main non-European ethnicities reported were Chinese, East Indian and Filipino ancestries, 70% which were foreign born (Statistics Canada, 2016). Therefore, the use of this ethnicity variable, although imperfect, was a scientific decision to try to capture this diversity, understanding the limitations. In order to help justify the importance of accounting for ethnicity in this analysis we have included the Stats Can reference cited above. The text now states: "Recent census data in Canada showed the main non-European ethnicities reported were Chinese, East Indian and Filipino ancestries (Statistics Canada 2016)."
2) Table 1 compares the sample vs. Ontario population and the authors have commented on the differences in results yet, not tested statistically. Can you add p-values to the Table to compare characteristics?	Yes, we have re-run the bivariate comparison to provide both standardized differences and p-values. Due to the large sample size of this study, all p-values are significant. However, the standardized difference shows the largest difference between population proportions is 0.6 in the youngest age group. As previously described, this is attributed to the primary care sample (young children see physicians for immunizations) vs. the whole Ontario pediatric population.
3) In the discussion (page 10 lines 10-14), please specify BMI-for-age for US children 2-20 years is compared to CDC 2000 growth charts.	We have specified the type of growth chart comparator that is used in the US. The text now states "In the United States, BMI-for-age for children 2-20 years is compared to CDC 2000 growth charts. Class II obesity is defined as a BMI >120% of the 95th percentile or a BMI ≥ 35 kg/m^2, whichever is lower, and class III is defined as a BMI >140% of the 95th percentile or a BMI ≥ 40 kg/m^2, whichever is lower."
4) Can the authors please comment on differences in characteristics between children who access primary care vs. those who do not in Ontario or Canada (see page 11 lines 24-26).	We have described differences between children that attend, and do not attend primary care across Ontario. The text now states: "In Ontario, 12% of children had no primary care physician visit billings over a two-year period. Children living in lower income neighbourhoods are more likely to have no primary care billings compared to children in higher income neighbourhoods.
Reviewer 2 **Sharma Atul -** University of Manitoba, Pediatrics and Child Health, Winnipeg, Man.	

Some of these differences are sizable, statistically significant, and known to be associated with differences in obesity rates (e.g. age, immigration status, income). For example, the proportion aged 0-4y is 48% in the sample vs 13% in the population, which is important, since obesity rates are substantially lower in this age group (c.f. reference 4). Immigrant numbers among children in their cohort were 1/4 of those in the general population (6.2 v 1.7%), which may be important given evolving population demographics and the impact of both immigration status and years since immigration on zBMI (Wahi et al, Can J Public Health, 2014). There was also a significant under-representation of lower socioeconomic status in the sample (for income quintile 1, proportions were 13 vs 19%, p< 10-15). In Canadian children, the association between lower SES and obesity has been well- described (Shields M, Findings from the Canadian Community Health Survey, 2005; Sharma and Rodd, Paediatrics & Child Health, 2017). Given an extensive literature on the effects of these factors on obesity rates, some discussion of the likely impact of selection biases seems warranted.	We agree with the comments about generalizability from Reviewer 2 and understand these factors affect zBMI. We have improved our discussion of the limited generalizability of this data in our limitations section, and how this selection bias may be affecting out prevalence estimates. The text now states: "The EMRALD study population has significantly more children <5 years, fewer immigrants, and an underrepresentation of children in the lower neighbourhood income quintiles. In Canada, these factors have been associated with obesity therefore our findings may be underestimates of the true prevalence."
A related question arises with regard to missing data: Twenty-four percent of patients were ineligible due to missing zBMI and 0.8% due to implausible zBMI This is a high missingness fraction. Did the missing group differ systematically?	We thank the reviewer for bringing up this point and have addressed in our previous comments to the editor. In addition to the editors' questions about reasons for missing data, we were able to assess how those with missing data were different on the demographic variables that do not change with time. There were no significant differences in sex. There were differences in ethnicity; a higher proportion of children with Chinese ethnicity (2.3% vs. 1.8%) and South Asian ethnicity (1.2% vs 0.9%) in the non-missing group compared to the missing group.
There are problems with this introductory paragraph. Firstly, reference (3) cites an MMWR Report based on WIC data, which discusses overall obesity rates in toddlers with no mention of severe obesity. And while reference (4) includes NHANES data from children aged 2-5y, it relies on CDC growth charts to characterize 3 overlapping classes: I: BMI > 95th percentile (upper limit normal ULN) II: BMI > lower of BMI=35 or 120% of ULN III: BMI > lower of BMI=40 or 140% of ULN Although class I obesity appears to increase in children aged 2-5y from 10.7% in 1999-2000 to 13.7% in 2015-16, both class II and III obesity were stable over 17 years at 1.8% and 0.2%, respectively. I would therefore suggest the authors review these references and perhaps re-write the introduction to better reflect their contents. When citing studies	We thank the reviewer for this comment, and have revised our introduction as suggested. The text now states: "However, national studies in the United States have shown conflicting trends depending on population studied. One study of low-income children 2 to 4 years enrolled in the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) reported a decrease in severe obesity from 2004 to 2014.(3) Another study of children 2 to 5 years using data from the National Health and Nutrition Examination Survey (NHANES), a nationally representative sample, showed an increase in the prevalence of class I obesity (>95^{th} percentile of BMI-for-age compared to the CDC 2000 growth charts) from 1999-2016, while both class II (>120% of the 95^{th} percentile) and class III (>140% of the 95^{th} percentile) remained stable.(4)"

based on CDC norms, care should also be taken to identify the standards being used. Using NHANES data, we have recently reported an increase in the population prevalence of obesity from 10.3% to 19.5% when CDC charts and definitions were replaced by their WHO analogues (Sharma and Rodd, JAMA Pediatrics, 2018), so definitions can seriously impact results.	
Nevertheless, inconsistent applications of WHO charts and definitions occur in several places. With regard to calculating WHO Z-scores, the authors cite reference (14): WHO Child Growth Standards based on length/height, weight and age. Acta Paediatrica (Oslo, Norway: 1992) Supplement. 2006; 450:76-85 However, this reference only describes the WHO MGRS (Multicenter Growth Reference Study) i.e. WHO growth standard for ages 0-5y. The WHO growth reference charts for ages 5-19y are based on a different North American (NCHS) dataset and were published separately (de Onis *et al*, WHO Bulletin, 2007).	We thank the reviewer for picking this up, and have added the appropriate reference to reflect the WHO methods used for children 5 to 19 years.
More importantly, neither reference addresses how to calculate "extreme" z-scores (outside of the range of -3 to +3), a technical point discussed at length in the WHO Methods and Development Manual (pp 324-6), which emphasizes the failure of the usual LMS formulae outside of the range -3 to 3. As Z-scores become more extreme, the usual formulae plateau asymptotically, and the WHO recommends the "SD23 adjustment method". This method uses the distance between Z = 2 and Z = 3 (or similarly -2 and -3) to scale more extreme scores. Given the focus on zBMI > 3, the specific method should be reported.	We have looked at the WHO Methods and Development Manual (https://www.who.int/childgrowth/standards/Technical_report.pdf?ua=1) to more adequately report on the specific method for accounting for the failure of the LMS method at the extremes. We have referenced the document and reported the method in the methods section. The text now states: "The WHO recommends an adjustment to the LMS method when examining data at the upper and lower extremes (<-3 and >+3) because the tails of the Box-Cox normal distribution are affected by extreme data points. This restricted LMS method, fixes the standard deviation between +2 SD and +3 SD to avoid making assumptions about the distribution of these extreme z-scores."
Given how error-prone EMR data are known to be, I appreciate their application of routine "data cleaning" algorithms prior to analysis (reference 15). I also note that they excluded "biologically implausible values" (BIV) for zBMI: values outside -5 and +6 were excluded However, this is *not* the WHO definition of BIV for zBMI, which excludes zBMI < -5 and zBMI > +5 (zBMI > 5 is actually quite rare in high quality research data). While I understand the desire to increase the number of subjects diagnosed with severe obesity, I am not sure that an arbitrary change in definition is warranted. If their results changed substantially with this alternate definition, I would think that both should be reported. In addition,	We thank the reviewer for the comments about the change of extreme BIV to +6. As part of Dr. Carsley's dissertation work the quality of EMR data was assessed. In this work, there was concern that the upper limit of +5 was artificially excluding children with true zBMI values. Further, we had heard from colleagues working in weight management for children that they were consistently seeing patients in this upper extreme (>5). Therefore a chart review of 225 patients with a zbmi between 5 and 10 was conducted to assess the number of true values, and true errors. Results are presented in the table below. The positive predictive value was just over 50%, meaning 1 out of 2 values were actually correct. In light of this, and because the focus of our study was the severe obesity group, we thought this was adequate evidence to move the cut-off to +6. However, because the data shown below has not been published yet, we accept this would appear as an arbitrary change. In order to keep our estimates conservative rather than overestimated, we have followed the reviewers advice and excluded those individuals between +5 and +6 (N=59), and re-run our analysis.

the WHO also provides BIV for height and weight Z-scores, which should also be checked, since discordant errors in numerator and denominator may yield plausible BMI Z-scores.

Table 2: Sensitivity, specificity, positive predictive value, and negative predictive value of each BIV cut-off >+5 and >+8

zBMI cut-off	Correct values	Incorrect values	Total	Sensitivity	Specificity	PPV	NPV
+5 to +6	38	37	75	63.3%	77.6%	50.7%	85.3%
>+6	22	128	150				
+5 to +7	56	59	115	93.3%	64.2%	48.7%	96.4%
>+7	<6	106	110				
+5 to +8	58	87	145	96.7%	47.3%	40.0%	97.5%
>+8	<6	78	80				

References:

Carsley S., Birken, CS., Parkin, PC., Pullenayegum E., Tu, K. Completeness and accuracy of anthropometric measurements in electronic medical records for children attending primary care. J Innov Health Inform. 2018 Mar 9;25(1):963.
Carsley S., Birken CS., Parkin PC., Pullenayegum E., Tu K. Accuracy of extreme BMI z-scores in children and adolescents. 2018 (Manuscript in preparation)
Dissertation: Using Electronic Medical Records to Examine Childhood Obesity Outcomes in Community-Based Primary Care
https://tspace.library.utoronto.ca/bitstream/1807/89753/1/Carsley_Sarah_E_201806_PhD_thesis.pdf

The text now states: “Height and weight data were cleaned using a standard set of data cleaning rules (15) and zBMI values outside -5 and +5 were excluded.”

Even more surprising is the decision to classify weight status in children < 5y by applying WHO definition for older children 5-19y. The table below summarizes the authors description for the standard WHO definitions of weight class by age (see also Table I in Rolland-Cachera *et al*, Int J of Ped Obesity 2011). Applying definitions intended for older children to infants and toddlers will exaggerate the severity of pathology by “shifting them to the right” and arbitrarily creates a new “severe obesity” category for young children who would otherwise be classified as obese (zBMI > 3).

We appreciate the concern about the label differences and have included supplemental files of all figures stratified by age group with the WHO labels. When using age groups that cross these age specific labels, it is difficult to report two different sets of labels especially when the numeric categories have not changed. Therefor for ease of communication when describing the study sample as a whole we use the older children definition which was mentioned in the methods (Page 6, line 4-5). There is also concern that the lack of label to adequately identify this group of young children may be reducing the opportunity for intervention.
We believe further work needs to be done on this issues and are open to further discussion to further characterise this issue.

The text now states: “There are several different classifications for weight status, which vary by age, country, and growth reference. The WHO definition of severe obesity begins at 5 years of age there is no category of severe obesity for children less than 5 years. In the United States, BMI-for-age for children 2-20 years is compared to CDC 2000 growth charts. Class II obesity is defined as a BMI >120% of the 95^{th} percentile or a BMI ≥ 35 kg/m^2, whichever is lower, and class III is defined as a BMI >140% of the 95^{th} percentile or a BMI ≥ 40 kg/m^2, whichever is lower.”

On page 6, line 5, I think they mean “multivariable” when they write “multivariate”. See Hidalgo and Goodman, *Multivariate or Multivariable*

We have corrected this to “multivariable”

Regression? American Journal of Public Health, Vol 103, 2013.	
In describing covariates for the multivariable regression model: Similarly, the Ontario Marginalization Index is a census-based, geographically-based index using postal code as a proxy for individual-level sociodemographics.(23) But it's not clear which of the ONMarg indices is being referred to i.e. material deprivation, residential instability, ethnic concentration, or dependency.	In this analysis we used the ONMarg Summary Score, described in this reference (reference: Matheson FI, Dunn JR, Smith KL, Moineddin R, Glazier RH. Development of the Canadian Marginalization Index: a new tool for the study of inequality. Canadian Journal of Public Health = Revue canadienne de sante publique. 2012;103(8 Suppl 2):S12-6.). We have clarified this in the text. The text now states "Similarly, the Ontario Marginalization Index is a census-based, geographically-based index using postal code as a proxy for individual-level socio-demographics; the summary score is a composite of four dimensions of material deprivation, ethnic concentration, dependency, and residential instability.(23)"
The results of the GEE linear regression for zBMI are clearly summarized in Table 3. While the inclusion of sex as a main effect speaks to the impact of sex on the intercept (mean zBMI), they make no mention of whether the slope of the zBMI vs time relationship also varied with sex, as previously described in analysis of pediatric data from the Canadian Health Measures Survey (Sharma and Rodd, CMAJ 2016), which included an interaction term for sex:time.	We thank the reviewer for identifying the potential for a sex:time interaction. We had the opportunity to rerun our analysis with this interaction term and it was not a significant term in our model. Therefore we did not include it in this updated analysis.
While GEE is a perfectly respectable method for dealing with repeated measures, a mixed model with a random subject effect would also provide an intra-class correlation to measure within-subject tracking over time. That said, I have no objection to this model. However, it would also be of interest to know whether the observed decline in zBMI over time (Figure 2 and Supplemental Figure 1) produced a decline in prevalence after adjustment for the other covariates, which could be tested using either GEE or a mixed effects logistic regression. It seems to me that a persistent decline in prevalence after adjustment for potential sociodemographic confounders (e.g. age distribution, immigration status, income quintile) would strengthen their conclusions regarding the nature and clinical importance of the observed trends.	We thank the reviewer for this suggestion. A mixed random effects model was discussed for this analysis, however since our main purpose was to report prevalence estimates and determine the effects of calendar time we elected to use GEE. Particularly because our sample size in the earlier years of this analysis is smaller and may result in an unstable model. We were able to perform this analysis using a multinomial GEE model to determine the association between calendar year and the odds of being in the overweight, obese, or severely obese categories compared to the normal weight group with adjustment for potential confounders (similar to the linear regression model). We have added the results of this analysis to our paper, and included figures demonstrating a similar decline in the probability of overweight, obese, and SO in 2015 compared to 2014. However, because of the loss of information in the categorization of the data some of the ORs are not significant, particular in the obese and SO categories where the sample sizes are smaller. Nonetheless, we feel this analysis supports our conclusion and strengths the overall study. In the Methods: The text now states "A second GEE model specified for a multinomial regression was performed using weight status as the dependent variable and using 2004 as the referent year." In the Results: The text now states "Figure 3 shows the odds of being overweight, obese, or severely obese decreased in almost all years after 2006. There was a significant decrease in the odds of being overweight in 2011-2015, lower odds of having obesity in 2013-2014, and lower odds of severe obesity in 2012 and 2014, compared to 2004. A stratified multinomial regression by age group was conducted and results are presented in Supplemental Figures 3-4. In 5 to 18 year olds, despite a similar decrease in odds of overweight, obesity, or severely obesity, the result was not significant."