

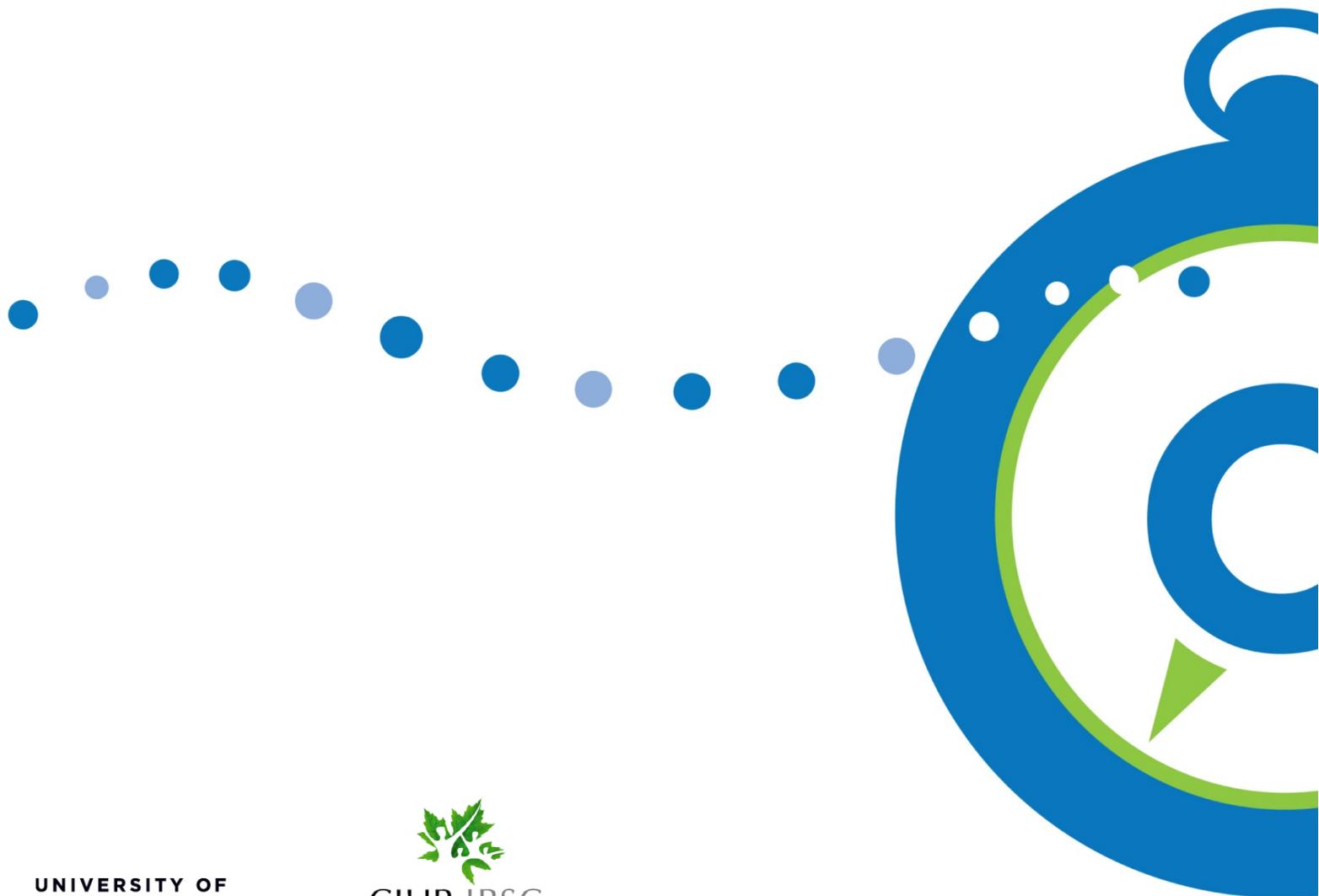


*Cohort study evaluating how changes in school programs, policies, and resources impact youth health behaviours*

## Methods for linking COMPASS student-level data over time

COMPASS Technical Report Series, Volume 1, Issue 2

April 2013



UNIVERSITY OF  
**WATERLOO**

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

University of Waterloo

Waterloo, Ontario

April 2013

©COMPASS

[www.compass.uwaterloo.ca](http://www.compass.uwaterloo.ca)



## ACKNOWLEDGEMENTS

Chad Bredin, BA<sup>1</sup>  
Scott Leatherdale, PhD<sup>2</sup>

1 - Propel Centre for Population Health Impact, University of Waterloo, Waterloo, ON Canada.

2 - School of Public Health and Health Systems, 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 ST. 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 ST. Leatherdale).

### Suggested citation:

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: [www.compass.uwaterloo.ca](http://www.compass.uwaterloo.ca)

### Contact:

COMPASS Research Team  
University of Waterloo  
200 University Ave West, BMH 1038  
Waterloo, ON Canada N2L 3G1  
[compass@uwaterloo.ca](mailto:compass@uwaterloo.ca)

## Table of Contents

Acknowledgements.....	ii
Introduction: .....	1
Methods.....	1
Testing our Self-Generated Code.....	2
Table 1 Examining the number of non-matches for each data linkage question in the validation study between T1 and T2 (n=204) .....	2
Refining our Self-Generated Code after the Validation Study.....	3
Discussion.....	3
References .....	4
<b>Appendix A</b> .....	5
<b>Appendix B</b> .....	6

## Introduction:

COMPASS is a longitudinal study (starting 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. 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 [1-4]. 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: [www.compass.uwaterloo.ca](http://www.compass.uwaterloo.ca)).

Given that COMPASS is a longitudinal study, we need to be able to follow our cohort of schools and students over time. While tracking the participating schools over time is not difficult, it can be more challenging to track students within those schools longitudinally. As such, within COMPASS, we needed to develop a simple yet robust method for tracking students over time. This procedure had to be done in a manner that protects student confidentiality, places minimal burden on respondents, is inexpensive and simple for COMPASS staff, and is effective for actually tracking students year to year.

This technical report provides details on the development and testing of the methodology used in COMPASS to link individual student-level data over time.

## Methods

As part of the COMPASS student-level data collection, all eligible students complete the COMPASS student questionnaire ( $C_q$ ) once annually, during class time. The  $C_q$  provides the COMPASS team with the individual-specific yet anonymous data required for linking the individual student data longitudinally. Consistent with previous research [5], the cover page of the  $C_q$  contains measures required to create a unique self-generated code for each respondent in a school based on a series of questions; the responses to these measures among each individual respondent do not change over time (i.e., the responses from each participating student in a school generates a unique identifier for him/her, but each participating student also has the same answers to his/her measures every year). This ensures the anonymity of the survey participants while still allowing COMPASS researchers to link each student's unique identifier data over multiple years.

## Testing our Self-Generated Code

Based on evidence from existing research [5], the COMPASS team created five questions that could be used to develop a unique self-generated code. The five questions developed were:

1. 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" ): \_\_\_\_
2. The first letter of the month in which you were born: \_\_\_\_
3. The last letter of your full **first** name: \_\_\_\_
4. The second letter of your **last** name: \_\_\_\_
5. The number of older brothers you have (alive and deceased): \_\_\_\_

For each of these questions, the answers (which should not change over the course of the study), can be combined to create a unique identifier for each student within a participating school. We then added these 5 questions to the cover sheet of the draft C<sub>q</sub> that was going to be used in the COMPASS validation study (the pilot study designed to test the reliability and validity of the C<sub>q</sub> core measures). This would allow us to test our ability to link the student data over time within a school using these five questions (refer to Appendix A for a copy of the C<sub>q</sub> cover sheet used in the COMPASS validation study).

As part of the COMPASS validation study [6], data were collected using a convenience sample of 204 students in grade 9 and 10 from four schools in Southwestern Ontario (Canada). Participants completed the C<sub>q</sub> during class time (~30 min) on two separate occasions. At time 1 (T1), staff administered the C<sub>q</sub> in classrooms using a common protocol and standardized instructions. After one week, the C<sub>q</sub> was re-administered to the same students using the same protocol (T2). We then examined how well we could match the 204 unique T1 identifiers to the T2 data using these five questions.

Among the 204 respondents with unique identifiers at T1, there was a perfect match for 65% (n=132) of them at T2 on all five measures; 31% (n=64) were matched on four out of five answers matched and the remaining 4% (n=8) were matched on three or less. As shown in Table 1, the majority of non-matches resulted from question 3 (18%), question 4 (9%), and question 5 (7%).

**Table 1 Examining the number of non-matches for each data linkage question in the validation study between T1 and T2 (n=204)**

Question Text	Number of non-matches (T1 to T2)
The first letter of your <b>middle</b> name (if you have more than one middle name use your first middle name, if you don't have a middle name use "Z" ): ____	9
The first letter of the month in which you were born: ____	2
The last letter of your full <b>first</b> name: ____	37
The second letter of your <b>last</b> name: ____	19
The number of <u>older</u> brothers you have (alive and deceased): ____	15

Since the matching rate of 65% was lower than we would consider ideal, we also examined the benefit of using a sixth question to assist with this linkage among the 35% or respondents where the linkage was problematic (i.e., the C<sub>q</sub> core question measuring the sex of the respondent: "Are you male or female?"). When we include the response of the respondents question about their sex to the T1 and T2 linkage, we ended up with an *overall* T1 and T2 match rate of 90%.

## Refining our Self-Generated Code after the Validation Study

In order to ensure robust data linkage over time in COMPASS, we decided it would be best to make some edits to the three questions that suffered the worst individual match rates in the validation study (Questions 3, 4 and 5).

Questions 3 and 4 were edited to reduce confusion in what the question was asking, especially among students with hyphenated names. As such, we made the following changes:

### *Question 3*

-changed from “The last letter of your full **first** name” to “The last letter of your full **last** name”.

### *Question 4*

-changed from “The second letter of your **last** name” to “The second letter of your full **first** name”.

Given that there was limited variability in the responses provided for Question 5 based on what we were measuring (i.e., the numerical responses only ranged from 0 to 3), we decided to use a different measure derived from existing evidence [5] that would provide a consistent response within individual students over time but also provide more variability between individual respondents (i.e., 26 different responses if we use a letter from the alphabet). As such, we made the following change:

### *Question 5*

-changed from “The number of older brothers you have (alive and deceased)” to “The first initial of your mother's **first** name (think about the mother you see the most)”.

In order to ensure as much variability between respondents as possible, we also decided to edit the response options for Question 2, “The first letter of the month in which you were born”. Instead of just using the eight different available letters pertaining to the first letters of the months of the year (J, F, M, A, S, O, N, D), we made this a numerical scale corresponding to the 12 different months of the year (January =1, February=2, March=3, etc.) and the wording of the question was changed to “The name of the month in which you were born”.

As such, the five new questions used to develop a unique self-generated code for COMPASS are:

1. 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" ): \_\_\_\_
2. The name of the month in which you were born: \_\_\_\_
3. The last letter of your full **last** name: \_\_\_\_
4. The second letter of your full **first** name: \_\_\_\_
5. The first initial of your mother's **first** name (think about the mother you see the most): \_\_\_\_

Refer to Appendix B for a copy of the C<sub>q</sub> cover sheet used in the COMPASS baseline data collection.

## Discussion

Given the short timeline between the validation study and the baseline data collection for COMPASS, we were not able to re-evaluate the impact that these changes to our tracking measures would have on

improving our T1 and T2 match rate of 90% from the validation study. Given that there are also C<sub>q</sub> measures about ethnicity, and grade that could potentially be used for improving linkage rates, we are confident in our ability to robustly track individual students within schools over time in COMPASS. As mentioned, our ability to match students over time is assisted in this study since we benefit from the ability to do our linkages within schools rather than across school (i.e., smaller units for the linkages make the linkages easier to perform and results in less potential risk of duplicate unique identifiers among students). To assist with the accuracy of our within school linkages, we also added a question to the C<sub>q</sub> following the validation study in which we ask “*Did you attend this school last year?*” (Yes, I attended the same school last year / No, I was at another school last year). If the answer is ‘no’, we know that there are no matching data from previous years for that student within that school. This will help us deal with the issue of students entering and exiting the study.

In a longitudinal study such as COMPASS, maintaining participants’ trust in the confidentiality of a questionnaire must be balanced with the ability to link data over multiple years. By utilizing simple yet confidential self-generated codes that are based on measures that do not change within individual student respondents over time, we feel that COMPASS has an effective and robust means for tracking individual study participants within schools over time.

## References

1. 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.
2. 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.
3. 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.
4. 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.
5. 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.
6. Leatherdale ST, Laxer RE: *Reliability and validity of the weight status and dietary intake measures in the COMPASS questionnaire: are the self-reported measures of body mass index (BMI) and Canada’s food guide servings robust?*. *Int J Behav Nutr Phys Act* 2013 10:42.



## Appendix B

### COMPASS student questionnaire cover sheet used in the baseline data collection (2012-13)



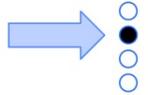
- This is **NOT** a test. All of your answers will be kept **confidential**. No one, not even your parents or teachers, will ever know what you answered. So, please be honest when you answer the questions.
- Mark only **one option per question** unless the instructions tell you to do something else.
- 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 <b>middle</b> 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 <b>last</b> letter of your full <b>last</b> name: ____	The <b>second</b> letter of your full <b>first</b> name: ____	The <b>first</b> initial of your <b>mother's</b> first name (think about the mother you see the most):____																																																																																																								
<table style="width: 100%; border: none;"> <tr><td>(A)</td><td>(J)</td><td>(S)</td></tr> <tr><td>(B)</td><td>(K)</td><td>(T)</td></tr> <tr><td>(C)</td><td>(L)</td><td>(U)</td></tr> <tr><td>(D)</td><td>(M)</td><td>(V)</td></tr> <tr><td>(E)</td><td>(N)</td><td>(W)</td></tr> <tr><td>(F)</td><td>(O)</td><td>(X)</td></tr> <tr><td>(G)</td><td>(P)</td><td>(Y)</td></tr> <tr><td>(H)</td><td>(Q)</td><td>(Z)</td></tr> <tr><td>(I)</td><td>(R)</td><td></td></tr> </table>	(A)	(J)	(S)	(B)	(K)	(T)	(C)	(L)	(U)	(D)	(M)	(V)	(E)	(N)	(W)	(F)	(O)	(X)	(G)	(P)	(Y)	(H)	(Q)	(Z)	(I)	(R)		<table style="width: 100%; border: none;"> <tr><td>①</td><td>January</td></tr> <tr><td>②</td><td>February</td></tr> <tr><td>③</td><td>March</td></tr> <tr><td>④</td><td>April</td></tr> <tr><td>⑤</td><td>May</td></tr> <tr><td>⑥</td><td>June</td></tr> <tr><td>⑦</td><td>July</td></tr> <tr><td>⑧</td><td>August</td></tr> <tr><td>⑨</td><td>September</td></tr> <tr><td>⑩</td><td>October</td></tr> <tr><td>⑪</td><td>November</td></tr> <tr><td>⑫</td><td>December</td></tr> </table>	①	January	②	February	③	March	④	April	⑤	May	⑥	June	⑦	July	⑧	August	⑨	September	⑩	October	⑪	November	⑫	December	<table style="width: 100%; border: none;"> <tr><td>(A)</td><td>(J)</td><td>(S)</td></tr> <tr><td>(B)</td><td>(K)</td><td>(T)</td></tr> <tr><td>(C)</td><td>(L)</td><td>(U)</td></tr> <tr><td>(D)</td><td>(M)</td><td>(V)</td></tr> <tr><td>(E)</td><td>(N)</td><td>(W)</td></tr> <tr><td>(F)</td><td>(O)</td><td>(X)</td></tr> <tr><td>(G)</td><td>(P)</td><td>(Y)</td></tr> <tr><td>(H)</td><td>(Q)</td><td>(Z)</td></tr> <tr><td>(I)</td><td>(R)</td><td></td></tr> </table>	(A)	(J)	(S)	(B)	(K)	(T)	(C)	(L)	(U)	(D)	(M)	(V)	(E)	(N)	(W)	(F)	(O)	(X)	(G)	(P)	(Y)	(H)	(Q)	(Z)	(I)	(R)		<table style="width: 100%; border: none;"> <tr><td>(A)</td><td>(J)</td><td>(S)</td></tr> <tr><td>(B)</td><td>(K)</td><td>(T)</td></tr> <tr><td>(C)</td><td>(L)</td><td>(U)</td></tr> <tr><td>(D)</td><td>(M)</td><td>(V)</td></tr> <tr><td>(E)</td><td>(N)</td><td>(W)</td></tr> <tr><td>(F)</td><td>(O)</td><td>(X)</td></tr> <tr><td>(G)</td><td>(P)</td><td>(Y)</td></tr> <tr><td>(H)</td><td>(Q)</td><td>(Z)</td></tr> <tr><td>(I)</td><td>(R)</td><td></td></tr> </table>	(A)	(J)	(S)	(B)	(K)	(T)	(C)	(L)	(U)	(D)	(M)	(V)	(E)	(N)	(W)	(F)	(O)	(X)	(G)	(P)	(Y)	(H)	(Q)	(Z)	(I)	(R)	
(A)	(J)	(S)																																																																																																										
(B)	(K)	(T)																																																																																																										
(C)	(L)	(U)																																																																																																										
(D)	(M)	(V)																																																																																																										
(E)	(N)	(W)																																																																																																										
(F)	(O)	(X)																																																																																																										
(G)	(P)	(Y)																																																																																																										
(H)	(Q)	(Z)																																																																																																										
(I)	(R)																																																																																																											
①	January																																																																																																											
②	February																																																																																																											
③	March																																																																																																											
④	April																																																																																																											
⑤	May																																																																																																											
⑥	June																																																																																																											
⑦	July																																																																																																											
⑧	August																																																																																																											
⑨	September																																																																																																											
⑩	October																																																																																																											
⑪	November																																																																																																											
⑫	December																																																																																																											
(A)	(J)	(S)																																																																																																										
(B)	(K)	(T)																																																																																																										
(C)	(L)	(U)																																																																																																										
(D)	(M)	(V)																																																																																																										
(E)	(N)	(W)																																																																																																										
(F)	(O)	(X)																																																																																																										
(G)	(P)	(Y)																																																																																																										
(H)	(Q)	(Z)																																																																																																										
(I)	(R)																																																																																																											
(A)	(J)	(S)																																																																																																										
(B)	(K)	(T)																																																																																																										
(C)	(L)	(U)																																																																																																										
(D)	(M)	(V)																																																																																																										
(E)	(N)	(W)																																																																																																										
(F)	(O)	(X)																																																																																																										
(G)	(P)	(Y)																																																																																																										
(H)	(Q)	(Z)																																																																																																										
(I)	(R)																																																																																																											

© COMPASS 2012



[serial]

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

