

# Barrera-Osorio et al 2011

Dimitrios & Eva

2023-03-05

## Contents

<b>1</b>	<b>Motivation</b>	<b>2</b>
<b>2</b>	<b>Data sources</b>	<b>2</b>
<b>3</b>	<b>Method</b>	<b>2</b>
<b>4</b>	<b>Descriptive statistics</b>	<b>4</b>
<b>5</b>	<b>Results</b>	<b>4</b>
<b>6</b>	<b>Conclusion</b>	<b>4</b>

# 1 Motivation

##Why is this research question relevant?

Education in Colombia and other middle-income countries face challenges such as high dropout rates among low-income students, and the reasons behind them, such as the high cost of education. Conditional cash transfers (CCTs) are an evidence-based intervention to increase participation in education. However, the authors highlight that there is little variability in the structure of programmes. The paper investigates if changes in the timing of payments affect the outcomes of interest: Attendance and re-enrollment. Optimising the structure of CCTs may contribute to improved education outcomes and reduce disparities in access to education.

## 1.1 What are the main hypotheses?

- The savings model will improve outcomes compared to the basic programme by relaxing possible savings constraints.
- The tertiary model will improve rates of graduation and tertiary enrollment compared to the basic programme by providing direct incentives for continuation of education.

# 2 Data sources

We investigated performing the replication with the data provided as part of the lecture. However, we soon discovered that the file did not contain all required variables, nor was there any meta data or other information about the variables in the dataset. A brief search revealed that the data and STATA scripts used to obtain the authors' results are freely available here.

For this project, we reference the following files: \* Data file: Public\_Data\_AEJApp\_2010-0132.dta \* STATA script for Table 3: Table\_03\_Attendance.do \* Meta data: AEJApp\_2010-0132\_Data\_ReadMe.pdf

## 2.1 Where does the data come from (country, time period, source)?

## 2.2 What are the key variables?

## 2.3 How are these measured?

# 3 Method

## 3.1 Research design

The research paper describes three interventions designed to improve attendance and educational outcomes for students in Colombia.

The first intervention ("basic") is similar to the PROGRESA/OPORTUNIDADES program, a conditional cash transfer program in Mexico that operated from 1997 to 2012. It pays participants 30,000 pesos per month (approximately US\$15) if the child attends at least 80% of the days in that month. Payments are made bi-monthly through a dedicated debit card, and students will be removed from the program if they fail to meet attendance targets or are expelled from school.

The second intervention, called the savings treatment, pays two-thirds of the monthly amount (20,000 pesos or US\$10) to students' families on a bi-monthly basis, while the remaining one-third is held in a bank account. The accumulated funds are then made available to students' families during the period in which students

prepare to enroll for the next school year, with 100,000 pesos (US\$50) available to them in December if they reach the attendance target every month.

The third intervention, called the tertiary treatment, incentivizes students to graduate and matriculate to a higher education institution. The monthly transfer for good attendance is reduced from 30,000 pesos per month to 20,000 pesos, but upon graduating, the student earns the right to receive a transfer of 600,000 pesos (US\$300) if they enroll in a tertiary institution, and after a year if they fail to enroll upon graduation.

All payments are based on reports provided to the Secretary of Education by the students' principals. Students will be removed from the program if they fail to meet attendance targets, fail to matriculate to the next grade twice, or are expelled from school.

The eligibility criteria for the basic and savings experiments were as follows:

- Children had to have finished grade 5 and be enrolled in grades 6 - 11.
- The children's families had to be classified into the bottom two categories on Colombia's poverty index (SISBEN).
- Only households living in San Cristobal prior to 2004 were eligible to participate.

The eligibility criteria for the tertiary education experiment were as follows:

- Children had to have finished grade 8 and be enrolled in grade 9 - 11.
- The children's families had to be classified into the bottom two categories on Colombia's poverty index, the SISBEN.
- Only households living in Suba prior to 2004 were eligible to participate.

The paper investigates differences in enrollment and graduation / progression to tertiary education for the three treatment groups compared to untreated controls. Randomization to treatment vs control group was stratified by location, school public vs private, gender and grade.

## 3.2 Analysis

### 3.2.1 What are the assumptions of the method?

For Model 1: From ChatGPT: Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. The assumptions about the data underlying linear regression are:

1. Linearity: There should be a linear relationship between the independent and dependent variables. This means that as the independent variable changes, the dependent variable changes proportionally.
2. Independence: The observations used in the regression analysis should be independent of each other. In other words, the value of one observation should not be influenced by the value of another observation.
3. Homoscedasticity: Homoscedasticity refers to the assumption that the variance of the dependent variable is constant across all values of the independent variable(s). This means that the spread of the residuals (the difference between the predicted value and the actual value) should be the same across all levels of the independent variable(s).
4. Normality: The dependent variable should be normally distributed at each level of the independent variable(s). This means that the residuals should be normally distributed and the distribution of the dependent variable should be symmetrical around the mean.
5. No multicollinearity: If there are multiple independent variables in the regression model, there should be no high correlation between these independent variables. This means that the independent variables should be unrelated to each other.

Violations of these assumptions can lead to biased or inefficient estimates of the regression coefficients and incorrect inferences about the relationship between the independent and dependent variables. Therefore, it is important to check these assumptions before interpreting the results of a linear regression analysis.

Fixed effect models: Used to control for unobserved factors that are constant over time at any level of analysis. The key is to identify the appropriate level of analysis for the fixed effect and include it in the model to account for the unobserved factors that affect the outcome variable. In this case, the school was chosen as the fixed effect, while standard errors are clustered within the individual. This reflects the multiple levels of clustering of effects (unobserved school characteristics and unobserved characteristics of individuals).

### 3.2.2 Are these assumptions plausible in this example?

Assumptions of linear regression: The authors do not discuss this.

1. Linearity:

### 3.2.3 Maybe: regression equation.

## 4 Descriptive statistics

- Describe sample

table1

## 5 Results

- Are these results plausible?
- How robust are the results to changing the sample?

## 6 Conclusion

- Replication of a research paper: How do results compare to results of research paper?

Source for clustered standard errors in R: <https://evalf21.classes.andrewheiss.com/example/standard-errors/>  
`fixef(lmer(Reaction ~ Days + (1|Subject) + (0+Days|Subject), sleepstudy))`