Research article

Completeness and accuracy of anthropometric measurements in electronic medical records for children attending primary care

Author address for correspondence:
Sarah Carsley
Child Health Evaluative Sciences
Hospital for Sick Children
Peter Gilgan Centre for Research and Learning
Toronto, ON M5G 0A4, Canada
Email: sarah.carsley@sickkids.ca

Sarah Carsley
Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada and Child Health Evaluative Sciences, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, Toronto, ON, Canada

Catherine S. Birken
Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada, Child Health Evaluative Sciences, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, Toronto, ON, Canada and Department of Paediatrics Medicine, University of Toronto, Toronto, ON, Canada

Patricia C. Parkin
Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada, Child Health Evaluative Sciences, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, Toronto, ON, Canada and Department of Paediatrics Medicine, University of Toronto, Toronto, ON, Canada

Eleanor Pullenayegum
Child Health Evaluative Sciences, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, Toronto, ON, Canada

Karen Tu
Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada, Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada, and Toronto Western Hospital Family Health Team, University Health Network, Toronto, ON, Canada

ABSTRACT

Background Electronic medical records (EMRs) from primary care may be a feasible source of height and weight data. However, the use of EMRs in research has been impeded by lack of standardisation of EMRs systems, data access and concerns about the quality of the data.

Objectives The study objectives were to determine the data completeness and accuracy of child heights and weights collected in primary care EMRs, and to identify factors associated with these data quality attributes.

Methods A cross-sectional study examining height and weight data for children <19 years from EMRs through the Electronic Medical Record Administrative data Linked Database (EMRALD), a network of family practices across the province of Ontario. Body mass index z-scores were calculated using the World Health Organization Growth Standards and Reference.

Results A total of 54,964 children were identified from EMRALD. Overall, 93% had at least one complete set of growth measurements to calculate a body mass index (BMI) z-score. 66.2% of all primary care visits had complete BMI z-score data. After stratifying by visit type 89.9% of well-child visits and 33.9% of sick visits had complete BMI z-score data; incomplete BMI z-score was mainly due to missing height measurements. Only 2.7% of BMI z-score data were excluded due to implausible values.

Conclusions   Data completeness at well-child visits and overall data accuracy were greater than 90%. EMRs may be a valid source of data to provide estimates of obesity in children who attend primary care.

Keywords: electronic health records, child, body mass index, data accuracy, obesity

INTRODUCTION

The use of electronic medical records (EMRs) in primary health care has recently improved in Canada with estimates of physician uptake increasing from 37% in 2009 to 75% in 2015.[1,2] Using primary care EMRs to measure, track and evaluate childhood obesity have been proposed for multiple reasons. First, children attend primary health care for well-child visits frequently in the first years of life.[3] This time period has also been proposed as a critical period when prevention and early intervention strategies may be most effective.[4] Second, it is a long-standing standard of care to measure children's weight and height (or length for children younger than 2 years) at these visits, offering a potentially robust source of growth data.[5,6] Finally, leveraging data from a previously established clinical infrastructure may be cost-effective[7] and has been proposed in previous health policy recommendations.[8]

To date, the use of primary care EMRs in research has been impeded by lack of standardisation of EMRs systems, data access and concerns about the quality of the data.[9,10] However, research networks, such as the Canadian Primary Care Sentinel Surveillance Network[11] and the Electronic Medical Record Administrative data Linked Database (EMRALD)[12] in Canada, and the Electronic Pediatric Research in Office Settings (ePROS) network in the United States, have attempted to move EMR research forward by acting as an intermediary between primary care providers and researchers. Multiple studies have been published using these data sources to estimate both child and adult obesity rates.[13–15] However, descriptions of data cleaning techniques are often insufficient to be reproduced, or are not reported.[16] As well, height and weight data are often subject to error because of multiple units that may be used to record the measurement (centimeters, meters, inches, feet, kilograms and pounds), multiple decimal places and reversal of measurements (recorded weight as height and vice versa). Therefore, methods for determining data quality within these EMR networks are necessary to ensure that the derived prevalence estimates are accurate. The overarching goal of this study was to assess the feasibility of using anthropometric data from primary care EMRs to generate prevalence estimates of childhood obesity. The primary objective of this study was to assess data quality by examining data completeness and accuracy. The secondary objective was to examine factors that may be associated with these data quality attributes in order to develop recommendations on best practices for using routine EMR data.

METHODS

Data source/study population

A cohort of children 0 to <19 years of age was identified from EMRALD, containing data from primary care family medicine practices in the province of Ontario.[17] As of 2016, 355 physicians within 41 practices using PS Suite, the largest market share vendor of EMR in the province, contributed data to this study.[18] EMRALD is housed at the Institute for Clinical Evaluative Sciences (ICES) which allows this research to comply with the Ontario Personal Health Information and Protection Act. This study was approved by the Sunnybrook Health Sciences and the Hospital for Sick Children research ethics boards.

All EMRALD pediatric patient records as of March 31, 2016 were included based on two levels of inclusion criteria: 1) physicians had to be using their EMR for a minimum of 2 years and 2) patients had to be rostered to an active EMRALD physician, be less than 19 years old as of March 2016 and have a valid identification number to link with the administrative databases at ICES. Exclusion criteria included patients in the newborn period (<28 days old) where potential complications from birth could increase the number of growth measurements. Variables extracted from the EMR were height/length, weight, date of measurement, sex, age at measurement and the number of years the physician has been using EMR.

Completeness

The primary outcome for this study was the presence of a complete set of data required to calculate age- and sex-standardised body mass index (BMI) z-score (zBMI) per primary care visit, including age at measurement, sex, height/length and weight. The presence of all four data points was recoded as a binary variable, representing a complete record. Data completeness was assessed for all visits as well as stratified by visit type, either well-child, sick visit or unknown.

Accuracy

Data accuracy was determined using two methods: 1) the proportion of biologically implausible values (BIVs) and 2) assessing potential invalid inliers using repeated measurements on the study subsample with ≥3 zBMI measurements. BIVs were defined by a validated algorithm[19] and the World Health Organization (WHO) Expert Committee[20] as z-scores for BMI-for-age < −5.0 to > +5.0, height-for-age < −6.0 to > +6.0, weight-for-age < −6.0 to > +5.0 (0–10 years of age), weight-for-length < −5.0 to > +5.0 (0–5 years of age).[21]

Measurement data extracted from EMRALD are numeric values with up to two decimal places and no associated units. Weight is stored by default in kilograms and height in centimeters. If the physician enters units in the EMR as pounds or feet/inches, the value is automatically converted to kilograms and centimetres when the data is extracted from the EMR. If no unit is written, it is assumed that the data are recorded as kilograms and centimetres.

To identify invalid inliers, defined as potential data errors within the clinically acceptable range, we investigated those subjects with ≥3 zBMI measurements. The mean and standard deviation of zBMI were calculated for each subject as well as the time interval between measurements. Standardised differences (SD) were calculated by subtracting the mean zBMI per subject from each measurement divided by the standard deviation. For children <12 months, we used a rule if one measurement was more than ±2.5 SDs (where SD is the SD of the individual's measurements) of a measurement within 3 months, or for children ≥12 months more than ±3 SD of a measurement within 6 months, it would be flagged as implausible. This rule, henceforth referred to as the 'invalid inlier rule', was tested on a subsample of patients with known correct and incorrect zBMI values (verified through chart review), and found a sensitivity to detect true errors of 100% and a specificity of 28.3%. Previous studies have applied similar rules to their study populations with repeated measurements.[16,22]

Factors associated with data quality

We assessed clinic, physician and patient level factors to determine what characteristics may have affected completeness or accuracy of data. Clinic level factors examined included the size of practice (number of rostered patients) and proportion of pediatric patients (number of pediatric patients divided by practice size). The physician level factor is years the physician has been using EMRs. It was previously shown that it takes approximately 2 years for a physician to adequately populate the EMR records for their practice.[12] Patient level factors examined were age and sex. Visit type was determined by linking growth data to the Ontario Health Insurance Plan (OHIP) database to determine the visit-code billed on the corresponding date of measurement. Well-child visits were coded using fee codes and diagnostic codes defined by ICES[3] and are available in Supplementary Table 1. Since standard of care is to measure children for both height and weight during well-child visits, the factors affecting data completeness analysis were modelled only for well-child visits.

Statistical analysis

Descriptive statistics were performed for all variables to determine distributions and to create the main outcome variables. Baseline characteristics between children in EMRALD and all Ontario children, from administrative data holdings at ICES, were compared. The data completeness variable was first created by identifying each record with a missing height, weight, age or sex. This data were then linked to the OHIP

Table 1 Baseline characteristics of EMRALD patients

Characteristics	EMRALD study sample	Rostered Ontario patients <19 years	All Ontario patients <19 years
Sample size	54,964	2,709,742	4,068,251
Patient			
Sex			
Male	49.5%	49.5%	48.9%
Female	50.5%	50.5%	51.1%
Age groups (as of 31 March 2016)			
<1	5.4%	-	-
1–4	33.6%	12.8%	15.4%
5–9	30.0%	18.3%	19.6%
10–14	20.9%	20.9%	19.9%
15–19	10.1%	24.2%	22.0%
Neighbourhood income quintile			
1 – Lowest income	15.6%	18.2%	19.6%
2	17.5%	18.7%	18.5%
3	21.2%	20.4%	19.5%
4	21.9%	21.9%	20.8%
5 – Highest income	22.5%	20.5%	19.7%
Unknown	684	0.4%	2.0%
Rurality			
Rural	18.7%	11.6%	10.6%
Urban	81.4%	88.3%	87.8%
Unknown		0.1%	1.6%
Physicians (N = 355)			
Years on EMR (mean)	7.44		
Range	0.2–25		
Practices (N = 41)			
Practice size (patient volume)			
Small < 5000	36.6%		
Medium 5000–10,000	34.2%		
Large > 10,000	29.3%		
Proportion of pediatric patients			
Mean %	18.2%		
Range	9.0%–28.0%		

database to identify which measurement records were taken at a well-child visit, sick visit or unknown. A large number of well-child visits occur during the first years of life due to scheduled immunisations;[23] therefore, the age by visit type interaction terms for data completeness were tested using a p-value for statistical significance of <0.05. Previously published rules to identify BIVs in EMRs[19] were applied, followed by the WHO flags for weight-for-age, height-for-age or BMI-for-age. All those patients with three or more zBMI measurements within −5 and +5 SD were assessed using the 'invalid inlier rule'. All records identified using these rules were recoded as

inaccurate values. Proportions and 95% confidence intervals (CIs) were calculated for data completeness and accuracy.

A generalised linear mixed model (GLMM) was used to examine the effect of patient, physician and practice factors on the data completeness and accuracy. The multiple repeated measurements of patients (level 3) were clustered within physician (level 2), within practice (level 1) required the use of a multilevel model to account for correlated data, which was specified as unstructured. A null model with random intercept was run initially to determine the amount of level-1 error variation attributable to the multiple practices providing data and an inter-class correlation coefficient was calculated.[24] A full model was run with all patient, physician and practice variables. All potential explanatory variables were selected *a priori* based on the literature and expert advice. All statistical calculations were performed using SAS Enterprise version 7.1 (SAS Institute, Cary, NC, USA).

RESULTS

A total of 54,694 children 0 to <19 years of age were identified from the EMRALD database contributing a total of 385,767 visits. Table 1 presents baseline characteristics of the patient, physician and practice level characteristics. The number of years physicians used the EMR ranged from 3 months to 25 years and the proportion of pediatric patients in a family practice ranged from 8% to 29%. Group practice size was categorised into small (<5000 patients), medium (5000–10,000) and large (>10,000) and accounted for 15 (36.6%), 14 (34.2%) and 12 (29.3%) practices, respectively. The EMRALD pediatric patients in this study were slightly younger than the overall Ontario pediatric population

and were similar in sex distribution and neighbourhood income quintile. In total, 70.7% of visits were from children aged 0–4 years and 29.3% from children and adolescents aged 5 to <19 years. 38,694 visits were excluded due to age <28 days old which included birth weights and children requiring multiple weight checks in the first month of life. However, only 201 patients did not have any visits after 1 month of age and were excluded.

Completeness

Overall, 66.2% (95% CI 66.1%–66.4%) of all primary care visits had a complete set of measurements to calculate a zBMI on 51,385 (93.5%) patients. Table 2 breaks down the missing data by variable type and age group for all visits. Missing height measurements accounted for the majority of incomplete data with 111,188 visits (32.0%, 95% CI 31.9%–32.2%) missing height; only 6044 (1.7%, 95% CI 1.7%–1.8%) visits were missing weight. When measurements were stratified by visit type 50.1% of growth measurements occurred at well-child visits, 35.2% occurred at sick visits, and 14.7% of records were at an unknown visit type. The proportion of complete data from well-child visits was 89% and 34% for sick visits.

Accuracy

Table 3 shows the proportion of inaccurate data based on each method to assess accuracy. In total, 6261 (2.72%) observations were determined to be likely inaccurate values. The first BIV identification using previously published cut-offs found 0.3% of weights and 0.2% of heights outside the biologically plausible range. The number of BIVs based on the WHO for BMI-for-age was 5661, 90.4% of all errors. The

Table 2 Data completeness by variable and age group

Data completeness	Total	Complete			Incomplete		
	N	N	%	95% CI	N	%	95% CI
Overall	347,073						
Height		235,885	68.0	(67.8–68.1)	111,188	32.0	(31.9–32.2)
Weight		341,029	98.3	(98.2–98.3)	6044	1.7	(1.7–1.8)
Sex		347,073	100	(100–100)	0	0.0	(0.0–0.0)
Age		347,073	100	(100–100)	0	0.0	(0.0–0.0)
Complete zBMI		229,841	66.2	(66.1–66.4)	117,232	33.8	(33.6–33.9)
By age group							
0–4 years	245,298						
Height		173,924	70.9	(70.7–71.1)	71,374	29.1	(28.9–29.3)
Weight		240,804	98.2	(98.1–98.2)	4494	1.8	(1.8–1.9)
Complete zBMI		169,430	69.1	(68.9–69.3)	75,868	30.9	(30.8–31.1)
5–19 years	101,775						
Height		61,961	59.5	(59.2–59.8)	39,814	38.2	(37.9–38.5)
Weight		100,225	96.2	(96.1–96.4)	1550	1.5	(1.4–1.6)
Complete zBMI		60,411	59.4	(59.1–59.7)	41,364	40.6	(40.3–40.9)

Table 3 Data accuracy by each BIV method and age group

Data accuracy	Total	Accurate			Inaccurate		
	N	N	%	95% CI	N	%	95% CI
Published cut-offs[19]	229,841						
Height		229,308	99.8	(99.8–99.8)	533	0.2	(0.2–0.3)
Weight		229,223	99.7	(99.7–99.8)	618	0.3	(0.3–0.3)
WHO BIV rules							
0–4 years	169,430						
Height/Length <-6 or >+6		167,203	98.7	(98.6–98.7)	2227	1.3	(1.3–1.4)
Weight <−6 or >+5		167,126	98.6	(98.6–98.7)	2304	1.4	(1.3–1.4)
BMI <−5 or >+5		165,560	97.7	(97.6–97.8)	3870	2.3	(2.2–2.4)
5–19 years	60,411						
Height <−6 or >+6		59,211	98.0	(97.9–98.1)	1200	2.0	(1.9–2.1)
Weight* <−6 or >+5	32,759	32,257	98.5	(98.3–98.6)	499	1.5	(1.4–1.7)
BMI <−5 or >+5		58,620	97.3	(96.9–97.2)	1791	3.0	(2.8–3.1)
Invalid inlier rule	196,327						
zBMI		196,241	99.9	(99.9–99.9)	86	0.04	(0.04–0.05)

*Only calculated for children <10 years

second examination of the data for invalid inliers included 31,495 patients that had ≥3 measurements contributing 196,327 observations. After applying the invalid inlier rule, only 86 (0.04%) additional errors were found.

Factors associated with data quality

Interaction terms for age and visit type were statistically significant ($p < 0.01$). Therefore, models assessing factors associated with completeness were restricted to well-child visits and stratified into age <5 years and ≥5 years. In children 0 to <5 years of age attending well-child visits, for every 1 year increase in age the odds for having both a height and a weight were 33% (95% CI: 30%–36%) higher (see Table 4). Conversely, in children 5–19 years for every 1 year increase

in age, the odds of having a complete height and weight was 2% (95% CI: 1%–4%) lower. Physicians that had been using the EMR longer had a marginally higher proportion of complete data. Larger clinics had 9% more complete data for every increase in 5000 patients in practice volume for younger children, however, 3% less complete data in older children. There was no significant difference between boys and girls. As well, examining the variation explained by practice cluster [level 1 null model inter-class correlation coefficient (ICC) = 14%–21%], there was significant variation in data completeness between practices for both age groups. Only the proportion of pediatric patients was statistically significant in the younger age group; however, clinical significance is debatable.

Table 4 Factors associated with data completeness (probability modelled is complete data point = yes)

Predictor	Age < 5 years Adjusted ORs (95% CI)	p-value	Age ≥ 5 years Adjusted ORs (95% CI)	p-value
Total sample size	159,390		26,188	
Null model (Level 1)	0.535		0.883	
ICC	13.9%		21.0%	
Patient level				
Age in years	1.33 (1.30–1.36)	<0.01	0.98 (0.96–0.99)	<0.01
Sex (ref = M)	1.01 (0.96–1.05)	0.81	1.02 (0.91–1.13)	0.86
Physician level				
Years on EMR	1.01 (1.00–1.01)	<0.01	1.03 (1.01–1.04)	0.01
Practice level				
Patient volume (for every 5000 patients)	1.09 (1.08–1.11)	<0.01	0.96 (0.91–0.99)	0.01
Proportion pediatric patients (0–19)	0.96 (0.93–0.98)	<0.01	0.99 (0.98–1.01)	0.23

Table 5 Factors associated with data accuracy (probability modelled is accurate data point = yes)

Predictor	Age < 5 years ORs (95% CI)	p-value	Age ≥ 5 years ORs (95% CI)	p-value
Total sample size	50,615*		22,100*	
Null model (Level 1)	0.8159		0.5575	
ICC	19.9%		14.5%	
Patient level				
Age in years	0.85 (0.81–0.89)	<0.01	1.11 (1.08–1.13)	<0.01
Sex (ref = *M*)	1.00 (0.87–1.16)	0.98	1.08 (0.92–1.27)	0.37
Physician level				
Years on EMR	0.97 (0.96–0.98)	<0.01	1.01 (0.99–1.04)	0.32
Practice level				
Patient volume (for every 5000 patients)	1.08 (1.05–1.11)	<0.01	1.08 (1.02–1.13)	<0.01
Proportion pediatric patients (0–19)	0.95 (0.94–0.96)	<0.01	1.00 (0.92–1.27)	0.99

*The generalised linear mixed model (GLMM) did not converge using the full sample. A random sample of 10,000 patients was selected

Due to the size of the data set and the number of random effects, the clustered data models lost stability and failed to converge when using the entire sample. Therefore, a subset of 10,000 patients was randomly selected in each age group and used to examine the effects of the covariates. In young children <5 years, for every additional year of age, the odds of having accurate data were 15% (95% CI: 11%–19%) *less* likely (see Table 5). On the other hand, in children 5–19 years for every year of additional age, the odds of having accurate data was 11% (95% CI: 8%–13%) *more* likely. Physicians that had been using EMR longer had a marginally lower proportion of accurate data points in the young age group, and no significant difference in the older age group. Larger practices had 8% increased data accuracy for every increase in 5000 patients in practice volume for both age groups. Practice variation explained a similar amount of the variance (level 1 null model ICC = 14.5%–19.9%) as the models examining data completeness.

DISCUSSION

This study examined the data quality of anthropometric measurements extracted from primary care EMRs in Ontario. Overall, data completeness was 66.2% and accuracy was 97.3%. Incompleteness was predominantly due to the high proportion of missing height data (32%). When we examined data completeness for measurements collected only at well-child visits, the proportion of complete records increased to 89%.

These results were similar to previous work on data completeness in EMRs. A study from Kaiser Permanente Colorado examined EMR data on children 3–17 years of age and reported 64% of patients had a BMI measurement at any primary care visit and >95% at well-child visits.[25] Accuracy of data recorded in EMRs was also consistent with the previous literature. One study in children 3–5 years of age used a similar multi-step data cleaning strategy and only found 2% of

data to be erroneous.[22] Another study replicated 11 different methods for identification of potential errors and found the prevalence of data errors ranged from 0.3% to 2.1%.[16]

The main factor that influenced data completeness and accuracy was child age. The direction and magnitude of the effect of age on data completeness at well-child visits changed when examining children separately by age group. This is likely due to the high number of well-child visits that occur in the first 2 years of life.[23] Primary care providers who see young infants more often in the early years may not complete both a length and weight if the child had been seen recently. Moreover, measuring length of a child less than 2 years requires appropriate equipment, such as a length board, which may be a barrier to a complete growth assessment.[26] Older children attending well-child visits in the 5–19 year age group had marginally higher data completeness. One reason may be because older children are less likely to attend well-child visits, height and weight measurements may be completed more often if the primary care provider had not seen the child in a longer time interval. Until recently the recommendations for growth monitoring applied to well-child visits only.[5,6] In 2015, the Canadian Task Force on Preventive Health Care changed the recommendation to performing both height and weight measurements at all visits for primary and secondary prevention of obesity.[27] Future analyses of EMR data will be able to assess the uptake of this recommendation.

Similar to the findings on data completeness, the effect of age on data accuracy was highly significant and differed by age group. One possible reason for this discrepancy is the tendency of measuring infants in pounds and ounces instead of kilograms. Identifying age as a determinant of data accuracy is important for future uses of EMR to develop new data cleaning algorithms that capture multiple unit conversions, especially for the youngest age group where data is most abundant. Despite the differences in age, the high proportion of accurate data was encouraging. One advantage of using

EMR data is the ability to examine multiple measurements on the same child.[16,28] This not only aids the data cleaning process by being able to examine measurements before and after a suspect value, but it allows researchers to examine how the same population of children can change over time, including into adulthood.

There were several limitations to this study. There may have been misclassification of the data accuracy outcome for several reasons. The lack of units for each numeric value for height and weight was problematic. Although most imperial system values were excluded in the assessment of BIVs, the prevalence of invalid inliers for subjects contributing only one or two measurements is unknown. In the WHO computer program, weight-for-age is not calculated beyond 10 years of age, making it harder to determine which outliers for zBMI data are from weight data in adolescents. The BIV cut-offs suggested by the WHO for calculation of zBMI may be too conservative and be incorrectly excluding those patients with extremely high zBMIs.[29] One previous study demonstrated the BIV cut-offs from the WHO underestimated obesity prevalence[30] and recently, the Center for Disease Control (CDC) changed their upper limit for BIVs from >+5 to >+8.[31] To the best of our knowledge, there are no validated or standardised rules on plausible changes over time to differentiate true errors from correct values.[32] More research is required to determine valid BIV cut-offs that can be used for large data sets that are becoming more available with improved health information databases. Lastly, the clinic size variable may have been underestimated for 13% of observations because in eight clinics not all physicians contribute data to the EMRALD network.

The results from this study raise important considerations of the feasibility of using growth data from EMRs for public health and surveillance purposes. Visit type and age were important determinants for whether or not measurements were complete, specifically height. Previous research has shown a difference in zBMI between well-child and sick visits[15]; we found mean zBMI from sick visits to be significantly higher than well-child visits. It may also be likely that children who attend regular well-child care are systematically different than those who only attend when sick. Therefore, it is important to acknowledge possible selection biases that may be introduced when using data collected in routine primary care.

Finally, our study population was skewed to be younger than Ontario rostered patients due to examining visits with growth data which are concentrated in children 0–4 years.

The next step to improving the quality of this EMR data should include developing more sophisticated data cleaning for efficiently maximizing the available data. This includes determining visit type through machine learning text classification of physicians' common 'short-hand' for indicating well-child visits in patient progress notes, validating correct BIVs for patients with severe obesity and determining implausible changes in height and weight over specific time intervals. Future research should develop and validate these data cleaning algorithms for large study populations in order for researchers to standardise techniques. However, despite the need for continuous evaluation of data quality, the current state of growth data was highly complete and accurate. EMRs are a good data source to characterise weight status in a large population of young children and may be useful in assessing uptake of recommendations or interventions related to childhood growth monitoring or obesity.

Acknowledgements

Funding

REFERENCES

1. Collier R. National physician survey: EMR use at 75%. *Canadian Medical Association Journal de l'Association Medicale Canadienne* 2015;187(1):E17–18.

2. Terry AL, Stewart M, Fortin M, Wong ST, Kennedy M, Burge F, et al. How does Canada stack up? A bibliometric analysis of the primary healthcare electronic medical record literature. *Informatics in Primary Care* 2012;20(4):233–40. Available from: https://doi.org/10.14236/jhi.v20i4.2. PMid:23890334.

3. Guttman A, Schultz SE and Jaakkimainen L. *Primary Care for Children.* Toronto, ON: Institute for Clinical Evaluative Sciences, 2006.

4. Woo Baidal JA, Locks LM, Cheng ER, Blake-Lamb TL, Perkins ME and Taveras EM. Risk factors for childhood obesity in the first 1,000 days: a systematic review. *American Journal of Preventive Medicine* 2016;50(6):761–79. doi: 710.1016/j.amepre.2015.1011.1012. Epub 2016 Feb 1022.

5. Greig A, Constantin E, Carsley S and Cummings C. Preventive health care visits for children and adolescents aged six to 17 years: The Greig Health Record – Executive Summary. *Paediatrics & Child Health* 2010;15(3):157–62. Available from: https://doi.org/10.1093/pch/15.3.157. PMid:21358896; PMCid:PMC2865953.

6. Rourke L, Leduc D, Constantin E, Carsley S and Rourke J. Update on well-baby and well-child care from 0 to 5 years: What's new in the Rourke Baby Record? *Canadian Family Physician* 2010;56(12):1285–90. PMid:21156890; PMCid:PMC3001919.

7. Wood GC, Chu X, Manney C, Strodel W, Petrick A, Gabrielsen J, et al. An electronic health record-enabled obesity database. *BMC Medical Informatics and Decision Making* 2012;12:45. Available from: https://doi.org/10.1186/1472-6947-12-45. PMid:22640398; PMCid:PMC3508953.

8. Healthy Kids Panel. *No Time to Wait: The Healthy Kids Strategy*. Toronto, ON: Ministry of Health and Long-Term Care, 2013.

9. Greenhalgh T, Potts HW, Wong G, Bark P and Swinglehurst D. Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method. *Milbank Quarterly* 2009;87(4):729–88. Available from: https://doi.org/10.1111/j.1468-0009.2009.00578.x. PMid:20021585; PMCid:PMC2888022.

10. Jensen RE, Chan KS, Weiner JP, Fowles JB and Neale SM. Implementing electronic health record-based quality measures for developmental screening. *Pediatrics* 2009;124(4):e648–54. Available from: https://doi.org/10.1542/peds.2008-3091. PMid:19786425.

11. Keshavjee K, Williamson T, Martin K, Truant R, Aliarzadeh B, Ghany A, et al. Getting to usable EMR data. *Canadian Family Physician* 2014;60(4):392. PMid:24733333; PMCid:PMC4046531.

12. Tu K, Widdifield J, Young J, Oud W, Ivers NM, Butt DA, et al. Are family physicians comprehensively using electronic medical records such that the data can be used for secondary purposes? A Canadian perspective. *BMC Medical Informatics and Decision Making* 2015;15:67. Available from: https://doi.org/10.1186/s12911-015-0195-x. PMid:26268511; PMCid:PMC4535372.

13. Biro S, Williamson T, Leggett JA, Barber D, Morkem R, Moore K, et al. Utility of linking primary care electronic medical records with Canadian census data to study the determinants of chronic disease: an example based on socioeconomic status and obesity. *BMC Medical Informatics and Decision Making* 2016;16:32. doi: 10.1186/s12911-12016-10272-12919.

14. Rigobon AV, Birtwhistle R, Khan S, Barber D, Biro S, Morkem R, et al. Adult obesity prevalence in primary care users: an exploration using Canadian Primary Care Sentinel Surveillance Network (CPCSSN) data. *Canadian Journal of Public Health* 2015;106(5):e283–89. doi: 210.17269/cjph.17106.14508.

15. Birken CS, Tu K, Oud W, Carsley S, Hanna M, Lebovic G, et al. Determining rates of overweight and obese status in children using electronic medical records: cross-sectional study. *Canadian Family Physician* 2017;63(2):e114–22. PMid:28209703; PMCid:PMC5395409.

16. Lawman HG, Ogden CL, Hassink S, Mallya G, Vander Veur S and Foster GD. Comparing methods for identifying biologically implausible values in height, weight, and body mass index among youth. *American Journal of Epidemiology* 2015;182(4):359–65. Available from: https://doi.org/10.1093/aje/kwv057. PMid:26182944; PMCid:PMC4528955.

17. Tu K, Mitiku TF, Ivers NM, Guo H, Lu H, Jaakkimainen L, et al. Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD). *The American Journal of Managed Care* 2014;20(1):e15–21. PMid:24669409.

18. OntarioMD. Core EMR specifications, 2015. Available from: https://www.ontariomd.ca/portal/server.pt/community/ontario_emr_specifications/current_emr_specifications/. Accessed 31 January 2017.

19. Smith N, Coleman KJ, Lawrence JM, Quinn VP, Getahun D, Reynolds K, et al. Body weight and height data in electronic medical records of children. *International Journal of Pediatric Obesity: An Official Journal of the International Association for the Study of Obesity* 2010;5(3):237–42.

20. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. *World Health Organization Technical Teport Series* 1995;854:1–452. PMid:8594834.

21. WHO. *AnthroPlus for Personal Computers Manual: Software for Assessing Growth of the World's Children and Adolescents*. Geneva, Switzerland: WHO, 2009. Available from: http://www.who.int/growthref/tools/readme_sas.pdf?ua=1. Accessed 31 March 2016.

22. Lo JC, Maring B, Chandra M, Daniels SR, Sinaiko A and Daley MF. Prevalence of obesity and extreme obesity in children aged 3–5 years. *Pediatric Obesity* 2014;9(3):167–75. Available from: https://doi.org/10.1111/j.2047-6310.2013.00154.x. PMid:23677690; PMCid:PMC3830709.

23. Guttmann A, Manuel D, Dick PT, To T, Lam K and Stukel TA. Volume matters: physician practice characteristics and immunization coverage among young children insured through a universal health plan. *Pediatrics* 2006;117(3):595–602. Available from: https://doi.org/10.1542/peds.2004-2784. PMid:16510636.

24. Snijders TA and Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling,* 2nd edition. Los Angeles, CA: Sage, 2012.

25. Estabrooks PA and Shetterly S. The prevalence and health care use of overweight children in an integrated health care system. *Archives of Pediatrics & Adolescent Medicine* 2007;161(3):222–7. Available from: https://doi.org/10.1001/archpedi.161.3.222. PMid:17339502.

26. Wood AJ, Raynes-Greenow CH, Carberry AE and Jeffery HE. Neonatal length inaccuracies in clinical practice and related percentile discrepancies detected by a simple length-board. *Journal of Paediatrics and Child Health* 2013;49(3):199–203. Available from: https://doi.org/10.1111/jpc.12119. PMid:23432733.

27. Canadian Task Force on Preventive Health Care. Recommendations for growth monitoring, and prevention and management of overweight and obesity in children and youth in primary care. *Canadian Medical Association Journal* 2015;187(6):411–21.

28. Nader PR, O'Brien M, Houts R, Bradley R, Belsky J, Crosnoe R, et al. Identifying risk for obesity in early childhood. *Pediatrics* 2006;118(3):e594–601. Available from: https://doi.org/10.1542/peds.2005-2801. PMid:16950951.

29. Ratcliff MB, Jenkins TM, Reiter-Purtill J, Noll JG and Zeller MH. Risk-taking behaviors of adolescents with extreme obesity: normative or not? *Pediatrics* 2011;127(5):827–34. Available from: https://doi.org/10.1542/peds.2010-2742.

30. Freedman DS, Lawman HG, Pan L, Skinner AC, Allison D, McGuire LC, et al. The prevalence and validity of high, biologically implausible values of weight, height, and BMI among 8.8 million children. *Obesity* 2016;24(5):1132–9. Available from: https://doi.org/10.1002/oby.21446. PMid:26991694; PMCid:PMC4846478.

31. CDC. A SAS Program for the 2000 CDC Growth Charts (ages 0 to <20 years). Available from: https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm. Accessed 31 January 2017.

32. Zemel BS. A commentary on the construction of weight velocity charts. *Nutrition in Clinical Practice: Official Publication of the American Society for Parenteral and Enteral Nutrition* 2009;24(5):651–53.