

Use of Self-Generated Identification Codes for Data Linkages in Highly Clustered Samples Involving Youth Populations

Authors:

Chidinma Nwankwo, Tiffany Waits, Avery Hennigar, Lourdes Fernandez, Amy Harris, William Waddell, Sam Bonelli, and Elizabeth Cavadel

Affiliation: Mathematica



Presenters



Tiffany Waits

(She/Her)

Principal Survey Researcher
Human Services, Mathematica



**Chidinma (Chichi)
Nwankwo**

(She/Her)

Researcher Analyst
Human Services, Mathematica





About Mathematica

Mathematica is an employee-owned, mission-driven company with trusted **expertise in data science, social science, technology, and advisory services across health, human services, and international development.**

We collaborate closely with federal agencies, state and local governments, foundations, businesses, universities, and professional associations to address pressing social issues, improve programs and policy, and drive equitable outcomes.



Study Challenge:

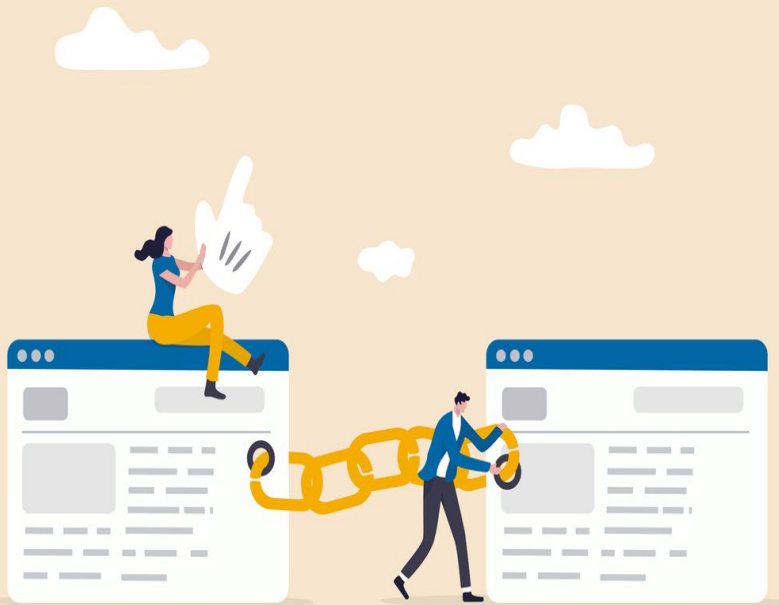
How do we link pre and post surveys at the individual level without using personally identifiable information?

- Need to link pre and post surveys to measure change over time
- No PII to allow for linking surveys
- Youth in schools and out-of-school settings
- Multi-mode survey based on program preference (paper or web)
- Youth highly clustered (grouped) into classes

Solution:

Ask a set of questions about respondent's characteristics on pre and post surveys to use for linking

- / Self-Generated Identification Codes (SGICs)
- / Researchers have used questions about respondents' characteristics to match data (Kristijansson et al., 2013) and to link responses between different time points (Calatrava et al., 2022).





SGIC questions were selected from literature

Common examples include variations on the following:

- First name (first and third letters)
- Last name (third letter)
- Date of birth (second digit of the day)
- Mother's first name (second and fourth letters)
- Father's first name (third letter)
- Paternal grandmother's first name (second letter)
- Color of respondent's eyes (choosing the initial of the color from a range of options).

Study Application:

Test SGIC method for linking in highly clustered settings with youth population

- / Apply method to youth population**
- / In highly clustered settings of classrooms**
- / Test 8 different questions across 2 survey versions**
 - 1 question was the same on both versions; a second question was the same with different response options
- / Each student received the same version survey and exact same questions at pre and post**



2 different sets of SGIC questions tested on 2 forms

1. You will be asked to take this survey twice, once before the program begins and once after the program ends. In order to protect your privacy, we will not collect any identifying information on this survey. Instead, your answers to the following 5 questions will be used to connect your pre-program survey responses to your post-program survey responses without us needing your name. It is important that the answers to these 5 questions stay the same between surveys.

/ SGIC Question Set 1

a. What is the first letter of your first name?

| | | | | |
|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 1 <input type="radio"/> A | 7 <input type="radio"/> G | 13 <input type="radio"/> M | 19 <input type="radio"/> S | 25 <input type="radio"/> Y |
| 2 <input type="radio"/> B | 8 <input type="radio"/> H | 14 <input type="radio"/> N | 20 <input type="radio"/> T | 26 <input type="radio"/> Z |
| 3 <input type="radio"/> C | 9 <input type="radio"/> I | 15 <input type="radio"/> O | 21 <input type="radio"/> U | |
| 4 <input type="radio"/> D | 10 <input type="radio"/> J | 16 <input type="radio"/> P | 22 <input type="radio"/> V | |
| 5 <input type="radio"/> E | 11 <input type="radio"/> K | 17 <input type="radio"/> Q | 23 <input type="radio"/> W | |
| 6 <input type="radio"/> F | 12 <input type="radio"/> L | 18 <input type="radio"/> R | 24 <input type="radio"/> X | |

d. What is the first letter of the name of the street where you live?

| | | | | |
|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 1 <input type="radio"/> A | 7 <input type="radio"/> G | 13 <input type="radio"/> M | 19 <input type="radio"/> S | 25 <input type="radio"/> Y |
| 2 <input type="radio"/> B | 8 <input type="radio"/> H | 14 <input type="radio"/> N | 20 <input type="radio"/> T | 26 <input type="radio"/> Z |
| 3 <input type="radio"/> C | 9 <input type="radio"/> I | 15 <input type="radio"/> O | 21 <input type="radio"/> U | 27 <input type="radio"/> |
| 4 <input type="radio"/> D | 10 <input type="radio"/> J | 16 <input type="radio"/> P | 22 <input type="radio"/> V | Prefer not to say/unknown |
| 5 <input type="radio"/> E | 11 <input type="radio"/> K | 17 <input type="radio"/> Q | 23 <input type="radio"/> W | |
| 6 <input type="radio"/> F | 12 <input type="radio"/> L | 18 <input type="radio"/> R | 24 <input type="radio"/> X | |

b. On what day of the month were you born?

| | | | | |
|---------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| 1 <input type="radio"/> 1 | 8 <input type="radio"/> 8 | 15 <input type="radio"/> 15 | 22 <input type="radio"/> 22 | 29 <input type="radio"/> 29 |
| 2 <input type="radio"/> 2 | 9 <input type="radio"/> 9 | 16 <input type="radio"/> 16 | 23 <input type="radio"/> 23 | 30 <input type="radio"/> 30 |
| 3 <input type="radio"/> 3 | 10 <input type="radio"/> 10 | 17 <input type="radio"/> 17 | 24 <input type="radio"/> 24 | 31 <input type="radio"/> 31 |
| 4 <input type="radio"/> 4 | 11 <input type="radio"/> 11 | 18 <input type="radio"/> 18 | 25 <input type="radio"/> 25 | |
| 5 <input type="radio"/> 5 | 12 <input type="radio"/> 12 | 19 <input type="radio"/> 19 | 26 <input type="radio"/> 26 | |
| 6 <input type="radio"/> 6 | 13 <input type="radio"/> 13 | 20 <input type="radio"/> 20 | 27 <input type="radio"/> 27 | |
| 7 <input type="radio"/> 7 | 14 <input type="radio"/> 14 | 21 <input type="radio"/> 21 | 28 <input type="radio"/> 28 | |

e. What is the first letter of your mother's first name?

| | | | | |
|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 1 <input type="radio"/> A | 7 <input type="radio"/> G | 13 <input type="radio"/> M | 19 <input type="radio"/> S | 25 <input type="radio"/> Y |
| 2 <input type="radio"/> B | 8 <input type="radio"/> H | 14 <input type="radio"/> N | 20 <input type="radio"/> T | 26 <input type="radio"/> Z |
| 3 <input type="radio"/> C | 9 <input type="radio"/> I | 15 <input type="radio"/> O | 21 <input type="radio"/> U | |
| 4 <input type="radio"/> D | 10 <input type="radio"/> J | 16 <input type="radio"/> P | 22 <input type="radio"/> V | |
| 5 <input type="radio"/> E | 11 <input type="radio"/> K | 17 <input type="radio"/> Q | 23 <input type="radio"/> W | |
| 6 <input type="radio"/> F | 12 <input type="radio"/> L | 18 <input type="radio"/> R | 24 <input type="radio"/> X | |

c. What is the second letter of your last name?

| | | | | |
|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 1 <input type="radio"/> A | 7 <input type="radio"/> G | 13 <input type="radio"/> M | 19 <input type="radio"/> S | 25 <input type="radio"/> Y |
| 2 <input type="radio"/> B | 8 <input type="radio"/> H | 14 <input type="radio"/> N | 20 <input type="radio"/> T | 26 <input type="radio"/> Z |
| 3 <input type="radio"/> C | 9 <input type="radio"/> I | 15 <input type="radio"/> O | 21 <input type="radio"/> U | |
| 4 <input type="radio"/> D | 10 <input type="radio"/> J | 16 <input type="radio"/> P | 22 <input type="radio"/> V | |
| 5 <input type="radio"/> E | 11 <input type="radio"/> K | 17 <input type="radio"/> Q | 23 <input type="radio"/> W | |
| 6 <input type="radio"/> F | 12 <input type="radio"/> L | 18 <input type="radio"/> R | 24 <input type="radio"/> X | |

/ SGIC Question Set 2

a. What is the first letter of your first name?

| | | | | |
|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 1 <input type="radio"/> A | 7 <input type="radio"/> G | 13 <input type="radio"/> M | 19 <input type="radio"/> S | 25 <input type="radio"/> Y |
| 2 <input type="radio"/> B | 8 <input type="radio"/> H | 14 <input type="radio"/> N | 20 <input type="radio"/> T | 26 <input type="radio"/> Z |
| 3 <input type="radio"/> C | 9 <input type="radio"/> I | 15 <input type="radio"/> O | 21 <input type="radio"/> U | |
| 4 <input type="radio"/> D | 10 <input type="radio"/> J | 16 <input type="radio"/> P | 22 <input type="radio"/> V | |
| 5 <input type="radio"/> E | 11 <input type="radio"/> K | 17 <input type="radio"/> Q | 23 <input type="radio"/> W | |
| 6 <input type="radio"/> F | 12 <input type="radio"/> L | 18 <input type="radio"/> R | 24 <input type="radio"/> X | |

d. In what month were you born?

| | |
|----------------------------------|-----------------------------------|
| 1 <input type="radio"/> January | 7 <input type="radio"/> July |
| 2 <input type="radio"/> February | 8 <input type="radio"/> August |
| 3 <input type="radio"/> March | 9 <input type="radio"/> September |
| 4 <input type="radio"/> April | 10 <input type="radio"/> October |
| 5 <input type="radio"/> May | 11 <input type="radio"/> November |
| 6 <input type="radio"/> June | 12 <input type="radio"/> December |

b. What is the first letter of your middle name?

| | | | | |
|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 1 <input type="radio"/> A | 7 <input type="radio"/> G | 13 <input type="radio"/> M | 19 <input type="radio"/> S | 25 <input type="radio"/> Y |
| 2 <input type="radio"/> B | 8 <input type="radio"/> H | 14 <input type="radio"/> N | 20 <input type="radio"/> T | 26 <input type="radio"/> Z |
| 3 <input type="radio"/> C | 9 <input type="radio"/> I | 15 <input type="radio"/> O | 21 <input type="radio"/> U | 27 <input type="radio"/> |
| 4 <input type="radio"/> D | 10 <input type="radio"/> J | 16 <input type="radio"/> P | 22 <input type="radio"/> V | No middle name |
| 5 <input type="radio"/> E | 11 <input type="radio"/> K | 17 <input type="radio"/> Q | 23 <input type="radio"/> W | |
| 6 <input type="radio"/> F | 12 <input type="radio"/> L | 18 <input type="radio"/> R | 24 <input type="radio"/> X | |

e. What color are your eyes? If they are more than one color, pick the color they are closest to.

| | |
|-------------------------------|-------------------------------|
| 1 <input type="radio"/> Green | 4 <input type="radio"/> Blue |
| 2 <input type="radio"/> Brown | 5 <input type="radio"/> Hazel |
| 3 <input type="radio"/> Black | 6 <input type="radio"/> Grey |

c. What is the second letter of your last name? Please select the response that includes the letter.

| |
|-----------------------------|
| 1 <input type="radio"/> A-E |
| 2 <input type="radio"/> F-J |
| 3 <input type="radio"/> K-O |
| 4 <input type="radio"/> P-T |
| 5 <input type="radio"/> U-Z |



How do the SGIC questions create a “unique code” for participants?

| SGIC questions from Form A | Example of student response | Numerical value of response | Site and Classroom Code | Final SGIC |
|--|-----------------------------|-----------------------------|-------------------------|---------------------|
| What is the first letter of your first name? | T | 20 | | |
| On what day of the month were you born? | 6 | 6 | | |
| What is the second letter of your last name? | A | 1 | | |
| What is the first letter of the street where you live? | R | 18 | | |
| What is the first letter of your mother's first name? | N | 14 | | |
| SGIC CREATED | | 20611814 | S1C4 | S1C120611814 |



Findings



- / Pre and post survey data linked for 83% of the students.**
 - To match, students needed to answer 3, 4, or 5 questions out of 5 the same between pre and post.
- / 55% of the students matched on all 5 questions.**
- / Some questions worked better than others.**
- / Missing data reduced ability to match.**



55% students matched on all 5 SGIC questions; 18% matched on 2 or less

| SGIC questions | Number of students | % Total matched | |
|-----------------------------|--------------------|-----------------|-----|
| Matched on all 5 questions | 28 | 55% | |
| Matched on 4 of 5 questions | 11 | 22% | 83% |
| Matched on 3 of 5 questions | 3 | 6% | |
| Matched on 2 of 5 questions | 3 | 6% | |
| Matched on 1 of 5 questions | 3 | 6% | |
| Matched on 0 questions | 3 | 6% | |



The number of surveys matched (≥ 3 identical SGIC questions) and the amount of missing data varied by classroom

| Class | Question Set | Pre-program survey respondents | Post-program survey respondents | Total students matched pre to post | Number missing | Match rate |
|--------------|--------------|--------------------------------|---------------------------------|------------------------------------|----------------|------------|
| Class 1 | Set 1 | 21 | 16 | 12/16 | 5 | 75% |
| Class 2 | Set 2 | 16 | 9 | 7/9 | 7 | 77% |
| Class 3 | Set 2 | 13 | 13 | 13/13 | 0 | 100% |
| Class 4 | Set 1 | 16 | 13 | 10/13 | 3 | 77% |
| Total | | 66 | 51 | 42/51 | 15 | 82% |



Some SGIC questions had higher match rates than others

| SGIC question | Total cases | Total cases matched | Pre/Post survey match rate |
|---|-------------|---------------------|----------------------------|
| In what month were you born? | 22 | 21 | 95% |
| What is the first letter of your first name? (c | 48 | 44 | 92% |
| What color are your eyes? If they are more than one color, pick the color they are closest to. | 22 | 21 | 91% |
| On what day of the month were you born? | 26 | 23 | 88% |
| What is the first letter of your middle name? Note included an option for No Middle Name | 22 | 18 | 82% |
| What is the first letter of your mother's first name? | 26 | 21 | 81% |
| What is the second letter of your last name? Please select the response that includes the letter. | 48 | 39 | 81% |
| What is the first letter of the name of the street where you live? | 26 | 20 | 77% |

What we learned:

Using SGICs to match individual data from two points is a viable option



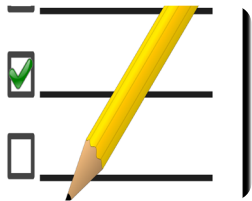
- / A clustered population improves the percentage of surveys that can be matched from pre to post.
- / Missing data from pre or post cannot be tracked or accounted for and reduces matching
- / The questions matter



What next?



Conduct validity check to confirm matches and track missing surveys



Assess differences by survey mode



Explore variations in question formatting and code



Develop and test SGIC questions for different populations and settings



Thank you!

