Ending the Musical Chairs Game in Higher Education: *

How a Data-driven Tool Improved Educational Outcomes in Colombia

Author:

Luis Omar Herrera-Prada[†]

Ibero-America Institute for Economic Research

Universität Göttingen

International Monetary Fund

June 28, 2023

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, Juan Carlos Muñoz, Juan Hernández, 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 Research Network Conference from the Colombian Central Bank 2022, LACEA 2022, MEA-SOLE 2022 meeting, Western Hemisphere Department Seminar at the International Monetary Fund, Development Economics Seminar at Universität Göttingen, and Public Policy Seminar at Universidad de los Andes - Colombia 2021 were invaluable. Latest version of this document is available in `https://herreraprada.com` . All views and errors are mine. *The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.*

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

Abstract

The System for the Prevention and Analysis of School Dropouts in Higher Education (SPADIES) was a pivotal tool in the Colombian Ministry of Education's efforts to address the challenge of increased enrollment of less prepared students in higher education. This software dashboard facilitated the collection, analysis, and visualization of student data, which enabled higher education institutions to prevent students from dropping out and also call back students who had already dropped out. Using a differences-in-differences approach, under the Callaway and Sant'Anna (2021)'s framework, this study finds that SPADIES reduced the probability of a student becoming a dropout by 7 basis points and increased the probability of graduating and graduating on-time, by 6 and 4 basis points, respectively. Although the impact may seem small, 7 basis points correspond to around 14,000 students, which almost double the size of the average higher education institution (HEI) in Colombia. Furthermore, SPADIES helped to decongest entry into the higher education system, reduce the burden of overpopulation on HEIs, and improve the efficiency of the enrollment process. These findings suggest that the implementation of a data-driven software dashboard can have significant positive effects on the quality and efficiency of higher education systems in developing countries.

1 Introduction

The dropout phenomenon in higher education poses a significant economic challenge due to its impact on the cost of training a qualified workforce for society. When students leave before completing their degree, it is more than just a waste of resources and a missed opportunity for the individual. However, it also undermines the goal of increasing productivity and efficiency in the labor force. To combat this issue, countries worldwide are exploring innovative solutions to keep students in school and increase enrollment, particularly in higher education. Between 2000 and 2010, Colombia experienced the "Crowding cohort"[1] phenomenon, which saw a surge in the number of less academically prepared students entering higher education due to policies aimed at increasing secondary graduation rates. This influx of students resulted in a decline in the quality of higher education institutions in Colombia (Orozco Silva, 2010; Herrera-Prada, 2013; Ferreyra et al., 2017) .

The higher education system in Colombia was not equipped to handle the surge in student enrollment during the period between 2000 and 2010, due to demographic growth and a policy aimed at increasing secondary enrollment and graduation rates. Starting in the mid-1990s and continuing through the late 2000s, secondary attainment increased due to a policy that promoted all students to the next grade level without considering academic merit. This approach left many students academically vulnerable, and they faced additional challenges such as economic depression and social-political turmoil, which further compounded their struggles in higher education. Colombia experienced its biggest economic depression (before COVID) in 1999, and also suffered the breakdown of peace talks in 2001, adding to the already challenging environment. All of these factors contributed to a decline in the quality of higher education institutions in the country. (ICFES, 2002; Orozco Silva et al., 2006; Orozco Silva, 2010).

Between 2002 and 2006, the Ministry of Education (MEN) designed a plan to improve educational outcomes in higher education, with a focus on increasing enrollment by providing financial and academic support to students. To respond to the increased demand, the tertiary system utilized existing resources, mainly infrastructure, and established a coordinated information system called SPADIES (System for the Prevention and Analysis of School Dropout in Higher Education)(Ministerio de Educación Nacional, 2008, 2009; Orozco Silva, 2010). SPADIES provided real-

[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).

time information on student enrollment, academic performance, peer quality, dropout rates, and graduation rates to higher education institutions, the MEN, and the public. Higher Education Institutions (HEIs) used SPADIES to identify at-risk students, and target them with different types of assistance, such as academic support, financial aid, and mental health services, to reduce the dropout rate (Ministerio de Educación Nacional, 2008, 2009, 2010).

This paper evaluates the impact of SPADIES on education outcomes and system efficiency in Colombia. The study uses SPADIES data from 1998 to 2017 and tracks the college paths of approximately 4 million students. Using a differences-in-differences approach, under the Callaway and Sant'Anna (2021)'s framework, where the treated are those students enrolled in an HEI with SPADIES installed.

I find that SPADIES reduced the dropout rate by 7 basis points and increased the probability of earning a degree and earning it on time by 6 bps and 4 bps respectively. These seemingly small effects translate into significant impact, with around 14,000 students saved from dropout, almost double the size of the average higher education institution in Colombia. I also find that SPADIES was most effective in reducing the dropout rate for males, students from public institutions, and low-income students. Furthermore, HEIs utilized SPADIES not only to prevent dropouts but also to re-engage students who had already dropped out, as evidenced by the increased number of transitions from absent and dropout status. This achievement has significant financial benefits for the population, as studies have shown that the average income of a graduate is at least 20% higher compared to that of a dropout. Also, the results suggest that SPADIES improved the efficiency of the higher education system by decongesting the entry of a growing student population, and by connecting students with HEIs and vice versa. SPADIES also facilitated the transition to a digital record-keeping system. My findings shed light on the potential for data-driven tools to improve educational outcomes in developing countries.

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

This section provides a comprehensive review of the literature that forms the foundation of this paper. Initially, the discussion will center on the literature concerning higher education's demand and supply sides, their influence on quality, and the progression of dropout analysis. Subsequently, the literature specific to the Colombian context will be presented. Furthermore, I will scrutinize the impact of technology on education, including the general analytical approaches identified in the literature. Finally, I will discuss the contribution of this study to the existing literature.

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.

Formal education, particularly higher education, plays a vital role in human capital formation and social mobility, according to canonical literature(Becker, 1962; Trow, 1974; Adams, 1984; Bank et al., 1990; ONU, 2013). However, the supply and demand dynamics of higher education lead to an unbalanced market due to access barriers, low enrollment rates, excess demand and/or supply, and low quality (Epple et al., 2006). This discrepancy arises due to factors such as a lack of public resources caused by an unexpected increase in applicants (Bound et al., 2009). The institutional factors influencing students' decision to enroll, continue, or drop out, such as the quality of education, support services, and peers, are also incorporated into cost-benefit analyses (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 student's integration or adaptation to the education system model (Tinto, 1975, 1982) and attrition as a set of conditions linked to the individual's socioeconomic factors, such as family conditions or academic performance during school (Bean, 1980, 1985).

In Colombia, initial studies on attrition in higher education focused only on a few institutions. Universidad Nacional de Colombia (2007) investigated the lag, graduation rate, and dropout rate in Colombia's largest public university and found that being a woman, especially 18 years old or younger, positively impacted the probability of obtaining a degree from any program. Financial aid or student loan programs decreased the dropout rate. Affirmative Action programs, such as unique admission mechanisms and alternative admission avenues through pre-university courses, were found to be essential in improving students' chances of staying in higher education and, consequently, enhancing access conditions and social equity (Sánchez et al., 2002). At the University of Antioquia, Castaño et al. (2006) studied the School of Engineering and the School of Economics' students in the second cohort of 1996 and found that being male, single, and over 18 years old increased the risk of dropping out. They also discovered that living with parents, having better academic performance, not working, having parents with a high level of education, and being female were characteristics associated with a decreased risk of dropping out. Public universities in Colombia implemented these studies to address the government's concerns about resource allocation in the educational system and viewed the dropout phenomenon as a waste of economic resources, human capital, and infrastructure (Cárdenas, 1996; Córtes et al., 2011; Facundo-Díaz, 2009). National-level research conducted by ICFES (2002) found that household financial conditions were the primary determinant of becoming a dropout student. However, Ministerio de Educación Nacional (2008) revealed that low academic skills (measured by the secondary school exit exam score), mismatch in career choice and skills, poor academic performance, and gender were the main reasons for the dropout rate. Finally, Herrera-Prada (2013) demonstrated that the "Crowding cohort" phenomenon, accompanied by a simultaneous increase in the average time needed to graduate, resulted in an overall decrease in the graduation rate in Colombia, despite a reduction in the dropout rate nationwide.

The impact of technology on education has been widely studied in various contexts, primarily focused on improving learning, teaching, research, and administrative systems (Tongkaw, 2013). There have been numerous studies on the influence of technology on new models of online learning (López-Pérez et al., 2011), motivation, and learning strategies (Valentín et al., 2013; Sailer et al., 2021), as well as on the use of information systems to improve learning (McGill and Klobas, 2009; Sari, 2014; Leong and Ibrahim, 2015). Additionally, there is research on the expansion of the educational system (Tongkaw, 2013; Rahman, 2020), the adoption of technology in learning environments (Lacka and Wong, 2021), and improving school attendance to enhance academic performance (Gomis-Porqueras et al., 2011). However, direct evidence of the use of technology at the administrative level for an entire country's higher education institutions, which can affect the coverage and quality of the higher education system, is lacking. The literature about this is quite limited, with the international literature focusing on small programs in regions of Germany and Chile and using econometric models (uplift models) to predict dropout rates (Berens et al., 2021) or design tailored anti-dropout programs (Olaya et al., 2020). However, survival, tailored, and uplift models have been employed in HEIs in Colombia to reduce the dropout rate since 2006, with no academic references other than the MEN reports on the SPADIES experiences.

This paper adds to the literature by revealing how SPADIES effectively achieved its policy goals. My analysis demonstrates how a data system that improved the flow of information among agents, coordinated policy and innovative policymakers developed tailored aid programs for pre-selected candidates, resulting in a new equilibrium with educational and social outcomes even better than the government had anticipated.

3 Context and Program Design

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

3.1 Context

Over the past 50 years, Colombia has undergone significant demographic changes, marked by large-scale migration from rural to urban areas. This influx of young people, seeking social and economic mobility, faced limited access to quality educational resources (Lucio and Serrano, 1992).

By the mid-1990s, the lack of high school capacity was apparent, prompting the government to promote students through primary and secondary school without academic restrictions. While this strategy increased secondary school coverage and completion rates, it also significantly declined education quality and skills (Orozco Silva et al., 2006; Orozco Silva, 2010; Herrera-Prada, 2013) .

The economic crisis of the late 1990s exacerbated this issue, revealing that few Colombians gained access to higher education, and those who did were at high risk of dropping out (Orozco Silva et al., 2006; Ministerio de Educación Nacional, 2008, 2010). Early 2000s studies revealed household economic situations were the primary barrier to higher education enrollment (ICFES, 2002). At this time, Colombia's gross college enrollment rate was only 20%, one of the lowest in the region (Ministerio de Educación Nacional, 2010; Ferreyra et al., 2017).

In response, the MEN implemented the "Educational Revolution" plan to boost educational attainment levels in higher education, which had lagged behind regional counterparts. The plan increased the enrollment rates from 20% in 2002 to over 40% in 2010 and surpassing 45% in 2016 (Orozco Silva et al., 2006; Ministerio de Educación Nacional, 2010, 2017). Public higher education institutions were primarily responsible for this surge, as evidenced by the near-zero or negative enrollment growth rates at private institutions in the early 2000s.

When the MEN introduced SPADIES, it provided timely information on enrollment, academic performance, peer quality, dropout rates, and graduation rates in the higher education market. Its primary objective was to reduce dropout rates by identifying at-risk students and offering targeted aid suggestions to HEIs. Additionally, SPADIES facilitated data collection on the higher education system and HEI performance indicators, which were publicly accessible. SPADIES became the standard for certifying the quality of programs and measuring improvements in access to resources in Colombia's higher education institutions.

3.2 Program Design

SPADIES is an information system that the MEN developed in 2005. The MEN created a software tool that could collect information on all students enrolled in higher education, allowing for data visualizations, statistical analysis, and reports on students at risk of dropping out. One module of SPADIES was public, enabling students, authorities, and the public to compare the performance of HEIs, and other departmental or national aggregates.

After obtaining sufficient good data, the MEN used the SPADIES database to estimate Duration Models and conduct focus group analyses nationwide to identify consistent behavior patterns among students who dropped out. All institutions had to have at least two delegates with permanent contact with the MEN every month to improve data quality. The MEN and an external consultant audited each month the reported data before being stored in the database to run the models. The results of the models were incorporated into the application to empower HEIs to identify at-risk students and take action to prevent them from dropping out programs prematurely.

The MEN visited each HEI to present the national and institutional estimates, install the software, explain the results, and train the staff to use the tool to identify and profile the population at risk of dropping out. The objective was to use this list to target 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.)

The MEN identified that the population most at risk of dropping out nationally was males with low household income and low academic skills, particularly those in associate programs or math-related programs. The MEN encouraged HEIs to design their first wave of aid programs targeting mainly this population (Ministerio de Educación Nacional, 2008).

Between 12 and 18 months after the software installation, the MEN made a follow-up visit ("Accompaniment") to verify the new outcomes and ensure the correct use of the application. By the end of 2013, all institutions had installed the software and received follow-up visits.

Beginning in 2008, the MEN organized multiple competitions focused on strategies to reduce dropout rates, utilizing the SPADIES to measure such rates and encourage innovation within the education sector. SPADIES enabled two significant milestones: (1) collecting information about the students at the individual level since 1998, and (2) the entry of many universities into the digital world. By the end of 2005, only a small percentage of institutions had electronic records, prompting the MEN's information systems to initiate a challenging migration process from paper to digital records. By 2017, approximately 60% of institutions could report complete data starting from 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 collect each HEI's information and add it to the system (the list of HEIs in each round is shown in Figure 1). Each round corresponded to a new group of HEIs with digital records in proper condition, enabling the process described above. However, the application of SPADIES in HEIs was not differentiated by the sex of the student or the sector, quality, or region of the HEIs (Figure 2). The fact that some HEIs are biased towards high-scoring and high-income students pursuing bachelor's degrees is evident; indeed, such students tend to enroll in better-resourced universities. Consequently, these HEIs were likelier to be enrolled in the earlier rounds due to their superior resources and higher data quality. In 2008, the MEN mandated that HEIs include the official dropout rate measured by SPADIES in the requirements for program quality certificates. Additionally, the MEN established a web portal through which the public could access critical statistics of individual HEIs or the entire higher education system, including information on the population, dropout rates, and graduation rates.

4 Data and Variables

This section provides an overview of the SPADIES database and its characteristics, followed by a description of the variables used in the program's definitions, including the new variables created for this study. Finally, the section explains the final database used in the empirical analysis.

4.1 SPADIES Database

The SPADIES database is the result of merging data from three sources: the MEN's National Information System for Higher Education (SNIES) database, the Colombian Institute for the Promotion of Higher Education (ICFES) database, and the Higher Education Institutions (HEIs) semestral report. The SNIES data is time-invariant and provides information on the characteristics of HEIs and their programs. The ICFES data is also time-invariant and collected at the end of secondary school through the Saber 11 exam. The HEIs' report updates student information every semester.

SPADIES uses a life history approach to collect its data, tracking only students who started college in 1998 or later. For this study, I use SPADIES data from 1998 to 2017, which includes approximately 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

The ICFES variables collected during the Saber 11 exam include the exam score, gender, year of birth, and household income. The Saber 11 exam score is a critical requirement for enrolling in higher education; thus, all students who enroll have a score. All other questions are optional for students, and this information is used for characterizing students and empirical analysis in this document.

SPADIES uses a standardized version of the Saber 11 test score. The MEN standardized the score as the ICFES used a different score range over time. Each student's percentile on the Saber 11 exam was assigned, resulting in a variable ranging from 1 to 100. Additionally, I created a dummy variable that takes a value of 1 if a student has a high score (above 90) and 0 otherwise, similar to the Ministerio de Educación Nacional (2008, 2010, 2017) methodology.

Gender and birth year are reported by both the ICFES database and the Freshmen report, while students report household income to the ICFES. If data for gender or birthday does not match, SPADIES uses the data from the ICFES unless it is missing.

SPADIES uses data from the SNIES database to obtain information on the characteristics of HEIs and their programs. The HEI characteristics include sector (public or private), category (universities or community colleges), and location. Program characteristics include the level (bachelor's - 4 or 5-year programs - or associate - 2 or 3-year programs) and field of knowledge.

Additionally, SPADIES has a variable that evaluates the quality of data reported per HEI, graded A, B, or C, based on the number of semesters reported and the level of detail provided per semester. A dummy variable was created to take a value of 1 if the HEI earned an A grade in the report, as recorded in the database in 2017.

4.2.2 SPADIES Time-Variant Variables

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

To track students' progress, SPADIES uses the Freshmen and Enrolled reports to identify students who enroll in a program and then tracks them until they are reported as Graduates. Suppose a student is not found in the Enrolled report or Graduated report. In that case, their status is updated to reflect their absence or dropout, depending on the number of semesters not enrolled. The statuses that SPADIES uses are:

1. **Graduated:** The Graduated category includes all individuals who successfully completed their higher education program. This category is further broken down into Graduated on time and Graduated late, encompassing individuals who completed their programs within the expected time frame and those who took longer, respectively. The expected graduation time was estimated as three years for associate programs and five years for professional programs. Therefore, $Graduated = Graduated\, on\, time + Graduated\, late$. So, for an easy understanding:
   (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).
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.

SPADIES updates a student's time in the system, academic performance, and report of aid received every semester. The status is assigned at the moment of data extraction from the database, which, in the case of this document, is 2017. In addition to the statuses provided by SPADIES, I created three new variables to measure transitions. I compare the time since the students first enrolled in college and the number of semesters reported in SPADIES. If there is a difference, it means that they left the school. If their current status is either graduated or active, it means that they had a transition, i.e., a 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 freshman during the period 1998 to 2012.

To avoid bias in dropout rates (which tend to be higher in the first few semesters) and graduation rates (which typically only occur after 4 or 8 semesters in associate and bachelor's programs, respectively), I made two changes to the original database containing 8 million students:

1. To ensure a mix of students at various stages of their program, I only included data for students who were reported as "active" since 2002. Specifically, this means they were reported as a "freshman" in any year before 2002 and appeared in the enrollment report for 2002.

2. To focus on cohorts with sufficient time to graduate, I only used data up to 2017 for students who were freshmen up to 2012.

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

I created a dummy variable to identify the time (year and semester) when SPADIES was installed in each Higher Education Institution (HEI). I used the semester of this visit as the start of the "treatment" of SPADIES on an HEI because during this visit, the institution received the first report from SPADIES of the students at-risk in their institution, and the HEI was able to access grants from the Ministry of Education (MEN) to create anti-dropout programs.

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

The final dataset is an unbalanced panel per individual-program-HEI and time, counting 4,131,302 students. It includes the students' gender, year of birth, household's income, the ICFES's test score, a dummy that takes the value of 1 if the program they are attending conducts to a bachelor level degree and it is 0 otherwise, a set of dummies that take the value of 1 depending on the assistance program (financial or academic) received and it is 0 if not received any assistance, a dummy that takes the value of 1 if their institution is public and it is 0 otherwise, a dummy that takes the value of 1 if their HEI is certified and it is 0 otherwise, a dummy that takes the value of 1 if their HEI is a main campus and it is 0 otherwise, a dummy that takes the value of 1 if the data from their HEI is A 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 ΔY_i if there was no treatment conditional to X. With this new assumption, the new DiD estimator becomes $\widehat{ATT}_*$ (Equation 6).

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

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

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 control 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 and 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) disaggregates 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 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 graduating 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 both financial and academic aids were well-targeted, but they were not enough. 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.

In line with 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.

Since the indicators are dynamically constructed depending on the cutoff time of the database, and these same statuses are dynamic over time, the final database cannot be cleanly used to estimate parallel trends. Therefore, the first part of this section shows the continuity of the averages of the indicators measured at the time of the database cutoff, analyzing the pre-SPADIES periods. However, restructuring the database can give a more formal presentation of the parallel trend assumptions.

To do so, I created a new database with a sample of students who never had contact with SPADIES for all HEIs, including students in all phases of their programs, which allows me a time horizon of up to 5 years before SPADIES. In Callaway and Sant'Anna (2021)'s terminology, these students are the "never treated." I then used the same SPADIES definitions to assign the status that these students would have had prior to the entry of SPADIES into their HEI. For this sample, I applied placebo treatments by assigning SPADIES-Placebo 1, 2, or 3 years prior to the actual entry of SPADIES into the HEI. As in reality, the treatment is assigned by HEIs (a total of 281 HEIs in the sample), but unlike in reality, the HEI assignment for each cohort was done randomly, maintaining the actual proportion of HEIs assigned to round 1 (27.3%) and 2 (19.6%). All HEIs were assigned, so the 53.1% remaining from the first two rounds are counted in Round 3 of the placebo. In total, I performed 100 different placebos, and for each treatment assignment, I used the classical DiD model and Callaway and Sant'Anna (2021)'s methodology to estimate the SPADIES-Placebo effect. The coefficients are shown in Figures 5 and 6. For all variables of interest, the results of the classical DiD model show a null effect of SPADIES-Placebo, which implies that the assumption of parallel trends is met. In the Callaway and Sant'Anna (2021) exercise, there is evidence of a null placebo effect for drop-outs and graduation, while for transitions, there is a small positive effect.

6.3 Main Results

In this section, I present the causal results for SPADIES estimated using Equations 7, 9, 11 and 12 by usign 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

According to Figure 7, SPADIES reduced the probability of dropping out by 7 basis points (bps). The Difference-in-Differences (DiD) estimators revealed that SPADIES had a positive impact on the probability of becoming a drop-out student, reducing it by 8 bps, as indicated by the Odds Ratio (OR), Inverse Probability Weighting (IPW), and Doubly Robust Inference (DRI) methods, as shown in Figure 7. Round 1 was the most effective round of SPADIES, with a reduction in the probability of dropping out, ranging from 9 bps to 12 bps. Six semesters after the installation of SPADIES, its impact reduced the probability of dropping out, ranging from 20 bps to 80 bps.

6.3.2 Graduation

As indicated in Figure 8, SPADIES increased the probability of graduating by at least 6 basis points (bps). Similar to the drop-out component, Round 1 was the most effective round, resulting in an increase in the probability of graduating, ranging from 9 bps to 11 bps. Six semesters after the treatment, SPADIES's impact led to an increase in the probability of graduating, ranging from 21 bps to 79 bps.

6.3.3 On-time Graduation

As shown in Figure 9, SPADIES increased the probability of graduating on time by 4 basis points (bps). The DiD estimators confirmed that SPADIES positively impacted the probability of graduating on time, especially for Round 1. Six semesters after the treatment, SPADIES increased the probability of graduating by 18 and 54 bps on time.

6.3.4 Transition from Absent

Figure 10 demonstrates that SPADIES increased the probability of transitioning from absent status by 3 basis points (bps). Notably, there was no significant difference in the probability of dropping out, graduating, or graduating on time across the different rounds. This result may be attributed to the short gap period required for the transition from absent status, which is only one semester. Across all rounds, there was a consistent increase in the number of transitions for students marked as absent, with no significant difference between rounds. The impact was similar for every semester, as all semesters after the installation of SPADIES showed a comparable rate of increase in the number of transitions for students marked as absent.

6.3.5 Transition from Drop-out

SPADIES has been shown to positively impact the probability of transitioning from drop-out status by 2 bps, as depicted in Figure 11. Notably, Round 4 had the best results among all rounds, potentially due to the shorter time since treatment and the HEIs' use of the list of students marked as drop-outs to call them back to the system. Although other rounds followed a similar approach, their effect decreased over time as the number of called-back students decreased, and new drop-out students required two semesters to be marked and traced.

It is important to note that once students are marked as absent, the HEI has the ability to call them back, making it difficult for them to transition to drop-out status after the implementation of SPADIES. As a result, the analysis conducted after the installation of SPADIES shows a clear improvement in drop-out transitions as HEIs become more adept at tracking and recalling their drop-out students. While the results per round indicate a decrease in the effect of SPADIES over time, the results per semester demonstrate that HEIs were able to improve their efficiency in tracking and bringing back drop-out students to school. However, this effect is found to be less pronounced over time, as evidenced by the results six semesters after the implementation of SPADIES.

6.3.6 Time Gap of Transition

The results in Figure 12 show that SPADIES increased the time gap during the transition by 0.6 semesters. This finding is closely linked to the transitions from dropouts, as the differential increase found in the results of Figure 11 can be attributed to the return of students who were previously marked as dropouts but brought back by the HEI with the help of SPADIES. These students had been outside the system for a significant period and would have remained so if not for using SPADIES. Therefore, the reported increase in the time gap is a positive outcome as it represents the return of students who may not have returned to school otherwise.

Students who had transitioned from dropout status before the implementation of SPADIES had already planned to return to school regardless of SPADIES; so it makes sense that the gap was shorter for them. On the other hand, the students who returned because of HEIs' targeting through SPADIES had been out of the system for many years, and the influx of these returning students led to the increase in the time gap. This gap is more significant in Round 4 because of the short time to incorporate new dropouts into the average of all transitions. Overall, the increase in the time gap during the transition reflects the successful re-entry of previously marked dropouts. The longer time gap is an indicator of the effectiveness of SPADIES in bringing these students back to the education system.

6.4 Disaggregated Results

In this section, I present the results of the analysis by dividing the sample based on time-invariant variables. The coefficients for each estimation method proposed for the DiD analysis are shown in Figures 13 to 16. The results indicate that SPADIES was more effective in reducing the probability of dropping out for males with low income from public HEIs. The only significant difference in results was observed based on the HEI sector, where public HEI students experienced a more considerable reduction in the probability of dropping out.

Regarding the probability of graduating, males with low income from public HEIs experienced an increase in the probability of graduating. However, a significant difference was observed between HEIs with and without quality certifications. Similar results, without significant differences among subpopulations, were found for the probability of graduating on time.

Analyzing the probability of transitioning from absent status, I found that students from non-certified HEIs and those in associate programs were more likely to transition from absent. In contrast, male 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 most significant population benefiting from SPADIES were males with low income from public HEIs, mainly in associate programs. This subset of the student population was also the most likely to have already dropped out and would not have returned if the HEIs did not use SPADIES to target them and call them back to school. This finding aligns with the Ministry of National Education's instructions to HEIs to target their aid programs towards this subset of the student population, according to Ministerio de Educación Nacional (2008, 2010).

7 Conclusions

The new millennium presented a significant challenge for Colombia in terms of education and the labor market. The students who reached higher education in the 2000s not only outnumbered their predecessors, but they were also significantly less prepared and had lower skills. This made them more vulnerable to becoming dropouts and caused a sharp deterioration in the quality of the entire higher education system. This issue also had an impact on enrollment and graduation rates.

Despite the scarcity of economic resources, the Ministry of Education in Colombia (MEN) designed a plan to address this education crisis. Various studies such as Orozco Silva et al. (2006); Orozco Silva (2010); Ministerio de Educación Nacional (2010); Herrera-Prada (2013); Ferreyra et al. (2017) have highlighted the severity of the situation and the need for urgent action. 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 to 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 could know in "real-time" statistics of the most critical outcomes for the higher education system.

The flow of information was crucial; not only were HEIs receiving training about how to operate the dashboard, and they were informed about strategies and protocols that other HEIs were successfully used to reduce the dropout rate. HEIs could 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 the 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, my analysis shows that 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, 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 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 becoming dropouts.

On the other hand, the peer effect from MacLeod and Urquiola, 2015, the increased information on 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 in transitions, which is undoubtedly due to the greater competition among HEI peers. Also, due to SPADIES, students could enroll HEIs with 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 contribute to the literature. Analyzing how the increased flow of information and competition leads to a redistribution of some students within the institutions and how it affects their future income would be an essential avenue for further research on increasing competition among HEIs and impacts on transfer students.

My analysis reveals that SPADIES resulted in higher student retention rates, on-time graduation rates, and overall graduation rates. It successfully reduced the dropout rate for all types of students, especially for the most at-risk population of males with low household income and low academic skills at public institutions, which make up 10.4% of the total population (see Figures 13 to 16).

My findings show that SPADIES reduced the probability of students becoming dropouts by 7 basis points (bps), equivalent to saving about 14,000 students from dropping out of the system. This is an impressive figure, considering that the average size of a higher education institution (HEI) in Colombia is 8,000 students. SPADIES also increased the probability of students earning their degree on time by 4 bps and earning their degree overall by 6 bps. SPADIES helped approximately 12,000 students earn their degrees, with 8,000 earning them on time.

Moreover, my analysis indicates that SPADIES increased the number of transitions from absent and dropout statuses by 3 bps and 2 bps, respectively. The average duration of a transition increased by 0.5 semesters. Non-certified HEIs effectively used SPADIES to bring back students from absent status. At the same time, certified public HEIs effectively used SPADIES to bring back students from dropout status. SPADIES also increased the average time gap during the transition by 0.6 semesters, which can be interpreted as positive because former dropout students who returned school thanks to SPADIES had a more extended period outside of school. Without SPADIES, these students might have never returned to school.

Many HEIs used information from SPADIES to target new aid programs. However, I found 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. At-risk students targeted with SPADIES had the highest academic vulnerabilities and were already at a higher risk of dropping out. SPADIES and aid support were not enough to prevent them from dropping out.

Finally, my analysis shows that SPADIES improved the efficiency of the higher education system by helping to decongest entry into the system of a ballooning student population. By reducing the time it takes students to graduate, SPADIES gradually reduced the burden of overpopulation on HEIs, which helped improve the enrollment rate. SPADIES also served as a bridge connecting students with information about HEIs and HEIs with information about students, allowing students to be more selective in their choice of attending an HEI and HEIs to mitigate the vulnerabilities of the new population arriving at 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 before the implementation of SPADIES. In summary, SPADIES was instrumental in breaking down the musical chairs game that was the higher education system in Colombia. SPADIES helped improve the efficiency and effectiveness of the higher education system in Colombia and changed the future of many students who would have otherwise dropped out.

These findings suggest that the implementation of a data-driven software dashboard can have significant positive effects on the quality and efficiency of higher education systems in developing countries.

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.

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

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.

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

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).

Epple, Dennis, Richard Romano, and Holger Sieg, "Admission, tuition, and financial aid policies in the market for higher education," *Econometrica*, 2006, *74* (4), 885–928.

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.

Gomis-Porqueras, Pedro, Jürgen Meinecke, and José A. Rodrigues-Neto, "New Technologies in Higher Education: Lower Attendance and Worse Learning Outcomes?," *Agenda - A Journal of Policy Analysis and Reform*, 2011, *18* (01).

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

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.

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

Lacka, Ewelina and T. C. Wong, "Examining the impact of digital technologies on students' higher education outcomes: the case of the virtual learning environment and social media," *Studies in Higher Education*, 2021, *46* (8).

Leong, Lam Wai and Othman Ibrahim, "Role of Information System (IS), Social Networking Technology (SNT) and WEB 2.0 for Improving Learning Outcomes: A Case of Malaysian Universities," *Procedia - Social and Behavioral Sciences*, 2015, *211.*

López-Pérez, M. Victoria, M. Carmen Pérez-López, and Lázaro Rodríguez-Ariza, "Blended learning in higher education: Students' perceptions and their relation to outcomes," *Computers and Education*, 2011, *56* (3).

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

MacLeod, W. Bentley and Miguel Urquiola, "Reputation and school competition," *American Economic Review*, 2015, *105* (11), 3471–3478.

McGill, Tanya J. and Jane E. Klobas, "A task-technology fit view of learning management system impact," *Computers and Education*, 2009, *52* (2).

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.

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.

Rahman, Muhammad Mofizur, "Impact of digital technology in higher education," *International Journal of Research in Business and Social Science (2147- 4478)*, 2020, *9* (5).

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

Sailer, Michael, Florian Schultz-Pernice, and Frank Fischer, "Contextual facilitators for learning activities involving technology in higher education: The Cb model," *Computers in Human Behavior*, 2021, *121.*

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.

Sari, Arif, "WITHDRAWN: Influence of ICT Applications on Learning Process in Higher Education," *Procedia - Social and Behavioral Sciences*, 2014, *116.*

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

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.

Tongkaw, Aumnat, “Multi Perspective Integrations Information and Communication Technologies (ICTs) in Higher Education in Developing Countries: Case Study Thailand.,” *Procedia - Social and Behavioral Sciences*, 2013, *93.*

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

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

Valentín, Alberto, Pedro M. Mateos, María M. González-Tablas, Lourdes Pérez, Estrella López, and Inmaculada García, “Motivation and learning strategies in the use of ICTs among university students,” *Computers and Education*, 2013, *61* (1).

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) Drop-out	(2) Graduated	(3) On-time graduation	(4) Transition from absent	(5) Transition from drop-out	(6) Time gap during transition	(7) Drop-out	(8) Graduated	(9) On-time graduation	(10) Transition from absent	(11) Transition from drop-out	(12) 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

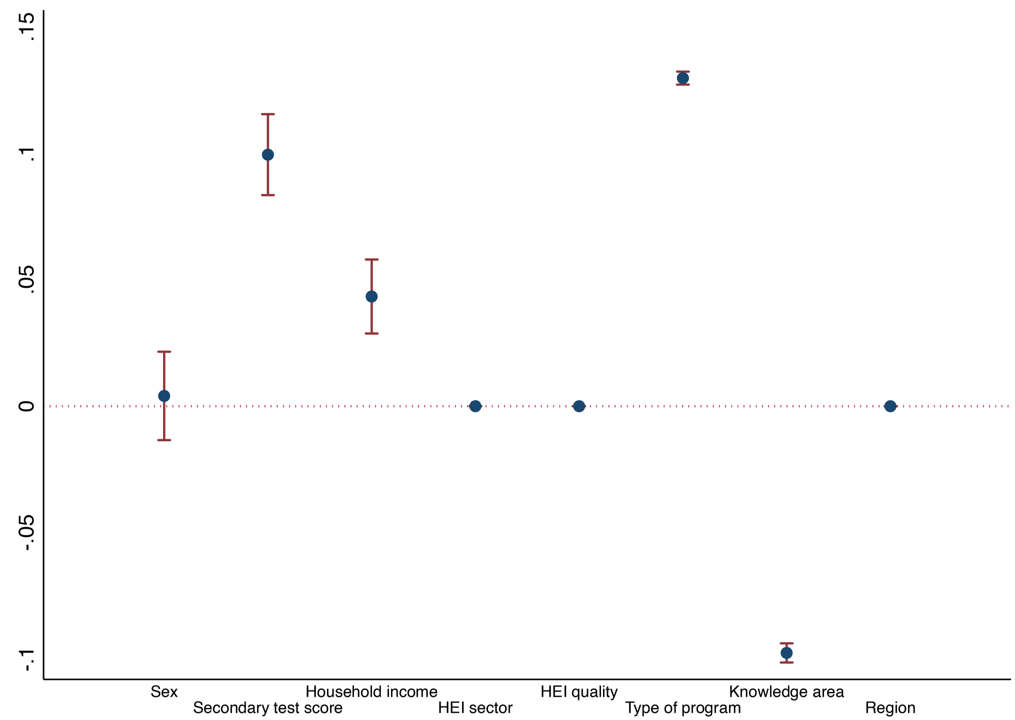
Note: The table shows the results for the estimations using Equation 1. SPADIES is a dummy of 1 if the student was enrolled in a semester after the SPADIES was installed in its HEI. Tenure in the program is expressed in semesters. Share of failed classes is estimated in t+1 as the ratio of the failed 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	

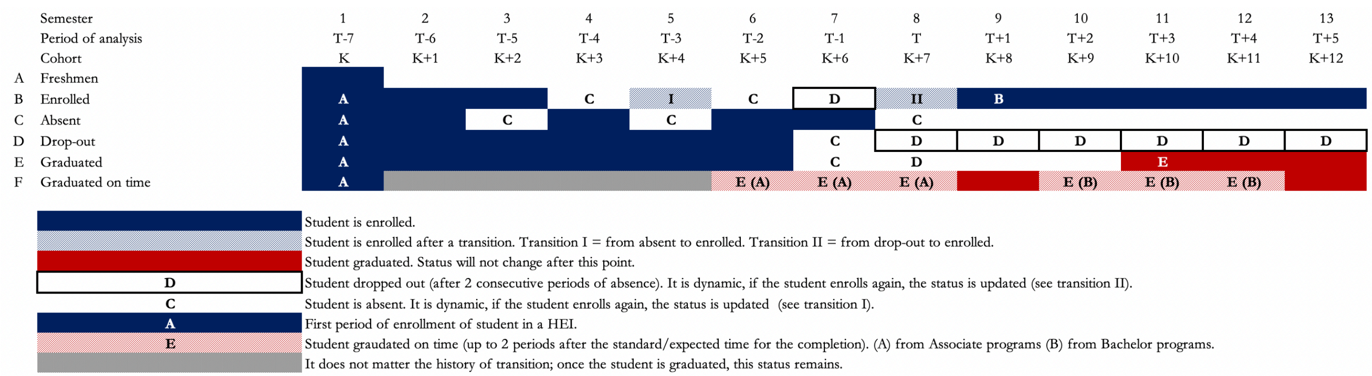
Notes: Five rounds -Round 1(2005-06), Round 2 (2006-07), Round 3 (2007), Round 4 (2008-09), and Round 5 (2010-11)- were necessary to complete all the HEIs' information into the system. The MEN and the Universidad de Los Andes visited and installed the dashboard in the main HEI; MEN expected that the 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



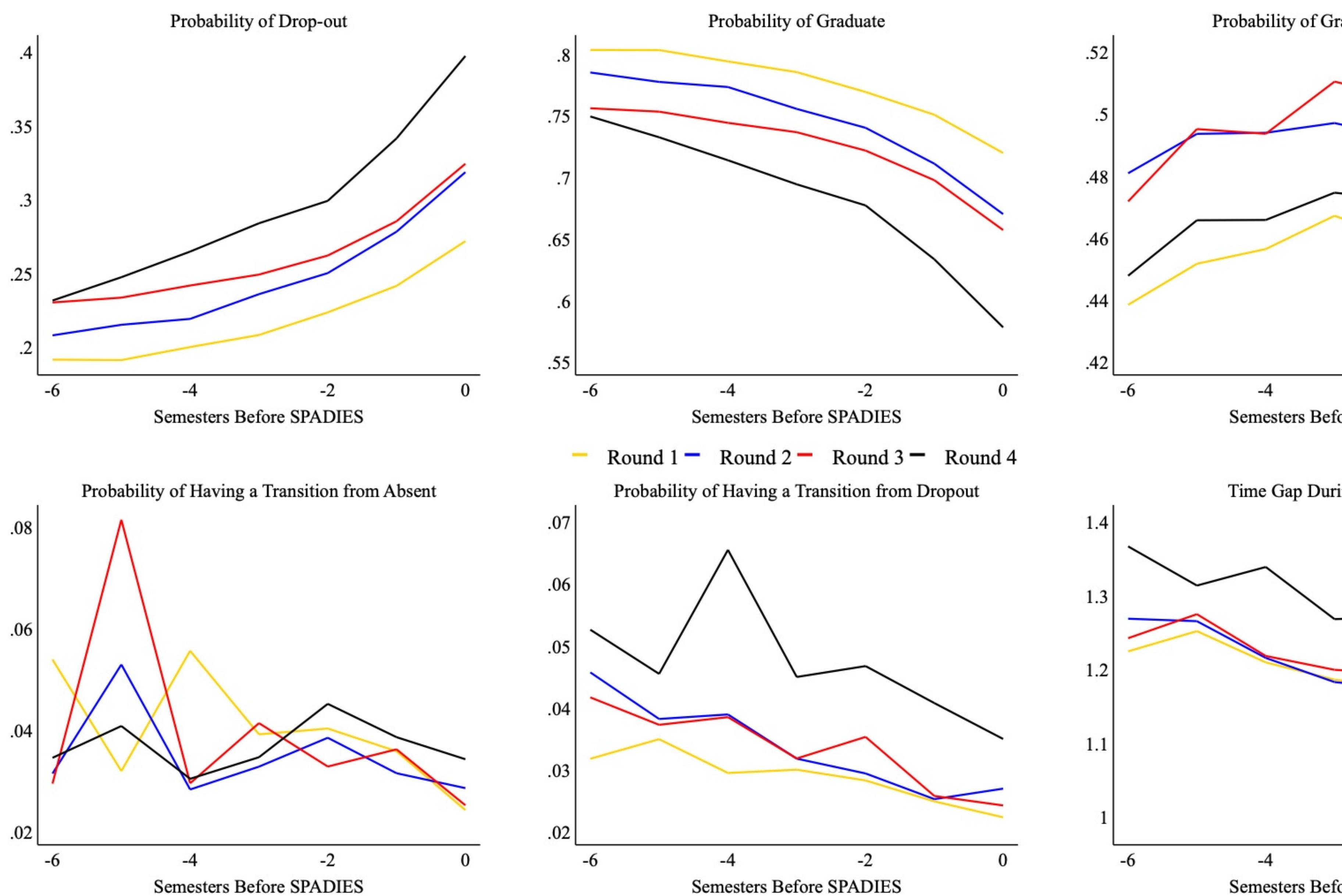
Notes: The figure shows the regression coefficients explaining the treatment variable using the students' main characteristics and HEIs. Whiskers show the 95% confidence interval. Sex is a dummy that is 1 if the student is female. The secondary test score is a dummy that is 1 if the student is in the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

Figure 3: SPADIES status analysis



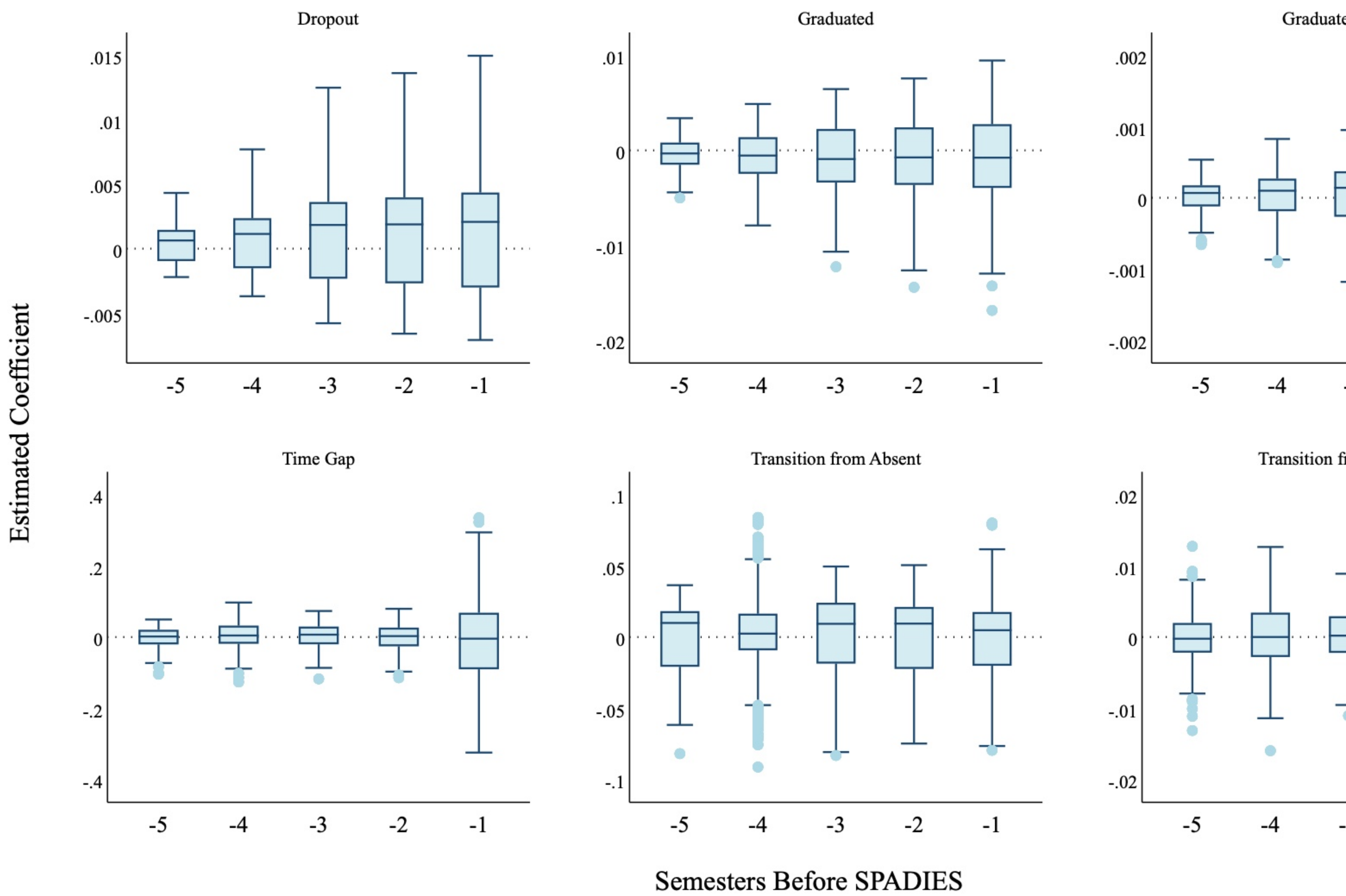
Notes: All the statuses in SPADIES are dynamic in time except by Graduated. Transition I shows an enrolled status once the student was Absent per 1 semester and returned to the system. Transition II shows that after being considered a dropout student the status was updated to "enrolled" once the student 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 Averages



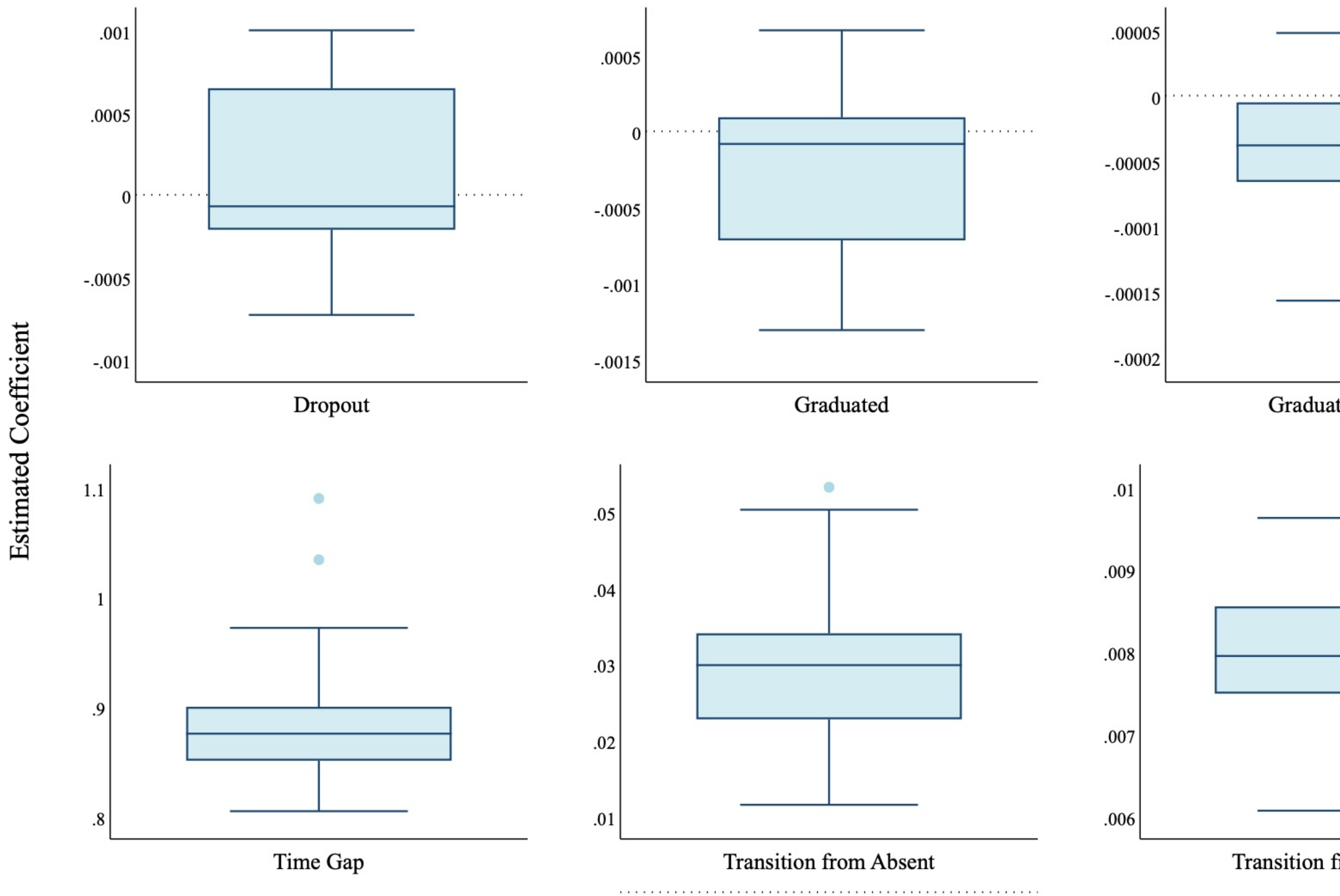
Notes: Chart reports the averages per round for the out put variables before SPADIES. Averages estimates using the full sample 4,131,302 individuals.

Figure 5: Parallel Trends - Classic DiD



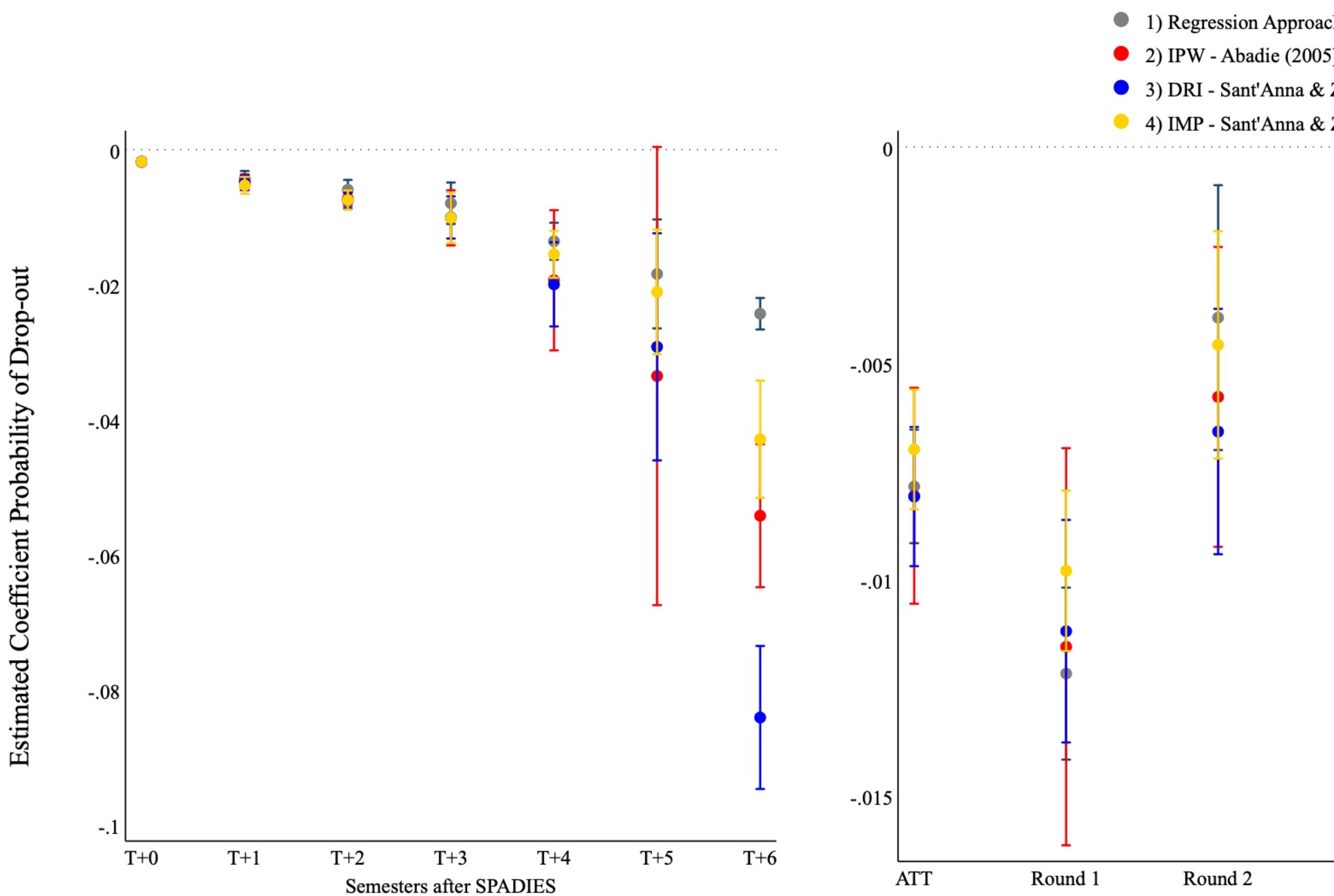
Notes: Chart reports estimated coefficients for the placebo test using the classic DiD methodology by using the placebo sample with 2,015,868 individuals.

Figure 6: Parallel Trends - Callaway and Sant'Anna (2021) ATT Estimation



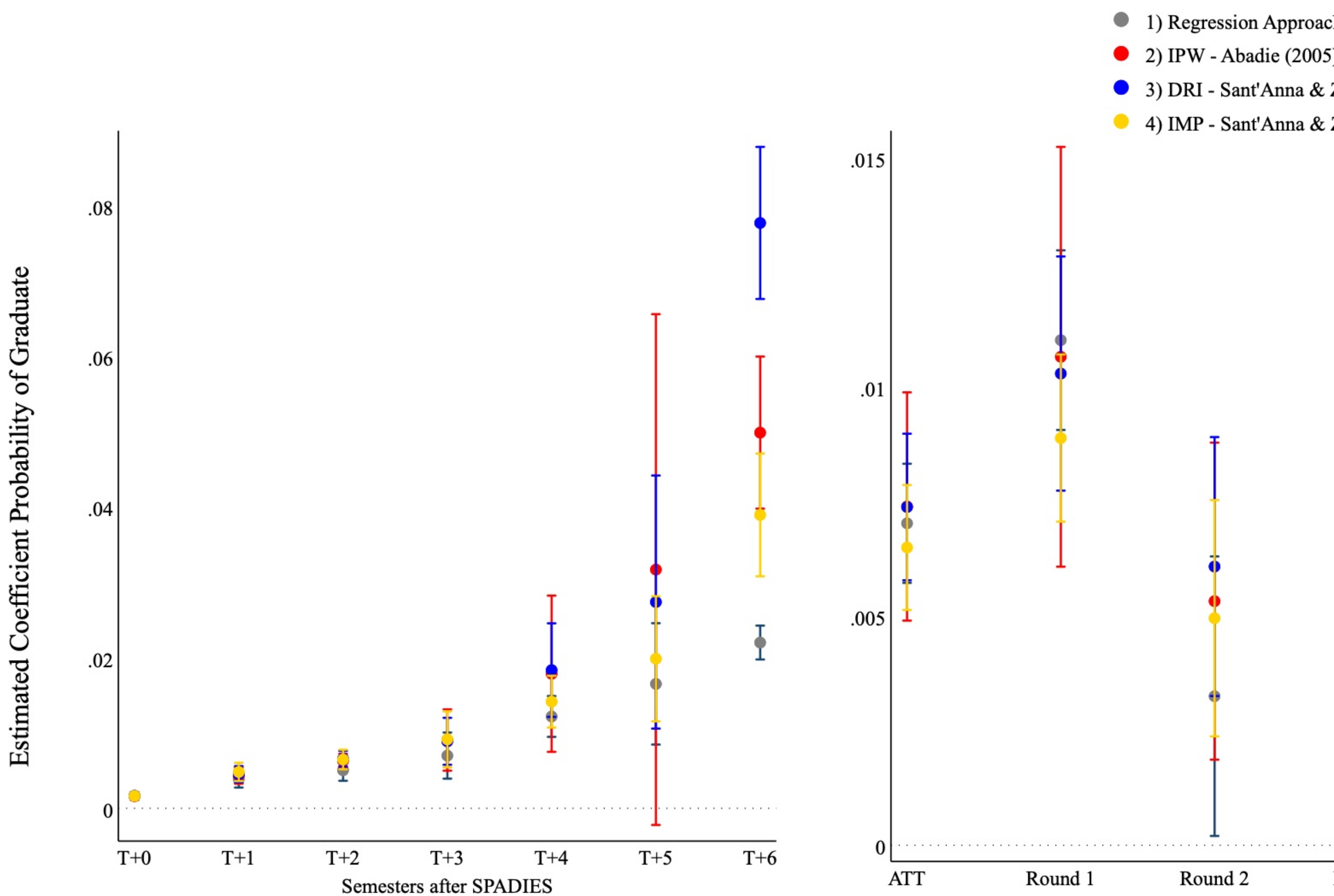
Notes: Chart reports estimated coefficients for the placebo test using the Callaway and SantAnna (2021) methodology by using the placebo sample with 2,015,868 individuals.

Figure 7: SPADIES ATT for the Probability of Dropping Out



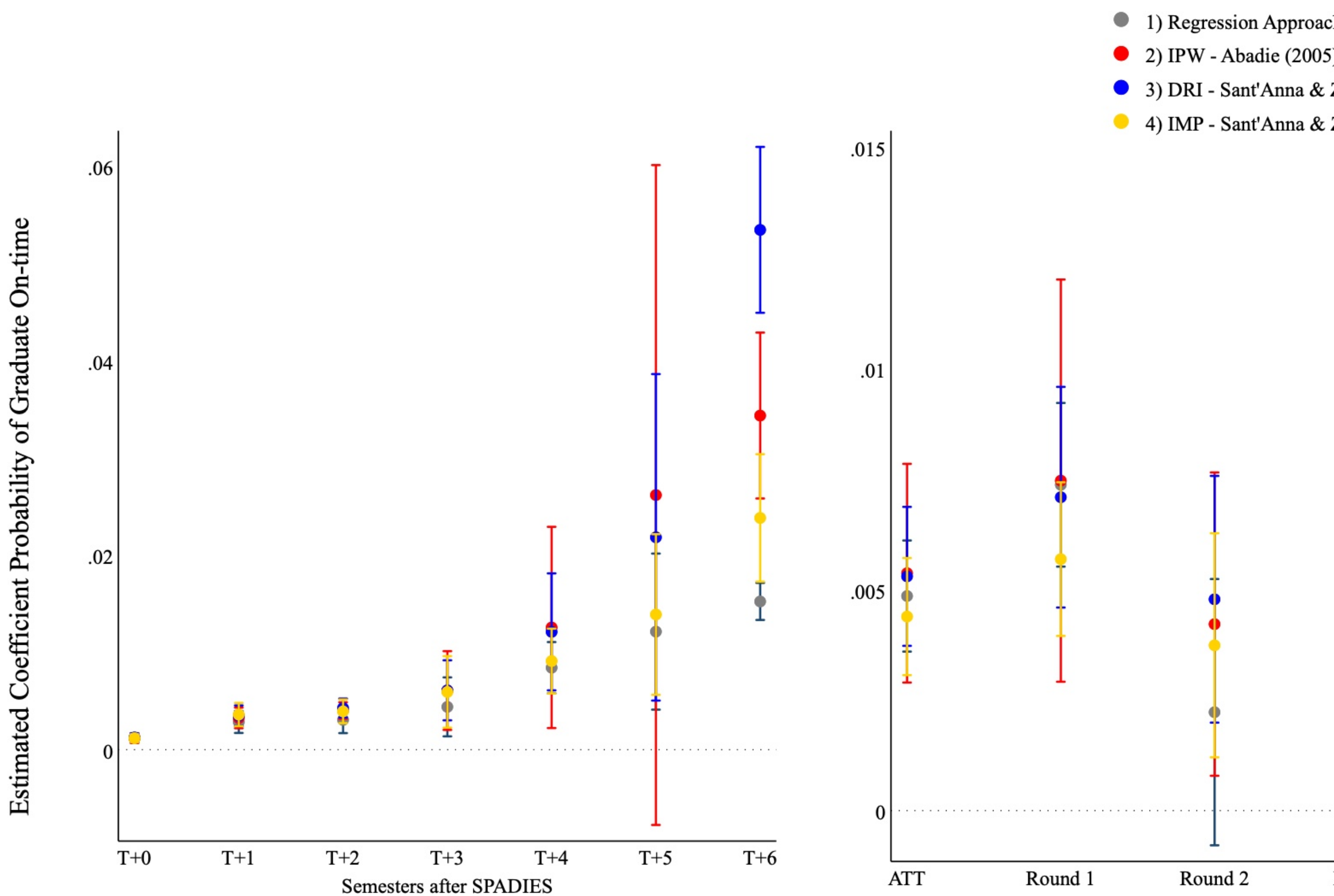
Notes: Chart reports the estimated coefficients for the probability of dropping out using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95%) using the full sample (4,131,302 individuals). The Chart is divided into two panels. On the left is an event analysis report and on the right panel a comparison of coefficients for the total and per round. Coefficients were estimated using Callaway and Sant'Anna (2021)'s framework using Rios-Avila et al. (2021) CSDID command in Stata.

Figure 8: SPADIES ATT for the Probability of Graduation



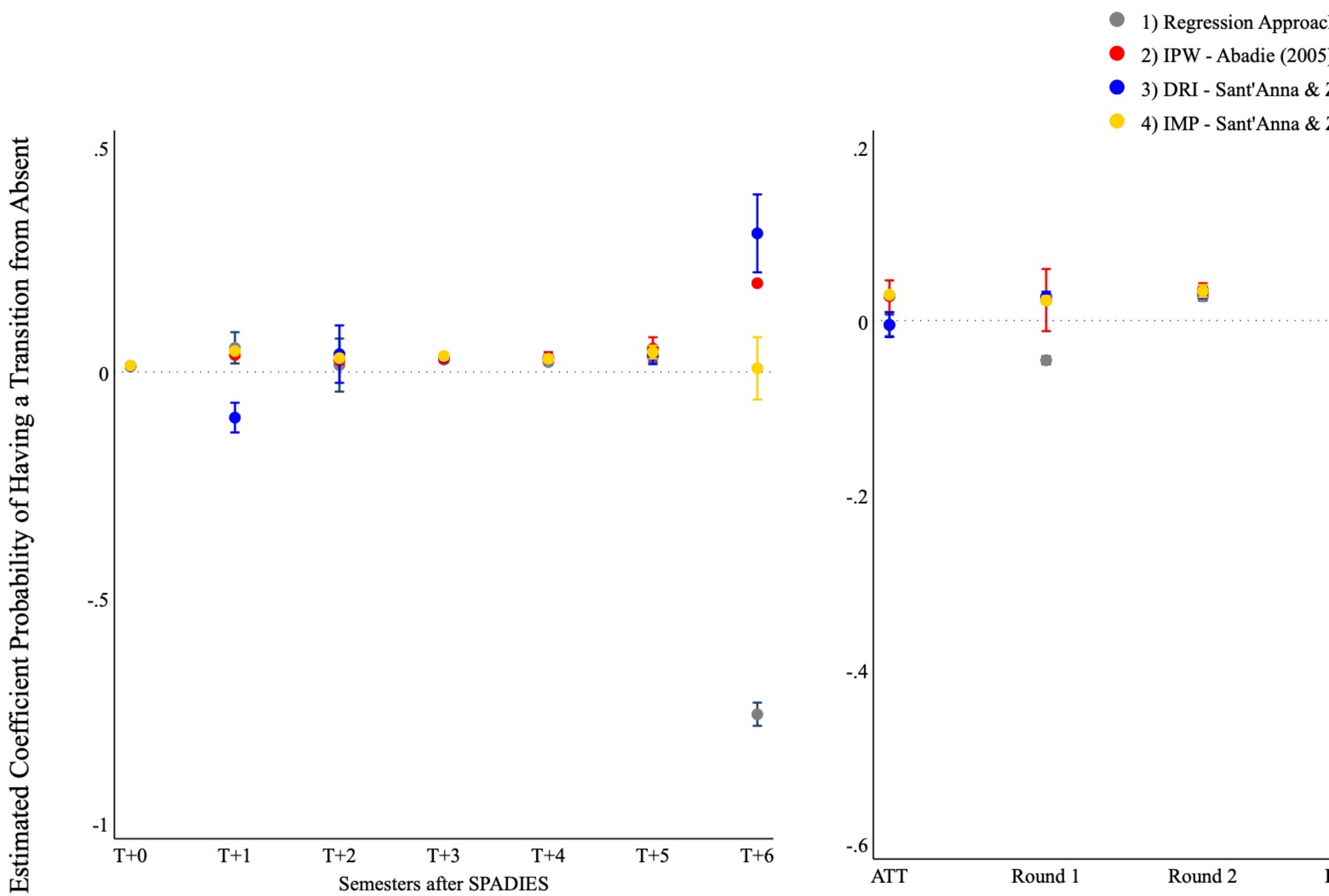
Notes: Chart reports the estimated coefficients for the probability of graduating using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95%) using the full sample (4,131,302 individuals). The Chart is divided into two panels. On the left is an event analysis report and on the right panel a comparison of coefficients for the total and per round. Coefficients were estimated using Callaway and Sant'Anna (2021)'s framework using Rios-Avila et al. (2021) CSDID command in Stata.

Figure 9: SPADIES ATT for the Probability of Graduating On-time



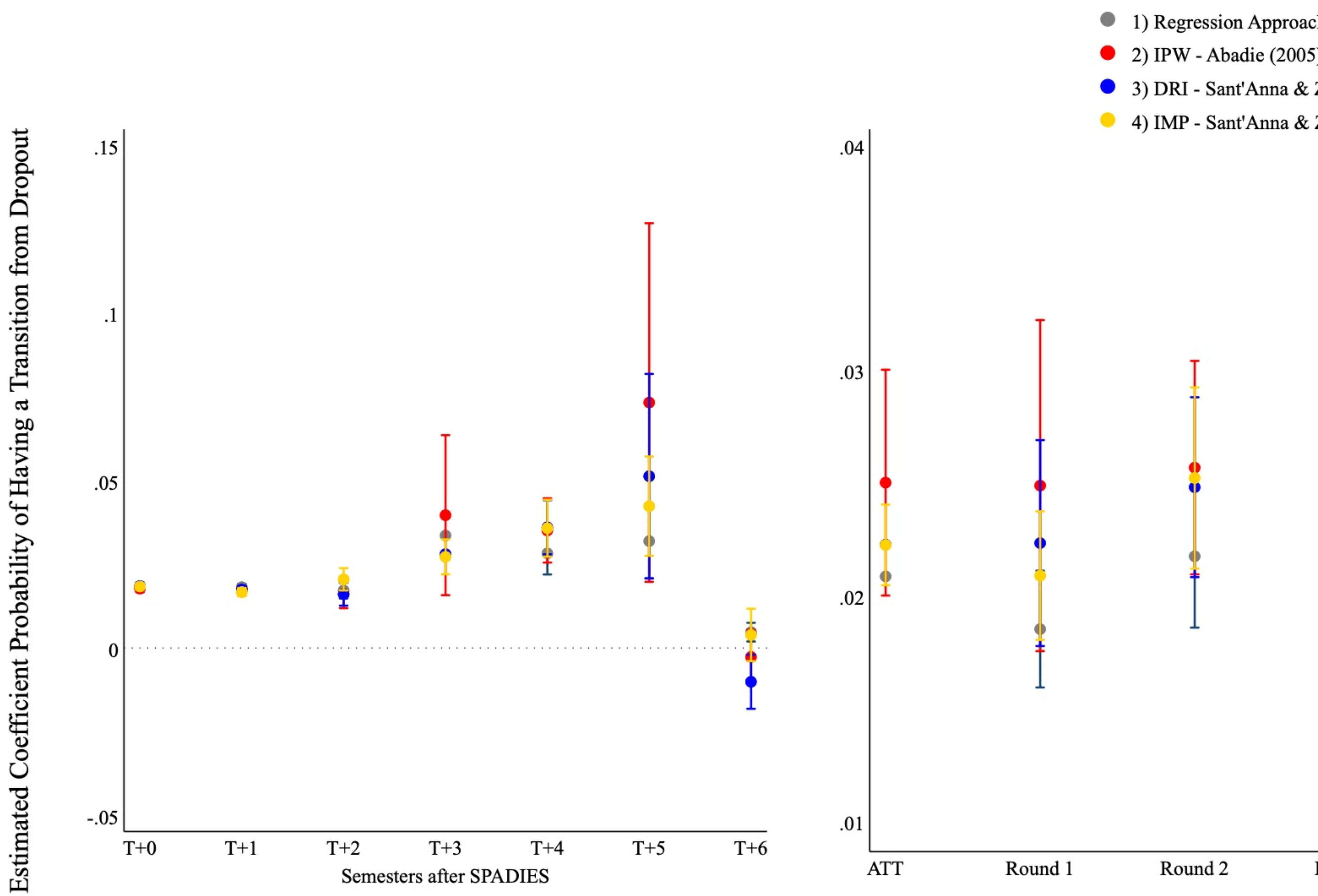
Notes: Chart reports the estimated coefficients for the probability of graduating on time using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95%) using the full sample (4,131,302 individuals). The Chart is divided into two panels. On the left is an event analysis report and on the right panel a comparison of coefficients for the total and per round. Coefficients were estimated using Callaway and Sant'Anna (2021)'s framework using Rios-Avila et al. (2021) CSDID command in Stata.

Figure 10: SPADIES ATT for the Probability of Having a Transition from Absent



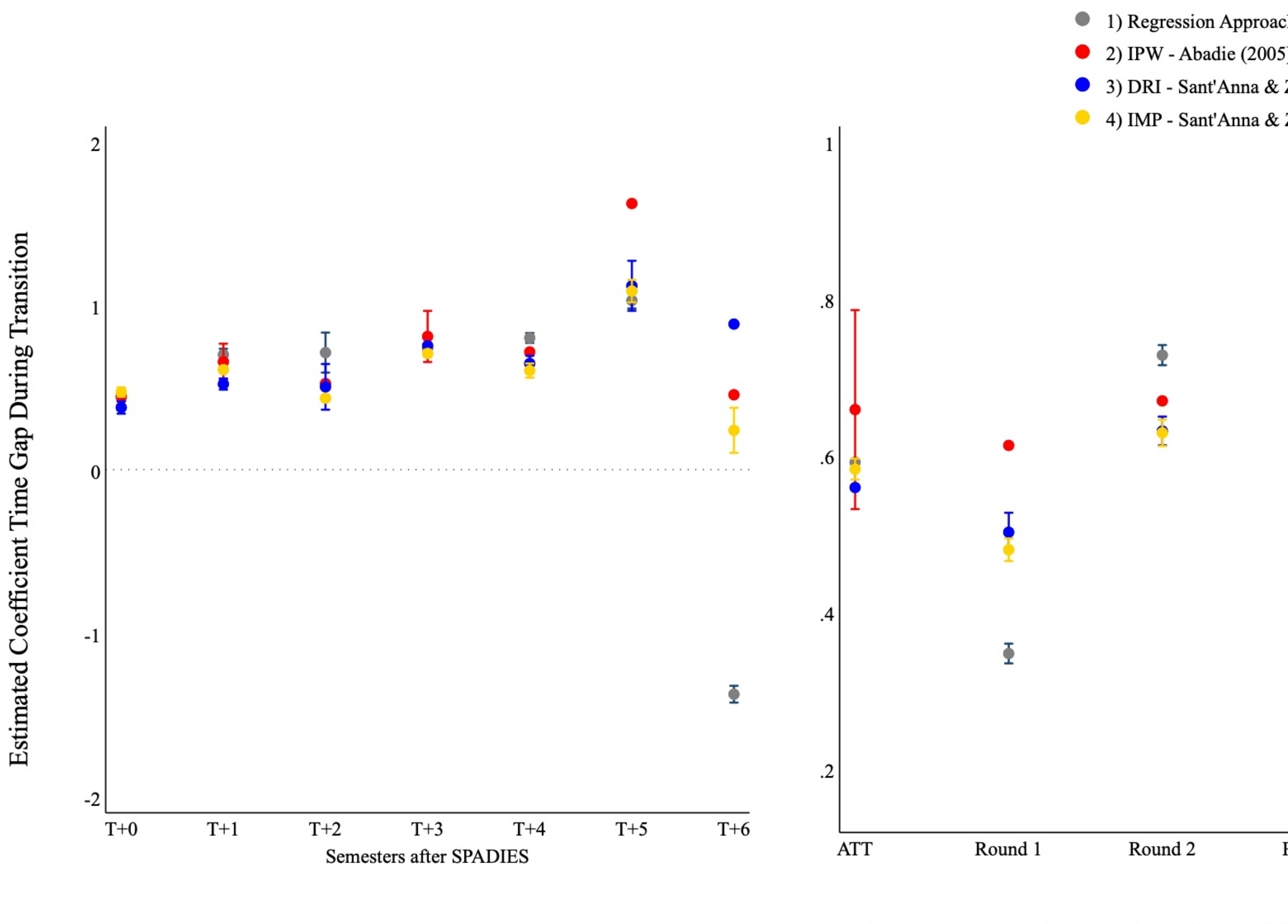
Notes: Chart reports the estimated coefficients for the probability of having a transition from Absentusing Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95%) using the full sample (4,131,302 individuals). The Chart is divided into two panels. On the left is an event analysis report and on the right panel a comparison of coefficients for the total and per round. Coefficients were estimated using Callaway and Sant'Anna (2021)'s framework using Rios-Avila et al. (2021) CSDID command in Stata.

Figure 11: SPADIES ATT for the Probability of Having a Transition from Drop-out



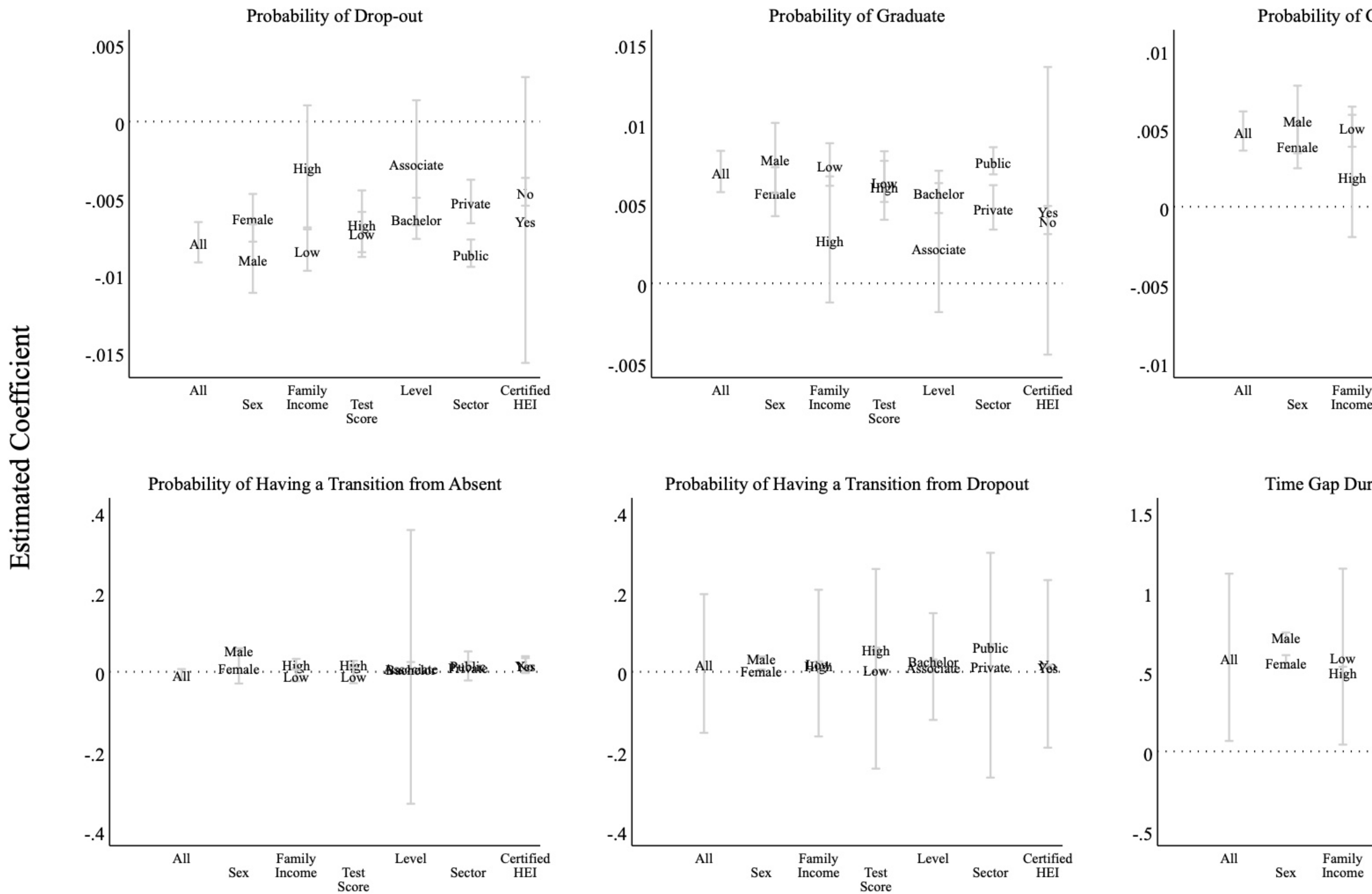
Notes: Chart reports the estimated coefficients for the probability of having a transition from drop-out using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95%) using the full sample (4,131,302 individuals). The Chart is divided into two panels. On the left is an event analysis report and on the right panel a comparison of coefficients for the total and per round. Coefficients were estimated using Callaway and Sant'Anna (2021)'s framework using Rios-Avila et al. (2021) CSDID command in Stata.

Figure 12: SPADIES ATT for the Time Gap During the Transition



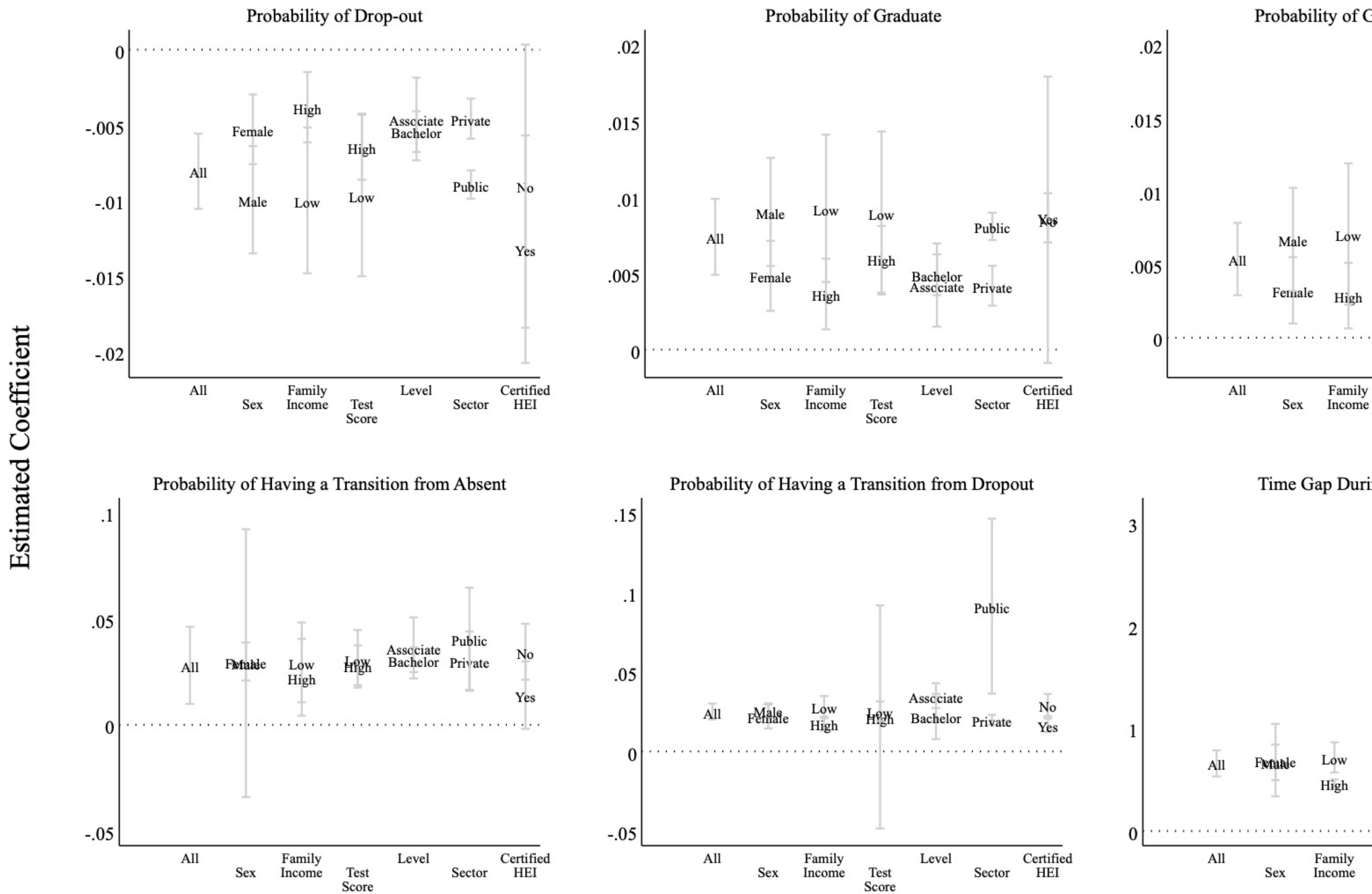
Notes: Chart reports the estimated coefficients for the time gap during transition using Equations 7 to 12 -OR, IPW, DRI, IMP- (whiskers at 95%) using the full sample (4,131,302 individuals). The Chart is divided into two panels. On the left is an event analysis report and on the right panel a comparison of coefficients for the total and per round. Coefficients were estimated using Callaway and Sant'Anna (2021)'s framework using Rios-Avila et al. (2021) CSDID command in Stata.

Figure 13: Disaggregated ATT Results Using OR Approach



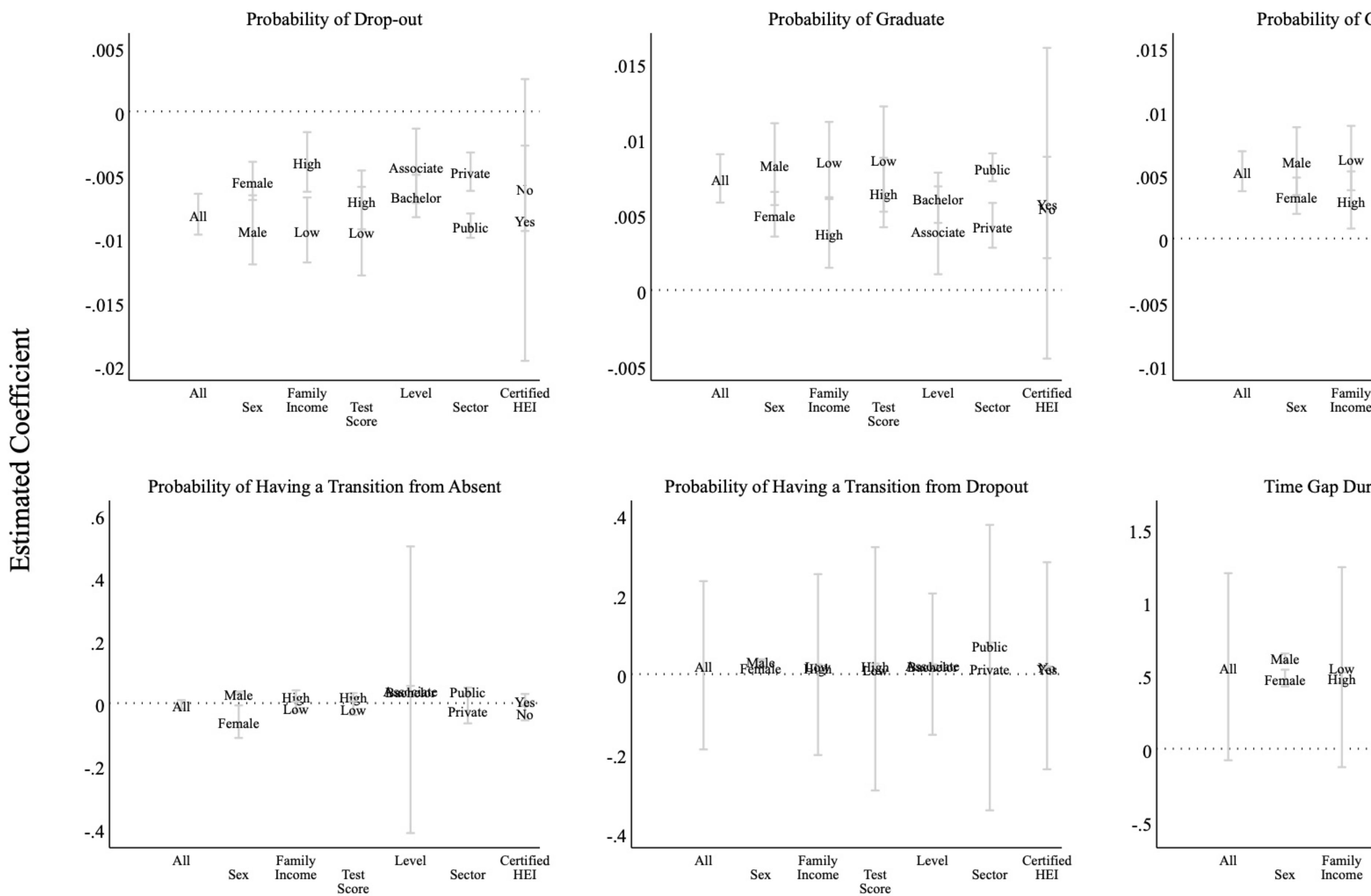
Notes: Chart reports the coefficients using Equation 2.7 -OR- (whiskers at 95%) for the estimations dividing the main sample into groups created using the time-invariant variables in the six outputs analyzed.

Figure 14: Disaggregated ATT Results Using IPW Approach



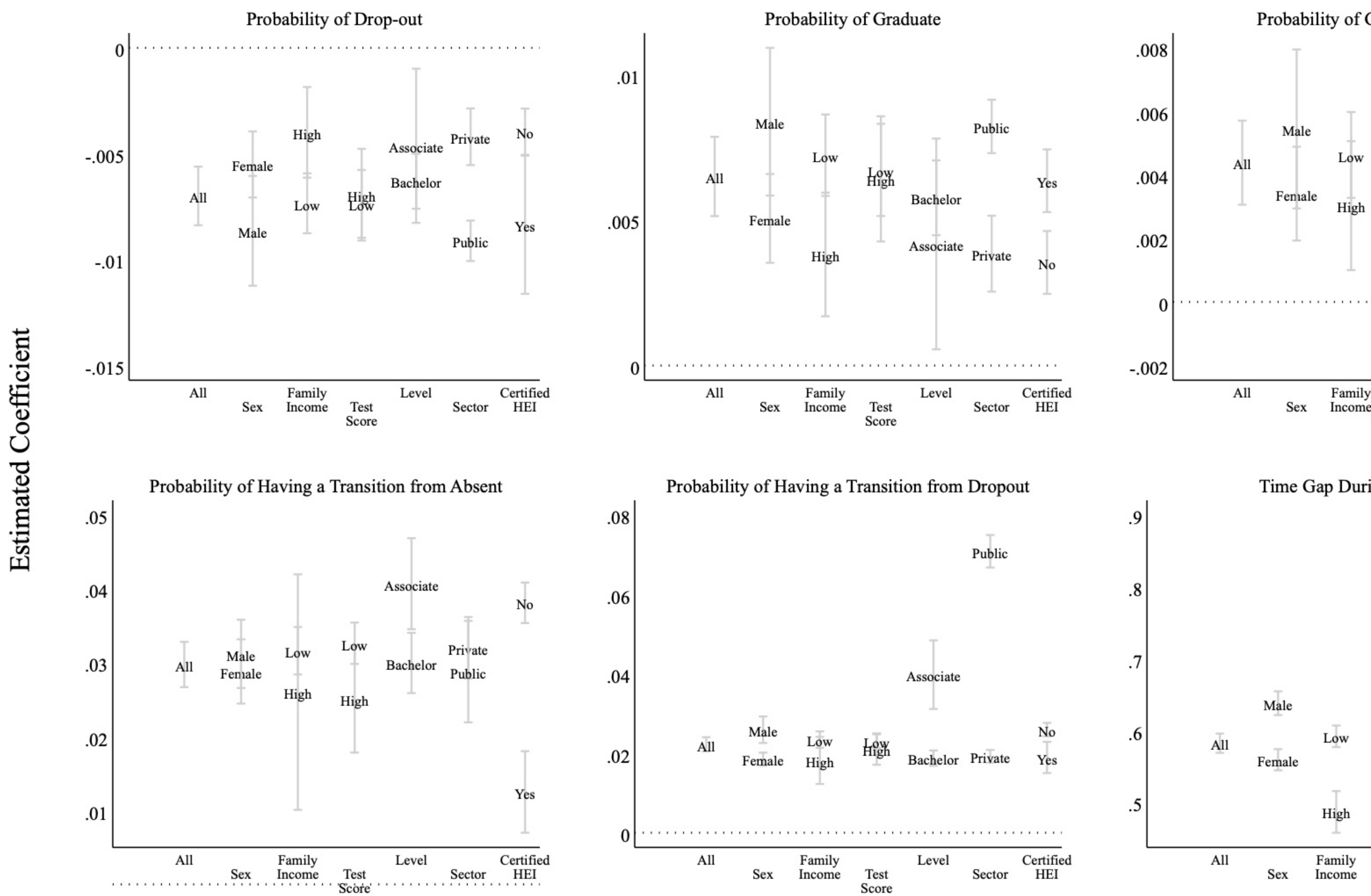
Notes: Chart reports the coefficients using Equation 2.9 -IPW- (whiskers at 95%) for the estimations dividing the main sample into groups created using the time-invariant variables in the six outputs analyzed.

Figure 15: Disaggregated ATT Results Using DRI Approach



Notes: Chart reports the coefficients using Equation 2.11 -DRI- (whiskers at 95%) for the estimations dividing the main sample into groups created using the time-invariant variables in the six outputs analyzed.

Figure 16: Disaggregated ATT Results Using IMP Approach



Notes: Chart reports the coefficients using Equation 2.12 -IMP- (whiskers at 95%) for the estimations dividing the main sample into groups created using the time-invariant variables in the six outputs analyzed.