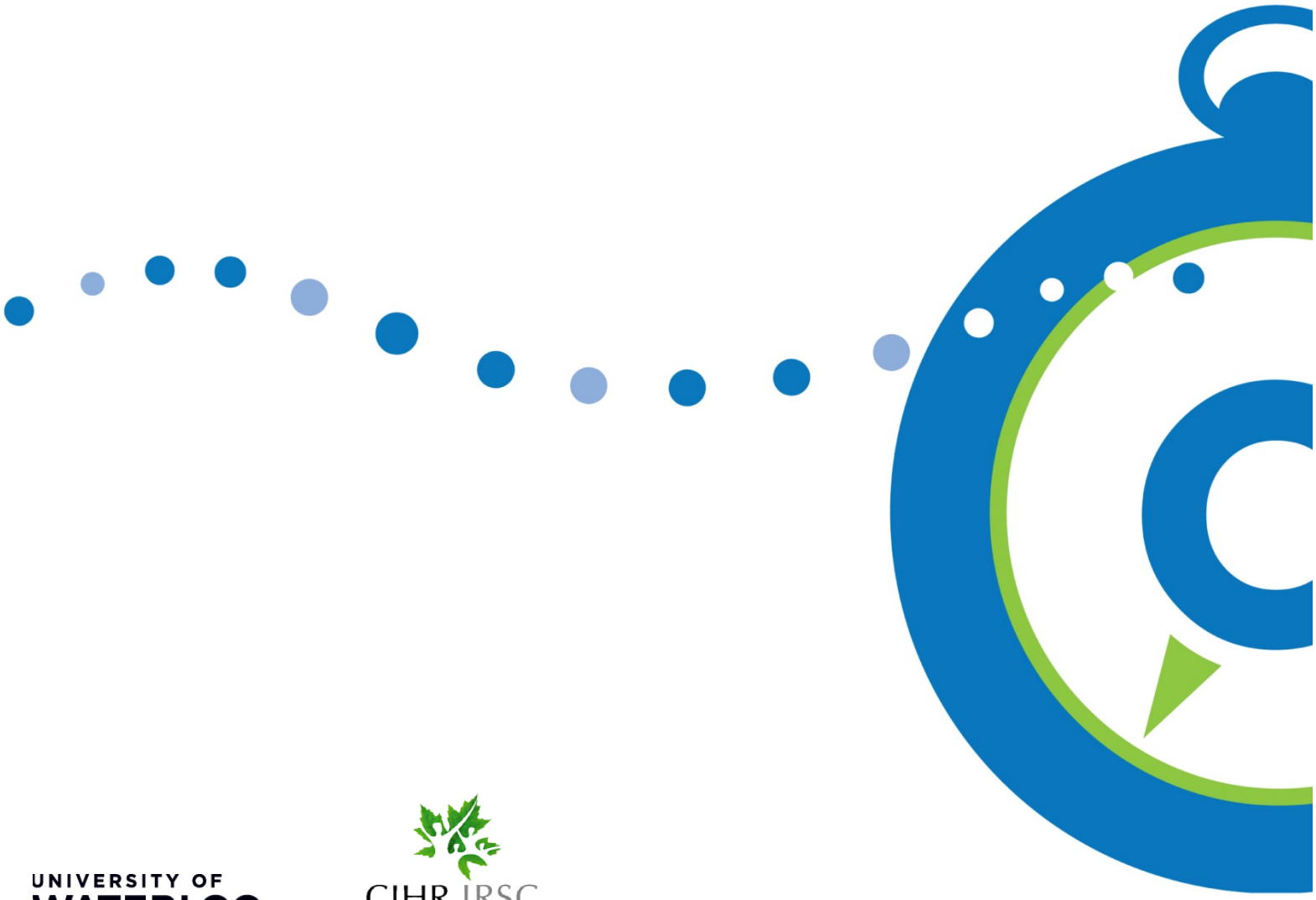
Cohort study evaluating how changes in school programs, policies, and resources impact youth health behaviours

Assessing longitudinal data linkage results in the COMPASS study

COMPASS Technical Report Series, Volume 3, Issue 4

August 2015



UNIVERSITY OF WATERLOO

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

University of Waterloo

Waterloo, Ontario

August 2015

uwaterloo.ca/compass-system

Acknowledgements

Wei Qian, MSc[1]
Katelyn Battista, MMath[2]
Chad Bredin, BA[3]
K. Stephen Brown, PhD[3]
Scott T. Leatherdale, PhD[1]

1 - School of Public Health and Health Systems, University of Waterloo, Waterloo, ON Canada.
2 - Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON Canada.
3 – Propel Centre for Population Health Impact, University of Waterloo, Waterloo, ON Canada.

Report funded by:

The COMPASS study was supported by a bridge grant from the Canadian Institutes of Health Research (CIHR) Institute of Nutrition, Metabolism and Diabetes (INMD) through the "Obesity – Interventions to Prevent or Treat" priority funding awards (OOP-110788; grant awarded to S. Leatherdale) and an operating grant from the Canadian Institutes of Health Research (CIHR) Institute of Population and Public Health (IPPH) (MOP-114875; grant awarded to S. Leatherdale).

Suggested citation:

Qian W, Battista K, Bredin C, Brown KS, Leatherdale ST. *Assessing longitudinal data linkage results in the COMPASS study*: Technical Report Series. 2015; 3(4). Waterloo, Ontario: University of Waterloo. Available at: https://uwaterloo.ca/compass-system/publications#technical

Contact:

COMPASS Research Team
University of Waterloo
200 University Ave West, BMH 1038
Waterloo, ON Canada N2L 3G1
uwaterloo.ca/compass-system

Table of Contents

Contents

Introduction

COMPASS is a longitudinal study (started in 2012-13) designed to follow a cohort of grade 9 to 12 students attending a convenience sample of Ontario secondary schools for four years to understand how changes in school environment characteristics (policies, programs, built environment) are associated with changes in youth health behaviours [1]. COMPASS originated to provide school stakeholders with the evidence to guide and evaluate school-based interventions related to obesity, healthy eating, tobacco use, alcohol and marijuana use, physical activity, sedentary behaviour, school connectedness, bullying, and academic achievement. COMPASS has been designed to facilitate multiple large-scale school-based data collections and uses in-class whole-school sampling data collection methods consistent with previous research [2-5]. COMPASS also facilitates knowledge transfer and exchange by annually providing each participating school with a school-specific feedback report that highlights the school-specific prevalence for each outcome, comparisons to provincial and national norms or guidelines, and provides evidence-based suggestions for school-based interventions (programs and/or policies) designed to address the outcomes covered in the feedback report (refer to: https://uwaterloo.ca/compass-system/).

One challenge associated with the COMPASS data is to link student-level data over years since these data are self-reported by students anonymously. COMPASS includes a series of questions in the student questionnaire (C_q) that are designed for linkage purposes only (see Appendix 1), and then uses the answers to these questions to create a unique code for each student in a school. This method was designed to be simple-to-complete and able to ensure students' anonymity while still allowing us to link each student's unique identifier data over multiple years [6-7]. The generated code allows us to link student-level data within each school using an algorithm developed by the COMPASS team, led by Brown. This linkage method was tested as part of the COMPASS validation study. The method was found to be robust and to produce sufficiently high linkage rates [6]. The linkage process has been completed for Year 1 (Y_1) and Year 2 (Y_2) data.

We have created a longitudinal sample of 11,049 students with the responses from Y_1 (2012/2013) and Y_2 (2013/2014). Provided with this longitudinal sample, users may ask how this sample represents the study population:

COMPASS contains both longitudinal and cross-sectional components; it is important to distinguish between them. The target population of our longitudinal sample covers students who are expected to attend Ontario high schools in both Y_1 and Y_2. This definition excludes most grade 12 students in Y_1 who would graduate from high school in Y_2.

It is also important to understand the potential sources of bias during the creation of the final sample, including bias due to convenience sampling, bias due to non-response (including absence, refusal, and drop-out), and bias due to linkage. The quasi-experimental design of COMPASS assumes the convenience sample does not introduce sampling bias. The potential bias due to linkage is the main interest of this document, and the response bias will be explored in another report. The linkage bias in terms of the dynamic trend is difficult to assess; instead, we evaluate the linkage bias in terms of a snapshot at Y_1.

This technical reports provides a detailed description of the linkage of Y_1 and Y_2 longitudinal student data, with the aim of helping data users understand the benefits and limitations of using these linked data.

Methods

Obtaining the Linked Sample

In Y_1, 30,147 grade 9 to 12 students were enrolled in the 43 participating schools and 24,173 of them (80.2%) completed the C_q. In Y_2, 29,945 grade 9 to 12 students were enrolled and 23,424 of them (78.2%) completed the C_q. Missing respondents resulted primarily from scheduled spares or absenteeism at the time of the C_q, and partially from student or parent refusal (see Table 2).

The longitudinal sample is created by linking Y_1 and Y_2 student responses to a six digit alpha-numeric code generated for each completed questionnaire using the responses to five specifically-designed questions along with the response to the question regarding the student's sex. Bredin and Leatherdale [6] provide more information on the creation of the identification questions. Within each school, Y_1 and Y_2 codes are compared by record. If the code for record A in Y_1 matches the code for record B in Y_2 on at least 5 out of 6 digits, A and B are considered to be a match. Note that students who answered "No" to the question "Did you attend this school last year?" in the Y_2 C_q are excluded from the linkage process.

Additional restrictions are then imposed to reduce false-linkage error. Using information from other questions in each record, the match is dissolved if:

- the difference in grade between Y_1 and Y_2 is less than zero or greater than one
- the difference in age is greater than two

In total, a two-year longitudinal sample of 11,049 students is created using the COMPASS Y_1 and Y_2 data.

Assessing the Quality of Linkage

The linkage process is subject to two types of errors: missing linkage error (matched pair is not identified) and false-linkage error (unmatched pair is identified as pair); see the shaded cells as shown in Table 1.

Table 1: Possibilities for matches and false matches

Actual	**Outcome**	
	Matched	**Unmatched**
Matched	True match (A)	False non-match (B) (Missing linkage error)
Unmatched	False match (C) (False-linkage error)	True non-match (D)

The false-linkage error rate is defined as

$$False\ Linkage\ Rate = \frac{C}{A + C} \times 100\%.$$

This false linkage error is difficult to evaluate since we cannot know the number of false matches (C) without a validation study; however, before the COMPASS survey, a validation study was conducted and data were collected from a convenience sample of 204 students [6] in which 132 matches were found and none of them were false matches. Thus, we may assume the false-linkage error is negligible, and later we will further verify this by looking at the consistency between Y_1 and Y_2 regarding student characteristics.

The matching rate is often used to measure the missing linkage error. The matching rate is defined as

$$Matching\ Rate = \frac{A}{A + B} \times 100\%$$

where A = 11,049 and B remains unknown. The denominator is the number of students who participated in both Y_1 and Y_2 C_q's and is unknown. We roughly estimate it from the number of Y_1 students or Y_2 students by subtracting the number of students who did not participate in both C_q's and thus were not expected to be linked.

Year 1 students not expected to be linked include:

a) students not participating in Y_1 C_q (5,672 students due to spares and absenteeism, 302 due to student or parent refusal)
b) students absent on the Y_2 C_q date (the Y_2 data show around 21.8% students were absent on the C_q date)
c) Grade 12 students graduating from the high school (5,669 grade 12 students in Y_1, and 283 linked to Y_2 grade 12)
d) students transferring out to other schools (this number is unknown)
e) students dropping out of school in Y_2 (Ontario 2012 high school student drop-out rate was 6.6% [8])

Year 2 students not expected to be linked include:

a) students not participating in the Y_2 C_q (6,192 due to spares and absenteeism, 329 due to student or parental refusal)
b) students absent on the Y_1 C_q date (the Y_1 data show around 18.8% students were absent on the C_q date)
c) Grade 9 students newly admitted into high school (6,342 grade 9 students in Y_2, and 12 remaining in grade 9 in Y_2)
d) students transferring in from other schools (this number is unknown)

Table 2 shows the breakdown of the number of students expected to be linked via the linking process. As shown, we have a raw matching rate of 80.5% for Y_1 and 79.6% for Y_2. As we mentioned before, this is a rough estimate, but it shows the linkage strategy worked well.

Table 2: Number of students linked and unlinked over two years

	Year 1	Year 2
Total Students Enrolled	**30,147**	**29,945**
Less: Missing Due to Spares and Absenteeism	- 5,672	- 6,192
Less: Missing Due to Student or Parent Refusal	- 302	- 329
Students Completing Survey	**=24,173**	**=23,424**
Less: Students in Grade 12	- 5,669	-
Plus: Students remaining in Grade 12	+ 283	-
Less: Students transferring out to other schools	.[1]	-
Percentage of students present on Y2 survey date	X 78.2%[2]	-
Percentage of not dropping out of schools in Y2	X 93.4%	-
Less: Students in Grade 9	-	- 6,342
Plus: Grade 9 students remaining in Grade 9 in Y2	-	+ 12
Less: Students transferred in from other schools	-	.
Percentage of students present on Y1 survey date	-	X 81.2%[3]
Total Students Expected to be Linked	**13,722**	**13,880**
Total Students Linked	**11,049**	**11,049**
Linkage Rate	*80.5%*	*79.6%*

To validate the accuracy of the linkage process in terms of false-linkage error, we examined characteristics of the matched students from Y_1 to Y_2. Tables 3 and 4 show the sex and grade distribution of matched students in Y_1 and Y_2. The majority of linked students provided consistent sex and grade information across both years. Fewer than 0.15% of matches have contradictory sex information, and no matches have contradictory grade information (a difference greater than 1 year). Only 3.0% of students reported staying in the same grade as the previous year, with the vast majority being grade 12 students. The consistency in the information for matched students suggests a very low false-linkage rate.

[1] Because we are unable to accurately quantify the percentage of students who move to a different school in a given year, we have not included this in the equation. The rate is likely significant, however, as 5.8% of 2012-13 participants in grades 10-12 reported in the C_q having not attended their current school the previous year.

[2] In Y2, 21.8% of students were absent for the C_q. Students in Y1 were assumed to have the same absentee rate in Y2.

[3] In Y1, 19.8% of students were absent for the C_q. Students in Y2 were assumed to have the same absentee rate in Y1

Table 3: Validating linkage accuracy using sex variables

Sex	Year 2			
Year 1	**Female**	**Male**	**Missing**	**Total**
Female	5782	5	31	5818
Male	10	5157	32	5199
Missing	16	16	0	32
Total	5808	5178	63	11049

Table 4: Validating linkage accuracy using grade variables

Grade	Year 2					
Year 1	**9**	**10**	**11**	**12**	**Missing**	**Total**
9	12	3979	0	0	6	3997
10	0	12	3727	0	8	3747
11	0	0	24	2982	5	3011
12	0	0	0	283	6	289
Missing	0	4	0	1	0	5
Total	12	3995	3751	3266	25	11049

As further validation of the sample, Table 4 shows the distribution of body mass index (BMI) for matched students in Y_1 and Y_2. Of the 11,049 matches, 7,722 had complete BMI information. Of the 7,722 matched students, 6,471 (83.8%) reported to be in the same BMI category across both years. Only 103 students (1.3%) reported moving by more than one weight category. Note that the BMI calculation requires complete height and weight information: Students not knowing or not reporting either one of these accounts for the lower-than-usual response rate.

Table 5: Assessing validity of linkage accuracy using BMI variable

BMI	Year 2				
Year 1	**Underweight**	**Heathy Weight**	**Overweight**	**Obese**	**Total**
Underweight	38	94	2	0	134
Heathy Weight	80	5266	388	47	5781
Overweight	4	358	796	126	1284
Obese	3	47	102	371	523
Total	125	5765	1288	544	7722

Tables 6 and 7 show the binge drinking and marijuana use status for the linked students in Y_1 and Y_2. Students are classified as either current, non-current, or never-users. A small but significant percentage of students reported contradictory responses; that is, students reporting to be current or non-current users in Y_1, but reporting having never used in Y_2. This amounts to 488 students (3.4%) for binge drinking and 208 students (1.9%) for marijuana use. While this result is initially surprisingly, it is similar to results often seen in other longitudinal studies where individuals' responses across time are compared. [9] More information on the classification is provided in the Substance Use section of the results.

Table 6: Binge-drinking status in Y1 and Y2: illustrating contradictory responses for alcohol-use over time

Binge Drinking	**Year 2**			
Year 1	**Never**	**Non-Current**	**Current**	**Total**
Never	4635	1560	589	6784
Non-Current	374	1250	849	2473
Current	104	387	1244	1735
Total	5113	3197	2682	10992

Table 7: Marijuana-use status in Y1 and Y2: illustrating contradictory responses for marijuana-use over time

Marijuana Use	**Year 2**			
Year 1	**Never**	**Non-Current**	**Current**	**Total**
Never	6898	946	522	8366
Non-Current	141	698	385	1224
Current	67	249	812	1128
Total	7106	1893	1719	10718

Results

Using Y_1 data and *excluding grade 12 students*, we compare linked respondents with non-linked respondents to show the potential bias for a group of selected variables. We break down the comparison by sex and grade; as a result, students with missing grade or sex information are also excluded. A total of 18,280 grade 9 to 11 students with complete grade and sex information are compared, 10,730 (58.7%) students are linked and 7,550 (42.3%) are not linked. Table 8 shows the distribution of the 18,280 students.

Table 8: Linkage rates by grade and sex

Sex		**Grade (Y1)**			
		9	**10**	**11**	**Total**
Total	Eligible Students	6270	6144	5866	18280
	Percentage Linked	63.6%	60.8%	51.2%	58.7%
Female	Eligible Students	3133	3099	2893	9125
	Percentage Linked	68.0%	64.6%	54.7%	62.6%
Male	Eligible Students	3137	3045	2973	9155
	Percentage Linked	59.2%	57.0%	47.8%	54.8%
Difference between genders		8.8%*	7.6%*	6.9%*	7.8%*

*: p-value <0.0001

The variables we selected to test for potential bias are grouped into five categories that represent the primary COMPASS study outcomes: obesity, physical activity, sedentary behaviour, substance use, and bullying and academics. For each variable analyzed, we compare the distributions of categorical variables using a Chi-square test or the means of continuous variables using a t-test separately for each sex and grade group. Students not reporting the information are excluded. A p-value of less than 0.05 is considered statistically significant in assessing whether differences exist between the linked samples and non-linked samples.

As a result of the analyses, obesity-related measures do not show significant differences between linked and non-linked samples. Significant differences are, however, consistently seen on measures of sedentary behaviour, substance use, and bullying and academics, and to a lesser degree on measures of physical activity.

Obesity

Obesity-related measures include a student's body mass index (BMI), as well as measures of whether students are receiving the Canada Food Guide recommended number of servings for each food group. The results showed no significant differences between the linked and non-linked samples on any of the obesity-related measures, with the exception of a significant difference in 'meats and alternatives' consumption for grade 10 males only.

Body Mass Index (BMI)

BMI is a measure of healthy body weight, calculated from a student's self-reported height and weight. Based on BMI scores, students are classified into four groups: Underweight, Healthy Weight, Overweight, and Obese, according to the BMI classification system defined by the World Health Organization. [10]

Of the 18,280 eligible students who completed the questionnaire in Y_1, 14,207 had complete BMI information. Of these students, 8,394 were linked and 5,633 were not linked. Table 9 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. No significant differences were observed between the linked and non-linked samples.

Table 9: Statistical evaluation of linkage rates for students responding to BMI questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF* = 3	p-value
Female	9	2188	1515	673	2.6	0.466
	10	2398	1581	817	4.8	0.186
	11	2310	1290	1020	1.6	0.654
Male	9	2313	1383	930	0.8	0.851
	10	2401	1430	971	6.0	0.111
	11	2417	1195	1222	2.5	0.472

*DF: degree of freedom

Table 10 shows the percentage of linked and non-linked students in each BMI class by grade and sex. Consistent with the results of the Chi-square tests, the distribution of BMI classification is very similar between linked and non-linked students in each sex and grade group.

Table 10: Comparing distribution of linked and non-linked data for BMI categories, by grade and sex

Female				Male			
Grade	**BMI**	**Linked**	**Non-Linked**	**Grade**	**BMI**	**Linked**	**Non-Linked**
9	Underweight	2	2.5	**9**	Underweight	2.2	1.9
	Heathy Weight	80	77.3		Heathy Weight	66.1	65.1
	Overweight	14.3	16.5		Overweight	20.8	21.2
	Obese	3.7	3.7		Obese	10.9	11.8
10	Underweight	1.6	2.2	**10**	Underweight	1.2	2.1
	Heathy Weight	81.2	78.8		Heathy Weight	67.6	64.8
	Overweight	13.5	13.7		Overweight	19.8	22.6
	Obese	3.7	5.3		Obese	11.5	10.6
11	Underweight	1.6	2.3	**11**	Underweight	2.3	1.6
	Heathy Weight	78.9	77.9		Heathy Weight	67.3	69.1
	Overweight	14.4	14.5		Overweight	20.6	19.6
	Obese	5.1	5.3		Obese	9.8	9.7

Fruit and Vegetable Consumption

Fruit and vegetable consumption is assessed according whether students received the Canada Food Guide recommended servings of fruits and vegetables in the previous day. Students are categorized according to whether they consumed at least the recommended number of servings, which is 7 servings for females and 8 servings for males. [11]

Of the eligible students, 17,795 completed the fruit and vegetable consumption question, including 10,536 linked students and 7,259 non-linked students. Table 11 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. No significant differences were observed between the linked and non-linked samples.

Table 11: Statistical evaluation of linkage rates for students responding to fruit and vegetable consumption questions, by grade and sex

Sex	**Grade**	**Total**	**Linked**	**Non-Linked**	**Chi-square** DF = 1	**p-value**
Female	9	3073	2096	977	0.5	0.499
	10	3053	1979	1074	2.1	0.145
	11	2831	1560	1271	0.7	0.411
Male	9	3029	1818	1211	3.5	0.060
	10	2941	1698	1243	0.0	0.930
	11	2868	1385	1483	0.8	0.379

Table 12 shows the percentage of linked and non-linked students in each category for fruit and vegetable consumption by grade and sex. Consistent with the results of the Chi-square tests, the results are similar between linked and non-linked students. The majority of students do not meet the recommended serving levels.

Table 12: Comparing distribution of linked and non-linked data for fruit and vegetable consumption variables, by grade and guideline categorization

Female				Male			
Grade	**Recommended Servings**	**Linked**	**Non-Linked**	**Grade**	**Recommended Servings**	**Linked**	**Non-Linked**
9	Does Not Meet	94.4	93.8	**9**	Does Not Meet	96.1	94.7
	Meets/Exceeds	5.6	6.2		Meets/Exceeds	3.9	5.3
10	Does Not Meet	94.3	93.0	**10**	Does Not Meet	95.6	95.6
	Meets/Exceeds	5.7	7.0		Meets/Exceeds	4.4	4.4
11	Does Not Meet	94.8	94.1	**11**	Does Not Meet	95.5	94.7
	Meets/Exceeds	5.2	5.9		Meets/Exceeds	4.5	5.3

Grain Product Consumption

Grain product consumption is assessed according whether students received the Canada Food Guide recommended servings of grain products (breads, cereals, rice, and pasta) in the previous day. Students are categorized according to whether they consumed at least the recommended number of servings, which is 6 servings for females and 7 servings for males. [11]

Of the eligible students, 17,794 completed the grain product consumption question, including 10,534 linked students and 7,260 non-linked students. Table 13 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. No significant differences were observed between the linked and non-linked samples.

Table 13: Statistical evaluation of linkage rates for students responding to grain consumption questions, by grade and sex

Sex	**Grade**	**Total**	**Linked**	**Non-Linked**	**Chi-square** DF = 1	**p-value**
Female	9	3072	2095	977	0.5	0.458
	10	3053	1979	1074	0.5	0.487
	11	2831	1561	1270	1.2	0.266
Male	9	3030	1818	1212	2.7	0.099
	10	2937	1696	1241	2.0	0.157
	11	2871	1385	1486	1.6	0.202

Table 14 shows the percentage of linked and non-linked students in each grain consumption category by grade and sex. Consistent with the results of the Chi-square tests, the results are similar between linked and non-linked students. The majority of students do not meet the recommended serving levels.

Table 14: Comparing distribution of linked and non-linked data for grain consumption variables, by grade and guideline categorization

Female				Male			
Grade	**Recommended Servings**	**Linked**	**Non-Linked**	**Grade**	**Recommended Servings**	**Linked**	**Non-Linked**
9	Does Not Meet	93.7	93.0	**9**	Does Not Meet	91.9	90.2
	Meets/Exceeds	6.3	7.0		Meets/Exceeds	8.1	9.8
10	Does Not Meet	93.9	93.3	**10**	Does Not Meet	91.2	89.6
	Meets/Exceeds	6.1	6.7		Meets/Exceeds	8.8	10.4
11	Does Not Meet	93.7	92.6	**11**	Does Not Meet	91.0	89.6
	Meets/Exceeds	6.3	7.4		Meets/Exceeds	9.0	10.4

Meats and Alternatives Consumption

Meat and meat alternative consumption is assessed according whether students received the Canada Food Guide recommended servings of meats and alternatives in the previous day. One serving of meat and alternatives includes cooked fish, chicken, beef, pork, or game meat, eggs, nuts or seeds, peanut butter or nut butters, legumes (beans), and tofu. Students are categorized according to whether they consumed at least the recommended number of servings, which is 2 servings for females and 3 servings for males. [11]

Of the eligible students, 17,786 completed the meats and alternatives consumption question, including 10,525 linked students and 7,261 non-linked students. Table 15 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. No significant differences were observed between the linked and non-linked samples, except for the male grade 10 group.

Table 15: Statistical evaluation of linkage rates for students responding to meats and alternatives consumption questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 1	p-value
Female	9	3073	2093	980	0.9	0.330
	10	3047	1977	1070	3.7	0.053
	11	2834	1563	1271	1.5	0.218
Male	9	3027	1812	1215	1.1	0.293
	10	2938	1697	1241	5.3	0.021
	11	2867	1383	1484	0.0	0.837

Table 16 shows the percentage of linked and non-linked students in each meats and alternatives category by grade and sex. Consistent with the results of the Chi-square tests, the results are similar between linked and non-linked students. A higher percentage of females than males meet the recommended number of servings.

Table 16: Comparing distribution of linked and non-linked data for meats and alternatives consumption variables, by grade and guideline categorization

Female				Male			
Grade	**Recommended Servings**	**Linked**	**Non-Linked**	**Grade**	**Recommended Servings**	**Linked**	**Non-Linked**
9	Does Not Meet	38.4	40.2	**9**	Does Not Meet	53.5	55.5
	Meets/Exceeds	61.6	59.8		Meets/Exceeds	46.5	44.5
10	Does Not Meet	36.2	39.7	**10**	Does Not Meet	50.3	54.6
	Meets/Exceeds	63.8	60.3		Meets/Exceeds	49.7	45.4
11	Does Not Meet	35.1	37.3	**11**	Does Not Meet	48.2	47.8
	Meets/Exceeds	64.9	62.7		Meets/Exceeds	51.8	52.2

Milk and Alternatives Consumption

Milk and milk alternatives consumption is assessed according whether students received the Canada Food Guide recommended servings of milks and alternatives in the previous day. Students are categorized according to whether they consumed at least the recommended number of servings, which is 3 servings for both females and males. [11]

Of the eligible students, 17,794 completed the milk and alternatives consumption question, including 10,535 linked students and 7,259 non-linked students. Table 17 shows the total responses by sex and

grade, as well as the p-values from the Chi-square tests. No significant differences were observed between the linked and non-linked samples.

Table 17: Statistical evaluation of linkage rates for students responding to milk and alternatives consumption questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 1	p-value
Female	9	3075	2096	979	0.0	0.891
	10	3051	1979	1072	0.3	0.574
	11	2832	1562	1270	1.7	0.193
Male	9	3026	1816	1210	1.9	0.174
	10	2939	1696	1243	0.5	0.494
	11	2871	1386	1485	0.7	0.413

Table 18 shows the percentage of linked and non-linked students in each milk and alternatives category by grade and sex. Consistent with the results of the Chi-square tests, the results are similar between linked and non-linked students. A higher percentage of males than females meet the recommended number of servings.

Table 18: Comparing distribution of linked and non-linked data for milk and alternatives consumption variables, by grade and guideline categorization

Female				Male			
Grade	**Recommended Servings**	**Linked**	**Non-Linked**	**Grade**	**Recommended Servings**	**Linked**	**Non-Linked**
9	Does Not Meet	57.9	57.6	**9**	Does Not Meet	38.9	41.4
	Meets/Exceeds	42.1	42.4		Meets/Exceeds	61.1	58.6
10	Does Not Meet	62.0	60.9	**10**	Does Not Meet	42.1	43.4
	Meets/Exceeds	38.0	39.1		Meets/Exceeds	57.9	56.6
11	Does Not Meet	61.7	64.1	**11**	Does Not Meet	42.9	44.4
	Meets/Exceeds	38.3	35.9		Meets/Exceeds	57.1	55.6

Physical Activity

The measure of students' physical activity levels showed varying results for significant differences between the linked and non-linked samples. Significant differences were only found for grade 9 females and grade 10 males, but linked samples consistently showed lower percentages of students meeting physical activity guidelines. Due to the consistency of these differences (regardless of significance), we decided the same test should be conducted simply by sex and simply by grade to ascertain if the differences were significant in a larger group break-down. Except for grade 11 students in these larger groups, the differences in the rate of meeting PA guidelines between linked sample and non-linked sample are significant.

Students are dichotomized according to whether or not they meet the Canadian Society for Exercise Physiology guidelines of at least 60 minutes of combined moderate and vigorous physical activity per day. [12] This is based on a student's reported number of minutes of spent doing vigorous and/or moderate physical activity in the last seven days.

Of the eligible students, 17,783 completed the physical activity questions, including 10,486 linked students and 7,297 non-linked students. The following table shows the total responses by sex and grade, as well as

the p-values from the Chi-square tests. Interestingly, the linked samples consistently showed lower percentages of students meeting physical activity guidelines, though this difference was only significant for grade 9 females and grade 10 males.

Table 19: Statistical evaluation of linkage rates for students responding to physical activity questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 1	p-value
Female	9	3051	2079	972	6.4	0.012
	10	3036	1969	1067	2.4	0.119
	11	2825	1551	1274	3.6	0.058
Male	9	3033	1808	1225	0.3	0.602
	10	2943	1693	1250	15.7	0.000
	11	2895	1386	1509	0.1	0.745

Table 20 shows the percentage of linked and non-linked students meeting physical activity guidelines.

Table 20: Comparing distribution of linked and non-linked data for physical activity variables, by grade and guideline categorization

Female				Male			
Grade	**PA Guideline**	**Linked**	**Non-Linked**	**Grade**	**PA Guideline**	**Linked**	**Non-Linked**
9	Yes	43.5	48.7	**9**	Yes	58.9	59.8
	No	56.5	51.3		No	41.1	40.2
10	Yes	39.8	42.9	**10**	Yes	54	61.4
	No	60.2	57.1		No	46	38.6
11	Yes	36.9	40.3	**11**	Yes	55.4	54.8
	No	63.1	59.7		No	44.6	45.2

Sedentary Behaviour

The measure of students' sedentary behaviour levels showed significant differences between the linked and non-linked samples by gender. Significant differences were found for females, with linked students reporting fewer minutes of daily sedentary behaviour, while no significant differences were found for males.

Sedentary behaviour is measured as the total number of minutes per day spent on: watching TV shows or movies, playing computer or video games, talking on the phone, surfing the internet, texting, messaging and emailing. To avoid over-reporting behaviours that are often conducted simultaneously, time spent texting or messaging is excluded from the final results. In addition, responses with total time exceeding 24 hours less time spent sleeping, and responses reporting the maximum time for each behaviour (9 hours and 45 minutes), are treated as erroneous and excluded from the analysis.

Of the eligible students, 17,584 completed the sedentary behaviour questions, including 10,427 linked students 7,157 non-linked students. Table 21 shows the total responses by sex and grade, as well as the p-values from the Satterthwaite t test. The linked samples showed significantly fewer minutes of daily sedentary activity for females in all grades, and no significant difference for males.

Table 21 Statistical evaluation of linkage rates for students responding to sedentary behaviour questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	t Statistic	p-value
Female	9	3034	2075	959	3.61	.0003
	10	2992	1963	1029	3.11	.0019
	11	2815	1549	1266	2.53	.0116
Male	9	2998	1793	1205	0.38	.7060
	10	2903	1673	1230	-0.14	.8865
	11	2842	1374	1468	-0.80	.4214

Table 22 shows the average number of minutes of sedentary behaviour per student per day, in each grade and sex category. Females have fewer daily minutes of sedentary behaviour than males, on average.

Table 22: Comparing mean minutes of sedentary behaviour, by grade and sex

Female			Male		
Grade	**Linked**	**Non-Linked**	**Grade**	**Linked**	**Non-Linked**
9	318	348	**9**	351	354
10	323	348	**10**	366	365
11	312	331	**11**	363	356

Substance Use

Measures related to substance use include students' smoking status, binge drinking status, and marijuana use. Significant differences were observed between the linked and non-linked samples across all substance use measures for all grades and genders.

Tobacco Use

Students' smoking status is derived using two survey questions:

1. Have you ever smoked 100 or more whole cigarettes in your like? (Yes/No)
2. On how many of the last 30 days did you smoke one or more cigarettes? (0, 1, 2-3, 4-5, …)

Students who answer yes to the first question and 1 or greater to the second question are classified as Current Smokers. Students who answer yes to the first question and 0 to the second question are classified as Non-Current Smokers. Students who answered no to the first question are classified as Never Smokers.

All of the 18,280 eligible students completed the smoking questions, including 10,730 linked students and 7,550 non-linked students. Table 23 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. The results showed significant differences between the linked and non-linked samples, with more linked students being Never Smokers and more non-linked students being Current Smokers.

Table 23: Statistical evaluation of linkage rates for students responding to tobacco-use questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 2	p-value
Female	9	3133	2130	1003	55.3	<0.001
	10	3099	2003	1096	33.3	<0.001
	11	2893	1581	1312	39.8	<0.001
Male	9	3137	1858	1279	58.0	<0.001
	10	3045	1736	1309	55.6	<0.001
	11	2973	1422	1551	48.4	<0.001

Table 24 shows the percentage of linked and non-linked students in each category by grade and sex. Across all grades and sexes, a higher percentage of non-linked students were considered Current Smokers and a higher percentage of linked students were considered Never Smokers.

Table 24: Comparing distribution of linked and non-linked data for tobacco-use variables, by grade and smoking status

Female				Male			
Grade	Smoker Status	Linked	Non-Linked	Grade	Smoker Status	Linked	Non-Linked
9	Never	98.9	95.0	9	Never	98.7	93.8
	Non-Current	0.2	0.2		Non-Current	0.1	0.8
	Current	0.9	4.8		Current	1.2	5.6
10	Never	97.5	93.4	10	Never	96.8	90.7
	Non-Current	0.3	1.0		Non-Current	0.6	0.8
	Current	2.2	5.6		Current	2.4	8.5
11	Never	96.3	80.7	11	Never	92.6	84.5
	Non-Current	0.8	1.4		Non-Current	0.6	1.7
	Current	2.9	7.9		Current	6.8	13.8

Binge Drinking

Students' binge drinking is classified according to their answers to the survey question "In the last 12 months, how often did you have 5 drinks of alcohol or more on any one occasion?" Students who answer "I have never done this" are classified as Never Binger Drinkers. Students who answer "I did not have 5 or more drinks on one occasion in the last 12 months" or "Less than once a month" are classified as Non-Current Binger Drinkers. Students who answer "Once a Month" or more frequently are classified as Current Binge Drinkers.

Of the eligible students, 18,203 completed the question on binge drinking, including 10,700 linked students and 7,503 non-linked students. Table 25 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. The results showed significant differences between the linked and non-linked samples, with fewer linked students classified as Current Binge Drinkers.

Table 25: Statistical evaluation of linkage rates for students responding to binge-drinking questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 2	p-value
Female	9	3125	2125	1000	48.4	<0.001
	10	3085	1994	1091	34.0	<0.001
	11	2880	1577	1303	32.1	<0.001
Male	9	3126	1856	1270	25.2	<0.001
	10	3032	1732	1300	34.9	<0.001
	11	2955	1416	1539	31.5	<0.001

Table 26 shows the percentage of linked and non-linked students in each category by grade and sex. Across all grades and sexes, a higher percentage of non-linked students were considered Current Binge Drinkers and a higher percentage of linked students were considered Never Binge Drinkers. The overall percentage of Current and Non-Current Binge Drinkers increases considerably as grade increases, for both females and males.

Table 26: Comparing distribution of linked and non-linked data for binge-drinking variables, by grade and binge-drinking status

Female				Male			
Grade	**Binge Drinker Status**	**Linked**	**Non-Linked**	**Grade**	**Binge Drinker Status**	**Linked**	**Non-Linked**
9	Never	76.4	66.2	**9**	Never	77.8	71.1
	Non-Current	16.4	19.7		Non-Current	15.4	17.4
	Current	7.2	14.1		Current	6.8	11.5
10	Never	59.7	50.0	**10**	Never	60.0	54.2
	Non-Current	24.0	26.2		Non-Current	23.6	20.8
	Current	16.2	23.7		Current	16.4	25.1
11	Never	45.1	37.1	**11**	Never	45.6	38.1
	Non-Current	31.8	30.9		Non-Current	26.6	24.5
	Current	23.1	32.0		Current	27.9	37.4

Marijuana Use

Students' marijuana use is classified according to their answers to the survey question "In the last 12 months, how often did you use marijuana or cannabis?" Students who answer "I have never used marijuana" are classified as Never Users. Students who answer "I have used marijuana but not in the last twelve months" or "Less than once a month" are classified as Non-Current Users. Students who answer "Once a Month" or more frequently are classified as Current Users.

Of the eligible students, 17,869 completed the question on marijuana-use, including 10,568 linked students and 7,301 non-linked students. Table 27 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. The results showed significant differences between the linked and non-linked samples, with fewer linked students classified as Current Users.

Table 27: Statistical evaluation of linkage rates for students responding to marijuana-use questions, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 2	p-value
Female	9	3085	2111	974	102.2	<0.001
	10	3053	1982	1071	73.1	<0.001
	11	2831	1559	1272	68.8	<0.001
Male	9	3048	1818	1230	85.7	<0.001
	10	2963	1708	1255	77.2	<0.001
	11	2889	1390	1499	88.5	<0.001

Table 28 shows the percentage of linked and non-linked students in each category by grade and sex. Across all grades and sexes, a higher percentage of non-linked students were considered Current Users and a higher percentage of linked students were considered Never Users. The overall percentage of Current and Non-Current Users increases considerably as grade increases, with more females considered Never Users.

Table 28: Comparing distribution of linked and non-linked data for marijuana-use variables, by grade and marijuana-use status

Female				Male			
Grade	Marijuana-Use Status	Linked	Non-Linked	Grade	Marijuana-Use Status	Linked	Non-Linked
9	Never	89.2	75.6	9	Never	87.2	75.0
	Non-Current	5.6	10.3		Non-Current	6.2	8.5
	Current	5.2	14.2		Current	6.6	16.4
10	Never	79.3	66.5	10	Never	75.6	62.0
	Non-Current	11.4	14.3		Non-Current	11.1	12.7
	Current	9.3	19.2		Current	13.3	25.3
11	Never	69.0	55.8	11	Never	64.0	47.2
	Non-Current	18.8	21.1		Non-Current	16.6	20.3
	Current	12.3	23.0		Current	19.4	32.4

Bullying and Academics

Bullying and academic-related measures include whether students have been bullied or have bullied others, how often students skip classes, and students' educational expectations. Significant differences were observed between the linked and non-linked samples across all bullying and academic measures for all grades and genders.

Being Bullied

Students are dichotomized according to whether or not they have been bullied by other students in the last 30 days, based on their answers to the question "In the last 30 days, in what ways were you bullied by other students?" Answers are recorded for all eligible students, with missing values recorded as "I have not been bullied in the last 30 days". Table 29 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. The results showed significant differences between the linked and non-linked samples, with fewer linked students reporting being bullied.

Table 29: Statistical evaluation of linkage rates for students responding to questions about being bullied, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 1	p-value
Female	9	3133	2130	1003	32.6	<0.001
	10	3099	2003	1096	8.0	0.0047
	11	2893	1581	1312	24.9	<0.001
Male	9	3137	1858	1279	5.5	0.0193
	10	3045	1736	1309	14.5	<0.001
	11	2973	1422	1551	8.6	0.0033

Table 30 shows the percentage of linked and non-linked students in each category by grade and sex. Across all grades, more females report being bullied than males.

Table 30: Comparing distribution of linked and non-linked data for being-bullied variables, by grade and bullied status

Female				Male			
Grade	**Bullied**	**Linked**	**Non-Linked**	**Grade**	**Bullied**	**Linked**	**Non-Linked**
9	No	75.3	65.5	**9**	No	80.8	77.4
	Yes	24.7	34.5		Yes	19.2	22.6
10	No	76.0	71.4	**10**	No	82.4	76.9
	Yes	24.0	28.6		Yes	17.6	23.1
11	No	79.1	71.0	**11**	No	82.2	77.9
	Yes	20.9	29.0		Yes	17.8	22.1

Bullying Others

Students are dichotomized according to whether or not they have bullied other students in the last 30 days, based on their answers to the question "In the last 30 days, in what ways did you bully other students?" Answers are recorded for all eligible students, with missing values recorded as "I did not bully other students in the last 30 days". Table 30 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. The results showed significant differences between the linked and non-linked samples, with fewer linked students reporting bullying others.

Table 31: Statistical evaluation of linkage rates for students responding to questions about bullying others, by grade and sex

Sex	Grade	Total	Linked	Non-Linked	Chi-square DF = 1	p-value
Female	9	3133	2130	1003	22.7	<0.001
	10	3099	2003	1096	6.8	0.009
	11	2893	1581	1312	9.5	0.002
Male	9	3137	1858	1279	9.7	0.002
	10	3045	1736	1309	10.8	0.001
	11	2973	1422	1551	9.5	0.002

Table 32 shows the percentage of linked and non-linked students in each category by grade and sex.

Table 32: Comparing distribution of linked and non-linked data for bullying-others variables, by grade and bullying status

Female				Male			
Grade	**Bullied Others**	**Linked**	**Non-Linked**	**Grade**	**Bullied Others**	**Linked**	**Non-Linked**
9	No	90.6	84.8	**9**	No	87.5	83.6
	Yes	9.4	15.2		Yes	12.5	16.4
10	No	88.5	85.2	**10**	No	86.1	81.7
	Yes	11.5	14.8		Yes	13.9	18.3
11	No	89.1	85.3	**11**	No	83.8	79.4
	Yes	10.9	14.7		Yes	16.2	20.6

Skipping Class

Students are categorized based on the number of classes they report skipping in the last four weeks. Students who report skipping 0-2 classes are categorized as Rarely/Never, students who report skipping 3-5 classes are categorized as Sometimes, and students who report skipping 6 or more classes are categorized as Often.

Of the eligible students, 17,841 completed the question on skipping class, including 10,539 linked students and 7,302 non-linked students. Table 33 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. The results showed significant differences between the linked and non-linked samples, with fewer linked students skipping more than two classes.

Table 33: Statistical evaluation of linkage rates for students responding to truancy questions, by grade and sex

Sex	**Grade**	**Total**	**Linked**	**Non-Linked**	**Chi-square** DF = 2	**p-value**
Female	9	3081	2104	977	50.1	<0.001
	10	3049	1976	1073	43.4	<0.001
	11	2837	1557	1280	54.1	<0.001
Male	9	3052	1810	1242	54.9	<0.001
	10	2935	1700	1235	53.3	<0.001
	11	2887	1392	1495	30.3	<0.001

Table 34 shows the percentage of linked and non-linked students in each category by grade and sex. Overall, higher grade students report skipping more classes. Across all grades, a higher percentage of linked students reported rarely or never skipping class, and a higher percentage of non-linked students reported skipping class sometimes or often.

Table 34: Comparing distribution of linked and non-linked data for truancy variables, by grade and frequency of skipping class

Female				Male			
Grade	**Skipping Class**	**Linked**	**Non-Linked**	**Grade**	**Skipping Class**	**Linked**	**Non-Linked**
9	Rarely/Never	96.3	90.7	**9**	Rarely/Never	97.6	92.0
	Sometimes	2.8	4.9		Sometimes	1.5	3.9
	Often	1.0	4.4		Often	0.8	4.1
10	Rarely/Never	95.0	88.9	**10**	Rarely/Never	94.9	87.9
	Sometimes	3.8	7.2		Sometimes	3.6	6.6
	Often	1.2	3.9		Often	1.5	5.6
11	Rarely/Never	91.7	83.8	**11**	Rarely/Never	90.9	84.2
	Sometimes	6.6	9.5		Sometimes	5.0	8.6
	Often	1.8	6.6		Often	4.1	7.2

Educational Expectations

Educational expectation is defined as the highest level of education students expect they will achieve. Expected education levels are categorized as High School Diploma or Less, College/Trade/Bachelor's Degree, Master's Degree or Higher, and Unsure.

Of the eligible students, 17,724 completed the question on educational expectation, including 10,478 linked students and 7,246 non-linked students. Table 35 shows the total responses by sex and grade, as well as the p-values from the Chi-square tests. The results showed significant differences between the linked and non-linked samples.

Table 35: Statistical evaluation of linkage rates for students responding to educational expectations questions, by grade and sex

Sex	**Grade**	**Total**	**Linked**	**Non-Linked**	**Chi-square** DF = 3	**p-value**
Female	9	3037	2073	964	18.2	<0.001
	10	3025	1962	1063	18.3	<0.001
	11	2832	1557	1275	33.3	<0.001
Male	9	3035	1808	1227	36.8	<0.001
	10	2919	1692	1227	34.9	<0.001
	11	2876	1386	1490	11.9	0.008

Table 36 shows the percentage of linked and non-linked students in each category by grade and sex. Generally, linked students have higher educational expectations, and fewer linked students expect to achieve only a high school diploma or less. Students in lower grades more often report they are unsure. Females report more often than males that they expect to achieve a master's degree or higher.

Table 36: Comparing distribution of linked and non-linked data for educational expectations variables, by grade and expected education achievement

Female				Male			
Grade	**Education Level**	**Linked**	**Non-Linked**	**Grade**	**Education Level**	**Linked**	**Non-Linked**
9	Unsure	30.8	27.3	**9**	Unsure	24.2	24.4
	High School or Less	6.5	9.8		High School or Less	5.4	11.3
	College/Bachelor	32.0	35.9		College/Bachelor	46.1	41.6
	Master or Higher	30.7	27.1		Master or Higher	24.3	22.7
10	Unsure	17.5	18.7	**10**	Unsure	14.8	15.8
	High School or Less	3.1	6.0		High School or Less	4.3	9.2
	College/Bachelor	42.1	42.2		College/Bachelor	52.0	51.1
	Master or Higher	37.3	33.0		Master or Higher	29.0	23.9
11	Unsure	12.0	13.3	**11**	Unsure	12.8	11.5
	High School or Less	2.8	5.6		High School or Less	4.7	7.7
	College/Bachelor	47.8	52.4		College/Bachelor	58.3	57.4
	Master or Higher	37.3	28.6		Master or Higher	24.2	23.4

Discussion

Despite decades of primary prevention efforts being targeted at improving the health of Canadian youth, those efforts in many domains seemed to be failing as evident by the current risk behavioural profile of Canadian youth [2]. Available evidence suggests that one of the major challenges inhibiting successful population prevention among youth in Canada was that no one was systematically collecting the necessary data to inform and evaluate prevention activities in a comprehensive or ongoing fashion. As such, the value of a longitudinal dataset such as COMPASS cannot be overstated. It is the ability to track a defined cohort of students and the schools they attend over time that allows COMPASS researchers to effectively evaluate the efficacy of natural experiments (programs and polices implemented in schools to improve student health), in ways that cross-sectional surveys cannot. Implementing policies and practises that have not been evaluated for effectiveness can potentially be a waste of time and valuable resources, while providing little or no improvement to student health (and at worst, can actually have a detrimental effect). Knowing what programs and policies work best, the populations for whom they work best, and the environments in which they work best, is paramount to implementing efficient and effective policies and practices that will have lasting impacts on youth health.

If the strength of a longitudinal dataset is the linkage of student data at multiple time points of a study, then the failure to link *all* student data is its limitation. While a significant portion of student data can be linked from one year to the next (~80% success rate), there is a smaller—but still significant—portion of student data that are not linked each year. As this report has illustrated, it does not appear to be a random collection of students whose data cannot be linked over time, but rather students who are more likely to drink, smoke, use marijuana, and be involved with bullying. Furthermore, those same students report skipping classes significantly more and are, therefore, more likely to be absent on a data collection day. If it is surmised based on these analyses that students who exhibit similar behaviours are more likely to skip school, then there should be concern that a specific subsample of a school population will be absent at any given time, as it introduces a level of bias to the data. This has two major implications:

First, researchers must account for differences in the linked vs non-linked data. If researchers wish to measure changes in eating habits, BMI scores, and other obesity-related outcomes over time, they can use the linked data without concern for in-school representativeness, knowing that there are no significant

differences between linked and non-linked student data. When measuring changes in substance-use behaviours, (tobacco-, alcohol-, and marijuana-use), however, researchers must account for the fact that a significant portion of students who report using these substances will not be included in the linked dataset (for example, smoking rates amongst linked students in grades 9 to 11 are so low that tracking any sort of behavioural change in that group over time is near-impossible). As such, measureable changes in behaviours will be more difficult to assess over time in these cases. Likewise, bullying, academic ambition, and—to a lesser degree—physical activity and sedentary behaviour data must also be used with some caution.

Second, knowing that on any given data collection day a larger proportion of substance-users than non-users will be absent from school—and will, therefore, not be included in any resulting datasets—suggests that it is likely that existing cross-sectional surveys (the current norm in surveillance research) are systematically under-reporting youth substance-use data. This is a further illustration of the value of the COMPASS system's design: Using passive permission protocols to minimize in-school sample bias [13], and being able to identify data bias by virtue of being longitudinal (even if it remains difficult to control for that bias), the COMPASS system is able to mitigate potentially serious shortcomings in the collection of student data in a way that cross-sectional studies cannot (especially those that utilize active consent protocols), and thus provide researchers with a more realistic idea of what student health behaviours actually are.

While this technical report has been created to illustrate which data show bias or not (so as to be a guide for data-users), there is additional work to be done to try to ascertain *why* these data are shown to be biased. It is interesting to note that while measures such as substance-use are both significantly and consistently biased, other measures are either consistent but not significantly biased (such as physical activity), or are significant but not consistently biased. When this linkage method is applied to the third wave of data, we will perhaps be better able to explain failed linkage based on how many students we are able to link in 2 out of 3 years, and how we may be able to infer for missing data.

References

1. Leatherdale ST, Brown KS, Carson, V, et al: *The COMPASS study: a longitudinal hierarchical research platform for evaluating natural experiments related to changes in school-level programs, policies and built environment resources*. BMC Public Health. 2014,14,331. doi:10.1186/1471-2458-14-331
2. Leatherdale ST, Burkhalter R: *The substance use profile of Canadian youth: exploring the prevalence of alcohol, drug and tobacco use by gender and grade*. Addict Behav 2012, 37:318-322.
3. Leatherdale ST, Manske S, Faulkner G, Arbour K, Bredin C: *A multi-level examination of school programs, policies and resources associated with physical activity among elementary school youth in the PLAY-ON study*. Int J Behav Nutr Phys Act 2010, 25;6. doi: 10.1186/1479 -5868-7-6.
4. Leatherdale ST, McDonald PW, Cameron R, Brown KS: *A multi-level analysis examining the relationship between social influences for smoking and smoking onset*. Am J Health Behav 2005, 29:520-530.
5. Leatherdale ST, Papadakis S: *A multi-level examination of the association between older social models in the school environment and overweight and obesity among younger students*. J Youth Adolesc 2011, 40:361 - 372.
6. Bredin C, Leatherdale ST. *Methods for linking COMPASS student-level data over time*. COMPASS Technical Report Series. 2013;1(2). Waterloo, Ontario: University of Waterloo. Available at: https://uwaterloo.ca/compass-system/publications#technical
7. Kearney K, Hopkins RH, Mauss AL and Weisheit RA: *Self-Generated Identification Codes for Anonymous Collection of Longitudinal Questionnaire Data*. The Public Opinion Quarterly, Vol. 48,No. 1 (Spring, 1984), pp. 370-378.
8. Statistics Canada. *Labour Force Survey 2012*. Ottawa: Statistics Canada, 2012.
9. Leatherdale ST, McDonald PW. *Are the Recommended Taxonomies for the Stages of Youth Smoking Onset Consistent with Youth's Perceptions of Their Smoking Status?* Canadian Journal of Public Health; Jul/Aug 2006; 97, 4; Research Library pg. 316
10. WHO. *Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee.* WHO Technical Report Series 854. Geneva: World Health Organization, 1995.
11. Health Canada: *Eating Well with Canada's Food Guide*. Minister of Health; 2011. http://www.hc-sc.gc.ca/fn-an/alt_formats/hpfb-dgpsa/pdf/food-guide-aliment/print_eatwell_bienmang-eng.pdf [accessed July 2015]
12. Canadian Society for Exercise Physiology. Canadian Physical Activity Guidelines for Youth – 12 to 17 years. 2013. http://www.csep.ca/CMFiles/Guidelines/CSEP-InfoSheets-youth-ENG.pdf.
13. Thompson-Haile A, Bredin C, Leatherdale ST. *Rationale for using an Active-Information Passive-Consent Permission Protocol in COMPASS*. COMPASS Technical Report Series. 2013;1(6). Waterloo, Ontario: University of Waterloo. Available at: www.compass.uwaterloo.ca

Appendix

Appendix A: Questions used to create self-generated code for the purpose of tracking data over time

Compass

- **This is NOT a test.** All of your answers will be kept **confidential**. No one, not even your parents or teachers, will ever know what you answered. So, please be honest when you answer the questions.
- Mark only **one option per question** unless the instructions tell you to do something else.
- Choose the option that is the **closest** to what you think/feel is true for you.

Please, use a pencil to complete this questionnaire

Please mark all your answers with full, dark marks like this:

START HERE

Please read each sentence below carefully. Write the correct letter, number, or word on the line and then fill in the corresponding circle.

The first letter of your **middle** name (if you have more than one middle name use your first middle name; if you don't have a middle name use "Z"):___	The name of the month in which you were born:_______	The **last** letter of your full **last** name:___	The **second** letter of your full **first** name:___	The **first** initial of your **mother's** first name (think about the mother you see the most):___
A J S	1 January	A J S	A J S	A J S
B K T	2 February	B K T	B K T	B K T
C L U	3 March	C L U	C L U	C L U
D M V	4 April	D M V	D M V	D M V
E N W	5 May	E N W	E N W	E N W
F O X	6 June	F O X	F O X	F O X
G P Y	7 July	G P Y	G P Y	G P Y
H Q Z	8 August	H Q Z	H Q Z	H Q Z
I R	9 September	I R	I R	I R
	10 October			
	11 November			
	12 December			

[serial]

University of Waterloo
200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1
Telephone: (519) 888-4567
uwaterloo.ca/compass-system