ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-249

The data revolution comes to higher education: Identifying students at risk of dropout in Chile

Paul T. von Hippel
University of Texas, Austin

Alvaro Hofflinger
Universidad de la Frontera

Enrollment in higher education has risen dramatically in Latin America, especially in Chile. Yet graduation and persistence rates remain low. One way to improve graduation and persistence is to use data and analytics to identify students at risk of dropout, target interventions, and evaluate interventions' effectiveness at improving student success. We illustrate the potential of this approach using data from eight Chilean universities. Results show that data available at matriculation are only weakly predictive of persistence, while prediction improves dramatically once data on university grades become available. Some predictors of persistence are under policy control. Financial aid predicts higher persistence, and being denied a first-choice major predicts lower persistence. Student success programs are ineffective at some universities; they are more effective at others, but when effective they often fail to target the highest risk students. Universities should use data regularly and systematically to identify high-risk students, target them with interventions, and evaluate those interventions' effectiveness.

VERSION: July 2020

Suggested citation: von Hippel, Paul T., and Alvaro Hofflinger . (2020). The data revolution comes to higher education: Identifying students at risk of dropout in Chile. (EdWorkingPaper: 20-249). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/2qfx-tq33

The data revolution comes to higher education: Identifying students at risk of dropout in Chile

Paul T. von Hippel
LBJ School of Public Affairs, University of Texas, Austin, Texas, USA
paulvonhippel.utaustin@gmail.com

Alvaro Hofflinger
Núcleo de Ciencias Sociales, Universidad de la Frontera, Temuco, Chile
alvaro.hofflinger@gmail.com

The data revolution comes to higher education: Identifying students at risk of dropout in Chile

Abstract

Enrolment in higher education has risen dramatically in Latin American, especially in Chile. Yet graduation and persistence rates remain low. One way to improve graduation and persistence is to use data and analytics to identify students at risk of dropout, target interventions, and evaluate interventions' effectiveness at improving student success. We illustrate the potential of this approach using data from eight Chilean universities. Results show that data available at matriculation are only weakly predictive of persistence, while prediction improves dramatically once data on university grades become available. Some predictors of persistence are under policy control. Financial aid predicts higher persistence, and being denied a first-choice major predicts lower persistence. Student success programs are ineffective at some universities; they are more effective at others, but when effective they often fail to target the highest risk students. Universities should use data regularly and systematically to identify high-risk students, target them with interventions, and evaluate those interventions' effectiveness.

Keywords: prediction, dropout, analytics, evidence, data

Introduction

Latin Americans' participation in higher education has broadened dramatically in recent years. Between 2000 and 2013 alone, the percentage of Latin America's 18-24-year-old population who attended higher education institutions doubled, from 21 to 43 percent. The growth rate has been even faster for women and for students from low-income families. The fastest growth has occurred in Chile, Peru, Bolivia, and Ecuador. Today approximately 20 million Latin American students attend 60,000 programs at 10,000 higher education institutions (Ferreyra, Avitabile, Botero, Haimovich, & Urzúa, 2017).

While rising participation stems in part from economic modernization and changing social norms, government policies and subsidies have also played a major role, especially in Chile. Since 2000, Chile has paid conditional cash transfers, or "fellowships," to incentivize poor or pregnant teenagers to complete high school. In 2003, Chile made high school completion mandatory. These reforms dramatically increased the percentage of poor students who completed secondary education and became eligible for higher education, Since 2005, Chile's central government has guaranteed student loans, and since 2015, Chile has offered free tuition to all students whose families came from the bottom 60 percent of the income distribution, expanding eligibility to the bottom 60 percent in 2018.

Although laudable, the impact of expanding access to higher education can be limited if many students drop out. The success of higher education policies depends not just on the number of students who enter the higher education system, but on the number who complete degrees. Yet higher education institutions graduate only about half of enrolled students in Latin America (Ferreyra et al., 2017), compared to about two-thirds of enrolled students in the United States (Tinto, 2012) . Students who drop out do not just squander their own investments in time, money, and opportunity cost. In a subsidized system, they also waste investments made by the government and taxpayers.

An emerging approach to reducing dropout is to use data for prediction, intervention, and evaluation (Baepler & Murdoch, 2010; de Freitas et al., 2015; Milliron, Kil, Malcolm, & Gee, 2017; Picciano, 2012). Data on student background, engagement, and performance can be used to fit predictive models that identify students at higher risk of dropout. Students flagged by these models as high risk can be targeted for interventions—ranging from a simple email to enrolment in a comprehensive student success program. New interventions can be developed to move the variables that risk models suggest are most predictive of dropout. The effectiveness of these interventions can be evaluated using much of the same data that was used for prediction—particularly if interventions are rolled out in a controlled way, using randomization, matching, or sharp eligibility thresholds. More effective interventions can be expanded; less effective interventions can be scaled back or retired. A continuing cycle of prediction, intervention, and evaluation can yield continuous improvements in student success.

The use of data for prediction, targeting, and evaluation has revolutionized industry after industry—from credit scoring and direct mail marketing in the 1960s to fraud detection and online search and advertising in the 1990s. Higher education has come late to the data revolution, which is surprising given the number of higher education workers with advanced data skills. Yet in the past few years hundreds of higher education institutions have begun using data for prediction, intervention, and evaluation. Some institutions develop their data infrastructure internally (Laude, Kaschner, Alvarado, & Connerat, 2017) while others partner with vendors such as Starfsh, EAB, or Civitas Learning (Milliron et al., 2017).

To date, the data revolution in higher education has been concentrated in the US and UK. Yet the revolution may have even greater potential in the developing world. Because citizens' educational attainment predicts national economic growth (Hanushek & Woessmann, 2010), many countries' development plans call for increasing the number of adults with higher education. Increasing the educational attainment of future parents is also one of the most effective ways to increase children's academic performance in the next generation (Hofflinger & von Hippel, 2017).

In this article, we demonstrate the potential of predictive models that highlight students at risk of dropout, and policies and programs that reduce dropout, at eight Chilean universities. Although the results vary across universities, all show that the student variables that are available at matriculation—including high school grades, entrance exam scores, and family income—are only weakly predictive of dropout. In most universities, dropout prediction improves dramatically once actual performance in university courses can be observed. Some predictors of persistence are under policy control. Financial aid predicts higher persistence, and being denied a first-choice major predicts lower persistence. Some student success programs are effective, but others are ineffective, and even effective programs often fail to target high-risk students who might benefit the most from them. The results highlight opportunities to better target and improve programs and policies to increase student success.

The Higher Education System in Chile

Until 1980, higher education in Chile was an elite system. Only 10 percent of adults 20-24 were enrolled in higher education. In a population with 1.1 million adults aged 20-24, there were only 110,000 students enrolled at just eight public and nine private, non-profit, mostly Catholic universities. But starting in 1981, midway through the military dictatorship of Augusto Pinochet (1973-1990), Chile's higher education system underwent radical change and expansion (Torche, 2005). Government funding, previously a fixed amount per student, was tied to rankings based on quality metrics including the student-instructor ratio, the percentage of instructors with PhDs, and research productivity (Arocena & Sutz, 2001; Manríquez, Mendoza, & Ramírez, 2015; Schiefelbein, 1990; Torche, 2005). New university campuses opened, and the whole university system was exposed to competition from a new for-profit, low-tuition sector of "professional institutes" and "technical training centres" (Bernasconi & Rojas, 2004). By 1991 there were 303 higher education institutions, most of in the new sectors; today there are 147 higher education institutions, but only 27 are traditional universities—18 public and 9 private.

Since Chile's return to democracy in 1990, the national government has implemented a series of reforms to broaden access to universities through state subsidies. In the 1990s, the government funded a limited number of zero-interest loans, direct from the government, to expand university access to students from low income families (Bernasconi & Rojas, 2004). (The second author attended university under that program.) Default rates were high, however, and the growth of the zero-interest loan program did not keep up with enrolment growth in higher education. In 2005, Chile's government began to guarantee fixed-interest private loans made by seven major banks to students who enrolled in accredited universities, technical training centres, or professional institutes (Mineduc, 2005). Participating banks could no longer request students' credit history or debt capacity, which had previously limited credit for students from low-income families. By 2010, one quarter of Chilean undergraduates received government-backed loans (World Bank, 2011). Although sceptics raised concerns that government guarantees would increase dropout, dropout rates have been lower among students with loan guarantees than among students who paid their own way (Santelices, Catalán, Kruger, & Horn, 2016).

An even more significant policy launched in 2015, when Chile passed the *ley de gratuidad*, giving free tuition to university students from families in the bottom half of the national income distribution. In 2018, eligibility expanded to the bottom 60 percent of the income distribution and to students at professional institutes and technical training centres. Approximately 250,000 of the 1.2 million students now enrolled in Chile's higher education system receive free tuition under the *ley de gratuidad* (Mineduc, 2017).

Figure 1 shows Chile's higher education enrolment trends since 1990. Both enrolment and equity have increased dramatically. Enrolment more than doubled in the 15 years after 1990, and doubled again, even more quickly, after government guarantees of private loans began in 2005. Most of the growth has been in the lower income quintiles. As recently as 2000, the higher education enrolment rate was over 5 times higher for students from the top income quintile than for students from the bottom quintile, but by 2016, the enrolment rate was only 1.6 times higher for the top quintile than for the bottom. Guaranteed loans and free tuition were major reasons for this convergence, although rising incomes and falling income inequality may also have played a role.

While government reforms have succeeded in broadening access, they have also raised concerns that the new, lower income students will have higher dropout risk, and the risk will be borne by taxpayers. Over 20 percent of university students and over 30 percent of students at professional institutes and technical training centres drop out within their first year. Although dropout *rates* are no higher today than they were in 2009, the total *number* of dropouts has increased with rising enrolments. To limit public liability, Chile's government cuts off free tuition from students who do not graduate within 3 to 5 years (depending on the major). This increases the incentive for institutions to encourage timely graduation.

A few studies have predicted higher education dropout in Chile. Significant academic predictors included high school grade point average, scores on the national higher education entrance exam (the PSU), and grades and number of courses taken during the first year of postsecondary education (Espinoza, 2015; Saldaña & Barriga, 2010). Financial aid was also a significant predictor of persistence (Santelices et al., 2016) , which to some extent validates the increasingly generous subsidies that Chile has offered to low-income students. Social psychological predictors include family psychological support (Himmel, 2002), students' loyalty to the institution (Rojas-Méndez et al., 2009), students' clarity of intentions regarding their major, and students' expectations regarding the value of their degree (Canales & De los Ríos, 2007). Many of these results echo findings from Europe and the US, where researchers have identified additional predictors, such as whether the student is satisfied with their major, and whether they have the ability to change if they are dissatisfied (Yorke & Longden, 2004). In Europe, working even part-time while in university is associated with lower persistence (Beerkens, Mägi, & Lill, 2011; Hovdhaugen, 2015), but in the US, working part-time is innocuous, but working full-time slows accumulation of credits (Darolia, 2014). Net of other variables, most demographic characteristics are not strong predictors of persistence; in the US, for example, black and Hispanic students are less likely to persist than white students on average, but at least as likely to persist as white students once high school grades, test scores, and income are controlled (Murtaugh, Burns, & Schuster, 1999) .

Despite this body of knowledge, "much of the research on student attrition has not been particularly helpful to those in the field who seek to develop and implement programs to improve retention and completion"(Tinto, 2012). Some variables, such as high school grades and family psychological support, are difficult or impossible for universities to control, and some, such as loyalty or engagement, are rarely measured in institutional data. Short of subsidizing tuition, which is expensive and has already been done by the Chilean government, it is not always obvious what a higher education administrator can do to increase student persistence.

In the rest of this article, we show when and how routinely collected institutional data can be used to identify high-risk students, target interventions, and evaluate intervention effects.

Data

We obtained data for approximately 28,000 students who enrolled at eight public universities in Chile, which were willing to provide information for this study. According to the the Ministry of Education's rankings of Chile's 27 public and private universities, two of the universities in our sample are in the top 10, and two are ranked near the bottom (Mineduc, 2018). These rankings affect universities' fiscal allocation from the central government as well as their reputation with potential students. Higher ranked universities can be more selective in which students they admit.

The students were in two cohorts; one cohort started university in March 2016, and one started in March 2017. In Chile, the school year runs from March through December, with a semester break in July, which is midwinter in the southern hemisphere.

We merged three different data sources. The first came from the Department of Evaluation, Measurement and Educational Records (DEMRE), which runs the national test required for admission to Chilean universities: the *Prueba de Selección Univeritaria* (PSU). DEMRE is roughly analogous to the College Board, which runs the Scholastic Achievement Test in the US. DEMRE data included variables available at the time of testing, application, admission, and enrolment. Demographic variables included students' gender, their parents' income and education levels, and the number of years elapsed since they completed high school (which correlated at 0.95 with student age). Academic variables included the type of high school that they attended (academic or vocational), their PSU scores, and their high school grade point average (GPA). To ease interpretation, we standardized GPA and PSU to have a mean of zero and a standard deviation (SD) of one. Enrolment variables included the student's major, and dummy variables indicating whether that major was their first, second, or lower choice. Since some majors had low enrolments, we grouped majors into six larger categories: natural sciences, social sciences, health, engineering, education, and law and accounting.

The second data source came from the Ministry of Education's Uniform Form for Student Socioeconomic Accreditation (FUAS), which is similar to the Free Application for Federal Student Aid (FAFSA) in the US. The FUAS contained the names of any fellowships that the student was awarded, including the *gratuidad*.

The third data source came from participating universities and contained variables summarizing students' early academic performance in university: their GPA from the first and second semester, and an indicator for whether they dropped out. We used the definition of dropout established by the Ministry of Education, which defined dropout as not enrolling in the same university for a third semester—i.e., for the first semester of the second year. Some students were missing GPA for the first or second semester, which indicated that they dropped out sooner. University 6 did not provide GPAs for semester 1.

Four of the eight universities also provided a dummy variable indicating which students participated in student success programs. Three universities provided a student success indicator for both cohorts, and one provided it for the 2017 cohort only. Although we lack details about specific programs, we do know something about their general characteristics. All success programs were free, and participation was voluntary. Some programs focused primarily on helping students adapt to university life, develop study skills, or manage test anxiety. Other programs focused on building basic skills in specific academic areas, usually in math, engineering, and natural sciences, and occasionally in social sciences as well.

Our data end at the start of year 2, but this not as limiting as it might appear. Most dropout happens early (Santelices et al., 2016), so accurate prediction and effective intervention must happen early as well.

Methods

While dropout is sometimes predicted with machine learning techniques, such as neural networks or random forests, simple logistic regression often predicts dropout nearly as well (Sorensen, 2016) or even better in some datasets (Aulck et al. 2016), at least when the number of good predictors is small. Logistic regression is also highly interpretable, which is helpful when analysts are familiarizing themselves with data, evaluating data quality, and communicating results to university leaders. (Classification trees are another simple and interpretable method.) Once a simple model is established, machine learning techniques may be useful to derive features and improve prediction, but the difference between using or not using a predictive model is much more important than which specific model is used.

In the 2016 cohort, we modelled students' probability of dropout using three different logistic regression models. Each was applicable at a different stage of students' university careers. The first model used data available at the start of the first semester. The second model included first semester GPA, which was available after the first semester for those students who finished it. The third model included second semester GPA, which was available after the first semester for those students who finished it.

We fit models separately to each university, estimating model parameters in the 2016 cohort and using those estimates to predict dropout in the 2017 cohort. In practice, this is how decision makers must use predictive models: data from past cohorts is used to predict dropout in future cohorts. Using different data for estimation and prediction also prevents us from overestimating the predictive accuracy of the model. Over-fitting is less of a danger when the number of predictors is small compared to the number of cases, but it is possible for a model fit to 2016 data to be less predictive in 2017.

We summarized the predictive validity of the models with a series of three *lift curves*. In practical terms, a lift curve shows what fraction X of students we would need to target with an intervention in order to reach some fraction Y of the students who would otherwise drop out. Each lift curve was calculated by first sorting students by descending dropout risk and then walking down the list and counting the students who later dropped out. For each student, the lift curve plots the percentage of all students that we have looked at so far (X) against the cumulative percentage of later dropouts (Y) that are included among those students.

Results

Some of the eight universities were more elite and selective than others. Table 1 illustrates the differences among the eight universities in descriptive statistics. Dropout rates ranged from 4 percent to 23 percent. Mean incomes were 42 percent higher at the university with the most affluent students than at the university with the least affluent students. Mean PSU scores were 1.3 SD higher at the highest-scoring university than at the lowest scoring university. The three universities with the highest PSU scores had the lowest dropout rates, though overall the correlation between PSU and dropout was only moderate (-0.4) as was the correlation between dropout and mean income (-0.4).

Predicting dropout in University 1

Let's begin by predicting dropout at University 1. Odds ratios for different predictors appear in the first three columns of Table 2a, in bold. Later, we'll compare results from other universities.

Prediction before semester 1

When students applied to University 1, several variables were significant predictors of dropout. The PSU exam score was one; holding other variables constant, a student with an exam score that is one SD higher had one-third lower odds of dropout ($p<.001$). High school GPA was also predictive, but only half as much as the exam score; holding other variables constant, a student with one SD higher high school GPA had one-sixth lower odds of dropout ($p<.10$).

Many demographic and background variables were not predictive once GPA and PSU score are controlled. In particular, gender, family income, and attending a vocational rather than academic high school were not significant net predictors of dropping out of university. Years since high school graduation, by contrast, was predictive: holding other variables constant, each additional year elapsed between the last year of high school and the first year of university is associated with 10 percent lower odds of dropout. This runs contrary to stereotypes that older, non-traditional students have higher dropout risk.

Students in the natural sciences had dramatically higher dropout risk than students in other majors. For students with similar exam scores and other variables, the odds of dropout were 3 to 4 times lower in social sciences, education, and health, and 8 times lower in law and accounting, than in the natural sciences. Students' major mattered more than whether it was their first choice. Compared to similar students in their first-choice major, otherwise similar students in their second or third choice major had one-eighth to one-quarter higher odds of dropout—a non-significant difference.

Generous fellowships predicted lower dropout risk. Compared to similar students without fellowships, students who received free tuition (*la gratuidad*) had one-third lower odds of dropout. Other fellowships were also associated with lower dropout odds, but not significantly.

Prediction after semester 1 and semester 2

When data on student performance in University 1 came in, prediction of dropout improved dramatically, and some variables that were predictive at entry faded in importance. By the end of semester 1, high school GPA and PSU scores faded to insignificance, and first semester GPA became more predictive than those variables ever were. Holding other variables constant, a student whose semester 1 GPA was one SD higher had three times lower odds of dropout. After semester 2, prediction improved further; semester 1 grades became less predictive, and semester 2 grades became more predictive than semester 1 grades ever were. Holding semester 1 grades constant, a student whose semester 2 GPA was one SD higher had five times lower odds of dropout.

Differences in dropout risk between majors persisted as university GPAs were added to the model. For example, natural science students remained more likely to drop out even after early university GPA was controlled. So the fact that natural science courses gave lower grades was not a complete explanation for the higher dropout risk of natural science students. Differences between older and younger students, and between students with and without fellowships also persisted, with similar coefficients but varying levels of significance.

Lift curves

Figure 2 uses lift curves to validate how well the models, which were fit to the 2016 cohort, predicted dropout in the 2017 cohort. When we sorted the students in descending order by their modeled dropout risk, the lift curve shows what percentage of future dropouts were in the riskiest 10 percent, riskiest 20 percent, and so on. From a university leader's point of view, the lift curve shows what fraction of students would need to be targeted with an intervention in order to reach some fraction of the students who would otherwise drop out.

If the model utterly failed to predict risk in the 2017 cohort, it would sort students in an order that was unrelated to their true risk of dropout. The lift curve would be diagonal ($Y = X$); it would slice Figure 2 in half, and the area under the curve (AUC) would be 50 percent. This diagonal lift curve, drawn for reference in Figure 2, would imply that, for example, we would have to intervene with 50 percent of all students to reach 50 percent of those at risk of dropout. Such a non-predictive risk score would be useless for targeting interventions at high-risk students. We might as well apply some broad, inexpensive intervention to every student.

There are three lift curves above the diagonal. One uses the risk score that relies on data available at the start of semester 1. One uses the risk score that can be calculated after semester 1 (including semester 1 GPA). And one uses the risk score that can be calculated at the end of semester 2 (including semester 2 GPA).

At the start of semester 1, the risk score had limited value. The AUC was just 64 percent, or 14 percent above chance. To reach 80 percent of potential dropouts, we would have had to target approximately 60 percent of all students for intervention. To reach even 50 percent of potential dropouts, we would have had to target approximately 30 percent of all students. While it would have been possible to target interventions using this risk score, the targeting would have been very imprecise. A case could still be made for applying apply a cheap intervention to everyone.

By the end of semester 1, when we could use the first semester GPA, the curve lifted considerably. The AUC rose to 78 percent, and we could reach 50 percent of potential dropouts by targeting just the 10 percent of students with the highest risk scores. But to reach 80 percent of potential dropouts, we would have needed to target 40 percent of students.

By the end of semester 2, when we could use second semester GPA, the curve lifted further. The AUC rose to 89 percent, and we could reach 80 percent of potential dropouts by targeting just the 10 percent of students with the highest risk scores. By the end of semester 2, however, 12-13 percent of students had already dropped out, and another 7-8 percent would not return for the start of semester 3. For many dropouts, the end of semester 2 would have been simply too late to intervene.

Predicting dropout at other universities

Figure 3 gives lift curves for all eight universities. As in University 1, in all eight universities, the model using data available before semester 1 was not very useful for predicting dropout. The AUC (given in Table 2) ranged from 55 to 64 percent, and to reach 50 percent of future dropouts administrators would need to target 35 to 45 percent of students for intervention.

After semester 1, though, prediction improved substantially. While none of the other universities had quite as high a lift curve as University 1, in six of the eight universities the AUC rose to at least 74 percent when the model used data from semester 1, and to at least 78 percent when the model used data from semester 2. Prediction was generally better after semester 2 than after semester 1, but in most universities, the improvement in prediction was much smaller than it was in University 1. Again, for many students, by the time semester 2 data become available it would have been too late to intervene.

The major exceptions to these patterns were Universities 2 and 6, where the predictive models were practically useless, with AUC of just 55 to 61 percent, even after semester 2 data became available. A possible explanation for this is at that both universities had very low dropout rates, just 4 to 6 percent, and University 6 was highly selective, with the highest exam scores and family incomes among participating universities. Dropout at selective universities with low dropout rate may be hard to predict if it is unrelated to the academic and financial variables in our data; instead, dropout at elite universities may be related to social, psychological, and institutional factors that are not available in our institutional data. It was not always hard, though, to make predictions in universities with low dropout rates. University 8, for example, also had just a 6 percent dropout rate, but prediction there was excellent, with an AUC of 82 percent after semester 1 and 89 percent after semester 2.

Table 2 gives odds ratios from the logistic regression models fit to the 2016 cohort at each university. The results vary across universities, which emphasizes the importance of fitting tailored models to each university and not assuming that the same model with the same coefficients will fit everywhere.

Nevertheless, the results did follow a few consistent patterns. Before semester 1, high school GPA and PSU scores generally predicted dropout about equally. In some universities, PSU scores were a slightly better predictor; in others, high school GPA was a slightly better predictor; in others, neither variable predicted dropout well at all. University GPA predicted dropout much better than high school GPA. After semester 1, semester 1 GPA was more predictive than high school GPA, and after semester 2, semester 2 GPA was more predictive than semester 1 GPA. These results strongly suggest that universities should update their estimates of dropout risk as new data come in, and not simply assume that the same students who had the highest risk at matriculation still have the highest risk a semester or year later.

Some strong predictors were under policy control. In five of the eight universities, fellowships were associated with significantly reduced odds of dropout, and in four, odds of dropout increased significantly for students who were denied their first-choice major. The results suggest that national policies that reduce costs for students can be beneficial, and that access to high-demand majors should not be restricted without good reason.

The results are equally remarkable for what they do not show. In not one university did the results show that low-income students had increased dropout risk when other variables are controlled. The results also did not show that female students have a consistent advantage. Although female students did have about one-quarter lower dropout risk than the male students, when other variables were controlled, six of the eight universities showed no significant relationship between gender and dropout risk. (At the other two universities, women had a lower net risk of dropout in one and a higher risk in the other.)

Did success programs reach and help high-risk students?

For four of the eight universities, our data included a variable indicating which first- year students were enrolled in student success programs to build academic skills or facilitate adaptation to university life. At these universities, between one- and two-thirds of first-year students participated in some kind of success program. But did those programs reach the students who had the highest dropout risk, according to our models?

At three of the four universities, the answer was no, according to Figure 4. At universities 3, 6, and 8, the distribution of dropout risk, estimated before semester 1, was practically the same for students who participated in student success programs as for students who did not.

University 1, by contrast, targeted its student success programs more effectively. At university 1, participants in success programs had a median estimated dropout risk of 10 percent, nearly twice nonparticipants' median risk of 6 percent. This suggests that University 1 did a plausible job of identifying higher-risk students and routing them into its success programs. Even so, improvement was possible, as there was substantial overlap between the risk distributions of participants and nonparticipants. Many students who were in success programs had lower risk scores than students who are not.

Why don't more high-risk students participate in success programs? And why do many students participate despite having little risk of dropout? One plausible explanation is that universities do not use a formal risk model to recruit students into success programs. Another plausible explanation is that most success programs are voluntary, and highly engaged students may seek them out even if they have little risk of dropout. A cliché about student success programs is that participants are often B students trying to get an A, rather than struggling students trying to avoid dropping out.

A more optimistic interpretation of Figure 4 is that success programs did succeed in reducing risk of dropout. According to this interpretation, program participants would have higher dropout risk if they did not participate, but thanks to success programs, participants have no higher risk of dropout than nonparticipants (except at University 1).

If this interpretation were correct, participation in student success programs would predict dropout risk net of the other variables in our predictive models. And in two of four universities, it did. According to Table 3, participation in success programs predicted a 30 to 40 percent reduction in the odds of dropout at Universities 1 and 3, but did not significantly change dropout risk at universities 6 and 8. These estimates controlled for variables from Table 2 that were available at the start of semester 1. When the model also controlled for first semester GPA, then participation in student success programs was not associated with dropout risk at any of the four universities. That is, success programs predicted persistence through semester 1, but not beyond.

To estimate the effect of success programs with greater rigor, we would need to know more about the programs and how students select into them. But taken at face value, our results suggest that success programs were somewhat effective at universities 1 and 3, but ineffective at universities 6 and 8. Note that universities 6 and 8 had dropout rates of just 6 to 8 percent, and it may have been difficult for a program to reduce dropout further. Note also that university 1, where programs were the most effective, was also the university that does the best job of getting higher-risk students to participate in success programs. Perhaps other universities could get better results if they also focused their student success programs on higher risk students.

Conclusion

Our results suggest that Chilean universities have an opportunity to use data and risk scores to improve student success. Risk scores can be used in two ways. The first is to target interventions. Three out of four institutions in our sample did not target their success programs to the highest-risk students. Despite the fact that approximately half of students participated in success programs, some high-risk students were overlooked, and some resources were wasted on low-risk students who would have persisted even if they did not participate. All success programs at the studied institutions were voluntary. Other institutions often target students using demographic characteristics such as income or gender, but we found demographics were only weakly predictive of dropout. Entrance exam scores and high school grades were more predictive, but not nearly as predictive as the first semester of university grades.

It would be much more effective for Chilean institutions to target interventions using a risk score, especially one that incorporated data from the first semester of university. Students with high risk scores should be targeted for interventions, while students with low risk scores can be left more or less alone. If institutions prioritized students by risk score, they could reach more potential dropouts with less expensive programs that enrol fewer students.

To use risk scores, Chilean university leaders will need analyses like the one in this paper on a regular and timely basis. Once first semester grades are in, institutions will need risk scores quickly, so that they can identify students for interventions well before the second semester gets underway. Universities can calculate risk scores internally, by engaging faculty or staff with the requisite skills. Or they can engage an external vendor to calculate risk scores for them.

With the right technology, it may be possible to intervene earlier than the end of first semester. If instructors use online gradebooks and online learning management systems, within the first two weeks of starting university, signs of low engagement and performance are already evident in online gradebooks, attendance sheets, and students forum activity (Civitas, 2016; Civitas Learning, 2016). If an institution harvests online data for timely prediction, it can intervene before the end of the first semester—even before the midterm.

Of course, better targeting will only improve outcomes if the university's interventions are effective. Evaluating the effectiveness of interventions is another major use for risk scores. There are several ways to use risk scores to evaluate interventions. One approach, illustrated in Table 3, is to use the variables that contribute to the risk score as covariates in a logistic regression that estimates the effect of an intervention. Another approach to match participating and nonparticipating students on the risk score, as well as matching them on a propensity score that predicts whether the student will participate in the intervention (Dong & Bulus, 2014; Leacy & Stuart, 2014). A third approach is to target interventions at students whose risk score exceeds a certain threshold, and then evaluate the intervention using a regression discontinuity analysis that compares intervention students just above the threshold with control students who are just below.

Institutions considering interventions do not need to limit their attention to student success programs. An intervention can be a student success program, but it can also be any institutional policy or behaviour with the potential to affect a student's probability of graduation. For example, many institutions restrict enrolment into certain majors, but our analysis suggests that this practice may increase dropout risk. Net of other predictors, students who are denied their first-choice major have about twice the odds of dropout as other students—and over 30 percent of students at the eight sampled Chilean universities are denied their first-choice major. While some students may be denied a major for good reason, others may be denied simply because a major is understaffed or wishes to seem exclusive. The question arises whether institutions can increase graduation rates simply by allocating more seats, or even more instructors, to the majors in highest demand.

Another major intervention is financial aid. Chile has invested heavily in guaranteed loans, fellowship, and tuition waivers to reduce the cost of higher education. Our results suggest that participation in these programs predicts lower dropout rates in Chile, as it does in the US (Villarreal (2017) for a review). However, Chile is already offering free tuition to students from the bottom 60 percent of the income distribution, and further progress clearly requires something beyond financial aid. We will only know what is possible when institutions start using data and analysis to target and evaluate a variety of interventions.

References

Arocena, R., & Sutz, J. (2001). Changing knowledge production and Latin American universities. *Research Policy*, *30*(8), 1221–1234. https://doi.org/10.1016/S0048-7333(00)00143-8

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. *ArXiv Preprint ArXiv:1606.06364*.

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, *4*(2), 17.

Beerkens, M., Mägi, E., & Lill, L. (2011). University studies as a side job: causes and consequences of massive student employment in Estonia. *Higher Education*, *61*(6), 679–692.

Bernasconi, A., & Rojas, R. C. (2004). *Informe sobre la educación superior en Chile, 1980-2003*. Editorial Universitaria.

Canales, A., & De los Ríos, D. (2007). *Factores explicativos de la deserción universitaria. Informe final proyecto Consejo Superior de Educación*. *Factores explicativos de la deserción universitaria: Informe final proyecto Consejo Superior de Educación*. CICESUniversidad de Santiago de ChileEstudios y documentos del Consejo Superior de Educación. Retrieved from https://www.cned.cl/file/1945/download?token=kq5f7fq_

Civitas. (2016). *Community insights: Emerging benchmarks and student success trends from across the civitas*. Austin, TX: Technical report, December. Retrieved from https://www.civitaslearningspace.com/introducing-community-insights/

Civitas Learning. (2016). *Community insights: Emerging benchmarks and student success trends from across the civitas*. Austin, TX: Technical report Volumen 1, Issue 2. Civitas learning.

Darolia, R. (2014). Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review*, *38*, 38–50.

de Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., … Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175–1188.

Dong, N., & Bulus, M. (2014). Prognostic Propensity Scores: A Method Accounting for the Correlations of the Covariates with Both the Treatment and the Outcome Variables in Matching and Diagnostics. In *Presented at the Society for Research in Educational Effectiveness, Washington, DC.*

Espinoza, Ó. (2015). El sistema de educación superior en Chile visto desde la perspectiva de la equidad: evidencias y recomendaciones. In C. Zuñiga, J. Redondo, M. Lopez, & E. Santa Cruz (Eds.), *Equidad en el sistema de educación superior de Chile: acceso, permanencia, desempeño y resultados* (pp. 73–120).

Ferreyra, M. M., Avitabile, C., Botero, J., Haimovich, F., & Urzúa, S. (2017). *At a Crossroads: Higher Education in Latin America and the Caribbean*. The World Bank.

Hanushek, E. A., & Woessmann, L. (2010). Education and economic growth. *Economics of Education*, 60–67.

Himmel, E. (2002). Modelos de análisis de la deserción estudiantil en la educación superior. *Revista Calidad de La Educación*, *17*, 91–108.

Hofflinger, A., & von Hippel, P. T. (2017). *Rising test scores in chile, 2002-2013: School choice, school resources, or family resources?* Retrieved from https://ssrn.com/abstract=3022262

Hovdhaugen, E. (2015). Working while studying: the impact of term-time employment on dropout rates. *Journal of Education and Work*, *28*(6), 631–651.

Laude, D. A., Kaschner, L. A., Alvarado, C. G., & Connerat, C. K. (2017). Strengthening success for students with multiple risk factors. In *The First Year of College: Research, Theory, and Practice on Improving the Student Experience and*

Increasing Retention. Cambridge University Press.
Leacy, F., & Stuart, E. (2014). On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Statistics in Medicine*, *33*(20), 3488–3508.
Manríquez, P., Mendoza, D., & Ramírez, K. (2015). Relación entre el aporte scal directo , la calidad del cuerpo docente y la producción cientí ca de las instituciones pertenecientes al Consejo de Rectores de las universidades chilenas. *Ran*, *1*(1), 39–52.
Milliron, M., Kil, D., Malcolm, L., & Gee, G. (2017). From Innovation to Impact: How Higher Education Can Evaluate Innovation's Impact and More Precisely Scale Student Support. *Planning for Higher Education*, *45*(4), 125–136.
Mineduc. Establece Normas para el Financiamiento de Estudios de Educación Superior (2005). Santiago, Chile: Ley 20027. Retrieved from https://www.leychile.cl/Navegar?idNorma=239034
Mineduc. (2017). Más de 257 mil jóvenes estudian con Gratuidad en la Educación Superior. Retrieved May 23, 2018, from http://www.gratuidad.cl/2017/06/01/mas-257-mil-jovenes-estudian-gratuidad-la-educacion-superior/
Mineduc. (2018). *Aporte Fiscal Directo: Decreto 7 del 12 de enero de 2018*. Retrieved from http://dfi.mineduc.cl/usuarios/MECESUP/File/2018/instrumentos/AFD/AFD2018DEC7-2018-95pc.pdf
Murtaugh, P., Burns, L., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education*, *40*(3), 355–371.
Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20.
Rojas-Méndez, J., Vasquez-Parraga, A., Kara, A. L. I., & Cerda-Urrutia, A. (2009). Determinants of student loyalty in higher education: A tested relationship approach in Latin America. *Latin American Business Review*, *10*(1), 21–39.
Saldaña, M., & Barriga, O. (2010). Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile. *Revista de Ciencias Sociales*, *16*(4), 616–628.
Santelices, M. V., Catalán, X., Kruger, D., & Horn, C. (2016). Determinants of persistence and the role of financial aid_ lessons from Chile. *Higher Education*, *71*(3), 323–342. https://doi.org/10.1007/s10734-015-9906-6
Schiefelbein, E. (1990). Chile: Economic incentives in higher education. *Higher Education Policy*, *3*(3), 21–26.
Sorensen, L. (2016). "Big Data" in Educational Administration: An Application for Predicting School Dropout Risk. In *Fall research conference, Association of Public Policy and Management*.
Tinto, V. (2012). *Completing college: Rethinking institutional action*. University of Chicago Press.
Torche, F. (2005). Privatization Reform and Inequality of Educational Opportunity : *Sociology of Education*, *78*(Mmi), 316–343.
Villarreal, M. U. (2017). Making ends meet: Student grant aid, cost saving strategies, college completion, and labor supply. In *Conference of the Association for Education Finance and Policy*. New York, NY.
World Bank. (2011). *Programa de crédito con aval del estado (CAE) de Chile: Análisis y evaluación. Sector de Educación de América Latina y el Caribe del Banco Mundial, Washington DC: Banco Mundial*. Retrieved from https://ciperchile.cl/wp-content/uploads/Informe-Programa-de-Crédito-con-Aval-

del-Estado-versión-espanol-2011-05-27.pdf
Yorke, M., & Longden, B. (2004). *Retention and student success in higher education*. McGraw-Hill Education (UK).

Tables

Table 1. Means of outcome and predictors

	University 1	University 2	University 3	University 4	University 5	University 6	University 7	University 8	Overall
Dropout rate	0.10	0.06	0.17	0.12	0.20	0.04	0.23	0.06	0.11
Standardized variables									
GPA, college semester 2	0.09	0.51	0.00	-0.34	0.15	-0.02	-0.35	0.15	0.00
GPA, college semester 1	0.17	0.06	-0.05	-0.30	0.14		-0.23	0.22	0.00
GPA, high school	0.10	-0.66	-0.02	-0.17	-0.32	0.49	-0.79	0.16	0.00
PSU exam	-0.26	-0.56	0.23	-0.06	-0.79	0.49	-0.36	0.09	0.00
Major									
Natural sciences	0.03		0.09	0.06	0.01	0.03		0.04	0.04
Social Sciences	0.13	0.20	0.10	0.15	0.17	0.11	0.16	0.20	0.14
Law or accounting	0.05	0.09	0.07	0.07	0.11	0.13	0.09	0.14	0.10
Engineering	0.56	0.46	0.41	0.39	0.27	0.55	0.75	0.32	0.48
Education	0.16		0.11	0.24	0.12	0.11		0.04	0.10
Health	0.08	0.26	0.22	0.10	0.32	0.07		0.25	0.14
Major was student's									
First choice	0.61	0.68	0.65	0.67	0.70	0.49	0.50	0.56	0.58
Second choice	0.19	0.14	0.18	0.18	0.13	0.28	0.22	0.24	0.21
Third choice	0.20	0.18	0.17	0.16	0.18	0.23	0.28	0.20	0.21
Student is female	0.45	0.57	0.47	0.53	0.55	0.47	0.39	0.58	0.49
Family income (million pesos/month)	0.45	0.60	0.55	0.57	0.56	0.63	0.51	0.60	0.57
Free tuition from national government	0.67	0.59	0.62	0.62	0.62	0.59	0.67	0.56	0.61
Other fellowship	0.17	0.19	0.19	0.19	0.15	0.20	0.19	0.18	0.18
Years since high school graduation	1.99	2.07	1.98	2.02	2.03	2.08	2.24	1.90	2.03
Vocational HS	0.28	0.38	0.10	0.15	0.28	0.18	0.24	0.13	0.20
Student in success program	0.67		0.64			0.40		0.61	
Students in both cohorts	4,040	1,423	3,227	2,761	1,697	6,831	3,236	5,219	28,434

Note. University 6 did not provide GPA for semester 1. Majors that are not offered at a university are blank. Student success program participation is limited to the 2017 cohort because one university did not report participation in 2016, and four universities did not report student success participation at all. All other statistics are pooled across the 2016 and 2017 cohorts.

Table 2a. Predictive modelling of dropout: Logistic regression, odds ratios, 2016 cohort, Universities 1-4

Predictor	University 1			University 2			University 3			University 4		
	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 1	After semester 2
GPA, college semester 2			0.17***			0.22***			0.25***			0.25***
			(0.03)			(0.08)			(0.03)			(0.04)
GPA, college semester 1		0.32***	0.61+		0.96	5.48***		0.27***	0.75+		0.34***	1.01
		(0.04)	(0.15)		(0.13)	(2.38)		(0.02)	(0.11)		(0.03)	(0.17)
GPA, high school	0.83+	1.10	1.36	0.74*	0.74*	0.77+	0.80**	1.01	1.02	0.88	0.99	1.05
	(0.08)	(0.15)	(0.30)	(0.10)	(0.10)	(0.11)	(0.06)	(0.09)	(0.13)	(0.08)	(0.10)	(0.16)
PSU exam	0.67***	0.95	1.35	0.91	0.91	0.96	0.80**	1.25*	1.17	0.90	1.12	1.13
	(0.07)	(0.13)	(0.30)	(0.18)	(0.18)	(0.19)	(0.06)	(0.12)	(0.16)	(0.10)	(0.13)	(0.20)
Major: natural sciences	3.20**	2.67+	3.24				1.35	0.81	0.61	1.69	0.89	0.87
(ref. social sciences)	(1.31)	(1.41)	(2.53)				(0.40)	(0.28)	(0.30)	(0.60)	(0.37)	(0.51)
Law or accounting	0.39	0.33	0.78	7.02*	7.12*	4.45+	0.92	0.59	0.51	0.74	0.93	0.42
	(0.25)	(0.28)	(0.90)	(6.07)	(6.18)	(3.99)	(0.28)	(0.21)	(0.27)	(0.27)	(0.38)	(0.24)
Engineering	1.93*	1.04	0.76	9.91**	9.96**	7.95**	1.95**	0.67	0.74	0.86	0.63	0.32**
	(0.52)	(0.40)	(0.46)	(7.36)	(7.40)	(5.94)	(0.44)	(0.19)	(0.30)	(0.21)	(0.18)	(0.14)
Education	1.03	1.01	1.09				1.22	1.27	1.07	0.69	1.49	0.46
	(0.34)	(0.46)	(0.82)				(0.34)	(0.41)	(0.48)	(0.19)	(0.48)	(0.23)
Health	0.88	0.75	0.44	2.02	2.04	1.88	1.04	0.62	0.67	0.40*	0.58	0.23+
	(0.38)	(0.44)	(0.45)	(1.73)	(1.75)	(1.61)	(0.28)	(0.20)	(0.32)	(0.17)	(0.27)	(0.19)
Major was student's second choice	1.16	1.02	0.94	1.84	1.81	1.66	1.48*	1.25	1.67+	1.71*	1.74*	1.96*
(ref. top choice)	(0.23)	(0.27)	(0.41)	(0.74)	(0.74)	(0.72)	(0.24)	(0.24)	(0.46)	(0.37)	(0.41)	(0.67)
	1.28	1.20	1.38	1.13	1.12	1.24	1.74***	1.79**	1.90*	1.30	1.33	1.42
Major was student's third choice or lower	(0.24)	(0.30)	(0.57)	(0.42)	(0.41)	(0.47)	(0.28)	(0.34)	(0.52)	(0.30)	(0.34)	(0.52)
Female (ref. male)	0.93	1.23	2.39*	0.52*	0.51*	0.49*	0.87	1.09	1.28	1.00	1.21	1.76+
	(0.16)	(0.28)	(0.89)	(0.16)	(0.16)	(0.16)	(0.12)	(0.17)	(0.29)	(0.18)	(0.25)	(0.51)
Family income (million pesos/month)	0.89	0.86	0.87	0.86	0.86	0.93	0.83	0.91	0.68	1.18	1.23	2.13**
	(0.17)	(0.22)	(0.36)	(0.30)	(0.30)	(0.33)	(0.14)	(0.18)	(0.21)	(0.21)	(0.24)	(0.58)
Free tuition from national government	0.62*	0.90	0.63	0.74	0.74	0.72	0.64*	0.61+	0.55	0.75	0.57*	0.74
(*la gratuidad*)	(0.14)	(0.28)	(0.32)	(0.30)	(0.30)	(0.30)	(0.14)	(0.16)	(0.21)	(0.20)	(0.16)	(0.33)
Other fellowship	0.72	1.06	1.18	0.71	0.71	0.78	0.66+	0.68	0.76	1.08	1.00	1.53
	(0.18)	(0.36)	(0.62)	(0.31)	(0.32)	(0.36)	(0.15)	(0.18)	(0.29)	(0.30)	(0.30)	(0.69)
Years since high school graduation	0.90*	0.82**	0.90	1.05	1.05	1.03	0.94+	0.85***	0.84**	1.03	0.96	0.95
	(0.05)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.07)
Vocational high school (ref. academic)	1.03	0.94	1.24	1.06	1.06	1.03	1.35	1.40	1.30	1.46+	1.61+	2.55*
	(0.18)	(0.22)	(0.48)	(0.31)	(0.31)	(0.32)	(0.26)	(0.32)	(0.43)	(0.33)	(0.41)	(0.96)
Area under lift curve (AUC)	64%	78%	89%	61%	61%	59%	60%	74%	78%	59%	76%	82%
Students in 2016 cohort	1,963	1,847	1,708	681	681	681	1,870	1,815	1,632	1,338	1,307	1,173

Note. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Standard errors in parentheses. Blanks mean that a major was not available at a university. At university 2, we merged education majors into health to avoid complete separation.

Table 2b. Predictive modelling of dropout: Logistic regression, odds ratios, 2016 cohort, Universities 5-8

	University 5			University 6		University 7			University 8		
Predictor	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 2	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 1	After semester 2
GPA, college semester 2			0.14***		0.83*			0.14***			0.83*
			(0.04)		(0.07)			(0.02)			(0.07)
GPA, college semester 1		0.16***	1.62				0.21***	0.93		0.30***	
		(0.03)	(0.50)				(0.02)	(0.17)		(0.03)	
GPA, high school	0.58***	0.91	0.94	1.09	1.09	0.82**	1.02	1.32*	0.94	1.18	1.09
	(0.06)	(0.14)	(0.21)	(0.14)	(0.15)	(0.06)	(0.09)	(0.18)	(0.09)	(0.18)	(0.15)
PSU exam	0.87	1.50*	0.98	1.01	1.07	0.71***	0.94	0.98	0.85+	1.10	1.07
	(0.11)	(0.27)	(0.25)	(0.11)	(0.12)	(0.06)	(0.10)	(0.17)	(0.08)	(0.15)	(0.12)
Major: social sciences	2.66	1.36	1.27	3.12*	3.08*				1.41	1.01	3.08*
(ref. natural sciences)	(2.42)	(1.44)	(1.94)	(1.40)	(1.38)				(0.43)	(0.45)	(1.38)
Law or accounting	1.17	2.39+	1.49	1.70	1.64	0.89	0.90	1.66	0.13***	0.09**	1.64
	(0.44)	(1.25)	(1.27)	(0.60)	(0.57)	(0.24)	(0.33)	(1.12)	(0.06)	(0.07)	(0.57)
Engineering	1.92*	0.79	1.34	0.97	0.85	1.79**	1.03	3.28*	0.54**	0.40***	0.85
	(0.57)	(0.35)	(0.95)	(0.29)	(0.26)	(0.32)	(0.25)	(1.66)	(0.10)	(0.11)	(0.26)
Education	2.71**	3.57*	5.14*	0.82	0.78				0.49+	0.56	0.78
	(0.91)	(1.82)	(4.13)	(0.33)	(0.31)				(0.21)	(0.34)	(0.31)
Health	1.33	0.67	0.62	1.34	1.34				0.58*	0.52+	1.34
	(0.39)	(0.29)	(0.45)	(0.52)	(0.53)				(0.15)	(0.20)	(0.53)
Major was student's second choice	0.88	1.16	1.19	1.07	1.07	1.44*	1.59*	1.55	1.16	0.91	1.07
(ref. top choice)	(0.24)	(0.47)	(0.67)	(0.23)	(0.23)	(0.21)	(0.31)	(0.45)	(0.21)	(0.25)	(0.23)
	1.74*	2.74***	2.64*	1.38	1.36	1.50**	1.84***	1.17	1.22	1.01	1.36
Major was student's third choice or lower	(0.37)	(0.84)	(1.18)	(0.30)	(0.29)	(0.20)	(0.33)	(0.33)	(0.24)	(0.28)	(0.29)
Female (ref. male)	0.76	0.98	1.40	1.18	1.18	0.94	1.30	1.97**	0.84	0.87	1.18
	(0.15)	(0.27)	(0.55)	(0.22)	(0.22)	(0.12)	(0.22)	(0.50)	(0.13)	(0.20)	(0.22)
Family income (million pesos/month)	0.99	0.82	0.75	0.91	0.94	0.90	1.38	1.17	0.98	1.34	0.94
	(0.25)	(0.29)	(0.38)	(0.20)	(0.21)	(0.16)	(0.32)	(0.47)	(0.15)	(0.28)	(0.21)
Free tuition from national government	0.63+	0.46*	0.61	2.12*	2.22*	0.43***	0.75	0.64	0.69+	1.00	2.22*
(*la gratuidad*)	(0.16)	(0.17)	(0.35)	(0.76)	(0.79)	(0.08)	(0.21)	(0.29)	(0.13)	(0.28)	(0.79)
Other fellowship	0.40**	0.33*	0.48	2.18*	2.23*	0.55**	0.82	0.52	0.92	0.81	2.23*
	(0.13)	(0.15)	(0.32)	(0.78)	(0.80)	(0.11)	(0.24)	(0.25)	(0.20)	(0.28)	(0.80)
Years since high school graduation	0.91+	0.91	0.98	0.84*	0.83*	1.00	0.93+	0.97	1.00	0.63***	0.83*
	(0.05)	(0.06)	(0.09)	(0.06)	(0.06)	(0.02)	(0.03)	(0.05)	(0.04)	(0.08)	(0.06)
Vocational high school (ref. academic)	1.22	1.67+	1.43	0.65	0.66	1.34*	1.31	1.83*	0.99	1.02	0.66
	(0.25)	(0.48)	(0.58)	(0.18)	(0.18)	(0.18)	(0.23)	(0.50)	(0.24)	(0.37)	(0.18)
Area under lift curve (AUC)	60%	74%	81%	55%	58%	59%	74%	80%	61%	82%	89%
Students in 2016 cohort	842	789	702	3,364	3,351	1,675	1,578	1,371	2,600	2,451	2,189

Note. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Standard errors in parentheses. University 6 did not provide GPAs for semester 1.

Table 3. Estimated effect of student success programs: Logistic regression odds ratios, Universities 1, 3, 6, and 8

	University 1			University 3			University 6			University 8		
Predictor	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 1	After semester 2	Before semester 1	After semester 1	After semester 2
Participation in student success programs	0.60*	1.03	0.85	0.69*	0.89	0.86	1.01	1.06	1.06	1.13	0.92	1.32
	(0.15)	(0.38)	(0.51)	(0.12)	(0.17)	(0.23)	(0.18)	(0.19)	(0.19)	(0.20)	(0.23)	(0.71)

Note. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Standard errors in parentheses.. All models control for the variables in Table 2. Data are from the 2016 cohort except in university 3 where we used the 2017 cohort because student success data for 2016 were unavailable.

Figures

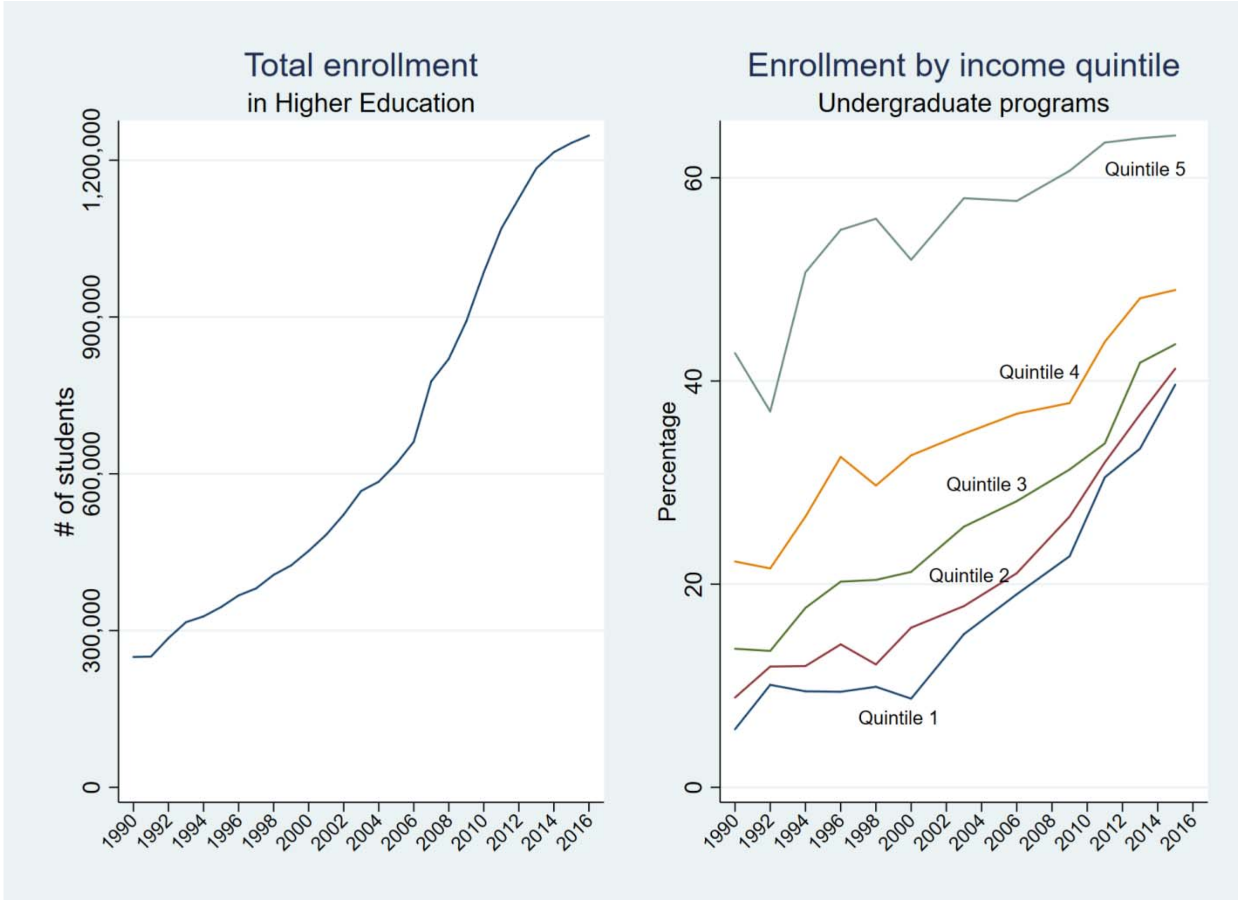

Figure 1. Total higher education enrolment and enrolment rates by income quintile, 1990-2016. Source: Authors' calculation from Chile's National Socioeconomic Characterization Survey (CASEN).

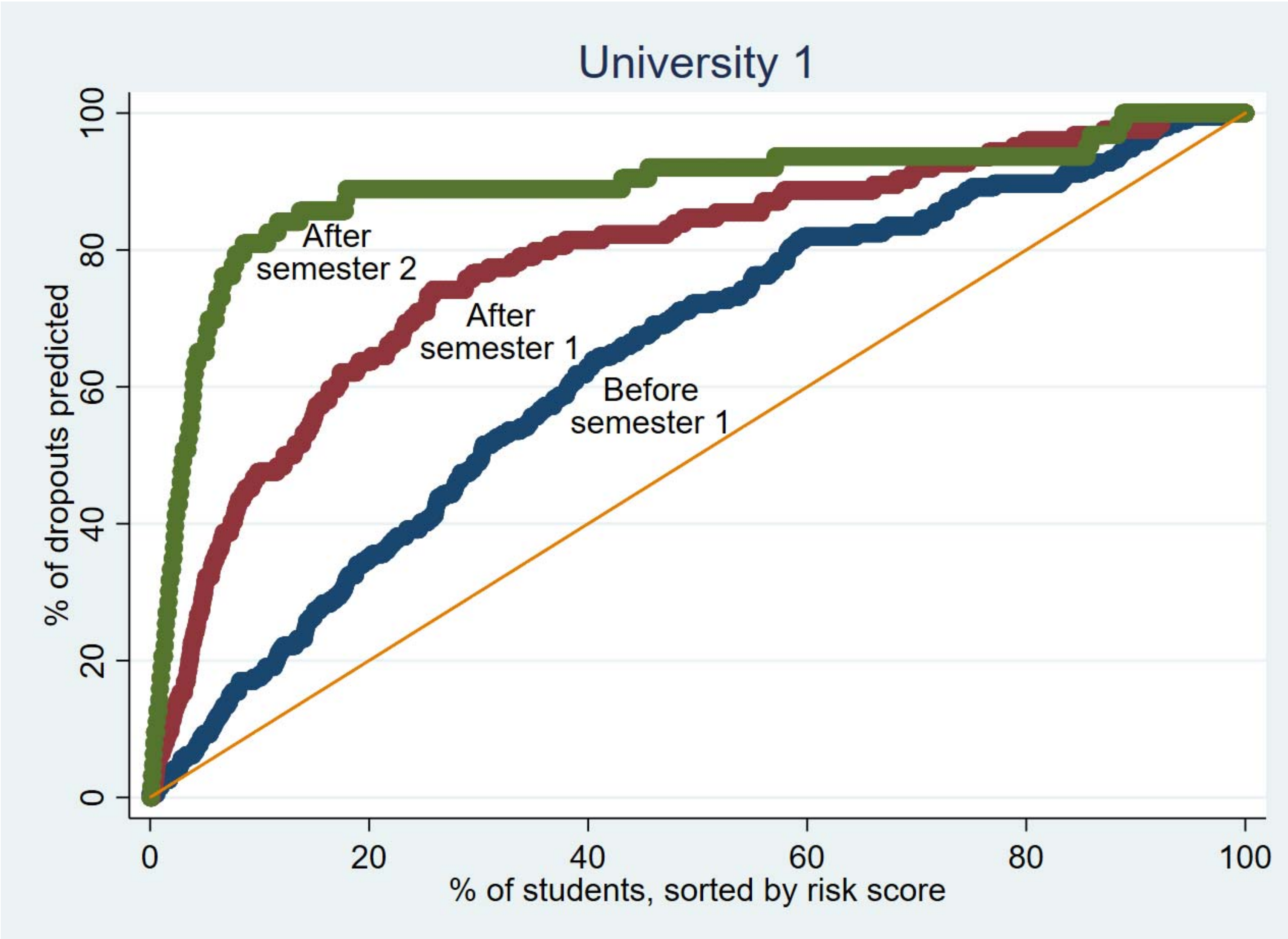

Figure 2. Lift curves for 2017 cohort in university 1. The diagonal reference line represents random predictions,

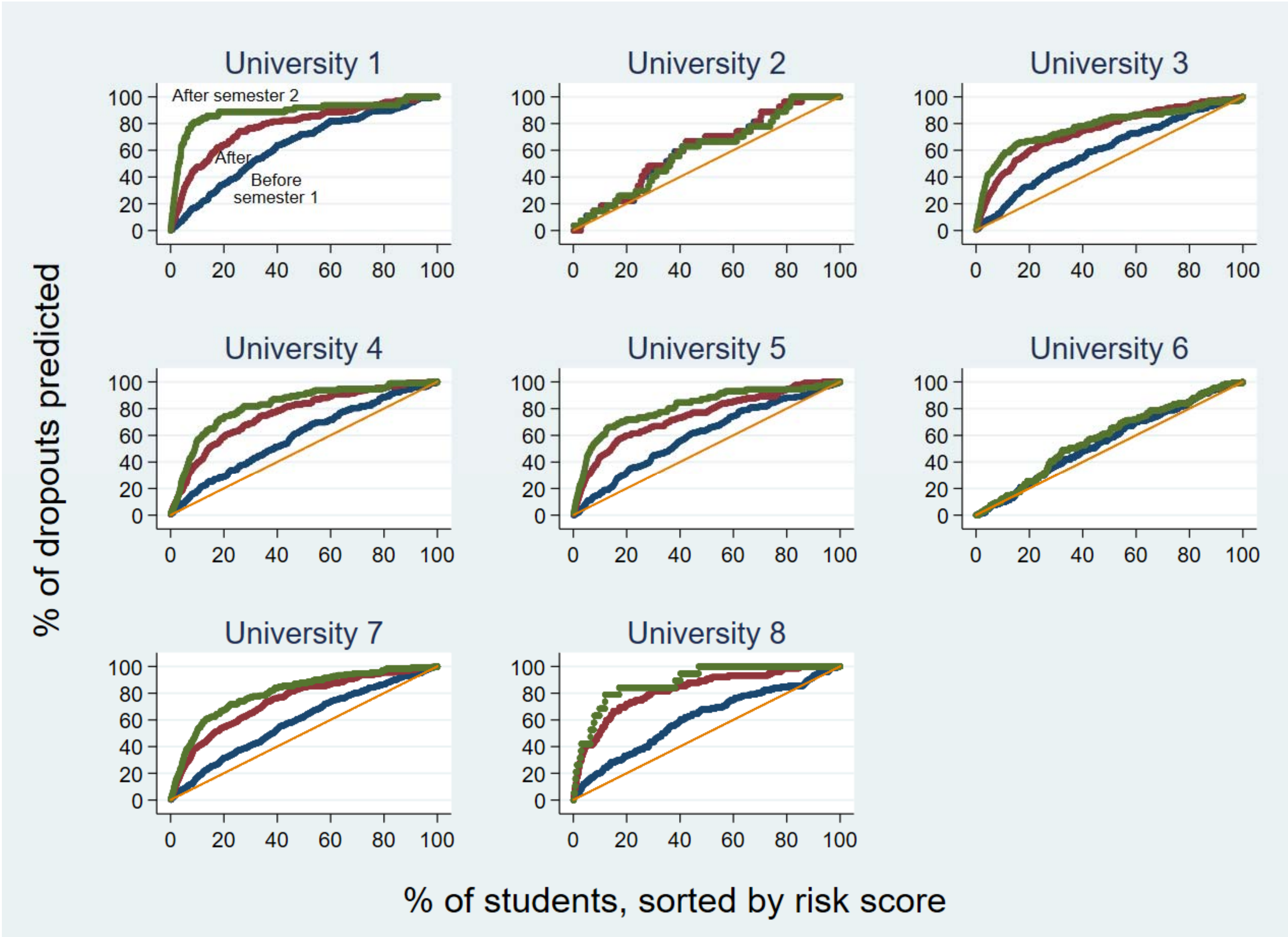

Figure 3. Lift curves for 2017 cohorts at all eight universities

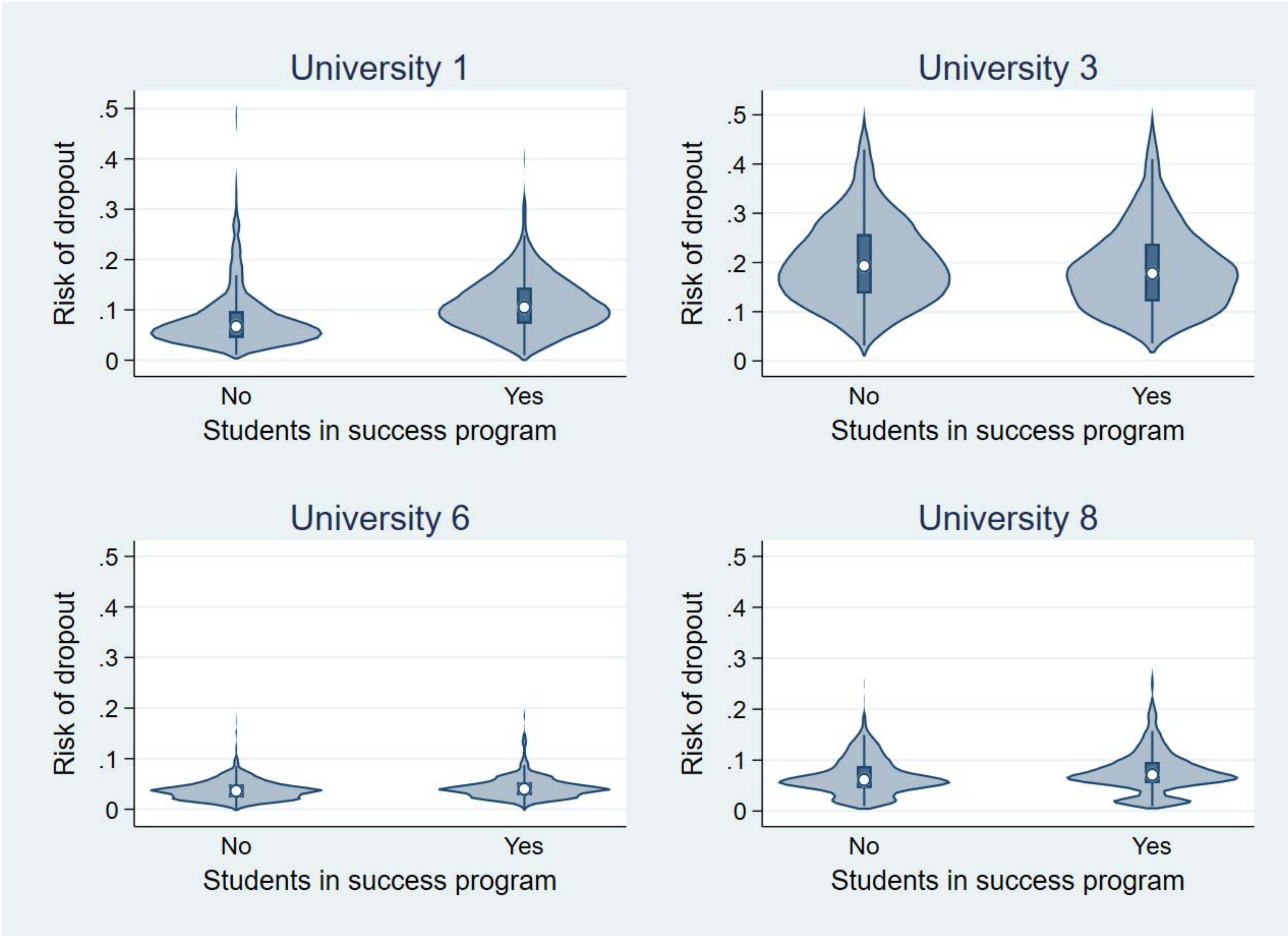

Figure 4. Distribution of risk scores, estimated before semester 1, for students in 2017 cohort who are in vs. out of student success programs. The figures are violin plots, which estimate the density of risk scores. Inside each violin plot is a boxplot; the box shows the interquartile range, and the white dot shows the median.