# Ending the Musical Chairs Game in Higher Education: *

## How a Software Dashboard (SPADIES) Unveiled Information that Reduced Drop-outs and Increased Graduation Rates in Colombia


Author:

Luis Omar Herrera-Prada†

*Ibero-America Institute for Economic Research*

*Universität Göttingen*

*International Monetary Fund*


June 20, 2022


**Keywords:** Dropout, Higher Education, Education, Graduation, On-time graduation

**JEL Classification Numbers: H40, H52, I23, I28, O33**



*I am thankful to Inmaculada Martinez-Zarzoso, Julián Cristia, Alexandra Sastoque, Nadmy Daniela Valbuena González, Sarah Belford, Juan Esteban Saavedra, Darío Maldonado, Andrés Córdoba, Elkin Garavito, Diana Durán, Paula Andrea Triana Romero, Román David Zárate, Román Andrés Zárate, Mónica Ospina, Fabio Sánchez, Gustavo Canavire, Fernando Rios-Ávila, SPADIES team 2005-2011 at Universidad de los Andes, and Ministry of Education. Comments received in the MEA-SOLE 2022 meeting, the Western Hemisphere Department' Seminar at the IMF, the Development Economics Seminar at Universität Göttingen, and Public Policy Seminar at Universidad de los Andes - Colombia were invaluable. Latest version of this document is available in `https://herreraprada.com` . All views and errors are mine.



†Corresponding Author: l.herreraprada@stud.uni-goettingen.de

**Abstract**

The new millennium brought with it a new challenge: Students reaching higher education were not only greater in number, they were also less prepared than their predecessors. This caused a sharp deterioration in the quality of the entire higher education system, which in turn affected enrollment and graduation rates. Colombia faced this challenge with severely scarce economic resources, but the Ministry of Education designed a plan to keep students in higher education. The plan increased the overall enrollment rate in higher education from 20% in 2002 to over 40% in 2010. One of the tools used to implement this plan was the System for the Prevention and Analysis of School Dropouts in Higher Education (SPADIES). This software dashboard facilitated the collection, analysis, and visualization of student data. This paper provides evidence that among other interesting spillover effects, SPADIES was vital to achieving higher educational attainment outcomes across the country. Using a differences-in-differences approach, under the novel Callaway and Sant'Anna (2021)'s framework, I find that SPADIES reduced the probability of a student becoming a drop-out by 7 bps and increased the probability of graduating and graduating on-time, by 6 bps and 4 bps, respectively. Although the impact may seem small, 7 bps correspond to around 14, 000 students, which almost double the size of the average higher education institution (HEI) in Colombia. I also find that higher education institutions used SPADIES not only to prevent students dropping out but also to call back students who had already dropped out. The number of transitions (change of status to enrolled) from absent and drop-out increased 3 bps and 2 bps, respectively, and the average duration of a transition increased 0.5 semesters, showing that HIEs succesfully brought students back to schools.

# 1 Introduction

The dropout phenomenon is one of the most costly and difficult problems in higher education. The costs associated with not obtaining the degree - wasted resources (structural and financial), the loss of efficiency in connecting graduates with the labor market, and the opportunity cost to dropouts - throw a stark spotlight on the need to reduce the dropout rate. Between 2000 and 2010, Colombia faced the "Crowding cohort" phenomenon[1] when the children of the "boomers" started to enter higher education, and this massive influx of less prepared students lead to a deterioration in the quality of higher education institutions in Colombia (Herrera-Prada, 2013; Ferreyra et al., 2017) .

In Colombia, the population that reached higher education between 2000 and 2010 was bigger than before because of demographic growth and the success of a policy to increase the secondary enrollment and graduation rate. However, the infrastructure of the higher education system was not prepared; it was essentially the same as it had been two decades ago Orozco Silva et al. (2006); Orozco Silva (2010). These students faced academic vulnerabilities because the policy that increased the enrollment rate in the mid-90s was based on promoting all the students to the next grade level regardless of academic merit. Furthermore, the children of the baby boomers faced economic depression, opposite conditions from what their parents had enjoyed. In the 1970s, when baby boomers were enrolling in higher education, Colombia had an economic boom from the coffee sector. In the early 2000s, when their children started to enter higher education, Colombia was suffering from the worst economic depression it had ever experienced (pre-COVID) and from a failed peace process dealing with the oldest guerrilla group on the continent, which created profound social-political turmoil and directed the use of the scarce resources to military action (Lucio and Serrano, 1992; ICFES, 2002; Orozco Silva, 2010).

Between 2002 and 2006, the Ministry of Education (MEN) designed a plan to improve educational outcomes in higher education and, specifically, to increase the enrollment rate by providing these cohorts with additional financial and academic support. The primary strategy to respond to the increase in demand was to use available resources from the public sector (mainly infrastructure), while at the same time creating a highly coordinated information system to track students and improve the overall efficiency of the system (Ministerio de Educación Nacional, 2008, 2009; Orozco

[1] "Crowding cohort" is a phenomenon when there is increased demand for fixed or reducing resources. This term was used for the first time in Bound and Turner (2006).

Silva, 2010). This is how the System for the Prevention and Analysis of School Dropout in Higher Education (SPADIES) was born. SPADIES provided "real time" information about student enrollment, academic performance, quality of peers, dropout rates, and graduation rates to authorities in higher education institutions (HEIs), MEN, and the general public. MEN promoted the use of SPADIES to help each HEI predict which students were at risk of dropping out, get their complete academic profile, and target them with different types of assistance to reduce the dropout rate. HEIs used three types of assistance: (1) academic assistance (e.g., tutoring or remedial classes), (2) financial assistance (e.g., scholarships), and (3) all other assistance, which ranged from food vouchers and transportation to mental health services (Ministerio de Educación Nacional, 2008, 2009, 2010).

This paper uses SPADIES data from 1998-2017 to track the college paths of about 4 million students to evaluate the impact of SPADIES on education and system efficiency outcomes. Using a differences-in-differences approach, under the novel Callaway and Sant'Anna (2021)'s framework, where the treated are those students enrolled in an HEI with SPADIES installed and controlling by household income, college academic performance, secondary school exit exam score, gender, region, and assistance received, I find causal evidence that SPADIES not only reduced the probability of dropping out by 7 basic points (bps), but also increased the probability of earning the degree by 6 bps and earning it on-time by 4 bps. I also find that the probability of having a transition (status change) from absent or dropout to enrolled increased by 3 bps and 2 bps, respectively. The average duration of a transiton increased 0.5 semesters. I also find SPADIES was most effective in reducing the dropout rate for males, students from public institutions, and low-income students. Finally, my results show that HEIs with SPADIES were able to bring students back to their programs from dropout status, particularly in public institutions. These students would not have returned to school in the absense of SPADIES. This explains the increase in the average time gap during the transition.

The following section presents a review of the literature. Section 3 details the SPADIES program design and the context in which it was created. Section 4 describes the dataset and variables. Section 5 presents the model specification. Section 6 discusses the results. Finally, Section 7 presents the conclusions and policy implications.

## 2 Literature Review

In this section, I review the literature that provides a foundation for the chapter. First, I will discuss the literature of the demand and supply sides for higher education, their implications on quality, and the dropout analysis evolution. Next, I will present some specific papers related to the Colombian context.

In 1999, the World Bank created the program "Education for All". One of its main objectives was to better monitor the educational indicators access, enrollment, and quality. The idea was to implement some of the lessons learned from developed countries' improvements in Higher Education Systems in developing regions like Latin America and to encourage the governments to track the results and indicators.

Canonical literature has identified that formal education and, more specifically, higher education plays a transformational role in human development and human capital growth. The formal education function provides individuals with the capacities and skills to enter into productive sectors and boost social mobility (Becker, 1962; Trow, 1974; Adams, 1984; Bank et al., 1990; ONU, 2013). Given the importance of human capital formation, complex dynamics are presented from the demand and supply conditions that end up in an unbalanced market due to barriers of access, low enrollment rates, excess demand and/or supply, and low quality (Epple et al., 2006). The discrepancy between supply and demand can be attributed to several reasons, including the lack of public resources due to an unexpected increase in applicants (Bound et al., 2009). The growth of the demand raises problems such as resource allocation and efficiency, attrition, late graduation or a lag in studies, and a decrease in the quality of the education (Bound and Turner, 2006). Subsequently, institutional factors that may influence the individual's decision to enroll or continue or not continue in a HEI, such as the quality of the peers or the education received and supports or services offered by the institution, are incorporated as a part of cost-benefit analysis (Tinto, 1975, 1982; Bank et al., 1990; MacLeod and Urquiola, 2015). The theoretical basis of the discussion around attrition is commonly classified into two parts: the first is the student's integration or adaptation to the education system model (Tinto, 1975, 1982), and the second explains attrition as a set of conditions linked to the individuals' socioeconomic factors, such as family conditions or academic performance during school (Bean, 1980, 1985).

However, the enrollment, dropout, and graduation rates and the quality of education depended on how students were sorted among HEIs; two main models developed by Epple et al., 2006 and MacLeod and Urquiola, 2015 clarify how this process happened. For this development, the information flow among agents was instrumental. Information provision, data access, and data mining have been at the center of the agenda to improve both education systems and their educational outcomes, especially in primary and secondary schools, over the past 25 years (Campbell and Levin, 2009; Asif et al., 2017; Allende et al., 2019; Kerr et al., 2020; Unesco, 2020).

In Colombia, the first analyses on attrition included only a couple HEIs. For example, Universidad Nacional de Colombia (2007) studied the lag, graduation rate, and dropout rate in Colombia's most important public university. The study demonstrates the importance of including contextual variables in the analysis. In general, they find that being a woman (especially 18 years old or younger) constitutes a significant characteristic that positively impacts the probability of obtaining a degree from any program. Financial aid or student loan programs decrease the dropout rate. In this sense, it is important to stress that the studies agree on the importance of including "Affirmative Action" programs, such as unique admission mechanisms and alternative admission avenues through pre-university courses. These programs aim at improving students' chances of staying in higher education and, as a consequence, improving access conditions and social equity (Sánchez et al., 2002). At the University of Antioquia, (Castaño et al., 2006) conducted a study on the School of Engineering and the School of Economics students in the second cohort of 1996. Using a duration model, including context variables, they found that being male, single, and over 18 years old is associated with a greater risk of dropping out. They also found that living with parents, having a better academic performance, not working, having parents with a high level of education and being female are characteristics associated with a decreased risk of dropping out. The studies implemented in public universities addressed the State's concern about the way resources were being used in the educational system, because the State understood the dropout phenomenon as a waste of economic resources, human capital (professors and staff), and infrastructure (Cárdenas, 1996; Córtes et al., 2011; Facundo-Díaz, 2009). Finally, at a national level, ICFES (2002)found that household financial conditions were the primary determinant for becoming a dropout student. However, Ministerio de Educación Nacional (2008) showed that the dropout rate's main reasons were low academic skills (measured by the secondary school exit exam score), mismatch in career choice and skills, poor academic performance, and gender. Finally, Herrera-Prada (2013) showed that the "Crowding cohort" phenomenon accompanied by a simultaneous increase in average time needed to graduate resulted in an overall decrease in the graduation rate in Colombia, even though the dropout rate decreased across the whole country.

However, the subject of analysis is the SPADIES program and not the anti-dropout models. As explained in detail in Section 3.2 SPADIES was a program that developed a computer application that allows users to predict the probability of an individual student dropping out and to monitor the efficiency of tailored anti-dropout programs for those students.

In this regard, the literature is quite limited. While the literature has developed in Colombia over the last 15 years, the only sources for this literature are reports from the Ministry of Education. Globally, the literature on programs like SPADIES and accompanying anti-dropout programs began in other countries after 2014, and they sought to tackle challenges that SPADIES had already overcome. Focused on small programs in regions of Germany or Chile, the international literature includes some econometric models (uplift models) to predict dropout rates (Berens et al., 2021) or design tailored anti-dropout programs (Olaya et al., 2020). SPADIES began developing uplift models in 2005 and enabled higher education practitioners to design and implement targeted anti-dropout programs across an entire country for more than 10 years.

This chapter contributes to the literature by revealing how SPADIES, the main mechanism used by the Colombian government to decrease the dropout rate, was effective in reaching the intended policy goals. My analysis will illustrate how a data system that enhanced the flow of information among agents, coupled with coordinated policy and creative policymakers, developed tailored aid programs to a listed pre selected candidates, and resulted in a new equilibrium with educational and social outcomes even better than the government had hoped for.

## 3 Context and Program Design

In this section, I present some facts and statistics about the Colombian context and the history behind the creation of SPADIES.

### 3.1 Context

In the last 50 years, Colombia has experienced dramatic demographic changes, including massive migration from rural to urban areas that brought both cheap labor and young people hoping for opportunities of social mobility to the city. The new generations were eager to progress in a country that was not prepared to provide them with adequate educational resources (Lucio and Serrano, 1992). In the mid-1990s, the lack of capacity in high schools was evident, but the government allowed students to advance through primary and secondary school without any academic restrictions. The promotion of students without restriction was successful, as coverage and completion rates in secondary schools increased, but the skills and quality of education significantly decreased (Orozco Silva et al., 2006; Orozco Silva, 2010; Herrera-Prada, 2013). The economic crisis of the late 1990s just accelerated and unveiled the inevitable: few Colombians made it to higher education, and those who did were likely to drop out early in their studies (Orozco Silva et al., 2006; Ministerio de Educación Nacional, 2008, 2010). Colombia started to analyze the phenomenon in the early 2000s and found that the main problem was the economic situation in students' households (ICFES, 2002). At that time, the nation's gross college enrollment rate was 20 percent, one of the lowest in the region(Ministerio de Educación Nacional, 2010; Ferreyra et al., 2017).

In response to this scenario, the Ministry of Education created a plan called "Educational Revolution" to increase educational attainment in Colombia, which had been lagging well behind its regional peers. The plan worked; the rate of higher education enrollment increased from 20% in 2002 to over 40% in 2010 to over 45% in 2016 (Orozco Silva et al., 2006; Ministerio de Educación Nacional, 2010, 2017). This process was primarily due to HEIs in the public sector, as evidenced by the fact that the enrollment rate of private HEIs was close to zero or negative in the first half of the 2000s. For the first time since 1974, the public sector enrolled more students than the private sector (Orozco Silva et al., 2006; Orozco Silva, 2010). This phenomenon becomes even more exceptional when examining the enrollment rate of growth. The pace of enrollment growth was lower than the annual average growth rate of the last 70 years (8.74%) (Orozco Silva et al., 2006). The question that then arises is: how did this overall enrollment increase happen if the pace was slower than the historical average? The answer has two components: first, a steady growth rate of just over 6% per year during the first decade (2000-2010) and, second, a simultaneous reduction in the number of students dropping out from their institutions. In contrast, the previous period saw a higher rate of enrollment but also a higher rate of dropouts.

When the MEN created SPADIES, it began providing "real time" information about enrollment, academic performance, quality of peers, dropout rates, and graduation rates to all the agents in this market. The dropout rate was the main target outcome, and each HEI was able to download a list of students at risk of dropping out with a complete profile for each student, including how to target aid. As well, SPADIES enabled the MEN to gather information on both the main characteristics of the overall higher education system and performance indicators for individual HEIs. The MEN aggregated this data and provided it to the public, and SPADIES became the standard for certifying institutions in Colombia for the quality of their programs and measuring improvements in access to resources.

### 3.2 Program Design

SPADIES is an information system conceived by the MEN at the end of 2004. The MEN commissioned the Universidad de los Andes to create a software tool that would collect information on all students enrolled in higher education and allowed for data visualizations, statistical analysis, and creation of reports on students at risk of dropping out. One module of SPADIES was public and allowed students, authorities, and the general public to make visualizations and analyses comparing the performance of HEIs or other departmental or national aggregates.

When the first data were obtained, the MEN's efforts were focused on understanding the factors that determined dropout probability. Consequently, the MEN and Universidad de Los Andes used the SPADIES database to estimate Duration Models and conduct a focus group analysis across the country. The results of these two analyses were later implemented in the software.

Once the database had enough information on each HEI to estimate the models and conduct temporal analyses of the cohorts, the MEN could identify consistent behavior patterns among the students who became dropouts. The MEN wanted to empower HEIs with this database so that they could identify those at-risk students and take action to prevent them exiting the programs prematurely. The MEN and the Universidad de Los Andes visited[2] each HEI to install the software,

[2] Each visit lasted between 4 and 6 hours. The participation of the University President in the visit was optional, but the participation of the Academic Vice-Pesident, the head of the Planning Department, and the Chief of the Systems Office was mandatory.

explain the results, and train the staff on the software. In the first part of the visit, the results of the national and institutional estimates were presented. The second part consisted of the installation of the software in at least one of the offices of the directors. After the installation, the final part of the meeting consisted of teaching HEI personnel how to use the tool to identify and profile the population at risk of dropping out. The objective was to use this list to target the different types of assistance to prevent students from dropping out. The assistance programs were sorted into three groups:

1. Financial aid (grants, scholarships, or any other financial support that did not charge a fee).
2. Academic aid (free tutorials or remedial courses).
3. Other type of aid (any other free aid given to the student different from those reported as financial or academic aid, such as mental health support, coaching, career guidance, etc.)

According to Ministerio de Educación Nacional (2008), the most at-risk population for dropping out at the national level was males with low household income and low academic skills, mainly in associate programs or math-related programs. The MEN encouraged HEIs to design the first wave of aid programs to target this population.

Between 12 and 18 months after installation, the MEN and the Universidad de Los Andes made one follow-up visit ("Accompaniment") to check the new outcomes and ensure the universities were correctly using of the application[3]. In 2011, all institutions had the software installed, and all of them had their follow-ups by the end of 2013. The Universidad de Los Andes was the consultant selected by MEN for the terms 2004-2011 and 2014-2017. Universidad EAFIT was in charge of the system between 2011 and 2013. After 2011, consultants only made follow-up visits, provided support to the MEN and HEIs improved the software and advanced research using the data for MEN. The consultants were required to have a team of people dedicated to answering questions and analyzing the information from each institution. All the institutions had to have at least two delegates with permanent contact with the MEN or the consultant every month. Data reported every week was audited by the consultant and MEN before being stored in the database.

[3]For example, some universities were forecasting the dropout rate of individual applicants to inform their selection process for admitting students. The MEN prohibited use of the application for this purpose and conducted the follow up visits to ensure HEIs were following the rules. Those actions were rapidly detected and stopped.

Starting in 2008, the MEN organized several contests on strategies to reduce dropout using SPADIES as a mechanism to measure the dropout rate and incentivize innovation in the education field. Two crucial milestones that SPADIES enabled were (1) collecting information at the individual level since 1998, which was a historical milestone for many institutions, and (2) the entry of many universities into the digital world. At the end of 2005, very few institutions had electronic records; the Ministry's information systems triggered an arduous migration process from paper to digital records. By In 2017, about 60% of institutions were able to report their complete data starting in 1998.

SPADIES' installation and training (this will become the starting point of the treatment in this analysis) were not randomly assigned because of the stark difference in the digital gap and quality of information of each HEI. Five rounds (Round 1 2005-06, Round 2 2006-07, Round 3 2007, Round 4 2008-09, and Round 5 2010-11) were necessary to complete the collection of each of the HEIs' information and add it into the system (The list of HEIs in each round are shown in Figure 1). Each round corresponded to a new group of HEIs having their digital records in shape in order to trigger the process mentioned above. However, there is no difference in the application of SPADIES in HEIs by the sex of the student, or the sector, quality, or region of the HEIs (Figure 2). The fact that some HEIs are biased toward high scoring and high income students pursuing bachelors degrees is evident; it is true that high income and high scoring students tend to go to better resourced universities. These HEIs were more likely to be enrolled in the first rounds because they had more resources and higher data quality.

Starting in 2008, the MEN required that HEIs include the official dropout rate measured by SPADIES in the requirements for the quality certificate of programs. The MEN also provided a web portal so that the public could access the main statistics of individual HEIs or the full higher education system, including characteristics of the population, dropout rates, and graduation rates.

## 4 Data and Variables

In this section, I first describe the SPADIES database and its characteristics. Next, I explain the variables that SPADIES uses, the program's definitions, and the new variables I created and included. In the last part of this section, I explain the final database I use in the empirical analysis.

## 4.1 SPADIES Database

The SPADIES database is obtained by merging data from three sources: MEN's National Information System for Higher Education (SNIES) database, the Colombian Institute for the Promotion of Higher Education (ICFES) database, and the HEIs semestral report.

The SNIES data is time-invariant and includes all the characteristics of the HEIs and their programs. The ICFES data is time-invariant as collected during the Saber 11 exam at the end of secondary school. The HEIs' report updates the information about its students every semester.

SPADIES uses the life history approach to collect its data. This means that SPADIES only tracks students who began college in 1998 or later. This chapter uses SPADIES data from 1998 to 2017, which includes about 8 million students. The dataset is an unbalanced panel per individual-program semester.

## 4.2 Variables

This subsection explains the variables that SPADIES has, how they are measured, and the new variables I created.

### 4.2.1 SPADIES Time-Invariant Variables

All the ICFES' variables collected during the Saber 11 exam include the Saber 11 exam score, gender, year of birth, and household income. The Saber 11 exam result is an essential requirement to enroll in higher education, so all the students who enroll in higher education have a score; all the other questions are optional for the students. This information is used for the students' characterization and the empirical analysis in this thesis. SPADIES uses a standardized version of the Saber 11 test score. As the ICFES used a different score range over time in the Saber 11 test score, the MEN required SPADIES' developers to standardized the score by assigning each student their percentile on the Saber 11 exam. The variable contains values from 1 to 100 depending on the student's percentile. I created an additional dummy that takes the value of 1 if a student has a high score (above 90) and 0 if otherwise (similar to Ministerio de Educación Nacional (2008, 2010, 2017)).

Gender and year of birth are reported by the ICFES database and by the Freshmen report. The household income is reported by the student to the ICFES.

SPADIES uses the data from the characteristics of the HEIs and their programs reported by SNIES. The HEIs characteristics include the sector (public or private), the category (Universities or Community Colleges), and their location. The program's characteristics include the level (Bachelor -4 or 5 year programs- or Associate -2 or 3 year programs-) and the field of knowledge.

Finally, SPADIES has a variable that qualifies the data reported per HEI. The grade depends on the number of semesters reported and the details of the data reported per semester. The grade could be A, B, or C; the highest grade is A. I created a dummy variable that takes the value of 1 if the HEI earned an A in the report, as recorded in the database in 2017.

### 4.2.2 SPADIES Time-Variant Variables

SPADIES received three main reports per semester from each HEI: Freshmen, Graduates, and Enrolled. Each HEI also provides the student ID, academic performance, information about financial or academic aid received, and program of study.

SPADIES looks for the students reported as Freshmen in the Enrolled report during every semester until they are reported as Graduates; if the student is not found in the Enrolled report or Graduated report, their status changes (to either absent or dropped out, depending on number of semesters not enrolled).

The statuses that SPADIES uses are:

1. **Graduated:** SPADIES defines this status as a person who earned a degree from a higher education program.; as the person was reported in the Graduates report. I divided this category into two groups:

    (a) Graduated on time (those who graduated within one year of the expected time of graduation).

    (b) Graduated late (those who graduated more than 1 year after their expected time of graduation).

    Graduated late and Graduated on time encompass all individuals that graduated from higher education. The expected graduation year was estimated as five years for professional programs and three years for associate programs. So, $Graduated = Graduated\,on\,time + Graduated\,late$.

2. **Dropout:** SPADIES defined dropout as a student that has not been reported in the system or Graduated after 2 or more consecutive semesters as of 2017.

3. **Absent** students are those who missed only one semester and are not reported as Graduated.

4. **Active** is any student taking classes as of 2017.

Definitions:

1. The **student's cohort** is the year and semester (YYYY-SS) in which the student enrolled as a freshman.

2. **Freshmen students** are those enrolled in their first semester in an HEI-program.

3. The **time in the system** is the account of semesters that a student is reported as enrolled; it is not the number of semesters since the start of its cohort (e.g., the student could have been absent one semester). SPADIES records a timestamp per semester in each student's record when it can find the student in the semestral enrollment report.

Therefore, every semester, the student's time in the system, academic performance, and the report of aid received are updated. The Status is assigned at the moment of the extraction of data from the database; in the case of this document, it is 2017. In addition to these statues from SPADIES, I created three new variables to measure the transitions. I compare the time since they first enrolled in the college and the numbers of semesters reported in SPADIES. If there is a difference, it means that they left the school. If their current status is graduated or active, it means that they had a transition, i.e. a period of time out of school before returning to school. There are two types of transition:

1. **Transition from absent,** means that in a period "T" the student was "Absent" and in "T+1" he became "Active" or "Graduated". The dummy variable following these transitions takes the value of 1 in "T+1" and 0 otherwise.

2. **Transition from dropout**, means that in a period "T" the student was a "Dropout" and in "T>=3" he became "Active" or "Graduated". The dummy variable following these transitions takes the value of 1 when the transition ends and 0 otherwise.

Finally, I created the variable **time gap of transition** that counts the amount of semesters during each transition; in the case of transitions from “Absent”, it is always 1, but for the transitions from dropout, it is always 2 or more semesters.

### 4.3 Final Data

SPADIES merges the time-invariant information from the ICFES database and the SNIES with the reported data of Freshmen, Enrolled, and Graduates every semester. The individual is defined as a student who has ever been enrolled in a program in an HEI as a freshmen during the period 1998 to 2012.

In the final database, I need to avoid any bias to drop-out (mainly in the first semesters) or graduation (only after 4 semesters in associate programs, or 8 semesters in bachellors) by making the following changes to the original database (8 million students):

1. Because SPADIES uses the life history approach, and in order to have periods with students in all stages of their program (from freshmen to seniors), I only use data for students reported as "Active" since 2002. To be "Active" in 2002 means the student was reported as a "Freshmen" in any year before (and including) 2002, and they should appear in the report of Enrollment in 2002.

2. I also need to have cohorts of students with enough time to graduate. It means, I use data up to 2017 but only for students who were freshmen until 2012.

After these two changes, the population in the database numbered 6,143,537 students.

I created a dummy that identified the time (Year - Semester) when SPADIES was installed in each HEI. I use the semester of this visit as the start of the "treatment" of SPADIES on an HEI, because during this visit, the insitution received the first report from SPADIES of the students at-risk in their institution, and the HEI was able to access to grants from MEN to create anti-dropout programs.

I included the unemployment rate estimated by the National Department of Statistics (DANE) by HEI department and year.

The final dataset with all variables is an unbalanced panel by individual-program-HEI and time, numbering 4,131,302 students in total. It includes the students' gender, year of birth, household income, ICFES's test score, bachelor degree track (a dummy takes the value of 1 if the program they are attending leads to a bachelor level degree and 0 if otherwise), assistance received (a set of dummies by assistance program, financial or academic, that take the value of 1 if received and 0 if not), institution type (a dummy that takes the value of 1 if their institution is public and 0 if not), HEI certification status (a dummy that takes the value of 1 if their HEI is certified and 0 if not), campus type (a dummy that takes the value of 1 if their HEI is a main campus and 0 if not), a dummy that takes the value of 1 if the data from their HEI has good quality and it is 0 otherwise, a set of dummies that take the value of 1 depending on the HEI's region (Bogotá, Valle del Cauca, Antioquía, and Atlántico), a set of dummies that take the value of 1 depending on the round of implementation their HEI was included in the program, a dummy that takes the value of 1 if the period of Enrollment is later than the starting time of the program in their HEI, the unemployment rate by their HEI's department and the year they were enrolled, the status in the system, the transition from absent, the transition from dropout, the time gap of transition, a dummy that takes the value of 1 if the student is a dropout student and 0 otherwise, a dummy that takes the value of 1 if the student is graduated and 0 otherwise, and a variable that accounts the number of transitions that a student had during their program.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

# 5 Model specification

## 5.1 The Basic Model

To measure the impact of SPADIES in different outcomes, I use the following equation:

$$Eq:1 \qquad Y = \beta_0 + \gamma_0 SPADIES_{it} + \gamma_1 SPADIES_{it} \times AA_{it}$$

$$+\gamma_2 SPADIES_{it} \times PFA_{it} + \gamma_3 SPADIES_{it} \times AA_{it} + \beta X_{it} + \epsilon_{it}$$

where $Y_{it}$ is a dummy that, depending on the model, measures one of five outcomes that are probabilities: the probability of dropping out, graduation, graduation on time, transitioning from absent, or transitioning from dropout. In the case of Time gap (the sixth outcome), $Y_{it}$ is a continuous numerical variable equal to the number of semesters of the transition. The variable of interest is $SPADIES_{it}$ , and it takes the value of one (1) if the individual is enrolled in a HEI in a period $t$ when SPADIES was already installed and zero (0) otherwise. I also include three interactions with a variable that takes the value of one (1) if student received aid in an specific time and zero (0) otherwise: SPADIES and academic aid (AA), SPADIES and public financial aid (PFA), and SPADIES and private financial aid (FA). I include these interactions to understand if the combination of SPADIES and the academic or financial aid improved education outcomes. The vector of controls $X_{it}$ is comprised of time variant and time invariant variables. The time variant variables include academic performance, occurrence of assistance if received and type (financial or academic), time that the student has been enrolled in the HEI (tenure), and departmental unemployment rate. The time invariant variables include a dummy for females, the year of birth, a dummy if the Saber 11 exam score is over the 90th percentile, a categorical variable for household income, and a set of dummies to indicate the region of the HEI that the student attends.

I will present two sets of results for Equation 1 using a Fixed Effects (FE) and a Random Effects (RE) framework. I use both models because they offer different advantages. While the RE model allows me to compare the results with previous literature as in ICFES (2002); Ministerio de Educación Nacional (2008, 2010) ), the FE model provides more consistent results to use as benchmark before the Differences in Differences (DiD) approach.

Recent literature (Sant'Anna and Zhao, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021) provides tools to correctly identify the causal effect of SPADIES on the six outcomes. Previously, it was not easy to correctly isolate the impact of SPADIES because SPADIES had five rounds, they were not randomly assigned, and the population could be very different among rounds. The main problem was that the treatment could be analyzed as five different treatments that overlapped in time. To estimate a causal inference, I am using the framework proposed by Callaway and Sant'Anna (2021). This DiD approach allows an identification, estimation, and inference of multiple time periods (up to 6 semesters after the treatment was applied). This approach also accounts for variation in treatment timing (in this case, the rounds) and potential differences in the treatments that allow me to hold the "parallel trends assumption" (PTA) only after conditioning on observed covariates in the period before treatment.

## 5.2 Canonical DiD

The basic DiD approach in the canonical format considers two periods and two groups (model 2X2). In the first period (T=0), the two groups are the same in terms of the treatment, as they do not receive any. In the second period (T=1), some of the individuals did receive the treatment creating the group called "treated" (SPADIES=D=1), while those that did not receive any treatment are called "controls" (SPADIES=D=0). So, if we assume that the treated group would follow its predetermined path given by its trend, in case of an absence of the treatment, any deviation from this trend is a causal effect of the treatment on the group. This deviation or difference is the Average Treatment effect on the Treated (ATT) (Equation 2). However, the $Y_{i,1}(0)|D_i = 1$ component is never observed, as it is unknown how the treated group would be in T=1 in the absence of treatment.

$$Eq : 2 \qquad ATT = E(\tau_i|D_i) = E(Y_{i,1}|D_i = 1) - E(Y_{i,1}(0)|D_i = 1)$$

As I do not know the path the treated group will follow in the absence of treatment, my best approach is to check the path of the control group. I assume that the path followed by the treated groups is parallel to the path followed by the control group. This assumption is known as the "Parallel Trend Assumption" (PTA) (Equation 3). This assumption is very strong and will be debated later, but by using it, we can re-estimate the ATT (Equation 4).

$$Eq : 3 \qquad E(Y_{i,1}(0) - Y_{i,0}|D_i = 1) = E(Y_{i,1} - Y_{i,0}|D_i = 0)$$

$$Eq : 4 \qquad \widehat{ATT} = E(Y_{i,1}|D_i = 1) - \widehat{E}(Y_{i,1}(0)|D_i = 1)$$

However, in practice, the empirical research usually faces designs with more than two periods or more than two treated groups. According with Callaway and Sant'Anna 2021, the solution has been to generalize the canonical approach by adding the groups and fixed effects to the specification . The debate about the correct specification has been growing in recent years, but the literature agrees that the standard Two-Way Fixed Effects (TWFE) approach may not be appropriate for the identification of treatment effects, in particular interpreting its results (Callaway and Sant'Anna, 2021) . As mentioned above, the PTA is hard to achieve, as treated and control groups are often not similar enough. To solve this, Sant'Anna and Zhao (2020) proposes to hold PTA for groups with the same pre-treatment characteristics X (Equation 5). Where $\theta(X)$ is the $\Delta Y_i$ if there was no treatment conditional to X. With this new assumption, the new DiD estimator becomes $\widehat{ATT}_*$ (Equation 6).

$$Eq:5 \qquad E(Y_{i,1}(0) - Y_{i,0}|D_i = 1, X) = E(Y_{i,1} - Y_{i,0}|D_i = 0, X) = \theta(X)$$

$$Eq:6 \qquad \widehat{ATT}_* = E(Y_{i,1}|D_i = 1) - \left[E(Y_{i,0}|D_i = 1) + \widehat{E}(\theta(X)|D_i = 1)\right]$$

### 5.3 Robust DiD Estimators for ATT

Now, using Rios-Avila et al. (2021)'s CSDID command in Stata, four types of DiD estimators are analyzed using Equation 6 as they present four different approaches to estimate the component $\widehat{E}(\theta(X)|D_i = 1)$ :

1. Regression Approach (OR). This approach estimates $E(\theta_i|D_i = 1)$ in two steps. The first step models $E(\theta_i|X) = \theta(X)$ as a function of X with data from the crontrol group only. The second step uses the predicted oucome for $\widehat{\theta}(x_i)$to estimate $E(\theta_i|D_i = 1)$. The ATT for this estimator is:

$$Eq:7 \qquad \widehat{ATT}_{OR} = E(\Delta Y_i|D_i = 1) - E(\widehat{\theta}(x_i)|D_i = 1)$$

2. Inverse Probability Weights (IPW) from Abadie (2005). In this method, the distribution of characteristics X for the control group is reorganized, so that the control group becomes more

similar to the treated group. To do so, it estimates a propensity score using a binomial model and then, using the predicted scores, estimates the inverse probability weights ω(x). The dependant variable in the propensity score is a marker for if the observation is part of the treatment group as a function of X.

$$P(D_i = 1|X) = F(X) \rightarrow \widehat{\pi}(X) = \hat{F}(X)$$

$$Eq\ :\ 8 \quad \omega(x_i) = (\widehat{\pi}(x_i))/(1 - \widehat{\pi}(x_i))$$

$$\Rightarrow \quad \widehat{E}(\theta_i|D_i = 1) = (E(\omega(x_i)\theta_i|D_i = 0))/E(D_i)$$

$$Eq : 9 \qquad \widehat{ATT_{IPW}} = E(\Delta Y_i|D_i = 1) - (E(\omega(x_i)\theta_i|D_i = 0))/E(D_i)$$

3. Doubly Robust Estimator (DRI) from Sant'Anna and Zhao (2020). The doubly robust estimators are a combination of the previous two estimators (OR and IPW). The model first uses the regression approach, and then it reshapes the groups using a propensity score estimation similar to the IPW approach. A propensity score is estimated using Equation 8, then $E(\theta_i|X)$ is modeled as a function of X and estimated using the weights obtained from Equation 8. See Equation 10.

$$Eq : 10 \qquad \theta_\omega(X) = Min \textstyle\sum_{i|D_i=0} \omega_i(x_i)(\theta_i - \theta(X_i))^2$$

$$Eq : 11 \qquad \widehat{ATT_{DRI}} = E(\Delta Y_i|D_i = 1) - E(\widehat{\theta_\omega}(x_i)|D_i = 1)$$

4. Improved Doubly Robust Estimator (IMP) from Sant'Anna and Zhao (2020). This estimator uses in the first step an approach similar to OR by estimating $E(\theta(X)|D_i = 1)$ using only control data a no weights. Then, it adds a correction Λ, calculating the weighted difference

between the predicted and the observed outcome in the control group. See Equation 12.

$$Eq:12 \qquad \widehat{ATT_{IPW}} = E(\Delta Y_i|D_i=1) - E(\widehat{\theta}(x_i)|D_i=1) - \Lambda$$

$$Where \quad \Lambda = E(\omega(x_i)\Delta Y_i|D_i=0)/E(\omega(x_i)|D_i=0) - E(\omega(x_i)\widehat{\theta}(x_i)|D_i=0)/E(\omega(x_i)|D_i=0)$$

## 5.4 Empirical framework

Callaway and Sant'Anna (2021) expands what was proposed previoulsy by Sant'Anna and Zhao (2020) and Abadie (2005). In particular, Callaway and Sant'Anna (2021) debates the application of DiD estimators when a variation in the timing of treatment existing, and they consider a natural generalization of the ATT to be a setup with multiple treatment groups and time periods. Callaway and Sant'Anna (2021) uses the average treatment effect for units who are members of a particular group $g$ at a particular time period $t$, that expressed in terms of the canonical form (Equation 2) is:

$$Eq:13 \qquad ATT(g,t) = E[Y_t(g) - Y_t(0)|G_g = 1]$$

The framework of Callaway and Sant'Anna (2021) incorporates OR, IPW, DRI, and IMP, fixes a group $g$ and allow variation in $t$, to understand how the proposed ATT evolves in time for a specific group. When this process is extended to all groups, they present the "group-time average treatment effect". In fact, the estimation performed by Rios-Avila et al. (2021)dissagregates the combinations of groups and times in multiple 2X2 models than then are aggregated per the fixed group $g$. After the process, an ATT and weights per period group allow consolidation of the ATT not only by group-time, but also by time (similar to an event analysis), by group (to analyze impacts per group and compare), and using a single robust consolidated estimator

It is important to note that groups can have different times. The population previously divided into two groups (treatment and control) is now sorted into three sets: treated, not yet treated, and control. Depending on the type of analysis, the framework can incorporate the "not yet treated" as controls or perform the estimation using only controls. In my analysis, as the individual is attached to a program in an HEI and cannot change its HEI or round, there is no potential "not yet treated“ group.

# 6 Results

The results section contains three parts: The first part presents the “Parallel Trend Assumption” (PTA) charts per outcome. I find similar trends in the all the periods before SPADIES was installed for the first 4 Rounds; although Callaway and Sant’Anna (2021) assumptions only require PTA stability in the pre-treatment period. The second part shows the event analysis for the four proposed estimates in the left panel and the total and group ATT per output in the right panel. For the main findings, I will focus my reading on the outcomes obtained by the IMP methodology(all the results are shown), which is the most robust and has the most conservative results. Finally, in the third part, I present the total ATT per sub-sample (according to time-invariant characteristics) per estimator type. The students in Round 5 account for only 0.3% of the entire system, they are not included in the PTA nor in the results analyses.

## 6.1 Regression Analysis

Table 3 reports the results for Equation 1. Analyzing the variable of interest SPADIES, I find that SPADIES reduced the probability of dropping out and increased the probability of graduating and graduating on time. According to these results, SPADIES also reduced the probability of transitioning from absent or drop-out and the time gap during the transition. Only the probability of dropping out results show similar levels comparing the FE and the RE frameworks. SPADIES Impacts are significantly higher in the FE framework. However, these results are not causal and are only a first approach to the results from SPADIES. I am presenting causal results later in this section. However, from these results, I can extract some relevant information about the determinants for the probability of dropping out, graduating, and transitioning. For example, except for the unemployment rate. all the expected signs are found in time-variant variables according to (ICFES, 2002; Ministerio de Educación Nacional, 2008; SPADIES, 2008; Ministerio de Educación Nacional, 2010). Those expected results are:

1. A decrease in the probability of dropping out, or an increase in the probability of graduat-

ing and graduating on time if: i) the tenure in the program increases, ii) students received academic aid or financial aid.

2. An increase in the probability of dropping out, or a reduction in the probability of graduating and graduating on time if the share of failed classes increased.

Economic theory and literature expect that the drop-out probability decreases if the unemployment rate is high. It happens because the students prefer to stay in school while finding favorable conditions in the labor market. However, in countries like Colombia, where the new population attending college has low household incomes, an increase in the unemployment rate could also mean that the household income was affected, and such students needed to pull out of school to support their households.

In the case of tenure in the program, one more semester in the system reduces the probability of dropping out and increases the probability of graduation and graduation on time. An increase in the "Share of failed classes" during the last semester increases the probability of dropping out and reduces the probability of graduating and graduating on time, similar to the findings in the prior literature (Ministerio de Educación Nacional, 2008, 2010; Herrera-Prada, 2013).

I find that receiving tutoring and mentoring reduces the probability of dropping out, but the interaction of tutoring and mentoring with SPADIES increases the probability of dropping out. This result can be explained because using SPADIES, HEIs targeted students with high academic vulnerabilities at a higher risk of dropping out, and the aid received was not enough to save them. The interaction of SPADIES with academic aid and public financial aid shows an increase in the transitions from absent and from dropout. In contrast, the interaction of SPADIES with private financial aid shows a reduction in the transitions from absent. A reduction in the transitions from absent without a corresponding increase in the transitions from dropout is a sign of the positive influence of SPADIES, as these forms of aid provided by the HEI are helping to prevent students from leaving school due to their financial situation. The results show that aids were well-targeted, but they were not enough. Significative results for interactions reflect the excellent targeting of the aid programs. In contrast, opposite expected signs signal that the aid was not enough to improve the odds for those students. Contrary to private financial aid that was more targeted on-demand per semester, public financial aid aimed to show benefits over the long run. The measure of success for public aid was whether or not the student graduated and the timeliness of their graduation. Also, the results support that public financial aid was targeted to skilled but financially vulnerable populations who will usually graduate on time if they do not drop out. This fact can explain the positive sign for the transition from dropouts receiving public financial aid, as students could get the degree. However, the increase in the transition gap and the positive sign in the interactions with academic aid and public financial aid suggest that students were brought back from dropout status to active status. SPADIES worked as a mechanism to look for the list of students already marked as dropouts and allowed the HEI to incentivize them to come back to school.

Next, analyzing the time-invariant results, I found that, compared with students in the rest of the country, the students in Bogota have more probability of dropping out and more probability of transitioning from either absent or drop-out to active less probability of graduating on time. Similar to ICFES (2002); Universidad Nacional de Colombia (2007); Ministerio de Educación Nacional (2008); SPADIES (2008) and Universidad Nacional de Colombia (2007), females have less probability of dropping out and more probability of graduating on time than their male counterparts. The younger the student is, the better the chances of not dropping out, but having a young age also decreases the probability of graduating or graduating on time. A younger age means fewer transitions each year, but the transition gap is bigger than older students. More household income or getting a higher score in the secondary test results in a reduction in the probability of being a drop-out student and increases the probability of graduating and doing it on time. Females, students with high scores on the secondary school exit exam, and high-income students have less probability of having a transition from absent or drop-out status.

[Table 4 about here.]

## 6.2 Parallel Trends

The assumption that allows unbiased estimation of the ATT for the SPADIES treatment is the existence of parallel trends. This assumption requires that the dependent variables follow the same trajectory prior to treatment in the treated groups and in the controls. Figure 4 presents the average of each of the outcomes for the last six semesters before the treatment of SPADIES. However, according to Callaway and Sant'Anna (2021), I only need a similar trend in the period pre-treatment (T= -1).

In terms of the probability of having a transition from absent, it is interesting that the trends in the periods pre-treatment are very similar between Rounds 1 and 3 and between Rounds 2 and 4. One possible explanation is that a transition from absent status only requires one semester, so this status is more sensitive to seasonality issues (e.g., wanting to work in the summer tourism months). Given the time of SPADIES installation, the cut in the data to establish the status of the student was mainly in the first half of the year for Rounds 1 and 3 and mainly in the second half of the year for Rounds 2 and 4. Depending on the HEI and time of the year, students could have different patterns if stopping their studies. Slopes for Rounds 2 and 4 are similar to Rounds 1 and 3 when I lag Rounds 1 and 3 by one semester. This effect is not as clear in terms of the probability of having a transition from drop-out, as the system requires students to be not active for more than 2 semesters; the potential seasonality would no longer affect this long period of time.

Finally, the trend for the time gap during transition is the same, regardless of the round. This trend suggests that students out of the system in periods pre-t-1 are structurally similar among Rounds and that HEIs could develop a plan to retain students in the semester pre-SPADIES, so they could show better results, as is evident by the hard decrease in the time gap line for all the rounds. This behavior did not affect the analysis as SPADIES revised the total population and not the flow for the last semester, as it is evident with the other variables analyzed.

## 6.3 Main Results

In this section, I present the causal results for SPADIES estimated using Equations 7, 9, 11 and 12 by uign the command CSDID from Rios-Avila et al. (2021). Results will be presented by the outcome in two panels: the left panel with the event analysis figure, and the right panel will present the coefficients of interest for the total of the program per Round.

### 6.3.1 Drop-out

SPADIES reduced the probability of dropping out by 7 bps (Figure 5). The results from the DiD estimators indicate that SPADIES had a positive impact on the probability of becoming a drop-out student by reducing it by 8 bps (OR, IPW, and DRI) (Figure 5). The most effective round of SPADIES, but also the one with the most time with which to see the effect, was Round 1 with a reduction in the probability of dropping out between 9 bps and 12 bps. Six semesters after the installation of SPADIES, its impact reached a reduction in the probability of dropping out between 20 bps and 80 bps.

#### 6.3.2 Graduation

SPADIES increased the probability of graduating by at least by 6 bps (Figure 6). Similar to the results in the drop-out component, the most effective round so far was Round 1, which saw an increase in the probability of graduating between 9 bps and 11 bps. Six semesters after applying the treatment, SPADIES's impact reached an increase in the probability of graduating between 21 bps and 79 bps.

#### 6.3.3 On-time graduation

SPADIES increased the probability of graduating on time by 4 bps (Figure 7). The results from the DiD estimators indicate that SPADIES had a positive impact on the probability of graduating on time, mainly for the Round 1. After six semesters, SPADIES increased the probability of graduating on time between 18 and 54 bps.

#### 6.3.4 Transition from absent

SPADIES increased the probability of having a transition from absent status by 3 bps (Figure 8). There is no significant difference between the rounds in the probability of dropping out, graduating, or graduating on time. The flatness in the results can be explained by the short gap period that the transition requires, i.e., one semester. All the rounds saw an increase in the number of transitions for students marked as absent at the same pace (no significant difference between rounds). The impact is similar for every semester. All the semesters after installation of SPADIES show about the same rate of increase in the number of transitions for students marked as absent.

#### 6.3.5 Transition from drop-out

SPADIES increased the probability of having a transition from drop-out status by 2 bps (Figure 9). This variable is the only one that shows one round having a better result than the others; Round 4 has the best results. This result can be explained because once the HEIs were trained, they used the list of students at risk to better target the aid programs, but they were also able to generate the list of students marked as drop-outs to call them back to the system. As the time since treatment is shorter for Round 4, the better results primarily reflect this call to drop-out students to return to school. Other Rounds did the same, but as they had more time to progress, the effect of the callback disappeared overtime once the number of called back students decreased, and the new drop-out students required two new semesters to be marked and traced. Now, remember that once students are reported as absent, the HEI could call them back, so it is not easy to reach the drop-out status after SPADIES was installed. Inline with this, the analysis after the installation of SPADIES reflects an improvement in drop-out transitions as HEIs are more experienced in tracking their drop-outs and calling them back to school. While the results per round showed that the effect decreased over time, the results per semester showed that HEIs were more efficient at tracking and bringing back students from drop-out status. However, this effect dilutes over time, as seen in six semesters after the treatment of SPADIES.

### 6.3.6 Time gap of transition

SPADIES increased the time gap during the transition by 0.6 semesters (Figure 10). This result is closely related to the transitions from dropouts, as the increase in the differentials I find in the results of Figure 9 are explained by the return to the classroom of students who were marked as dropouts and brought back by the HEI. These students, who were outside the system, had several periods outside the system and would have remained outside the system if not for the utilization of SPADIES. The increase in the time gap reported is positive because it represents the return of students who would not have come back to school otherwise. Those students who had transitioned from dropout status before the implementation of SPADIES were students who planned to return regardless of SPADIES, so the time gap was not as long. These students who returned because HEIs targeted them through SPADIES had been out of the system for a significant number of years; the influx of these returning students led to the increase in the time gap. Likewise, this gap is more significant in Round 4 due to the short time to incorporate new dropouts into the average of all transitions.

### 6.4 Dissagregated Results

In the last part of this section, I report the results by dividing the sample using the time-invariant variables. Figures 11 to 14 show the coefficient for each of the estimations methods proposed for the DiD analysis. In terms of the probability of dropping out, the results suggest that SPADIES was more effective for males with low income from public HEIs. Only the results by the HEI sector (public or private) are significantly different; those students in public HEIs saw a more considerable reduction in the probability of dropping out.

In terms of the probability of graduating, this population (males with low income from public HEIs) experienced an increased in the probability of graduating. However, there is a significant difference between those HEIs with quality certifications and those without. Similar results, but without significant difference among categorthe subpopulations analized are found in regards to the probability of graduating on time.

In analyzing the probability of transitioning from absent status, I find that those students from non-certified HEIs and students in associate programs were more likely to transition from absent. In contrast, those males students from associate programs in public HEIs were more likely to have a transition from drop-out.

In terms of the time gap during transitions, the student population that most benefited population from SPADIES were the males with low income from public HEIs, and mainly in associate programs; these are also the students who were most likely to have already dropped out and would not have returned had the HEIs not used SPADIES to target them and call them back to school. This subset of the student population was expected to be the most benefited by SPADIES as HEis were instructed to target their aid programs to them, according to Ministerio de Educación Nacional (2008, 2010).

## 7 Conclusions

The new millennium brought a complex educational and labor market challenge for Colombia – the children of the boomers reached higher education. They were not only greater in number than previous generations, but they were also significantly less prepared and had lower skills than their predecessors, putting them at higher risk of becoming dropouts (Orozco Silva et al., 2006; Ministerio de Educación Nacional, 2010; Ferreyra et al., 2017).It caused a sharp deterioration in the quality of the entire higher education system, affecting enrollment and graduation rates. (Ministerio de Educación Nacional, 2010; Orozco Silva, 2010; Herrera-Prada, 2013). Moreover, Colombia faced this challenge with severely scarce economic resources, but the Ministry of Education designed a plan to address the education crisis.(Ministerio de Educación Nacional, 2010; Orozco Silva et al., 2011).

Colombia developed a strategy that would use dynamic system tools to respond to the challenge; they called this strategy the "Educational Revolution"(Orozco Silva, 2010; Ministerio de Educación Nacional, 2010). One of the MEN's main tools was SPADIES, a software application to collect higher education student data with a user-friendly interface that would help institutions target at-risk students for anti-dropout assistance (Ministerio de Educación Nacional, 2008). In fact, the requirement to participate in this "revolution" modernized the protocols and records for many HEIs that, when used with SPADIES's tools provided information to the different agents in this market: authorities in the MEN and HEIs, prospective and current students, employers, and the public at large (Ministerio de Educación Nacional, 2010, 2017). All the agents were able to know in "real-time" statistics of the most important outcomes for the higher education system.

The flow of information was crucial; not only were HEIs receiving training about how to operate the dashboard, but they were also informed about strategies and protocols that other HEIs were successfully used to reduce the dropout rate. HEIs were able to compare themselves to their peers on the main educational outcomes. Students were aware of what was happening and the relative standing of their schools, as the dropout rates were reported widely in media, and these rates were also used to promote the programs and the HEIs. The students with the greatest need, both active and dropouts, became the main target of various new programs and aid. Thus, no matter the equilibrium model used, whether the maximization of income or quality from Epple et al. (2006) or the effect of the peers from MacLeod and Urquiola (2015), my analysis shows that the information provided by SPADIES directly impacted the educational outcomes rates. Improvements in outcomes are evidence of the poor quality of information pre-SPADIES.

New information provided by SPADIES enabled the Colombian higher education system to reach a new, more efficient, and more socially desirable equilibrium. Any of the recent equilibrium models explains the path to this new equilibrium. On the one hand, in Epple et al. (2006)– where a set of HEIs maximize quality and other profits, and the students attend HEIs depending on their skills and financial capabilities – the students choose where to study depending on the institution's quality, the cost of the tuition, and the admission policies. HEIs with high quality will be more selective, where selection can be by price, admission policies, or both. Reputation matters. So for high-quality HEIs, SPADIES will impact the selectivity; HEIs will accept only those students with a lower risk of becoming dropout students, targeting to receive students with higher Saber 11 scores and/or high household income. For the low-quality HEIs, SPADIES information becomes a promotional tool that HEIs could use to attract more students. SPADIES also worked to promote their social job; they could show how many and how efficient their aid programs were. Finally, students will attend HEIs where their chances of getting the degree are worth the risk of risk of becoming dropouts.

On the other hand, the peer's effect from MacLeod and Urquiola (2015) the increased information of HEI quality triggered greater competition among HEIs, so they increased the number of programs to provide aid to at-risk students, targeted their resources more effectively, and better-traced students throughout their college experience. One impressive result of the MEN's use of SPADIES was the increase of transitions, and it is no doubt due to the greater competition among HEI peers. Also, as a result of SPADIES, students were able to enroll HEIs that had better aid programs to reduce dropout rates or increase graduation rates. The competition also resulted in some HEIs making it easier to attain the degree (e.g., they removed the fee required to attain the degree). These changes reduced the barriers to graduation, particularly in low-quality institutions, and allowed more students to graduate. The data, however, does not allow me to track those students who transferred from one HEI to another. Future research exploring the migration between HEIs would substantially contribute to the literature. Analyzing how the increased flow of information and competition lead to a redistribution of some students within the institutions and how it affected their future income would be an essential avenue for further research on increasing competition among HEIs and impacts for transfer students.

Overall, my research shows that SPADIES resulted in higher student retention and higher rates of on-time graduation and graduation overall. SPADIES achieved its primary purpose in reducing the dropout rate for all types of students, particularly for the population most at-risk of dropping out – males with low household income and low academic skills at public institutions -10,4% of the total population- (see Figures 10 to 14).

SPADIES reduced the probability of becoming a dropout student by 7 basic points (bps); this reduction is equivalent to about 14,000 students saved from becoming dropouts in the system. This figure is impressive when considering that the average size of an HEI in Colombia is 8,000 students. Equally important, SPADIES increased the probability of earning the degree by 6 bps and earning it on time by 4 bps. It means that SPADIES helped the equivalent of about additional 12,000 students earn their degree, of which 8,000 earned the degree on time.

I find that SPADIES increased the number of transitions from absent and from dropout by 3 bps and 2 bps, respectively. The average duration of a transition increased by 0.5 semesters. The results also showed that non-certified HEIs were very good at using SPADIES to bring back students from absent status, while certified public HEIs were very good at using SPADIES to bring back students from dropout status. In addition, SPADIES increased the average time gap during the transition by 0.6 semesters. I interpret this result as positive because those who had transitioned from dropout status before SPADIES implementation were students who planned to return (i.e., students with lower transition times on average), while those who returned to school from dropout status thanks to SPADIES had a more prolonged period outside of school. These former dropout students had perhaps been condemned to never return to school if not for SPADIES. The return of these students to the education system after SPADIES increased the overall time gap.

Many HEIs used information from SPADIES to target new aid programs. Interestingly, I find that receiving tutoring and mentoring reduced the probability of dropping out, but the interaction of tutoring and mentoring with SPADIES increased the probability of dropping out. Because the at-risk students targeted with SPADIES time were those with the highest academic vulnerabilities, they were already at a higher risk of dropout, and SPADIES plus the aid support were not enough to save them from becoming dropouts (see Table 3). Despite the previous result, the increase in transitions and the time gap during transitions suggest that SPADIES was used to target students in aid programs and track students who became dropouts and call them back to school (see Table 3 and Figures 7-9).

Finally, I find that SPADIES improved the higher education system's efficiency because it helped decongest entry into the system of a ballooning student population. SPADIES helped more students to graduate and to do so on time. By reducing the time it takes students to graduate, SPADIES gradually reduced the burden of overpopulation on HEIs, which helped improve the enrollment rate. HEIs were allowed to serve more new students than they could have otherwise. SPADIES served as a bridge connecting students with information about HEIs and HEIs with information about students. Students could be more selective in their choice of attending an HEI, and HEIs were better able to mitigate the vulnerabilities of the new population that were arriving to higher education.

SPADIES also had two unintended spill-overs; it helped dropout students return to school and brought the higher education system into the digital era. Many HEIs were still using paper records prior to the implementation of SPADIES.

SPADIES was instrumental in breaking down the musical chairs game that was the higher education system in Colombia. By providing information to the various agents in the marketplace, SPADIES improved the system's efficiency and changing the future of many students who would have otherwise dropped out.

## References


**Abadie, Alberto**, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, jan 2005, *72* (1), 1–19.

**Adams, Jack E.**, "Affirmative Action in Higher Education: A Sourcebook," *The Journal of Higher Education*, 1984, *55* (1), 113.

**Allende, Claudia, Francisco Gallego, and Christopher Neilson**, "Approximating the Equilibrium Effects of Informed School Choice," *Working Paper*, 2019.

**Asif, Raheela, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider**, "Analyzing undergraduate students' performance using educational data mining," *Computers and Education*, 2017.

**Auguste, Byron, Paul Kihn, and Matt Miller**, "Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching," Technical Report, Mckinsey and Company 2010.

**Bank, Barbara J., Ricky L. Slavings, and Bruce J. Biddle**, "Effects of Peer, Faculty, and Parental Influences on Students' Persistence," *Sociology of Education*, 1990.

**Barro, Robert J.**, "Human capital: Growth, history, and policy - A session to honor Stanley Engerman: Human capital and growth," *American Economic Review*, 2001, *91* (2), 12–17.

**Bassi, Marina, Matías Busso, Sergio Urzúa, and Jaime Vargas**, "Desconectados Habilidades, educación y empleo en América Latina," 2012, p. 36.

**Bean, John P**, "Dropouts and turnover: The synthesis and test of a causal model of student attrition," *Research in Higher Education*, 1980.

_ , "Interaction Effects Based on Class Level in an Explanatory Model of College Student Dropout Syndrome," *American Educational Research Journal*, 1985.

**Becker, Gary**, "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 1962, *70* (S5), 9.

**Berens, Johannes, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff**, "Early Detection of Students at Risk. Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods," *SSRN Electronic Journal*, 2021.

**Bound, John and Sarah Turner**, "Cohort Crowding: How Resources Affect Collegiate Attainment.," 2006, *12424.*

**_ , Brad Hershbein, and Bridget Terry Long**, "Playing the admissions game: Student reactions to increasing college competition," in "Journal of Economic Perspectives" 2009.

**_ , Michael Lovenheim, and Sarah Turner**, "Understanding the decrease in college completion rates and the increased time to the baccalaureate degree," *Ann Arbor, Mich.: Population Studies Center, University of Michigan Institute for Social Research*, 2007, pp. 1–91.

**Braxton, John M**, "Student Success in College: Creating Conditions that Matter," *Journal of College Student Development*, 2007.

**Cabrera, Alberto, Linda Serra Hagedorn, Amaury Nora, and Ernest Pascarella**, "Differential impacts of academic and social experiences on college-related behavioral outcomes across different ethnic and gender groups at four-year institutions," *Research in Higher Education*, 1996.

**_ , María Castañeda, and Amaury Nora**, "College Persistence: Structural Equiations Modeling Test of Integrated Model of Student Retention," *The Journal of Higher Education*, 1993, *64* (2), 123–139.

**Callaway, Brantly and Pedro H.C. Sant'Anna**, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

**Campbell, Carol and Ben Levin**, "Using data to support educational improvement," *Educational Assessment, Evaluation and Accountability*, 2009.

**Cárdenas, Ernesto**, "Estudio de la deserción estudiantil en programas de ingeniería en la Universidad Nacional de Colombia." Tesis de maestria en dirección universitaria, Universidad de los Andes 1996.

**Castaño, Elkin, Santiago Gallón, Karoll Gómez, and Johanna Vásquez**, "Análisis de los factores asociados a la deserción y graduación estudiantil universitaria," *Lecturas de Economia*, 2006, *65*, 9–36.

**Córtes, Hernán, Luis Gallego, and Gerardo Rodríguez**, "The engineering faculty today: an approach towards consolidating academic indicators," *Ingeniería e Investigación*, 2011, *31* (1).

**Cruces, Guillermo**, "Inequality in education: evidence for Latin America," *Falling Inequality in Latin ...*, 2014.

**de Roux, Nicolás and Evan Riehl**, "Disrupted academic careers: the returns to time off after high school," 2019.

**Epple, Dennis, Richard Romano, and Holger Sieg**, "Admission, tuition, and financial aid policies in the market for higher education," 2006.

**Facundo-Díaz, Ángel**, "Análisis sobre la deserción en la educación superior a distancia y virtual: el caso de la UNAD - COLOMBIA," *Revista de investigaciones UNAD*, 2009, *8* (2).

**Ferreyra, María Marta, Ciro Avitabile, Javier Botero Álvarez, Francisco Haimovich Paz, and Sergio Urzúa**, *Momento decisivo: La educación superior en América Latina y el Caribe* 2017.

**Gentsch, Kerstin**, "How admission policy shapes college access: Evidence from two sectors." PhD dissertation 2016.

**Glewwe, Paul W., Eric Hanushek, Sarah D Humpage, and Renato Ravina**, "School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010," 2011.

**Goodman-Bacon, Andrew**, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, dec 2021, *225* (2), 254–277.

**Hanushek, Eric and Ludger Woessmann**, "Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries," *Economic Journal*, 2006, *116* (115), C63–C76.

— **and** — , “Schooling, educational achievement, and the Latin American growth puzzle,” *Journal of Development Economics*, 2012, *99* (2), 497–512.

**Hastings, Justine S and Jeffrey M. Weinstein**, “Information, School Choice, and Academic Achievement: Evidence from Two Experiments,” *Quarterly Journal of Economics*, 2008, *123* (4), 1373–1414.

**Herrera-Prada, Luis Omar**, “Determinantes de la tasa de graduación y graduación a tiempo en la educación superior de Colombia 1998-2010,” *Coyuntura Económica*, 2013, *XLIII* (1), 143.177.

— , “The Economic Consequences of Setting Foot in a College in Colombia,” 2021.

**ICFES**, “Estudio de la deserción estudiantil en la educación superior en Colombia.,” Technical Report, Universidad Nacional - ICFES, Bogotá 2002.

**Kerr, Sari Pekkala, Tuomas Pekkarinen, Matti Sarvimäki, and Roope Uusitalo**, “Post-secondary education and information on labor market prospects: A randomized field experiment,” *Labour Economics*, 2020.

**Lucio, Ricardo and Mariana Serrano**, *La Educación Superior: Tendencias y Políticas Estatales.* 1992.

**MacLeod, W. Bentley and Miguel Urquiola**, “Reputation and school competition,” 2015.

**Ministerio de Educación Nacional**, “Análisis de determinantes de la deserción en la educación superior colombiana con base en el SPADIES.,” Technical Report, Ministerio de Educación Nacional - Universidad de los Andes, Bogotá 2008.

— , “Deserción estudiantil en la educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos para su prevención.,” Technical Report, Bogotá 2009.

— , “La Revolución Educativa 2002 - 2010. Informe de gestión.,” Technical Report, Bogotá 2010.

— , “Boletín Educación Superior,” Technical Report, Bogotá 2017.

**Mizala, Alejandra and Miguel Urquiola**, “School markets: The impact of information approximating schools’ effectiveness,” *Journal of Development Economics*, 2013.

**Olaya, Diego, Jonathan Vásquez, Sebastián Maldonado, Jaime Miranda, and Wouter Verbeke**, "Uplift Modeling for preventing student dropout in higher education," *Decision Support Systems*, 2020, *134* (May), 113320.

**ONU**, "Objetivos de Desarrollo del Milenio," *Naciones Unidas*, 2013, p. 64.

**Orozco Silva, Luis Enrique**, *La Política de Cobertura: eje de la revolución educativa, 2002-2008.*, Bogotá: Ediciones Uniandes, 2010.

_ **, Alberto Roa Valero, and Luis Carlos Castillo Gómez**, "La Educación Superior en Colombia," 2011, pp. 1–83.

_ **, Javier Medina Vásquez, María Pérez Piñeros, and Alberto Roa Valero**, "Informe Colombia," in "Proyecto Informe Sobre la Educación Superior en Iberoamérica" 2006.

**Park, Kang H and Peter M Kerr**, "Determinants of academic performace: A multinomial Logit Approach," *Journal of Economic Educations*, 1990, *21*, 101–111.

**Portes, Pedro R. and Spencer Salas**, "The dream deferred: Why multicultural education fails to close the achievement gap. A cultural historical analysis," *Cultura y Educación*, 2007.

**Porto, Alberto and Luciano Di Gresia**, "Rendimiento de estudiantes universitarios y sus determinantes," *Revista de Economía y Estadística*, 2004.

**Rios-Avila, Fernando, Brantly Callaway, and Pedro H.C. Sant'Anna**, "csdid: Difference-in-Differences with Multiple Time Periods in Stata," 2021.

**Saavedra, Juan and Carlos Medina**, "Formación para el Trabajo en Colombia *," 2012.

**Sánchez, Fabio, Margarita Quirós, Carlos Reverón, and Alberto Rodríguez**, "Equidad Social En El Acceso Y Permanencia En La Universidad Pública Determinantes Y Factores Asociados," 2002, *7191*, 1–48.

**Sant'Anna, Pedro H.C. and Jun Zhao**, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, nov 2020, *219* (1), 101–122.

**SPADIES**, "Reporte Modelo SPADIES al MEN," Technical Report, Universidad de los Andes, Bogotá 2008.

**St. John, Edward P., Alberto Cabrera, Amaury Nora, and Eric H. Asker**, "Economic Influences on Persistence Reconsidered," in "Reworking the Student Departure Puzzle," Vanderbilt University Press, oct 2016, pp. 29–47.

**Stratton, Leslie S., Dennis M. O'Toole, and James N. Wetzel**, "A multinomial logit model of college stopout and dropout behavior," *Economics of Education Review*, 2008, *27* (3), 319–331.

**Sun, Liyang and Sarah Abraham**, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, dec 2021, *225* (2), 175–199.

**Tinto, Vincent**, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Review of Educational Research*, 1975.

_ , "Limits of Theory and Practice in Student Attrition," *The Journal of Higher Education*, 1982.

**Trow, Martin**, "Problems in the Transition from Elite to Mass Higher Education," *International Review of Education*, 1974, *18*, 61–82.

**Unesco**, "Using data to improve the quality of education," 2020.

**Universidad Nacional de Colombia**, *Cuestión de Supervivencia. Graduación Deserción y Rezago*, Bogotá: Beta Impresores Ltda, 2007.

Table 1: Students description

| Variable | Obs. | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|---|
| Drop-out rate | 4,131,302 | .478 | .5 | 0 | 1 |
| Graduation rate | 4,131,302 | .275 | .447 | 0 | 1 |
| On-time graduation rate | 4,131,302 | .222 | .416 | 0 | 1 |
| Transitions | 4,131,302 | .055 | .228 | 0 | 1 |
| Transitions from absent | 4,131,302 | .029 | .167 | 0 | 1 |
| Transition from drop-out | 4,131,302 | .026 | .159 | 0 | 1 |
| Transitions time gap | 4,131,302 | .932 | 1.15 | 0 | 29 |
| Tenure in program | 4,131,302 | 4.97 | 3.42 | 1 | 35 |
| Share of failed classes | 4,131,302 | .119 | .244 | 0 | 1 |
| Received tutoring or mentoring | 4,131,302 | .121 | .327 | 0 | 1 |
| Received financial aid | 4,131,302 | .254 | .435 | 0 | 1 |
| Female | 4,131,302 | .502 | .5 | 0 | 1 |
| Year of birth | 4,131,302 | 1988 | 5.95 | 1960 | 1998 |
| Secondary test score | 4,131,302 | 61.8 | 28.4 | 1 | 100 |
| Students with secondary test score >= 90 | 4,131,302 | .202 | .402 | 0 | 1 |
| Household income | 4,131,302 | 1.74 | 1.33 | 0 | 9 |
| Unemployment rate | 4,131,302 | 11 | 2.58 | 5.87 | 22.3 |

Source: ICFES-HEIs. The Unemployment rate from DANE (National statistical office).

Table 2: Higher Education System data description

| Variable | Obs. | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|---|
| Public institution | 4,131,302 | .421 | .494 | 0 | 1 |
| High quality institution | 4,131,302 | .308 | .462 | 0 | 1 |
| Main institution | 4,131,302 | .656 | .475 | 0 | 1 |
| Good data report | 4,131,302 | .964 | .187 | 0 | 1 |
| Institution located in Bogota | 4,131,302 | .415 | .493 | 0 | 1 |
| Institution located in Valle del Cauca | 4,131,302 | .065 | .247 | 0 | 1 |
| Institution located in Antioquia | 4,131,302 | .15 | .357 | 0 | 1 |
| Institution located in Atlantico | 4,131,302 | .056 | .229 | 0 | 1 |
| HEI from 1 round of implementation including sub-locations | 4,131,302 | .273 | .445 | 0 | 1 |
| HEI from 2 round of implementation including sub-locations | 4,131,302 | .196 | .397 | 0 | 1 |
| HEI from 3 round of implementation including sub-locations | 4,131,302 | .231 | .421 | 0 | 1 |
| HEI from 4 round of implementation including sub-locations | 4,131,302 | .297 | .457 | 0 | 1 |
| HEI from 5 round of implementation including sub-locations | 4,131,302 | .003 | .058 | 0 | 1 |

Source: ICFES-HEIs.

Table 3: SPADIES rounds description

| | Round | | | | | Total |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | |
| Drop-out rate | .457 | .482 | .464 | .505 | .485 | .478 |
| Graduation rate | .313 | .245 | .274 | .262 | .187 | .275 |
| On-time graduation rate | .231 | .2 | .227 | .226 | .18 | .222 |
| Transitions = 1 | .055 | .047 | .053 | .062 | .042 | .055 |
| Transitions from absent | .031 | .027 | .029 | .029 | .025 | .029 |
| Transition from drop-out | .024 | .02 | .023 | .034 | .017 | .026 |
| Transitions time gap | .976 | .908 | .92 | .919 | .78 | .932 |
| Received tutoring or mentoring | .085 | .3 | .087 | .064 | .079 | .121 |
| Any kind of Aid = 1 | .228 | .351 | .244 | .218 | .451 | .254 |
| Female = 1 | .474 | .513 | .519 | .507 | .452 | .502 |
| Year of birth | 1988 | 1988 | 1987 | 1987 | 1988 | 1988 |
| Secondary test score | 73 | 60 | 60.8 | 53.5 | 56.5 | 61.8 |
| Students with secondary test score >= 90 | .362 | .168 | .177 | .099 | .119 | .202 |
| Household income | 1.99 | 1.68 | 1.73 | 1.56 | 1.65 | 1.74 |
| Unemployment rate | 11.7 | 10.5 | 10.8 | 11 | 10.2 | 11 |
| HEI public sector = 1 | .547 | .423 | .34 | .371 | .198 | .421 |
| High quality institution = 1 | .541 | .314 | .324 | .08 | .007 | .308 |
| Main institution (campus) = 1 | .717 | .617 | .559 | .702 | .455 | .656 |
| Good data report = 1 | .987 | 1 | .99 | .909 | 0 | .964 |
| Institution located in Bogota | .337 | .43 | .398 | .493 | .226 | .415 |
| Institution located in Valle del Cauca | .117 | .006 | .068 | .055 | 0 | .065 |
| Institution located in Antioquia | .196 | .181 | .101 | .123 | .266 | .15 |
| Institution located in Atlantico | .082 | .068 | .015 | .052 | .229 | .056 |
| Observations | 1,092,960 | 783,459 | 925,378 | 1,315,802 | 13,703 | 4,131,302 |
| Share of total | 26.46 | 18.96 | 22.40 | 31.85 | .33 | 100 |

Source: ICFES-HEIs. The Unemployment rate from DANE (National statistical office).

Table 4: Main Results Panel Regression Approach

| | Fixed Effects | | | | | | Random Effects | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | Drop-out | Graduated | On-time graduation | Transition from absent | Transition from drop-out | Time gap during transition | Drop-out | Graduated | On-time graduation | Transition from absent | Transition from drop-out | Time gap during transition |
| SPADIES | -0.014*** | 0.007*** | 0.007*** | -0.012*** | -0.018*** | -0.158*** | -0.013*** | 0.001*** | 0.004*** | -0.003*** | -0.000*** | -0.097*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Tenure in program | -0.000*** | 0.002*** | 0.001*** | 0.001*** | 0.003*** | -0.029*** | -0.002*** | 0.003*** | 0.002*** | -0.001*** | -0.001*** | -0.020*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Share of failed classes | 0.010*** | -0.008*** | -0.007*** | 0.007*** | 0.010*** | -0.083*** | 0.013*** | -0.011*** | -0.009*** | 0.031*** | 0.025*** | -0.012*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Received tutoring or mentoring | -0.023*** | 0.032*** | 0.017*** | -0.010*** | -0.008*** | -0.019*** | -0.021*** | 0.029*** | 0.016*** | -0.004*** | 0.005*** | -0.015*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.005) | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.004) |
| Received financial aid | -0.002*** | 0.005*** | 0.002*** | -0.005*** | -0.005*** | 0.010*** | -0.003*** | 0.005*** | 0.002*** | -0.004*** | -0.005*** | 0.012*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Unemployment rate | 0.004*** | 0.002*** | 0.001*** | 0.002*** | 0.006*** | 0.042*** | 0.004*** | 0.004*** | 0.002*** | 0.000*** | 0.001*** | 0.025*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Institution located in Bogota | | | | | | | 0.010*** | -0.025*** | -0.020*** | 0.003*** | 0.001*** | -0.042*** |
| | | | | | | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Female | | | | | | | -0.092*** | 0.076*** | 0.075*** | -0.003*** | -0.004*** | 0.014*** |
| | | | | | | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Year of birth | | | | | | | -0.003*** | -0.015*** | -0.011*** | -0.001*** | -0.001*** | 0.001*** |
| | | | | | | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Students with secondary test score >= 90 | | | | | | | -0.132*** | 0.142*** | 0.083*** | -0.001*** | -0.003*** | 0.062*** |
| | | | | | | | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.001) |
| Household income | | | | | | | -0.015*** | 0.003*** | 0.005*** | -0.001*** | -0.001*** | -0.004*** |
| | | | | | | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Academic aid x SPADIES | 0.027*** | -0.025*** | -0.014*** | 0.009*** | 0.008*** | 0.062*** | 0.026*** | -0.023*** | -0.014*** | 0.004*** | -0.005*** | 0.035*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.005) | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.004) |
| Private financial aid x SPADIES | -0.002*** | -0.004*** | -0.001*** | -0.002*** | -0.000 | 0.007*** | -0.002*** | -0.006*** | -0.002*** | 0.002*** | -0.001*** | 0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Public financial aid x SPADIES | 0.003*** | 0.007*** | 0.007*** | 0.004*** | 0.007*** | 0.095*** | 0.003*** | 0.011*** | 0.010*** | 0.001*** | -0.000 | 0.086*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Constant | 0.236*** | 0.434*** | 0.313*** | 0.016*** | -0.041*** | 0.756*** | 5.722*** | 30.369*** | 21.880*** | 1.243*** | 1.382*** | -2.156*** |
| | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.003) | (0.082) | (0.081) | (0.074) | (0.015) | (0.013) | (0.113) |
| Observations | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 | 22,453,352 |
| Number of id | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 | 4,131,302 |
| $R^2$ | 0.0142 | 0.103 | 0.0179 | 0.000554 | 0.000754 | 0.00676 | 0.0307 | 0.103 | 0.0417 | 0.00294 | 0.00453 | 0.00716 |
| Dependent variable mean | 0.478 | 0.275 | 0.222 | 0.055 | 0.029 | 0.026 | 0.478 | 0.275 | 0.222 | 0.055 | 0.029 | 0.026 |

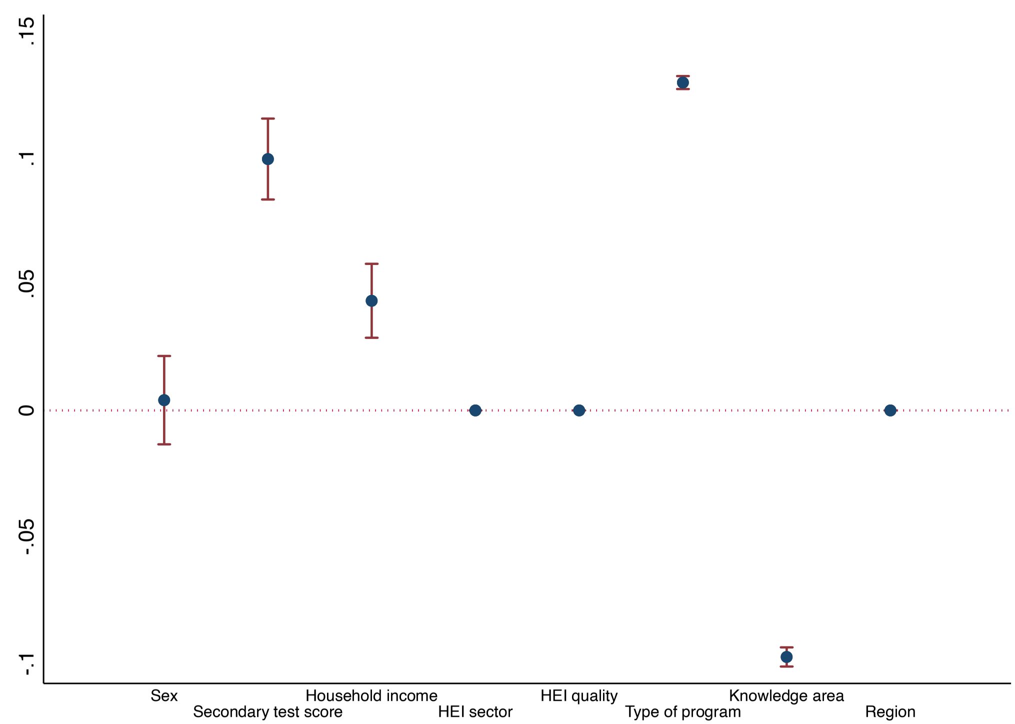
Note: The table shows the results for the estimations using Equation 1. SPADIES is a dummy of 1 if the student was enrolled in a semester after the SPADIES was installed in its HEI. Tenure in the program is expressed in semesters. Share of failed classes is estimated in t+1 as the ratio of the failed classes and total classes reported by the HEI in t. The HEI reports academic and Private financial aid to SPADIES. The unemployment rate is the semestral average for the region where the HEI is located. MEN reports the location of the HEI in the HEI's directory. Female is a dummy that takes the value of 1 if the student report being female at the moment of the secondary test score. Year of birth is reported by the student at the moment of the secondary test. Students with a secondary test score >=90 is a dummy that is 1 if the student has a score higher than 90. Household Income is an increasing categorical variable reported by the student at the moment of the secondary test. Source: SPADIES.

Figure 1: HEIs distribution by round

| | Main Campus | Other Campuses |
|---|---|---|
| Round 1 | 1101 1111 1121 1201 1202 1203 1204 1209 1301 1701 1704 1707 1710 1711 1712 1713 1719 1735 1801 1803 1804 1805 1812 1813 2711 3201 | 1102 1103 1104 1124 1125 1210 1219 1220 1221 1222 1223 1702 1705 1708 1716 1723 1730 1731 1802 1817 1829 2737 9122 9125 |
| Round 2 | 1106 1110 1113 1117 1119 1120 1218 1703 1706 1709 1715 1725 1729 1815 2302 2704 2707 2712 2721 2727 2744 2746 2749 2805 2811 2812 2813 2815 2829 3115 3117 3302 3705 3803 3830 4102 4801 4808 | 1107 1108 1109 1123 1724 1732 1822 2841 3116 |
| Round 3 | 1105 1112 1114 1205 1206 1208 1213 1214 1714 1722 1726 1728 1806 1823 1824 1827 1828 1830 1832 1833 2104 2209 2702 2708 2710 2713 2719 2723 2724 2725 2728 2732 2823 2825 2832 2847 2850 3301 3702 3710 3713 4101 4111 4702 4711 4714 4721 4726 4832 5802 | 1215 1216 1733 1807 1808 1809 1810 1811 1834 |
| Round 4 | 1115 1118 1122 1207 1212 1217 1718 1720 1734 1814 1818 1825 1826 1831 1835 2102 2106 2110 2114 2206 2207 2208 2211 2301 2701 2709 2715 2720 2730 2731 2733 2736 2738 2739 2740 2741 2745 2747 2748 2810 2818 2820 2824 2827 2828 2830 2831 2833 2834 2837 2838 2840 2842 2848 2849 3102 3103 3104 3107 3204 3703 3706 3712 3715 3716 3718 3719 3720 3725 3801 3805 3806 3807 3808 3809 3810 3811 3812 3817 3819 3820 3821 3822 3824 3826 3827 3828 3829 3831 3833 3834 4106 4107 4108 4109 4110 4112 4701 4705 4708 4709 4716 4719 4727 4803 4806 4810 4811 4813 4817 4818 4822 4825 4826 4827 4829 4835 4837 5801 9101 9116 9117 9119 9120 9121 9124 9126 9128 9131 | 1717 1816 1819 1820 |
| Round 5 | 2743 2836 2902 2903 2906 3114 3303 3724 3802 4812 9102 9127 9129 9132 9899 9903 | |

Notes: Five rounds -Round 1(2005-06), Round 2 (2006-07), Round 3 (2007), Round 4 (2008-09), and Round 5 (2010-11)- were necessary to complete all the HEIs' information into the system. The MEN and the Universidad de Los Andes visited and installed the dashboard in the main HEI; MEN expected that the main HEI shared SPADIES with its other campuses."Other campuses" were included in the same round that its parent HEIs; they are an extension of one of the main HEIs in other regions (e,g. Universidad Nacional de Colombia code 1101 is the main public national university, located in Bogotá, and it is the parent of 1102 that is the campus located in Medellín. In some cases, the "Other campuses" administration is autonomous, and in other cases, it administration depends directly on the main campus. There is not a rule about this).

Figure 2: SPADIES assignation balance




Notes: The figure shows the regression coefficients explaining the treatment variable using the students' main characteristics and HEIs. Whiskers show the 95% confidence interval. Sex is a dummy that is 1 if the student is female. The secondary test score is a dummy that is 1 if the student is in the top 10

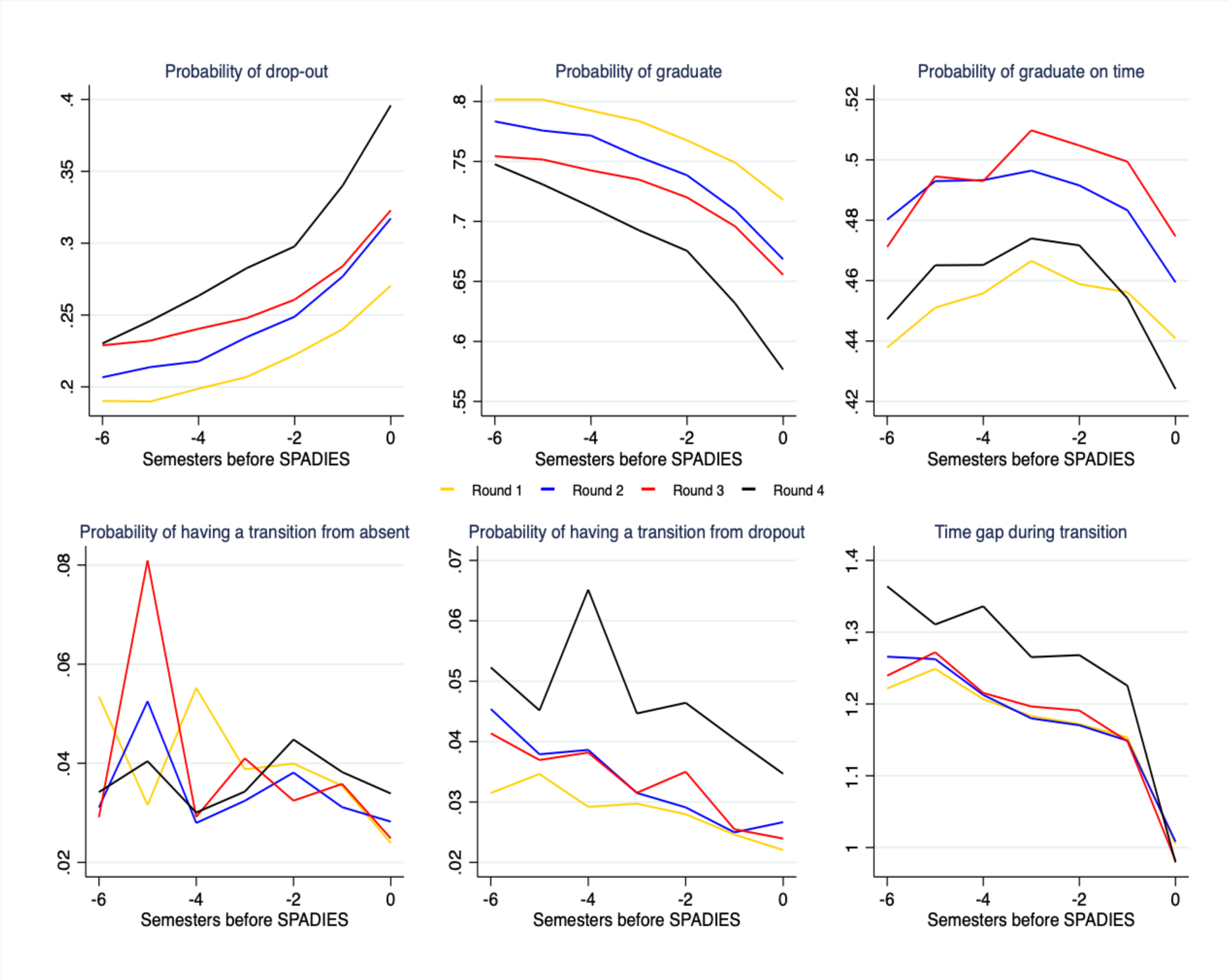
Figure 3: SPADIES status analysis

| | Semester | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Period of analysis | T-7 | T-6 | T-5 | T-4 | T-3 | T-2 | T-1 | T | T+1 | T+2 | T+3 | T+4 | T+5 |
| | Cohort | K | K+1 | K+2 | K+3 | K+4 | K+5 | K+6 | K+7 | K+8 | K+9 | K+10 | K+11 | K+12 |
| A | Freshmen | | | | | | | | | | | | | |
| B | Enrolled | A | | | C | I | C | D | II | B | | | | |
| C | Absent | A | | C | | C | | | C | | | | | |
| D | Drop-out | A | | | | | | C | D | D | D | D | D | D |
| E | Graduated | A | | | | | | C | D | | | E | | |
| F | Graduated on time | A | | | | | E (A) | E (A) | E (A) | | E (B) | E (B) | E (B) | |

| | |
|---|---|
| | Student is enrolled. |
| | Student is enrolled after a transition. Transition I = from absent to enrolled. Transition II = from drop-out to enrolled. |
| | Student graduated. Status will not change after this point. |
| D | Student dropped out (after 2 consecutive periods of absence). It is dynamic, if the student enrolls again, the status is updated (see transition II). |
| C | Student is absent. It is dynamic, if the student enrolls again, the status is updated (see transition I). |
| A | First period of enrollment of student in a HEI. |
| E | Student graudated on time (up to 2 periods after the standard/expected time for the completion). (A) from Associate programs (B) from Bachelor programs. |
| | It does not matter the history of transition; once the student is graduated, this status remains. |

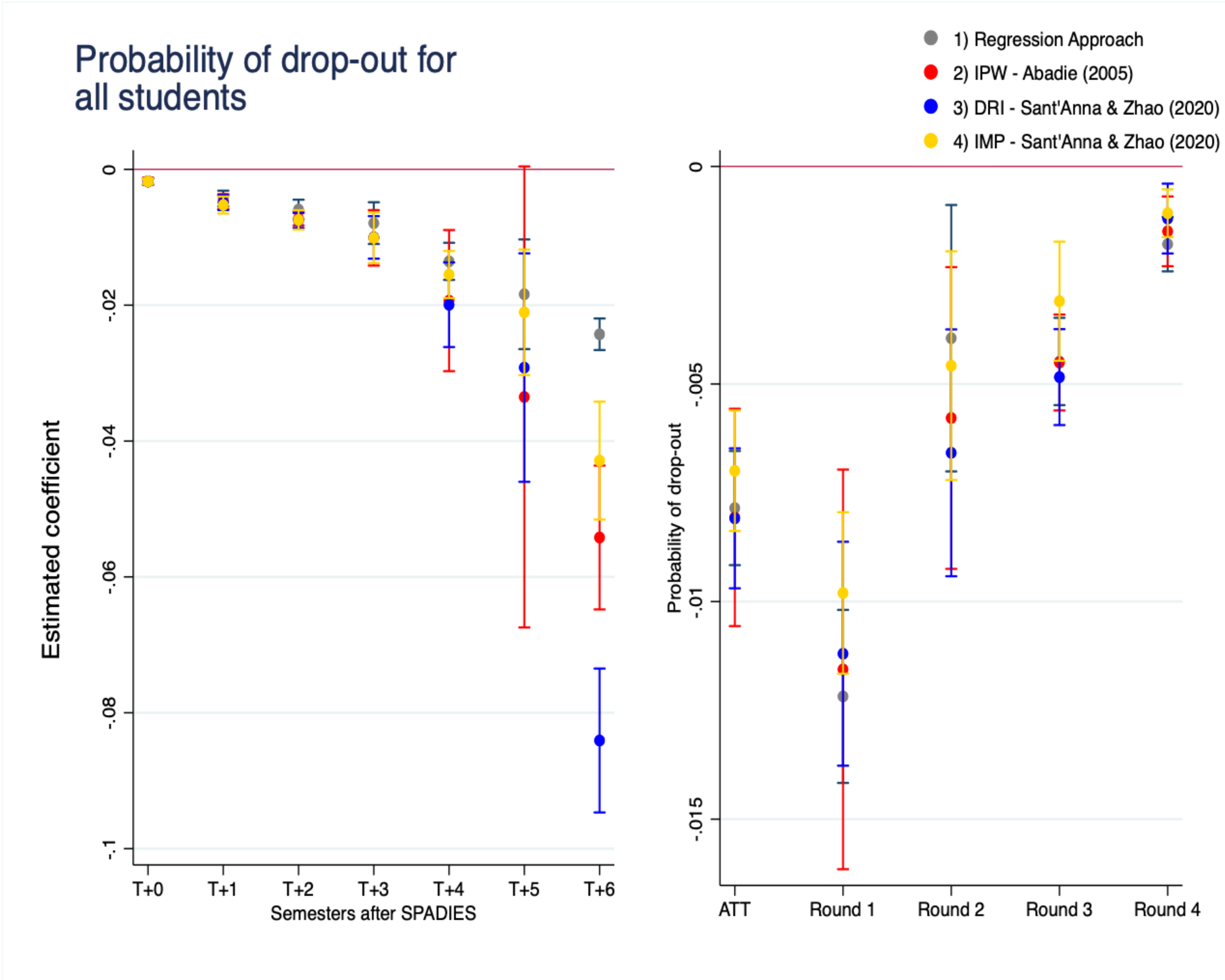
Notes: All the statuses in SPADIES are dynamic in time except by Graduated. Transition I shows an enrolled status once the student was Absent per 1 semester and returned to the system. Transition II shows that after being considered a dropout student the status was updated to "enrolled" once the student returned to the system. Letters from "A" to "F" show the status of the student. Each row shows an example of the status in T. "E(A)" if graduated from a tech and "E(B)" from a professional program. Letters in cells are reference to examples of dynamic statuses in time. Time is recorded as the times an student is enrolled; i.e. for cohort K, if our analysis is in period "T", an student who has been enrolled all the semesters will be in 8th semester, but none of the examples account this. In the example "B Enrolled" the student will account 5th, as the student has been enrolled in 5 semesters and considered absent in 2 semesters and dropout in 1. In the example "E Graduated" the student will account 6 semesters (1 as Absent and 1 as dropout).

Figure 4: Parallel Trends



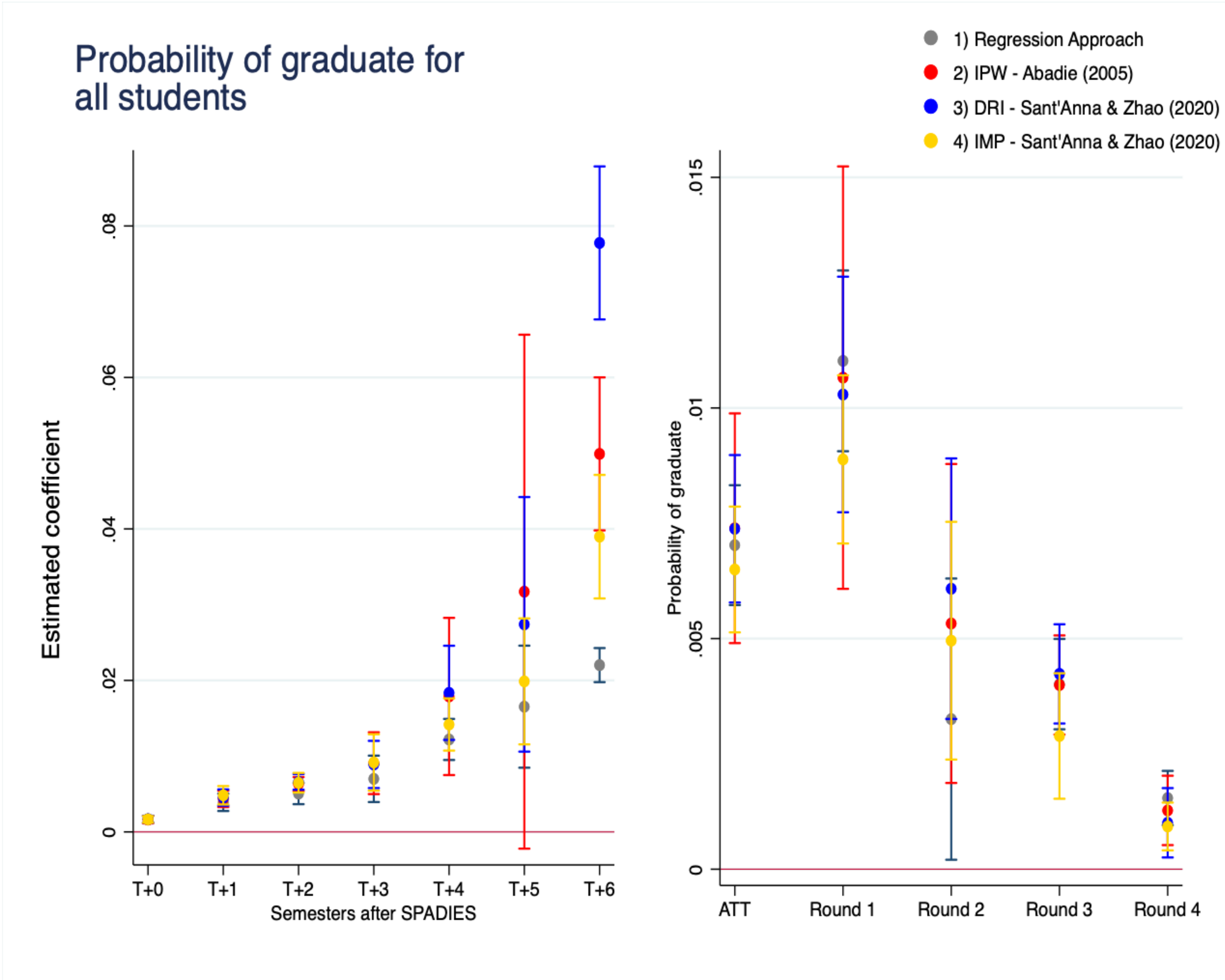

Notes: Chart reports the averages per round for the out put variables before SPADIES. Averages estimates using the full sample for rounds 1 to round 4 (4,117,599 individuals). Data for Round 5 (13,703 individuals) is not shown.

Figure 5: SPADIES ATT for the probability of dropping out



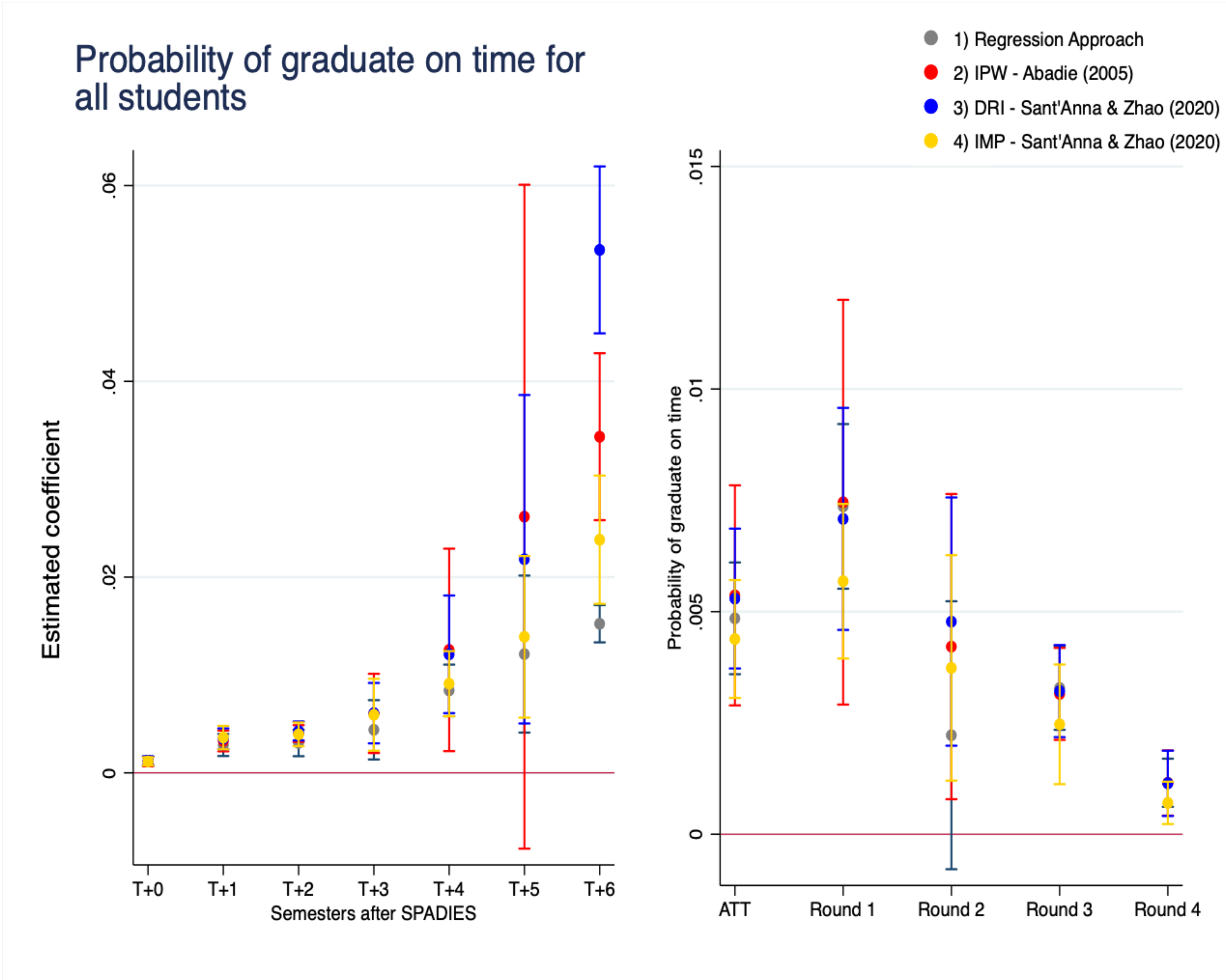

Notes: Chart reports the estimated coefficients for the probability of dropping out using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95/

Figure 6: SPADIES ATT for the probability of graduation



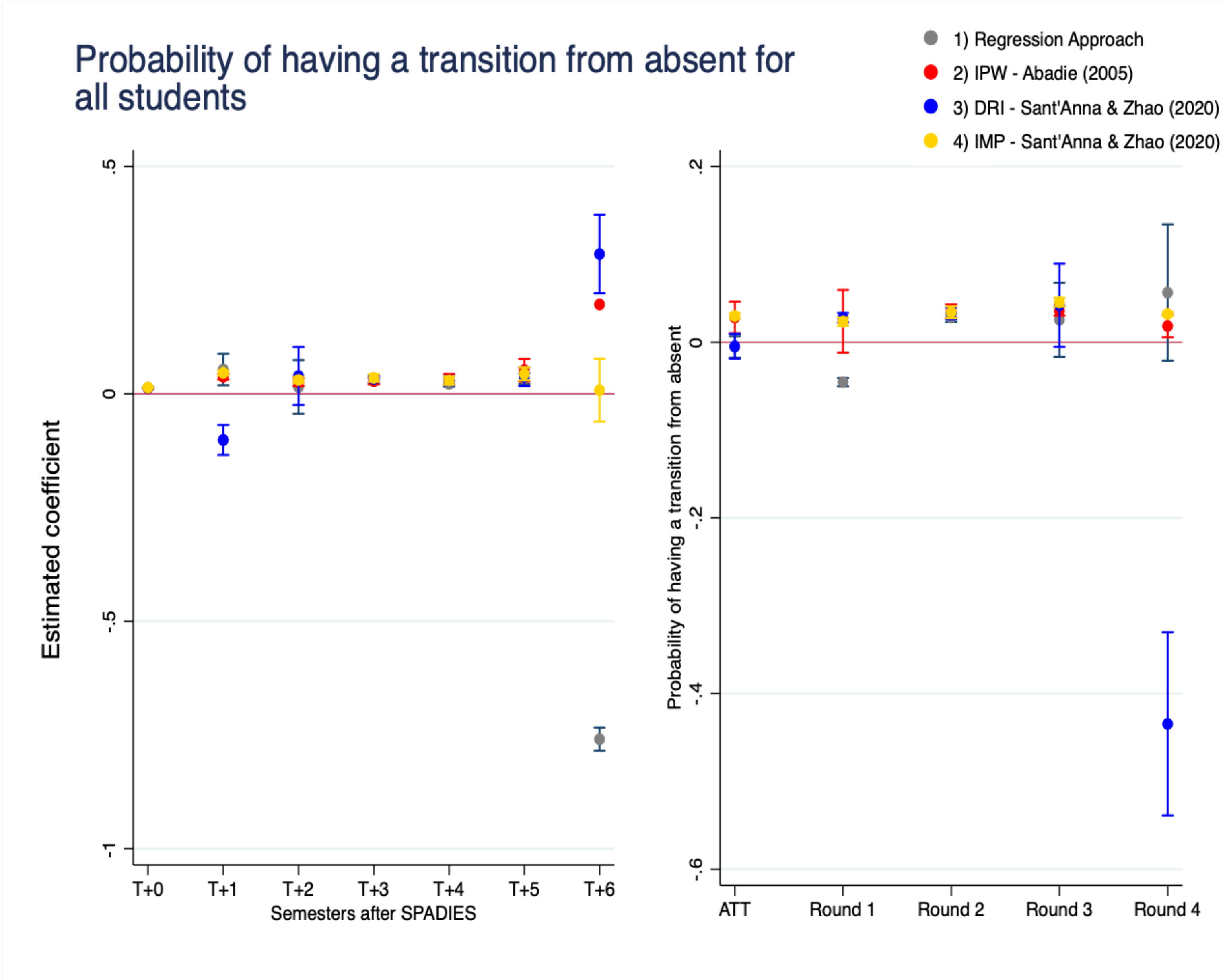

Notes: Chart reports the estimated coefficients for the probability of graduating using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95/

Figure 7: SPADIES ATT for the probability of graduating on time



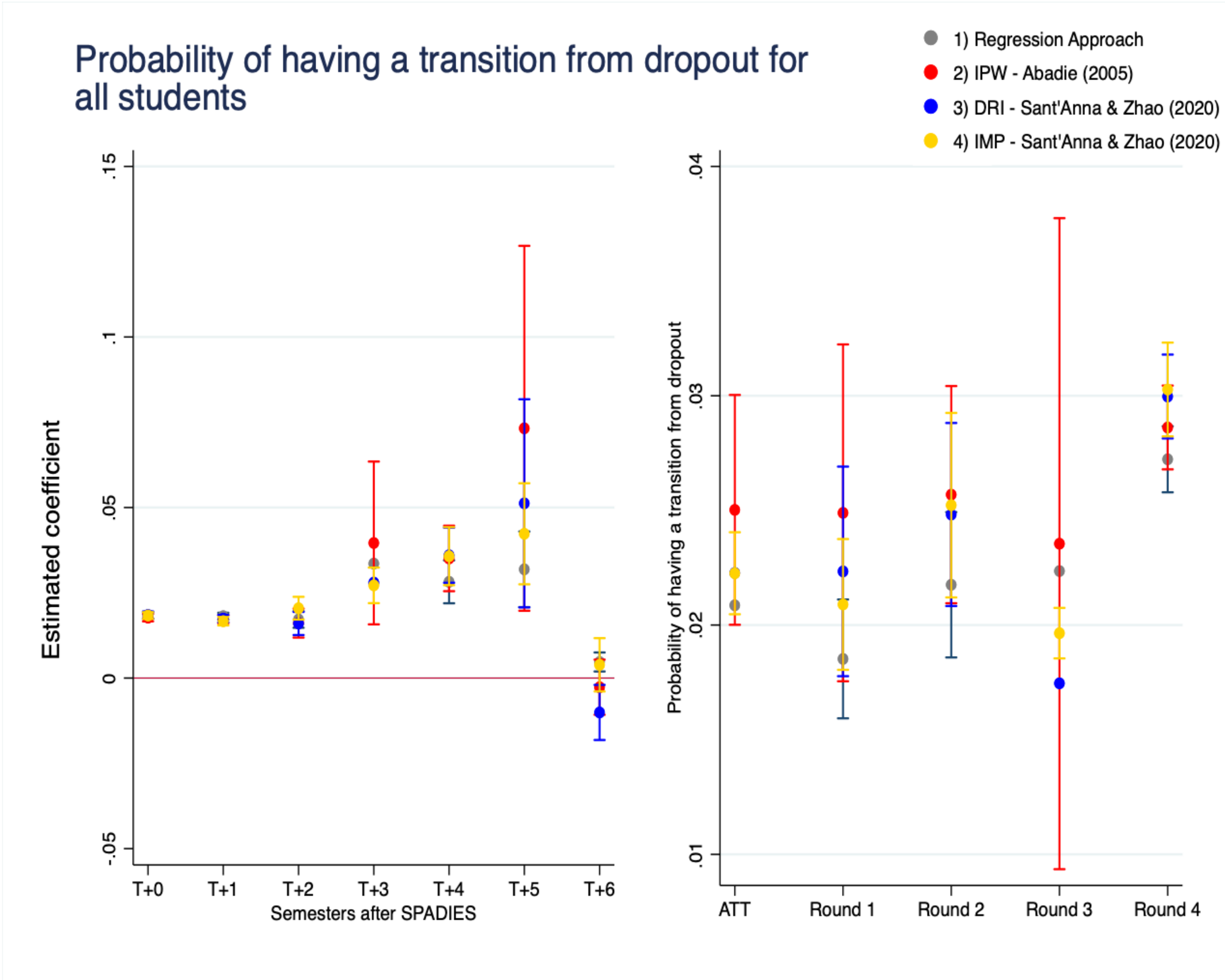

Notes: Chart reports the estimated coefficients for the probability of graduating on time using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95/

Figure 8: SPADIES ATT for the probability of having a transition from Absent



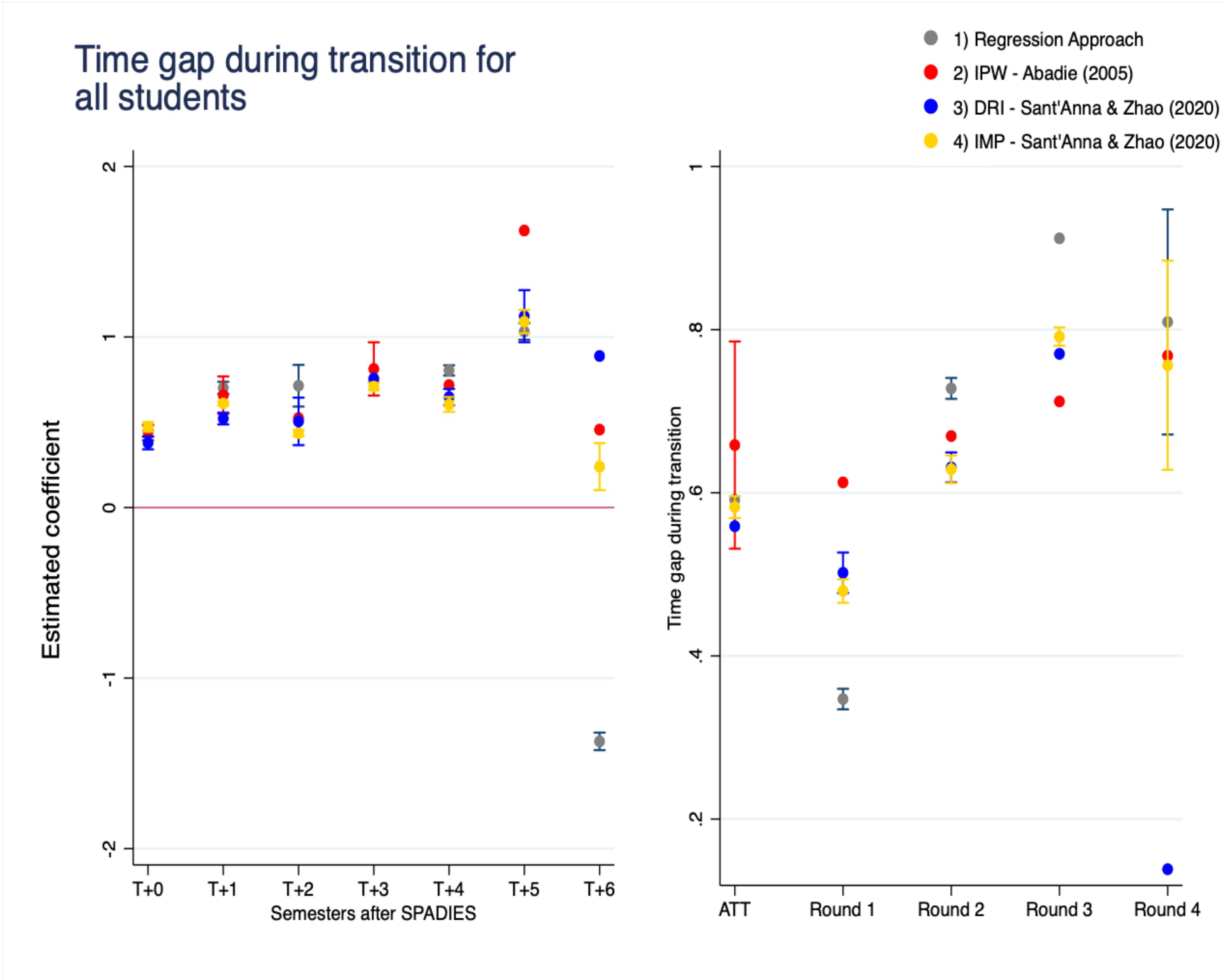

Notes: Chart reports the estimated coefficients for the probability of having a transition from Absentusing Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95/

Figure 9: SPADIES ATT for the probability of having a transition from Drop-out



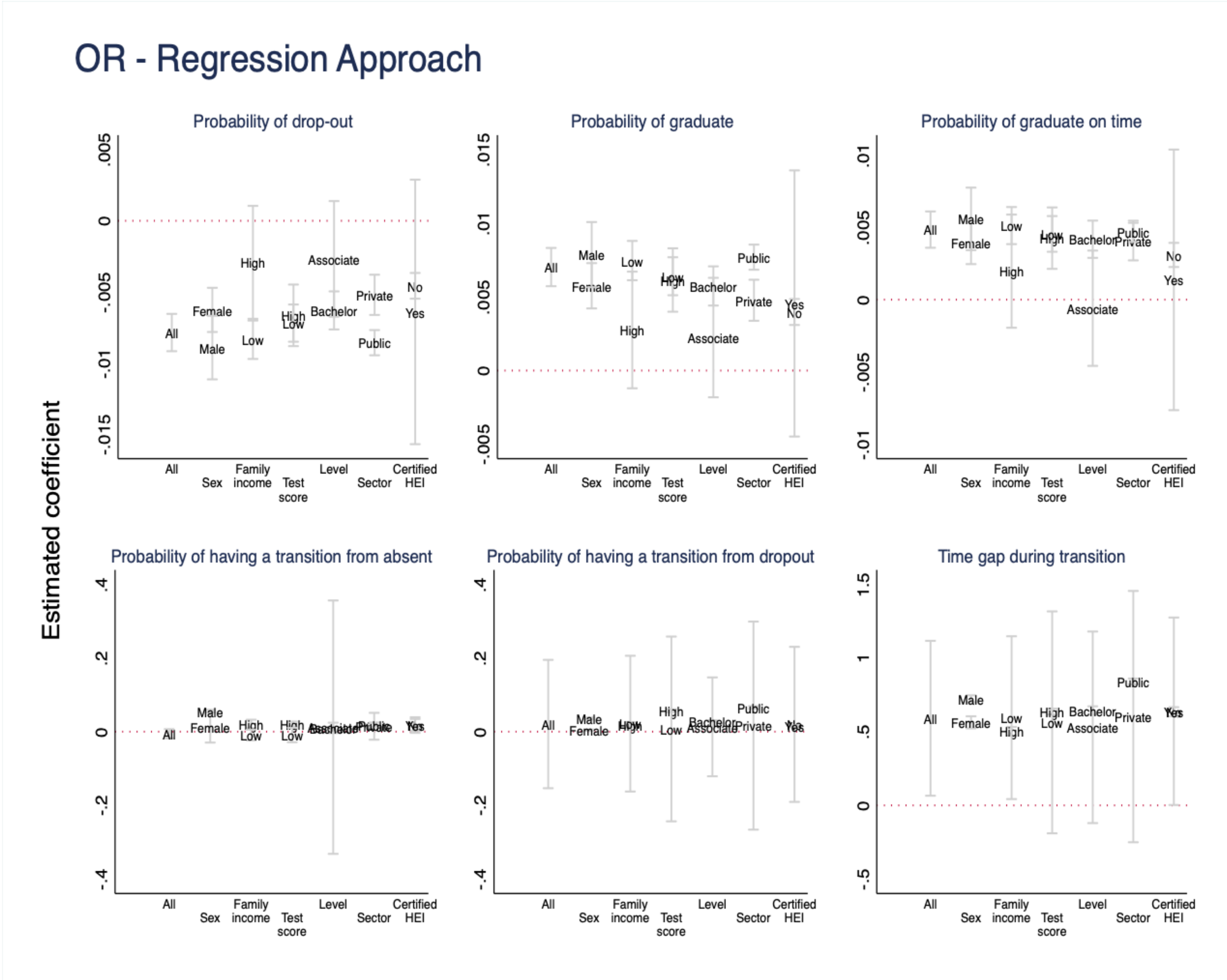

Notes: Chart reports the estimated coefficients for the probability of having a transition from drop-out using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95/

Figure 10: SPADIES ATT for the time gap during the transition



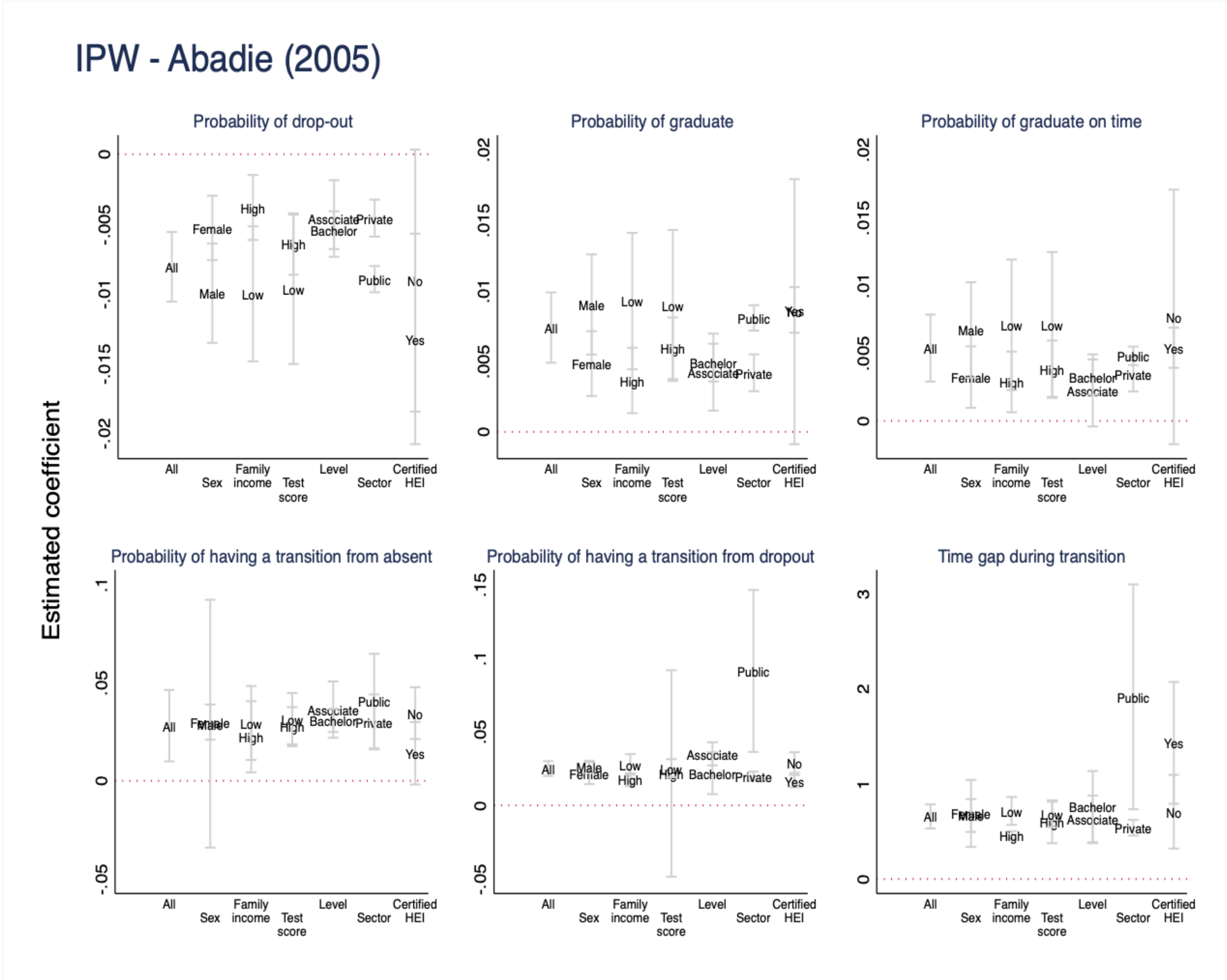

Notes: Chart reports the estimated coefficients for the time gap during transition using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95/

Figure 11: Dissagregatted ATT results using OG approach



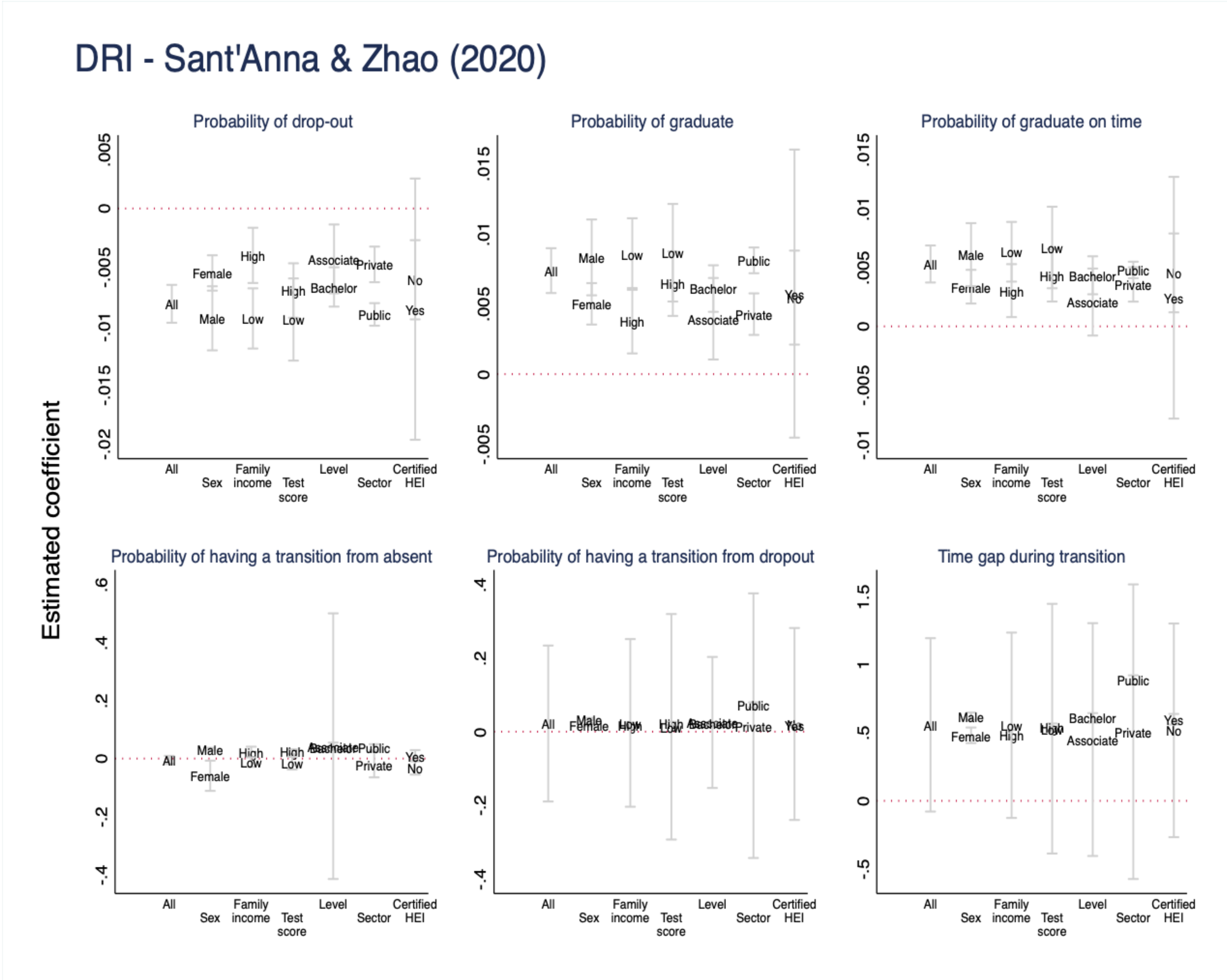

Notes: Chart reports the coefficients using Equation 7 -OR- (whiskers at 95/

Figure 12: Dissagregatted ATT results using IPW approach



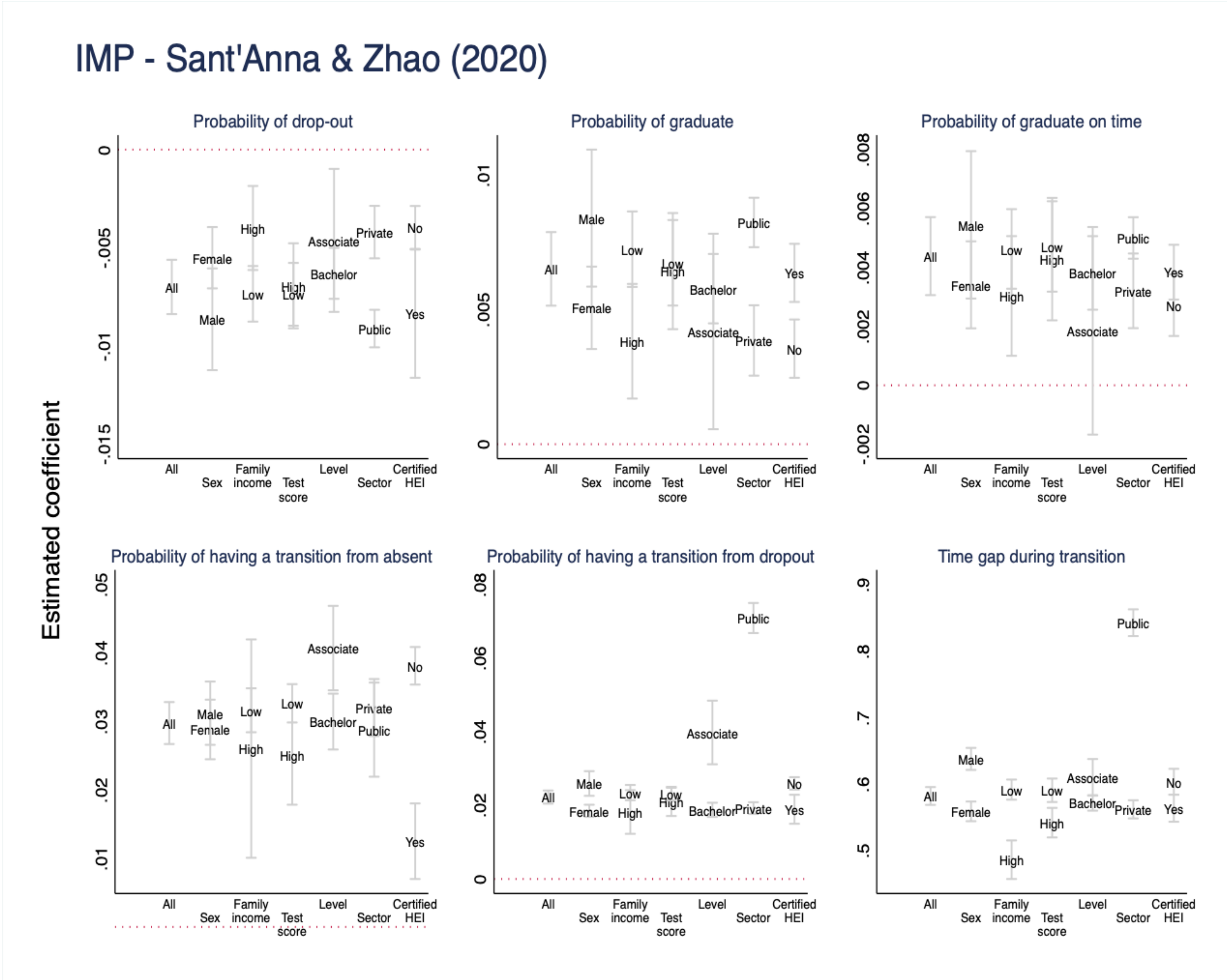

Notes: Chart reports the coefficients using Equation 9 -IPW- (whiskers at 95/

Figure 13: Dissagregatted ATT results using DRI approach




Notes: Chart reports the coefficients using Equation 11 -DRI- (whiskers at 95/

Figure 14: Dissagregatted ATT results using IMP approach




Notes: Chart reports the coefficients using Equation 12 -IMP- (whiskers at 95/