ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 21-438

Bringing Transparency to Predictive Analytics: A Systematic Comparison of Predictive Modeling Methods in Higher Education

Kelli A. Bird
University of Virginia

Benjamin L. Castleman
University of Virginia

Zachary Mabel
The College Board

Yifeng Song
University of Virginia

Colleges have increasingly turned to predictive analytics to target at-risk students for additional support. Most of the predictive analytic applications in higher education are proprietary, with private companies offering little transparency about their underlying models. We address this lack of transparency by systematically comparing two important dimensions: (1) different approaches to sample and variable construction and how these affect model accuracy; and (2) how the selection of predictive modeling approaches, ranging from methods many institutional researchers would be familiar with to more complex machine learning methods, impacts model performance and the stability of predicted scores. The relative ranking of students' predicted probability of completing college varies substantially across modeling approaches. While we observe substantial gains in performance from models trained on a sample structured to represent the typical enrollment spells of students and with a robust set of predictors, we observe similar performance between the simplest and most complex models.

VERSION: July 2021

Suggested citation: Bird, Kelli A., Benjamin L. Castleman, Zachary Mabel, and Yifeng Song. (2021). Bringing Transparency to Predictive Analytics: A Systematic Comparison of Predictive Modeling Methods in Higher Education. (EdWorkingPaper: 21-438). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/hd2e-7e02

Bringing Transparency to Predictive Analytics:

A Systematic Comparison of Predictive Modeling Methods in Higher Education

Kelli A. Bird[1]

Benjamin L. Castleman[1]*

Zachary Mabel[2]

Yifeng Song[1]

Abstract

Colleges have increasingly turned to predictive analytics to target at-risk students for additional support. Most of the predictive analytic applications in higher education are proprietary, with private companies offering little transparency about their underlying models. We address this lack of transparency by systematically comparing two important dimensions: (1) different approaches to sample and variable construction and how these affect model accuracy; and (2) how the selection of predictive modeling approaches, ranging from methods many institutional researchers would be familiar with to more complex machine learning methods, impacts model performance and the stability of predicted scores. The relative ranking of students' predicted probability of completing college varies substantially across modeling approaches. While we observe substantial gains in performance from models trained on a sample structured to represent the typical enrollment spells of students and with a robust set of predictors, we observe similar performance between the simplest and most complex models.

Acknowledgements

We are grateful for our partnership with the Virginia Community College System and in particular Dr. Catherine Finnegan. We are grateful to financial support from the Lumina, Overdeck, and Heckscher Family Foundations. Any errors are our own.

1 University of Virginia
2 The College Board
* Corresponding Author

I. Introduction

Predictive analytics have become increasingly common in the education sector. Colleges and universities use predictive analytics for various purposes, ranging from identifying students who might default on their loans to targeting alumni who are likely to give generously to the institution (Ekowo & Palmer, 2016). The most common use of predictive analytics, however, is to identify students at risk of failing courses or dropping out (Alamuddin, Rossman, & Kurzweil, 2019; Milliron, Malcolm, & Kil, 2014; Plak et al, 2019), and to direct various student success strategies (e.g., intrusive advising, additional financial aid) to these students. Numerous contextual factors have motivated institutions to turn towards predictive analytics. While enrollment rates have increased steadily over the last decade and socioeconomic inequalities in college participation have narrowed (US Department of Education, 2019), completion rates remain relatively stagnant and socioeconomic disparities persist and have widened over time (Bailey & Dynarski, 2011; Chetty et al., forthcoming). Students are borrowing a record amount of money to fund their postsecondary education -- total student debt now exceeds $1 trillion -- with default rates highest among students who drop out before finishing their degree (Bastrikin, 2020; Looney & Yannelis, 2015). In light of these trends, state and federal policy makers have put increasing pressure on institutions to increase completion rates.

Despite this increased pressure, at broad-access institutions attended by most undergraduates, the level of resources available to invest in completion strategies has declined considerably over time as states have reduced their appropriations to public higher education (Deming & Walters, 2017; Ma et al, 2017). The use of predictive analytics in higher education has the potential to increase efficiency in how scarce resources are allocated by targeting students who may benefit most from additional intervention. Adoption of predictive analytics strategies has been broad and rapid; a third of all institutions have invested in predictive analytics and collectively spend hundreds of millions of dollars on technology that utilizes predictive analytics (Barshay & Aslanian, 2019).

For efficiency gains to be realized from predictive analytics, though, predictions from underlying models must be accurate, stable, and fair. However, in most cases researchers and college administrators have little to no ability to evaluate predictive analytics software on these dimensions, as most predictive analytics products used in higher education are proprietary and operated by private. This lack of transparency creates multiple risks for institutions and students. Models may vary substantially in the accuracy with which they identify at-risk students, which can lead to inefficient and ineffective investment of institutional resources. Furthermore, biased models can lead institutions to intervene disproportionately with students from underrepresented backgrounds and may reinforce existing psychological barriers that students encounter, including feelings of social isolation and anxiety (Walton & Cohen, 2011).

In this paper, we address the lack of transparency in predictive analytics in higher education by systematically comparing two important dimensions of predictive modeling. First, we compare different approaches to sample and variable construction and how these affect model accuracy. We focus in particular on how two analytic decisions affect model performance: (1) random truncation of a current cohort sample to align to the enrollment length distribution of historic cohorts and (2) the inclusion of term-specific and more complexly-specified variables (e.g., a variable measuring the trend in students' GPA over time). Second, we investigate how the choice of modeling approach, ranging from methods many institutional researchers would be familiar with, such as OLS regression and survival analysis, to more complex approaches like tree-based classification algorithms and neural networks, impacts model performance and the stability of student predicted scores (i.e., “risk rankings").

We examine these features of predictive modeling in the context of the Virginia Community College System (VCCS), which consists of 23 community colleges in the Commonwealth of Virginia. We have access to detailed student records for all students who attended a VCCS college from 2000 to the present. Community colleges serve numerous functions, including targeted skill development, broader workforce readiness, terminal degree production, and preparation to transfer to a four-year institution. Each of these functions have different potential measures of success. In this paper, we focus in particular on the outcome of whether students graduate with a college-level credential within six years of initial entry to systematically compare predictive modeling strategies.

Our analysis yields several primary conclusions. First, while models are very consistent in whether they predict whether a given student graduates, they vary in how they rank a given student’s predicted probability of graduating. For instance, among students that the OLS model rank in the bottom decile of the probability of completing college, only 60 percent also ranked in the bottom decile according to the XGBoost approach. This lack of consistency in student ranking holds across the distribution of risk. This result suggests that the notion of relative “risk” is not stable and can be quite sensitive to the modeling strategy used. For institutions that use predictive modeling to intervene with a targeted subset of students, such as students at greatest risk of dropout, different models are likely to identify different students for intervention.

Second, predictive models that leverage randomly truncated samples, term-specific predictors, and more complexly-specified variables have higher performance than models trained on samples without truncation or with basic variables (e.g., cumulative credits completed) that may be more readily available to higher education administrators and institutional researchers. This suggests there are gains to complexity in sample and variable construction, whether institutions pursue that work internally or through an external vendor. Finally, in terms of modeling approach, we do not observe substantial increases in accuracy from more complex models. All models we compare have high levels of accuracy in predicting whether a student will graduate or not.

We contribute to the evidence base on the efficacy of predictive analytics in higher education in this paper in several ways. Ours is the first paper of which we are aware that systematically evaluates and compares the performance of different sample and variable construction approaches and modeling strategies in an applied setting. In doing so, we bring transparency to a practice that is increasingly common but frequently opaque in higher education. Our findings also elucidate the tradeoffs to common modeling decisions and the contexts in which the expected returns to sophisticated machine learning methods (over and above conventional regression-based models) are largest. Finally, we discuss important questions around the ethics, cost, and efficacy of using predictive analytics that higher education administrators and researchers may want to consider in determining their approach to predicting student success.

II. Conceptual Model

To develop a conceptual model of how administrators at broad-access institutions use predictive analytics, we draw on several reports that collectively provide case studies of how dozens of institutions have incorporated predictive analytics into their practice (APLU, 2016; Burke et al, 2017; Ekowo & Palmer, 2016; Kemplin, Grand, & Ramos, 2018; Paterson, 2019; Stark, 2015; Treaster, 2017). In this review, we observe two broad commonalties in predictive analytics usage. First, nearly all institutions' use of predictive analytics is in response to two interwoven contextual factors: (1) Increasing pressure on institutions to increase success rates, including shifts in state public financing for higher education towards outcomes-based funding allocations; and (2) Declining overall state appropriations towards public higher education, which result in institutions have fewer resources to allocate towards college success strategies and interventions. These combined factors increase pressure on institutions to target scarce resources as efficiently as possible to achieve meaningful improvements in success outcomes. Second, while most institutions use data to inform broad institutional practice, the predictive analytics applications are primarily geared towards targeting individual student outreach, primarily through faculty or advisor intervention.

Institutions apply predictive analytics across the life cycle of students' engagement with the institution. For instance, predictive analytics have become increasingly commonplace in enrollment management and financial aid packaging as broad-access institutions have become increasingly reliant over time on tuition as a primary source of revenue. Institutions like Wichita State University use models to target recruitment and marketing investments to students most likely to apply and matriculate, enroll, and succeed at the institution (Ekowo & Palmer, 2016). Institutions like Jacksonville State University and University of Texas - Austin use models to inform aid allocations, respectively directing scholarships to students who are predicted to stay enrolled at the institution (rather than transfer elsewhere) or to students who are predicted to drop out absent additional financial support (Ekowo & Palmer, 2016; Paterson, 2019).

Institutions also use predictive analytics to identify courses in which academic performance is predictive of later success at the institution, and to target interventions to students who are predicted to struggle in those courses. For instance, the University of Arizona learned from a predictive model that students who earn a C in introduction English composition have a lower probability of graduating, and allocated additional academic supports to such students (Treaster, 2017).

By far the most common use of predicted analytics reported in these case studies is to identify students at risk of dropping out before completing their degree. Georgia State University has received substantial attention for its use of predictive analytics to identify students who were struggling academically and to provide them with additional support. Like many other institutions that use predictive analytics to identify students at risk of withdrawal prior to completion, Georgia State partnered with a private company (EAB) to develop an algorithm using student-level administrative data from numerous historic cohorts. Other institutions like Temple University developed their own predictive analytics models and "early alert" systems to identify at-risk students. Across the institutions featured in the case studies we reviewed, most leveraged the "early alerts" generated by predictive models to either trigger proactive outreach from academic advisors to students, or to encourage faculty to reach out to students in their classes who were struggling to succeed (Ekowo & Palmer, 2016). At some institutions, like the University of North Carolina-Greensboro, administrators group students into deciles of predicted risk of withdrawal and target more intensive interventions to students with the highest risk ratings (Klempin, Grant, & Ramos, 2018).

These common uses of predictive analytics by administrators at broad access institutions rest on the assumption that the underlying prediction models--whether for enrollment management or to target student success interventions--are producing student-level risk predictions that are accurate, stable, and fair. In the remainder of the paper, we investigate the extent to which these assumptions hold across predictive modeling approaches.

III. Empirical Strategy

A. Data

The data for this study come from VCCS system-wide administrative records over the summer 2007 through spring 2019 academic terms. These records include detailed information about each term in which a student enrolled, including their program of study, courses taken, grades earned, credits accumulated, financial aid received, and degrees earned. The records also include basic demographic information, including gender, race, and parental education. Finally, we observe all credentials awarded by VCCS colleges beginning in 2007. In addition to VCCS administrative records, we also have access to National Student Clearinghouse (NSC) graduation and enrollment records. NSC data allows us to observe all enrollment periods and postsecondary credentials earned at non-VCCS institutions from 2004 onward.

B. Outcome Variable Definition

We focus on predicting the probability a student completes any college-level credential within six years. For simplicity, we refer to our outcome as "graduation" throughout the paper. Based on this outcome definition, 34.1 percent of students in our sample graduated. We choose to focus on the outcome of graduation rather than dropout because dropout is more ambiguous and difficult to define, particularly in the community college context. For instance, if a student leaves for a few semesters, it is unclear whether they "stop out" and plan to return to college at a later date or have dropped out with no plans to return. Within our sample, 37.7 percent of students leave VCCS for at least one non-summer term and later return to higher education (either to VCCS or a non-VCCS institution); 23.3 percent of students leave for at least one full year and later return to higher education.

While all VCCS credentials are designed to be completed in two years or less if the student is enrolled full-time, prior research has shown that only 16 percent of certificate earners and only five percent of Associate degree earners graduate within two years (Complete College America, 2014). We focus on graduation within six years because we consider credential completion from both VCCS and non-VCCS institutions, some of which are four-year institutions students transfer to after their enrollment at VCCS.

Sample Construction

Our sample consists of students who enrolled at a VCCS college as a degree-seeking, non-dual enrollment student for at least one term, with an initial enrollment term between summer 2007 and summer 2012 (the last cohort for whom we can observe six years of graduation outcomes). We provide additional details on our sample definitions in Appendix 1.

For each student in our sample, we observe their information for the entire six-year window after their initial enrollment term. While in all of our models we use the full six years of data to construct the outcome measure, we test two different approaches to constructing model predictors. First, using data from students initially enrolled between summer 2007 and summer 2012, we construct a sample using all information from initial enrollment through the term when the student earned his/her first college-level credential, or the end of the six-year window, whichever comes first. The primary concern with this approach to predictor construction is that models fitted using all available data for historical cohorts of students may not be generalizable to currently enrolled students, whose enrollment spells do not extend to the full six years or to credential attainment. Therefore, in our second approach, also using data from students initially enrolled between summer 2007 and summer 2012, we construct a historical sample of students using a random truncation procedure that resembles the distribution of enrollment lengths for currently enrolled students.

The first two columns of Table 1 show the distribution of the number of terms since initial VCCS enrollment for students enrolled in fall 2012 (the most recent fall term in our sample). In the first row, we see that 33 percent of students enrolled at a VCCS institution in fall 2012 first enrolled in that term, and their enrollment length is therefore equal to one term. In our second approach to predictor construction, we randomly truncate the data in the full sample to resemble the distribution of enrollment lengths among fall 2012 enrollees. For example, we randomly assign 33 percent of students from the training and validation samples to have an enrollment length of one -- in other words, for those students, we only use their first term of data to construct their model predictors, regardless of how long they were actually enrolled at VCCS. Columns 3 and 4 of Table 1 show the full distribution of enrollment lengths in the truncated training and validation samples, which are described below. The modal length of enrollment is one term, but there is substantial variation across students. For example, 17 percent of students have an enrollment length of four terms. We discuss in more detail the motivation and steps for our approach to sample truncation in Appendix 1.

Our resulting analytic sample consists of n = 331,254 students, which we randomly divide into training (90 percent) and validation sets (10 percent).[2] The training set is used to construct and fine-tune predictive models, while the validation set is held out throughout the model construction and tuning process, and is used to evaluate the performance of the final prediction model. This division is standard practice in predictive modeling work to ensure that the model is evaluated on "unseen" data and therefore free of bias due to model overfitting.[3] We further discuss the summary statistics of the full analytic sample in Appendix 1.

C. Predictor Construction

In addition to exploring how different sample constructions affect model performance, we investigate how the incorporation of predictors with differing degrees of complexity affects model performance. We first test models that use simple, non-term-specific predictors most readily available to higher education administrators and researchers. These predictors include demographic information (e.g., race/ethnicity, parental education) along with a set of cumulative measures up to a student's last observed term (overall or within the randomly truncated observation window), such as cumulative GPA and the share of all attempted courses the student completed. Second, we examine how model performance changes with the inclusion of additional non-term-specific predictors that are more complex to construct, such as the number of terms and quality of non-VCCS institutions a student attended before VCCS, and the standard deviation of a students' term GPA in all previous enrolled terms. Third, we investigate how model performance is affected by the further inclusion of simple term-specific predictors, such as term-specific GPA, credits attempted, and the share of attempted credits earned. Finally, we include more complexly specified term-specific predictors, including academic and financial aid information such as term-specific credits withdrawn, 200-level credits attempted, the amount of financial aid received, and enrollment intensity at non-VCCS institutions. Appendix 2 provides a full list of the predictors we test, organized by the sequence in which we test their inclusion in the prediction models.

D. Predictive Models

We use six different but commonly used estimation strategies in the social and computational sciences to predict the probability of credential attainment within six years (Attewell & Monaghan, 2015; Hand, Mannila & Smyth, 2001; Herzog, 2006): Ordinary Least Squares

(OLS), Logistic Regression, Cox Proportional Hazard (CPH) Survival Analysis, Random Forest, Gradient boosted machines (XGBoost) and Recurrent Neural Networks (RNN).

OLS, Logistic regression, and CPH are models commonly used by researchers in all areas to perform predictive modeling tasks, due to their ease of implementation and interpretation. We include OLS and Logistic due to user familiarity, fast run times, and high degree of interpretability of output.[4]

CPH is one the most commonly used methods of survival analysis in the social sciences when the goal is to predict not only whether, but when the likelihood of an event will occur.[5] Although our goal in this paper is to predict whether students will complete college at any point within a six year window and not the timing of completion within that window, we include CPH among the estimation strategies we compare because survival analysis methods may be familiar to institutional researchers who are considering using predictive analytics in higher education. As we discuss in further detail in Appendix 3, two limitations should be considered when comparing the performance of CPH models to the performance of the other estimation strategies we employ. First, we exclude time-varying predictors from CPH models because the assumptions required for their inclusion (i.e., for each currently enrolled student, we must impute the values of all time-varying predictors in all future, unobserved terms over the six-year window), are extremely strong. Nevertheless, it remains possible that model performance would improve with the inclusion of time-varying predictors. Second, although survival analysis models can address complications associated with time-censored data using alternative approaches to random sample truncation (e.g., inclusion of model parameters to account for unobserved heterogeneity), we only estimate the CPH model on the randomly truncated sample. We do so because, as the results in Table 1 show, applying a model trained on a non-truncated sample of previously enrolled students to generate out-of-sample predictions for currently enrolled students raises questions of model generalizability that alternative approaches may not address. In addition, using the randomly truncated training sample allows for more interpretable inter-model comparisons since our primary predictions from all other models are derived using the randomly truncated sample. Compared to OLS, Logistic regression, and survival analysis, tree-based methods (Random Forest and XGBoost) and neural network models (RNN) are less commonly used in the field of education, in part because they are more complicated to implement.[6] However, they generally exhibit superior predictive performance because they more easily allow for capturing nonlinear and interactive relationships between the outcome and predictors. The basic building blocks of tree-based methods are decision trees, which flexibly identify patterns (sometimes quite complex) between the outcome of interest and the predictors (Breiman et al, 1984; James et al, 2013). However, because decision trees are highly sensitive to the sample and set of predictors included in building the tree, individual decision trees typically are not generalizable (i.e., they do not perform well on unseen data). We address this limitation through the use of two tree-based ensemble models, Random Forest and XGBoost. We describe additional detail and considerations for the implementation of tree-based methods in Appendix 3.

Neural networks are a class of predictive modeling techniques whose model architecture resembles the network of biological neurons. Neural networks make predictions using highly complex patterns between inputs and the outcome of interest using a sequential "layering" process. Recurrent Neural Networks (RNN) are a special type of neural network that sequentially transmit information of time-dependent inputs through "recurrent" layers. Although RNN models can exhibit strong performance in complicated, sequence-dependent prediction tasks, they are especially complex and computationally demanding.

As we show below, the most accurate models use the full set of predictors described above (i.e., both basic and complex non-term and term-specific predictors). The base models we test thus include the full set of predictors.[7] For the OLS, Logit, Random Forest and XGBoost models, we also rank-order the predictors based on their "importance" -- i.e., their explanatory power in predicting the probability of graduation within six years. We provide additional details about the predictor importance measures in Appendix 3.

F. Model Comparison and Evaluation Methods

Our aim is to compare the accuracy and stability of the predictions generated from the six different prediction methods that we tested. To make these comparisons, we calculate five evaluation statistics on the validation sample for each model:[8]

- C-statistic: a measure of "goodness of fit" of predictive models. Specifically, the c-statistic is equal to the probability that a randomly selected student who actually graduated has a higher predicted score than a randomly selected student who did not graduate.
- Precision: a measure capturing how often a model's positive prediction is correct. Specifically, the precision value is equal to the share of students the model classifies as graduates (predicted positives) who actually graduated (true positives), i.e., $Actual\ Graduates / Predicted\ Graduates$.
- Recall: a measure capturing a model's ability to correctly classify actual graduates as predicted graduates. Specifically, the recall value is equal to the share of actual graduates that the model correctly predicts will graduate, i.e., $Predicted\ Graduates / Actual\ Graduates$.

- F1-score: a measure that accounts for the inherent tradeoff between precision and recall as the prediction score threshold used to distinguish students classified as graduates versus non-graduates changes. Mathematically, the F1-score is equal to the harmonic mean of precision and recall (i.e., $2 * [(precision * recall)/(precision + recall)]$) and ranges from 0 to 1, with higher values denoting stronger model performance.
- Rank order of predicted scores: Models may have very similar overall performance but generate inconsistent predictions for a given student, especially in terms of relative risk. For every combination of model pairs, we therefore calculate the magnitude of change across each student's predicted score percentile in model A and model B. We then report summary statistics of within-student distributional changes in predictions across models.

III. Results

A. Full vs. truncated sample

We first compare the model performance of the full training sample to the model performance using the truncated training sample for models that only include simple non-term-specific predictors as well as models that include both simple and complex non-term-specific predictors.[9] We present the results in Figure 1. For all models, we observe an increase of 0.03 - 0.09 in c-statistic values for the truncated training sample compared to the non-truncated training sample. Furthermore, the performance of Random Forest and XGBoost models on the non-truncated training sample using only simple non-term-specific predictors (Panel A) are comparable to the performance of OLS and Logit models on the truncated training sample using all the non-term-specific predictors (Panel B). This suggests that it is possible to achieve strong model performance with the simplest approaches to sample and variable construction; however, doing so requires more sophisticated modeling approaches.[10] We also report the c-statistic values for comparing model performance of different sample construction methods in column 1-4 of Appendix Table A1.

B. Complexity of variable construction

Having demonstrated the improvement in model performance from using truncated samples that more closely resemble current enrolled students, we now turn to assessing the impact of model performance when simple versus more complexly-specified predictors are used to predict graduation. Figure 2 shows that models that only include 14 basic non-term-specific predictors produce relatively informative and reliable predictions. OLS and Logit models generate c-statistics between 0.81-0.82; the CPH model produces a c-statistic of 0.84, and Random Forest and XGBoost models yield c-statistics between 0.85-0.86. However, we observe that adding more complexly-specified non-term-specific predictors to the models, for a total of 61 non-term-specific models, meaningfully improves the performance of all five models. Across all models the c-statistic values increase by 0.03 - 0.04. We further examine how adding simple term-specific predictors that are commonly utilized by institutions, such as the number of credits attempted and term GPA, influences model performance. Model performance improves slightly across all five models with the addition of basic term-specific predictors, and the OLS and Logit models improve most (with increases in c-statistic values of 0.02 - 0.03 versus less than 0.02 across all other models). Lastly, we examine changes in model performance with the further addition of more complexly-specified term-specific predictors, such as the number of 200-level credits attempted in each term. Those term-specific predictors result in minimal improvement to model performance. The marginal increase in c-statistic value is no greater than 0.002 across all six models when complexly-specified term-specific predictors are included in the estimation procedure. We conclude that even the simplest variable construction can lead to reasonably informative and reliable predictions of graduation. At the same time, there is value in constructing more complexly-specified non-term-specific predictors and simple term-specific predictors to optimize model performance. We also report the c-statistic values for comparing model performance of different variable construction methods in Appendix Table A1.

C. Model Accuracy

In this section, we present three additional model accuracy statistics beyond the c-statistic (precision, recall, F1-score) for both graduates and non-graduates to further investigate model performance. For this analysis, we compare the performance of "base" models across all six modeling choices, all of which are trained and validated on the same randomly truncated samples and include the full set of non-term-specific and term-specific predictors (331 predictors). We present the results of this analysis in Figure 3. The first set of bars replicates in graphical form the c-statistic values reported in Figure 2. The c-statistics are very similar across the six models, ranging from 0.884 for the OLS model to 0.903 for the XGBoost model. These fairly high c-statistics are not particularly surprising, given both the large sample size and detailed information we observe about students in the sample. It is somewhat surprising, however, that the c-statistic for a basic model such as OLS, which requires no model tuning in the base version, is nearly as high as the c-statistic for the XGBoost model, which is much more labor- and computing-intensive. To put this result in context, within our validation sample of approximately 33,000 students, the XGBoost model accurately predicts the graduation outcome for 681 additional students compared to OLS. The most computationally intensive model, RNN, actually has a slightly lower c-statistic than XGBoost.[11]

Figure 3 shows that the precision and recall values are also very similar across the six models, though the non-graduation precision and recall values are significantly higher than the graduation precision/recall values: graduation precision and recall respectively range from 71-75 percent and 71-80 percent; non-graduation precision and recall respectively range from 85-89 percent and 84-87 percent. This difference is driven by the fact that the graduation rate of the validation sample is fairly low at 34.1 percent.[12] Since the most common use of predicted scores in higher education is to identify students who are at risk of withdrawal prior to graduation, we expect the non-graduation recall values to be of greatest salience to researchers and college administrators developing interventions based on predicted scores. Interestingly, while the XGBoost model outperforms the other five models in terms of every other evaluation metric in Figure 3, the OLS model has a higher non-graduation recall value than the XGBoost model, and the Logistic and XGBoost models have the same recall value.

Finally, Figure 3 shows the graduation and non-graduation F1-score for both the "graduated" and "did not graduate" outcomes. Because there can be a tradeoff between precision and recall, the F1-score is used to provide a more consistent comparison of model performance that factors in both dimensions of model performance.[13] Overall, the F1-scores are highest for the XGBoost model. While the graduation F1-score follows a similar pattern to the c-statistic, with the OLS model having the lowest F1-score (0.729) followed by the CPH (0.741), Logistic (0.742), Random Forest (0.743), RNN (0.758) and XGBoost (0.772) models, we see that the ranking of non-graduation F1-Score is slightly different, with Random Forest model performing worst in relative terms (0.857), followed by the CPH (0.858), Logistic (0.864), OLS (0.865), RNN (0.866) and XGBoost (0.876) models. In practical terms, the difference in non-graduation F1-scores between the XGBoost and Random Forest models results in 246 fewer actual graduates predicted not to graduate in the validation sample (Type I errors) and 520 fewer actual non-graduates incorrectly classified as graduates (Type II errors).

As discussed above, we anticipate researchers and college administrators to be most interested in identifying students at risk of *not* graduating. Therefore, in all subsequent results, we report the c-statistic and non-graduation F1-scores associated with each model. However, for parsimony, we focus our discussion on the c-statistic values, which are easier to interpret directly and with which researchers and college administrators are likely more familiar.

Taken together, the results thus far demonstrate that the base models perform very similarly in terms of how accurately they predict the probability of graduating or not graduating from college, despite varying considerably with respect to their computational complexity and familiarity to researchers and practitioners.

D. Consistency of risk rankings

We now turn to the question of how consistent the base models are in assigning risk rankings to students. We first examine in Figure 4 the consistency with which the models rank students on the binary outcome of graduating or not graduating. Across model pairs (e.g., comparing OLS to Random Forest), we observe high degrees of consistency in whether the models predict that a particular student will or will not graduate. For instance, 91.3 percent of students are assigned the same outcome when predictions are derived from XGBoost or OLS models. All rates of consistency across model pairs exceed 90 percent.

Still, the high consistency rates we observe in Figure 4 may mask differences in risk rankings within the two possible predicted outcomes (graduate or not graduate). We therefore examine in Figure 5 the consistency of students' risk rankings.[14] Each density plot in Figure 5 shows a comparison between two model pairs. For each plot, the x-axis represents the difference in percentile ranking for a given student across the two models. For example, if a student's predicted score was in the 10th percentile in Model A but in the 15th percentile for Model B, then their value would be equal to five. The vertical dotted lines represent the 25th and 75th percentiles of the difference in predicted score percentile; the diamonds represent the 10th and 90th percentiles. The OLS and Logistic models appear to generate the most similar percentile rankings for a given student: The 25th and 75th percentiles of the difference in predicted score percentile are -2 and 2 percentile points, respectively. Logistic and CPH models also generate quite similar percentile rankings, with the 25th and 75th percentiles of the difference in predicted score percentile being -3 and 2 percentile points, respectively. However, the differences in percentile ranking across all other model pairs are more substantial, with 31 percent of students moving at least 10 percentiles, and with 7 percent of students moving at least 20 percentiles.

Institutions may vary in which students they target for proactive outreach and intervention along the distribution of predicted risk. Some colleges may take the approach of targeting students at highest-risk, while others may focus on students in the middle of the risk distribution if the risk factors for those students are perceived to be more responsive to intervention. In Figure 6, and in Appendix Figures A1-A7, we thus compare the consistency with which a given student is assigned to each risk decile across model pairs based on their predicted probability of graduation. The three panels of Figure 6 examine changes in risk decile assignment across model pairs using the bottom, third, and fifth deciles as reference points, respectively. Appendix Figures A1-A7 show analogous results using all other deciles as the reference points. To illustrate the degree to which risk assignments fluctuate, Figure 6 also reports into which decile students not consistently assigned to the bottom decile fall. As the first plot shows, among students with OLS-derived predicted values in the bottom decile, 86 percent are also assigned predicted values in the bottom decile and 14 percent are assigned values in the second decile when predictions are generated by Logistic modeling; the same rate of consistency occurs between Logistic and CPH models. However, discrepancies are more pronounced across all other model pairs. We observe the next highest rate of consistency with respect to the OLS and CPH model comparison: 78 percent of students predicted to be in the bottom decile by the OLS model are predicted to be in the bottom decile by the CPH model, while 21 percent and 1 percent of students are respectively assigned to decile two and three based on the CPH model predictions. Some model pairs (e.g., Random Forest versus RNN) assign half of students in the bottom decile to a different decile. As the second two plots show, when we compare the consistency of students' predicted scores between model pairs using the third and fifth deciles as reference points, the share of students assigned to the same risk decile across models is even lower. Taken together, the results in Figure 6 and Appendix Figures A1-A7 demonstrate that the relative ranking of students based on predicted score is quite sensitive to modeling choice and instability is observed along the entire distribution of predicted risk.

Despite the instability in relative risk rankings, Figure 7 shows that the share of students assigned to the bottom and third decile who do not graduate is similar across all six base models. This indicates that the models perform similarly well at sorting non-graduates into the bottom and third decile of the risk ranking distribution, but which students are assigned to those deciles differs. This arises because all the models perform similarly at predicting risk in the bottom third of the risk distribution; as a result, we are not able to make value judgments about the differences in model-derived risk rankings, despite the non-trivial instability in risk rank ordering across models. By comparison, Figure 7 shows that the share of students who did not graduate assigned to the fifth decile varies more across the six base models, ranging from 82.9 percent for the Logit model to 86.6 percent for the XGBoost model. The differences in model-derived risk rankings between the regression models and the more sophisticated prediction methods are partly explained by the increased prediction accuracy of the more sophisticated methods for students on the margin of not graduating.

Part of the movement across risk deciles is also likely attributable to the fact that, while the models exhibit similar levels of accuracy, they assign different levels of importance to the predictors to generate predictions. Figure 8 shows the degree of overlap of the top 20 percent of predictors based on their feature importance across model pairs.[15] While the level of overlap is relatively high between the regression-based (62 percent) and tree-based models (77 percent), the cross-family pairs share fewer than 35 percent of the most important predictors in common. In sum, our analysis shows that students' predicted risk of not graduating can vary meaningfully across modeling strategies. For researchers and administrators, this instability means that modeling selection can significantly impact which students receive outreach and support if resource constraints prohibit colleges from intervening with all students predicted not to graduate. We discuss the practical implications of these results in Section IV.

E. Models with a reduced set of predictors or a reduced sample size

As we describe in the empirical strategy section above, we incorporate 331 predictors into the base models. Furthermore, after exploring the complexity of variable construction in section III.B, we concluded that the performance of models is largely unaffected by the exclusion of complexly-specified term-specific predictors from the base models. In Section III.B we also showed that models experience more significant reductions in performance when simple term-specific predictors and complexly-specified non-term-specific predictors are excluded. In Appendix 5 we further investigate changes to model performance when restricting the set of predictors by examining the stability of risk rankings across the base models and models that include fewer predictors. The results of that analysis reveal that excluding predictors that have negligible impact on model performance (such as the complexly specified term-specific predictors) only leads to a modest change in the risk rank ordering of students, with the OLS and Logistic regression models exhibiting greater stability in risk rankings than other models. We also find that excluding all term-specific predictors leads to more significant changes in the rank ordering of students, with the tree-based methods exhibiting greater stability in risk rankings than the regression methods. The tree-based methods generate more stable risk rankings in this context because they exhibit better prediction accuracy than regression methods when term-level predictors are excluded from the prediction models.

We also tested how the base models perform in much smaller settings, limiting the data to one medium-sized VCCS college and separately to a 10% random sample of the data. We find that, despite the significant reductions in sample size, models applied to smaller samples perform similarly well compared to the base models in larger samples.[16] However, once again we find that the risk rank ordering of students changes substantially in smaller versus larger samples. This is especially true for the tree-based methods. We discuss these results in more detail in Appendix 6.

F. Preliminary investigation of bias in predictive models

While a full investigation of potential bias within predictive models--and potential strategies to mitigate that bias--is beyond the scope of this paper, we do provide a preliminary exploration of potential bias given the common concern that predictive modeling in education may be biased against subgroups with historically lower levels of academic achievement or attainment (see, for example, Ekowo & Palmer, 2016). To illustrate this issue, Figure 9 shows the actual graduation rates of students in our validation sample, by gender, race/ethnicity, Pell status, age, and first-generation status. We see that many historically disadvantaged groups -- including Black and

Hispanic students, Pell recipients, first generation college goers, and older students -- have significantly lower graduation rates compared to their counterparts. Including these types of demographic characteristics in predictive models can result in historically disadvantaged subgroups being assigned a lower predicted probability of graduation, even when members of those groups are academically and otherwise identical to students from more privileged backgrounds.[17] Removing demographic predictors is an intuitive approach to addressing concern of bias in predictive models; researchers and administrators might reason that, without demographic predictors in the model, students with the same academic performance backgrounds would be assigned the same predicted score, regardless of race, age, gender, or income. Furthermore, some state higher education systems and individual institutions face legal obstacles or political opposition to including certain demographic characteristics in predictive models (Baker, 2019; Blume & Long, 2014). We therefore examine how excluding demographic predictors affects the performance and student-specific risk rankings of the base models.

Figure 10 compares the c-statistic and non-graduation F1-score values of the base models with models that exclude the following demographic characteristics: race/ethnicity, gender, Pell eligibility, age, and first-generation status. Despite the strong relationship between these demographic characteristics and graduation shown in Figure 9, the accuracy of all the models is virtually unchanged (the performance metrics all change by less than one percent) when demographic characteristics are excluded. This occurs because many of the non-demographic predictors that remain in the model are highly correlated with both student demographic characteristics and the probability of graduation. We show this explicitly by identifying the top-20 predictors in terms of feature importance from the XGBoost model that excludes demographic characteristics.[18] We then compare the mean values of those predictors for Black versus non-Black students, and for older (age 25 and up) versus younger students. Table 2 shows there are large and statistically significant differences between Black and non-Black students and between older and younger students across nearly all 20 predictors. For example, in row 2 of Table 2, Black students have a cumulative GPA of 2.13 on average compared to 2.63 among non-Black students; the difference of 0.51 grade points is significant at the one percent level.[19] In other words, even when race is not incorporated into prediction models explicitly, the results still reflect the factors that drive race-based differences in educational attainment seen in Figure 9. While full exploration of potential bias in predictive modeling is beyond the scope of this paper, we view this as an important area for further study. We also provide a detailed discussion of the effect of removing demographic predictors from base models on the movement of students across the distribution of risk rankings in Appendix 5.

IV. Discussion

In an era when colleges and universities are facing mounting pressure to increase completion rates, yet public funding for higher education is being cut, institutions have embraced predictive analytics to identify which students to target for additional support. We evaluated the performance of different approaches to sample and variable construction and to different modeling approaches to better understand the tradeoffs to modeling choices. Perhaps the most salient finding from our analysis is that, for a given student, the notion of "risk" is not stable and can vary meaningfully across the modeling strategy used. This instability is most pronounced when compared tree-based and neural network modeling approaches, and among students with more moderate risk of withdrawal prior to completion. For instance, across model pairs, fewer than 70 percent of students assigned a risk rating in decile 3 by one model were also assigned to decile 3 by the other model.

The evidence in this study does suggest that institutions would realize important gains in model accuracy through thoughtful sample and predictor construction. In general, more sophisticated tree-based models differentiate between graduates and non-graduates more accurately than simpler regression-based models, although the gains in accuracy are small. More complex models also generate student risk rankings whose ordering is more sensitive to modeling choices, such as which predictors are included in the models or which institutions or students are included in the sample.

Given these findings, a natural question is under what conditions should colleges consider using tree-based versus regression-based models for targeting purposes. In technical terms, our results suggest that sophisticated machine learning approaches offer a slight advantage when colleges use predictions to target students broadly. The subset of students flagged for intervention is not likely to change considerably in those circumstances, even when different modeling choices produce moderate changes to student risk rankings. Our results also suggest that the value of using tree-based prediction models increases when institutions have limited choice over modeling decisions (e.g., due to legal restrictions over the inclusion of student attributes or because of data limitations). Alternatively, rank order stability becomes more consequential when colleges can only target a small subset of students for additional support; in such cases, we find that OLS and Logistic regression models have a comparative advantage.

There are a broader set of questions that are important for institutions to consider when making decisions about using predictive analytics in higher education. Regardless of modeling approach, there are numerous important ethical considerations. One relates to the bias issue; as we show above, students from underrepresented groups are likely to be ranked as less likely to graduate regardless of whether demographic measures are included in the models. On the positive side, this could lead to institutions investing greater resources to improve outcomes for traditionally-disadvantaged populations. But there is also the potential that outreach to underrepresented students could have unintended consequences, such as reinforcing anxieties students have about whether they belong at the institution. This could exacerbate existing equity gaps within institutions (Barshay & Aslanian, 2019; Walton & Cohen, 2011). There are also important ethical questions around the data elements that institutions incorporate into their predictive models, and whether students are aware of and would consent to these uses of data (Brown & Klein, 2020). For instance, researchers at the University of Arizona use ID swipes to monitor student movement around campus, including when students depart from and return to their dorms (Barshay & Alisanian, 2019). While these measures have the potential to contribute meaningfully to model accuracy, they raise important issues around student privacy that higher education administrators should actively consider.

A second question is whether the benefits of predictive modeling outweigh the costs. To inform this question, we conduct a back-of-the-envelope benefit-cost calculation, which we describe in more detail in Appendix 8. In the context of a community college with 5,000 students, our estimates of model accuracy imply that using a more advanced prediction method like XGBoost would translate into the institution correctly identifying an additional 64 at-risk (i.e., non-graduating) students compared to OLS. If realizing this improvement requires the purchase of proprietary predictive modeling services, the average cost to colleges is estimated to be $300,000.[20] This implies an average cost per additional correctly identified at-risk student of $4,688. While this is solely a back-of-the-envelope calculation, we believe it nonetheless illustrates the importance of higher education leaders critically evaluating whether the gains from more sophisticated approaches to predictive analytics are likely to be greater than what could be realized from alternative investments of those resources.

A final question is whether predictive analytics are actually resulting in more effective targeting of and support for at-risk students in higher education. While few studies to date have examined the effects of predictive analytics on college academic performance, persistence, and degree attainment, the three experimental studies of which we are aware find limited evidence of positive effects for at-risk students (Alamuddin, Rossman, & Kurzweil, 2019; Milliron, Malcolm, & Kil, 2014; Plak et al, 2019). More research is needed to understand the role of predictive analytics in improving institutional performance. One challenge to identifying the impacts of predictive analytics on student outcomes is that it is easy to conflate the targeting value of predictive modeling with the efficacy of interventions built around its use. The slightly positive or null effects found in previous studies may reflect that predictive models convey limited information about students upon which institutions can act. Alternatively, even if predictive models contain actionable information, coupling data analytics with ineffective interventions could conceal the targeting value of predictive analytics. One approach to isolating the targeting value of predictive modeling is to examine whether intervention effects vary by model-generated predictions. To our knowledge prior research has not examined this question and it merits attention in future work. More work is also needed to understand the extent to which predictive modeling in higher education suffers from algorithmic bias and whether that diminishes the efficacy of predictive modeling for historically underserved groups.

In conclusion, the findings in this paper reveal that institutional leaders should carefully consider the intended uses for predictive modeling in their local context before choosing to invest in expensive predictive modeling services.

References

Abu-Mostafa, Y. S. & Magdon-Ismail, M. & Lin, H. T. (2012). “Learning From Data”. http://amlbook.com/

Alamuddin, R., Rossman, D., & Kurzweil, M. (2019). Interim Findings Report: MAAPS Advising Experiment. New York: Ithaka S+R.

Association of Public & Land-Grant Universities (APLU, 2016). “Congress Should Lift the Ban on Student Level Data in HEA Reauthorization.” Policy Brief: https://www.aplu.org/library/congress-should-lift-the-ban-on-student-level-data-in-hea-reauthorization/File.

Attewell, P., & Monaghan, D. B. (2015). Data Mining for the Social Sciences. Oakland, CA: University of California Press.

Bailey, M. J., & Dynarski, S. M. (2011). Inequality in postsecondary education. In G. J. Duncan & R. J. Murnane (Eds.), Whither Opportunity (pp. 117–132). New York: Russell Sage Foundation.

Baker, D. J. (2019). Pathways to Racial Equity in Higher Education: Modeling the Antecedents of State Affirmative Action Bans. *American Educational Research Journal*, 56(5), 1861–1895. https://doi.org/10.3102/0002831219833918

Banko, M. & Brill E. (2011). “Scaling to very very large corpora for natural language disambiguation”, Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. https://dl.acm.org/doi/10.3115/1073012.1073017

Barshay, J., & Aslanian, S. (2019). “Under a watchful eye: Colleges are using big data to track students in an effort to boost graduation rates, but it comes at a cost”. APM Reports. Access on May 4, 2020 at: https://www.apmreports.org/story/2019/08/06/college-data-tracking-students-graduation.

Bastrikin, A. (2020). Student loan debt statistics. Last accessed on May 13, 2020 at https://educationdata.org/student-loan-debt-statistics/#.

Blume, G. H., Long, M. C. (2014). Changes in levels of affirmative action in college admissions in response to statewide bans and judicial rulings. *Educational Evaluation and Policy Analysis*, 36, 228–252

Breiman L, Friedman JH, Olshen R, Stone C (1984) Classification and regression trees. Wadsworth & Brooks, Pacific Grove, https://www.taylorfrancis.com/books/9781315139470

Burke, M., A. Parnell, A. Wesaw, & K. Kruger (2017). “Predictive Analysis of Student Data: A Focus on Engagement and Behavior.” The National Association of Student Personnel Administrators (NASPA), Inc: https://www.naspa.org/images/uploads/main/PREDICTIVE_FULL_4-7-17_DOWNLOAD.pdf

California Office of the Governor. (2019). Governor's Budget Summary -- 2019-20: Higher Education. Last accessed on May 15, 2020 at http://www.ebudget.ca.gov/2019-20/pdf/BudgetSummary/HigherEducation.pdf.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). “Semantics Derived Automatically from Language Corpora Contain Human-Like Biases.” Science, 356(6334), 183–86.

Che Z, Purushotham, Cho K, Sontag D, Liu Y (2018) Recurrent Neural Networks for Multivariate Time Series with Missing Values. Sci Rep 8, 6085, https://doi.org/10.1038/s41598-018-24271-9

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (forthcoming). Income segregation and intergenerational mobility across colleges in the United States. The Quarterly Journal of Economics

Chouldechova, A. (2017). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. Big Data, 5(2), 153–63.

Complete College America (2014). “Four-Year Myth: Make College More Affordable. Restore the Promise of Graduating on Time.” Washington, DC: Complete College America. Retrieved from https://eric.ed.gov/?id=ED558792

Corrin, W., Sepanik, S., Rosen, R., & Shane, A. (2016). Addressing early warning indicators: Interim impacts from the Investing in Innovation (i3) Evaluation of Diplomas Now. New York: MDRC.

Deming, D. J., & Walters, C. R. (2017). The impact of price caps and spending cuts on U.S. postsecondary attainment (NBER Working Paper No. 23736). Cambridge: National Bureau of Economic Research.

Dougherty, K. J., & Natow, R. S. (2015). The Politics of Performance Funding for Higher Education: Origins, Discontinuations, and Transformations. Baltimore: Johns Hopkins University Press.

Ekowo, M. & I. Palmer. “The Promise and Peril of Predictive Analytics in Higher Education: A Landscape Analysis.” New America Policy Paper, October 2016. https://www.newamerica.org/education-policy/policy-papers/promise-and-peril-predictive-analytics-higher-education/

Faria, A.-M., Sorensen, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). Getting Students on Track for Graduation: Impacts of the Early Warning Intervention and Monitoring System after One Year (REL 2017–272). Washington, DC: American Institutes for Research.

Fereshteh, J., S. Tarnpradab, & K.A. Hua (2019). “Syntactic recurrent neural network for authorship attribution.” https://arxiv.org/abs/1902.09723

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, Inc.

Hand, D., Mannila, H., Smyth, P. (2001). Principles of Data Mining. Cambridge, Ma: MIT Press.

Herzog, S. (2006). Estimating student retention and degree completion time: Decision trees and neural networks vis-a-vis regression. *New Directions for Institutional Research*, 131(Fall), 17-33. https://doi.org/10.1002/ir.185

Inside Higher Ed & Hanover Research (2020). Responding to the COVID-19 crisis: A survey of college and university presidents. Washington, DC: Inside Higher Ed.

Iwamasa, D., & Thrasher, M. (2019). An Analysis of Instructional Expenditures in U.S. Public Higher Education from 2004 through 2015. Planning for Higher Education, 47(3), 28-37.

James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning: with applications in R. Springer, New York, http://faculty.marshall.usc.edu/gareth-james/ISL/

Jiang, H. & Nachum, O. (2019). "Identifying and Correcting Label Bias in Machine Learning." ArXiv, https://arxiv.org/abs/1901.04966

Karpathy A (2015), The Unreasonable Effectiveness of Recurrent Neural Networks. Andrej Karpathy Blog, http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Klempin, S.C., M. Grant, & M. Ramos (2018). "Practitioner Perspectives on the Use of Predictive Analytics in Targeted Advising for College Students." CCRC Working Paper No. 103: https://ccrc.tc.columbia.edu/publications/practitioner-perspectives-predictive-analytics-targeted-advising.html

Lichtenberger, E. l., & Dietrich, C. (2017). The community college penalty? Examining the bachelor's completion rates of community college transfer students as a function of time. Community College Review, 45(1), 3-32

Looney, A. & Yannelis, C. (2015). A crisis in student loans? How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults. Brookings Papers on Economic Activity. Washington, DC: Brookings Institution.

Ma, J., Baum, S., Pender, M., & Welch, M. (2017). Trends in College Pricing. New York: The College Board.

Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and Action Analytics: Three Case Studies to Consider. Research & Practice in Assessment, Winter, 70–89.

Mills, M. (2011). Introducing survival and event history analysis. SAGE Publications Ltd https://www.doi.org/10.4135/9781446268360.

Murakami, K. (May 15, 2020). State cuts grow deep. Washington, CD: Inside Higher Ed.

Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366, 447-453.

Paterson, J. (2019). "What a predictive analytics experiment taught 11 colleges about sharing data." Higher Ed Dive, April 18, 2019: https://www.highereddive.com/news/what-a-predictive-analytics-experiment-taught-11-colleges-about-sharing-dat/552986/

Plak, S., Cornelisz, I., Meeter, M., & van Klaveren, C. (2019). Early warning systems for more effective student counseling in higher education: Evidence from a Dutch field experiment. Amsterdam: Amsterdam Center for Learning Analytics and VrijeUniversiteit Amsterdam.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? Education Finance and Policy, 6(1), 43–74.

Ramaswami, M., & Bhaskaran, R. (2009). A study on feature selection techniques in educational data mining. *Journal of Computing*, 1(1), 7–11. https://arxiv.org/abs/0912.3924

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. Journal of Applied Psychology, 104(10), 1207–1225. https://doi.org/10.1037/apl0000405

Shapiro, D., Dundar, A., Wakhungu, P.K., Yuan, X., Nathan, A, & Hwang, Y. (2016, September). Time to Degree: A National View of the Time Enrolled and Elapsed for Associate and Bachelor's Degree Earners (Signature Report No. 11). Herndon, VA: National Student Clearinghouse Research Center.

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Stark, T. (2015). "Leveraging Analytics in Community Colleges." *EDUCAUSE Review*, September 14, 2015: https://er.educause.edu/articles/2015/9/leveraging-analytics-in-community-colleges

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology, 69(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Suzgun, M., Y. Belinkov, & S.M. Sieber (2019). "On Evaluating the Generalization of LSTM Models in Formal Languages". *Proceedings of the Society for Computation in Liguistics (SCiL)*, pp 277-286. https://doi.org/10.7275/s02b-4d91

Treaster, J.B. (2017). "Will You Graduate? Ask Big Data." *The New York Times*, February 2, 2017, https://www.nytimes.com/2017/02/02/education/edlife/will-you-graduate-ask-big-data.html?_r=1

U.S. Department of Education (2019). The condition of education: College enrollment rates. Last accessed on May 13, 2020 from https://nces.ed.gov/programs/coe/indicator_cpb.asp.

U.S. Department of Education. (2020). College Scorecard. Last accessed on May 13, 2020 at https://collegescorecard.ed.gov/data.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. Science, 331(6023), 1447-1451.

Zhang, B. H. & Lemoine, M & Mitchell, M. (2018). "Mitigating Unwanted Biases with Adversarial Learning", Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. https://arxiv.org/abs/1801.07593

Notes

1. Among students who earn a credential within 6 years in our sample, 31 percent earn their credential from a non-VCCS institution. An additional 18 percent of graduates earn a credential from both a VCCS and non-VCCS institution within six years. For our earliest cohort of students (those initially enrolled during the 2007-08 academic year), we observe 78.4 percent of all eventual degree completions through the last term of data available (spring 2018) within six years of initial enrollment. And while a sizeable share of VCCS students intend to transfer to a four-year institution before earning their VCCS credential and bachelor's degrees are typically designed to be completed within four years, over half of bachelor's degree-seeking students take more than four years to graduate (Shapiro et al, 2016); time to bachelor's degree is longer for community college transfer students (Lichtenberger and Dietrich, 2017).

2. 90/10, 80/20, 70/30 are all typical ratios used to split samples into training/validation sets. The smaller the validation set, the more likely measurement error will degrade the evaluation of model performance. At the same time, a smaller validation set increases the size of the training set, which enables development of more informative prediction models. In the context of this study, because over 30,000 students are included in the validation sample based on the 90/10 ratio, the validation sample is sufficiently large for evaluating model performance reliably and allows us to include more observations in the training sample to maximize prediction precision.

3. In other words, all predictive models have the possibility of fitting the training set well but not performing equally well on the unseen data, which is caused by the model tendency to pick up the idiosyncrasies/noises from the finite training set during the model fitting procedure. So, it is necessary to withhold part of the full data as the validation set to avoid overestimating model performance.

4. OLS, also known as a linear probability model in the context of a binary outcome variable, may not conform with all theoretical assumptions of a classification model (e.g., the predicted scores are not bound to fall between zero and one). Still, it is the predictive model that typically requires the least computing power and offers the highest degree of interpretability.

5. Discrete Time Survival Analysis (DTSA) methods would also be appropriate since we observe data in term intervals. However, we employ CPH to model graduation as a function of continuous time because it is easy to implement, widely used in the field, and from a practical perspective, the predictions generated from discrete-time and continuous-time methods are virtually identical in most applications (Mills, 2011; Singer & Willett, 2003).

6. For example, implementation of these methods requires execution of model tuning and cross-validation procedures. We follow conventional standards of practice in machine learning for tuning and cross-validation and discuss those procedures in detail in Appendix 3.

7. Following convention, we explored feature selection for the regression and tree-based models as a pre-processing step with the goal of removing potentially irrelevant predictors that could diminish model performance. However, model performance did not improve as the number of predictors decreased in the feature selection routine, which suggests there are essentially no noisy predictors present in the full-predictor model.

8. We provide a more detailed description of each evaluation statistic in Appendix 4.

9. RNN models are not applicable for this analysis because time-dependent predictors are excluded from the models used for testing full versus truncated samples. Furthermore, we do not estimate CPH models using the non-truncated sample due to the particular sample construction procedures we employed for survival analysis modeling. We refer the reader to the section on CPH modeling in Appendix 3 for further details.

10. We do not test the comparison of full vs. truncated sample construction for models that include term-specific predictors, because when using the non-truncated training sample, there is not a reliable and robust way of imputing term-specific predictor values in unobserved terms for observations in the validation sample. Furthermore, even though we could apply missing value imputation methods to the validation sample, this would not resolve the fact that the distribution of enrollment durations for students in non-truncated samples do not resemble those of currently enrolled students. As a result, we expect that non-truncated samples with imputed term-level predictors would perform worse than truncated samples, as is observed in the case of models that only use non-term-specific predictors.

11. Our hypothesis as to why RNN does not significantly outperform the simpler models in this application, while in other applications it often does, is that the average sequence length per student (i.e. the number of actively enrolled terms) is too low to benefit from the sequential structure of the RNN model. One-third of students in the training sample have only one time step; 60 percent of have fewer than three time steps; and 79 percent have fewer than five time steps. Prior research has found that increased sequence length in the training sample leads to improved prediction accuracy of RNN models (Suzgun, Belinkov, & Shieber, 2019; Jafariakinabad, Tarnpradab, & Hua, 2019).

12. Given the relatively skewed distribution of the graduation outcome, we tested whether upweighting the observations of actual graduates improved model performance. It did not.

13. For the same model, precision and recall move in opposite directions as the threshold of predicted scores used to categorize students as either at-risk or not at-risk changes. For instance, the non-graduation value increases as the threshold increases, because more actual non-graduates will be correctly identified. At the same time, non-graduation precision will decrease because the higher threshold will predict that more actual graduates will not graduate. For example, the

Random Forest model has the lowest value of graduation precision and the middle values of graduation recall and graduation F1-score.

14. In Appendix Table A2, we report Person's and Spearman's rank correlation coefficients across the models. The correlations range from 0.92 to 0.99, indicating a relatively high level of consistency in rank orderings across the models and the full distribution of risk rankings. However, as shown in Figure 5, the correlations mask non-trivial differences in percentile rankings between model pairs for some students.

15. Feature importance measures the contribution of each predictor to the construction of predicted probabilities. The CPH and RNN models are excluded from the results in Figure 8 because those prediction methods do not generate feature importance measures.

16. Due to the pattern of results we observe across the regression and tree-based models, and given the substantial time required to fit and fine-tune the RNN models, we did not perform this additional analysis for the RNN model.

17. This source of bias would likely result in students from historically disadvantaged groups being *more* likely to be identified as at-risk of not graduating and targeted for additional resources. While that might appear to benefit students from historically disadvantaged groups, increased intervention could be detrimental if, for example, outreach from college administrators reinforces students' anxieties about their potential for college success and thus increases their probability of dropout (Steele & Aronson, 1995; Walton and Cohen, 2011). More broadly, this type of bias would also result in a less efficient distribution of scarce institutional resources to support students.

18. We focus on the top-20 predictors in terms of feature performance from the XGBoost model because that model demonstrates the highest overall level of accuracy.

19. In Appendix Table A3, we further show that there is almost complete overlap (92-94 percent) in terms of the predictors with highest feature performance between the base models and models that exclude demographic characteristics. This reinforces that excluding demographic characteristics makes very little change to the risk levels assigned to different groups of students.

20. This cost is reported by James Wiley, a technology analyst with Eduventures, in Barshay and Aslanian (2019).

Table 1: Distribution of enrollment length for fall 2012 enrollees and truncated training and validation samples

Enrollment length	Fall 2012 enrollees	Truncated training sample	Truncated validation sample
(1)	(2)	(3)	(4)
1	0.333	0.333	0.3331
2	0.0433	0.0433	0.0432
3	0.0757	0.0756	0.0756
4	0.1696	0.1696	0.1696
5	0.0263	0.0264	0.0264
6	0.0456	0.0456	0.0456
7	0.0973	0.0973	0.0973
8	0.0152	0.0152	0.0152
9	0.0298	0.0298	0.0298
10	0.0591	0.0591	0.0591
11	0.01	0.01	0.01
12	0.0177	0.0177	0.0177
13	0.034	0.034	0.034
14	0.006	0.006	0.006
15	0.0107	0.0107	0.0107
16	0.0225	0.0225	0.0225
17	0.0042	0.0042	0.0042
N	115,413	298,139	33,115

Notes: enrollment length refers to the number of terms since initial VCCS enrollment, including Fall, Spring, and Summer terms, and including terms in which the student was not enrolled. The truncated training and validation samples include data up through each student's randomly assigned enrollment length in order to construct predictors. See text for more details.

Table 2: Racial and Age differences in the 20 most important features

Predictor	Black	Non-black	Difference		Age 25+	Less than 25	Difference	
Slope of term GPA	-0.169	-0.136	-0.033	***	-0.149	-0.143	-0.007	
Cumulative GPA	2.126	2.631	-0.505	***	2.801	2.364	0.438	***
Slope of term-level number of credits attempted	-0.521	-0.504	-0.017		-0.352	-0.581	0.23	***
% of attempted credits that were withdrawn	0.118	0.086	0.032	***	0.082	0.1	-0.019	***
% of attempted credits that were 200-level courses	0.197	0.223	-0.026	***	0.233	0.209	0.025	***
% of attempted credits that were developmental courses	0.226	0.135	0.091	***	0.151	0.162	-0.011	***
% of attempted credits that were completed	0.694	0.81	-0.116	***	0.843	0.752	0.091	***
Total grant dollars received in first year	2001	1219	781.3	***	1432	1414	18.31	
Standard deviation of term-level share of attempted credits that were withdrawn	0.161	0.127	0.034	***	0.121	0.141	-0.021	***
Credits attempted in first Fall term	9.058	10.064	-1.006	***	7.979	10.592	-2.613	***
Standard deviation of term-level share of attempted credits that were completed	0.225	0.164	0.061	***	0.13	0.2	-0.07	***
Term-level GPA in first Fall term	2.408	2.759	-0.351	***	3.142	2.494	0.648	***
Credits attempted in first Spring term	9.472	10.154	-0.683	***	8.406	10.631	-2.225	***
Term-level GPA in first Spring term	2.365	2.724	-0.359	***	3.166	2.436	0.73	***
Credits attempted in second Fall term	6.24	7.272	-1.032	***	5.318	7.73	-2.412	***
Term-level GPA in second Fall term	2.344	2.654	-0.31	***	3.012	2.447	0.565	***
Credits attempted in first Spring term	6.961	6.658	0.304	**	6.333	6.852	-0.519	***
Credits attempted in first Summer term	3.287	2.864	0.423	***	4.168	2.535	1.633	***
Total grant dollars received in second year	2603	1394	1210	***	2213	1445	767.8	***
Term-level GPA in second Spring term	2.473	2.724	-0.252	***	3.111	2.501	0.609	***

Notes: this table shows the differences of the top 20 predictors based on feature performance from the XGBoost model. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 1: Model performance (c-statistic) under different sample construction methods

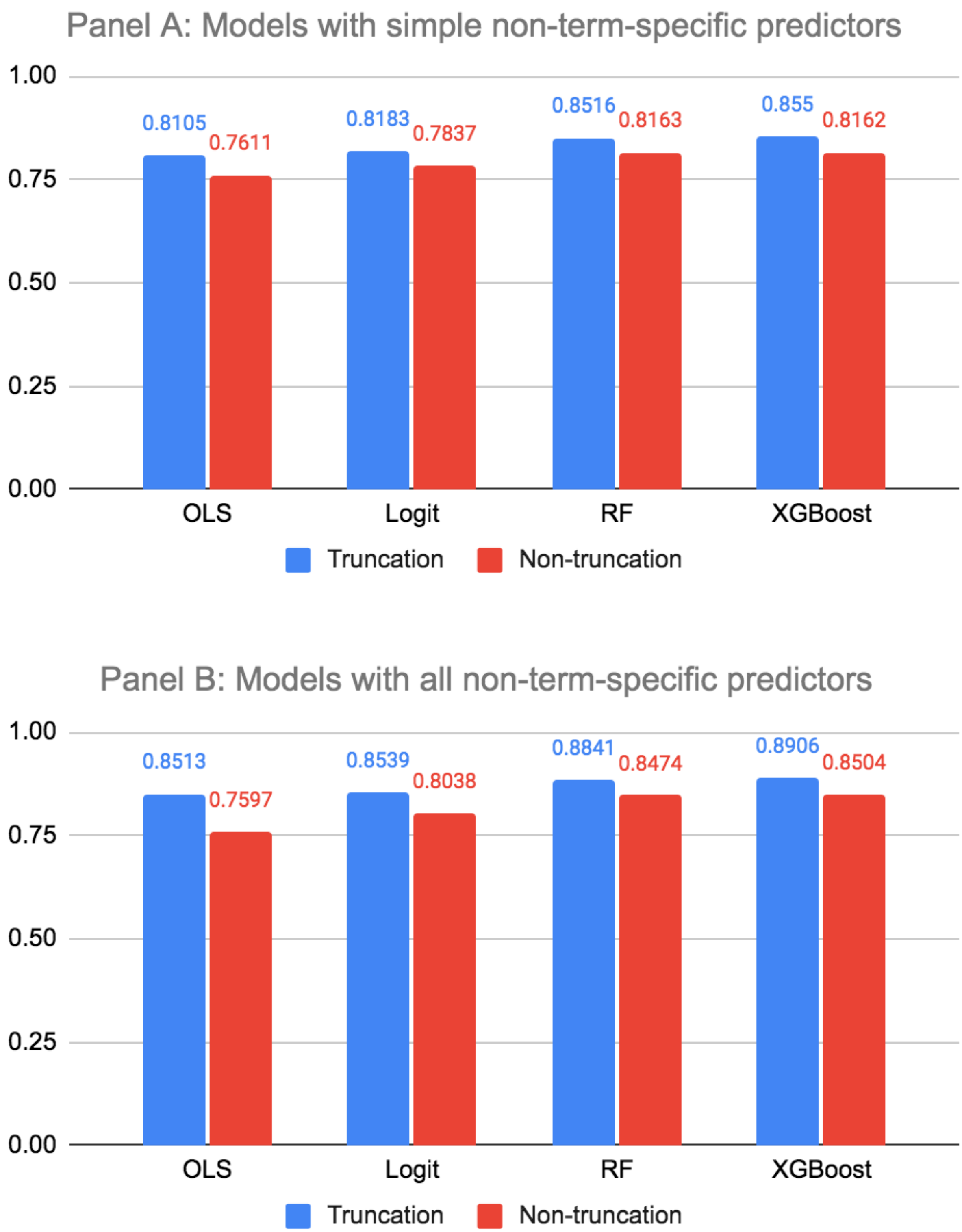


Figure 2: Model performance (c-statistic) under different predictor construction methods

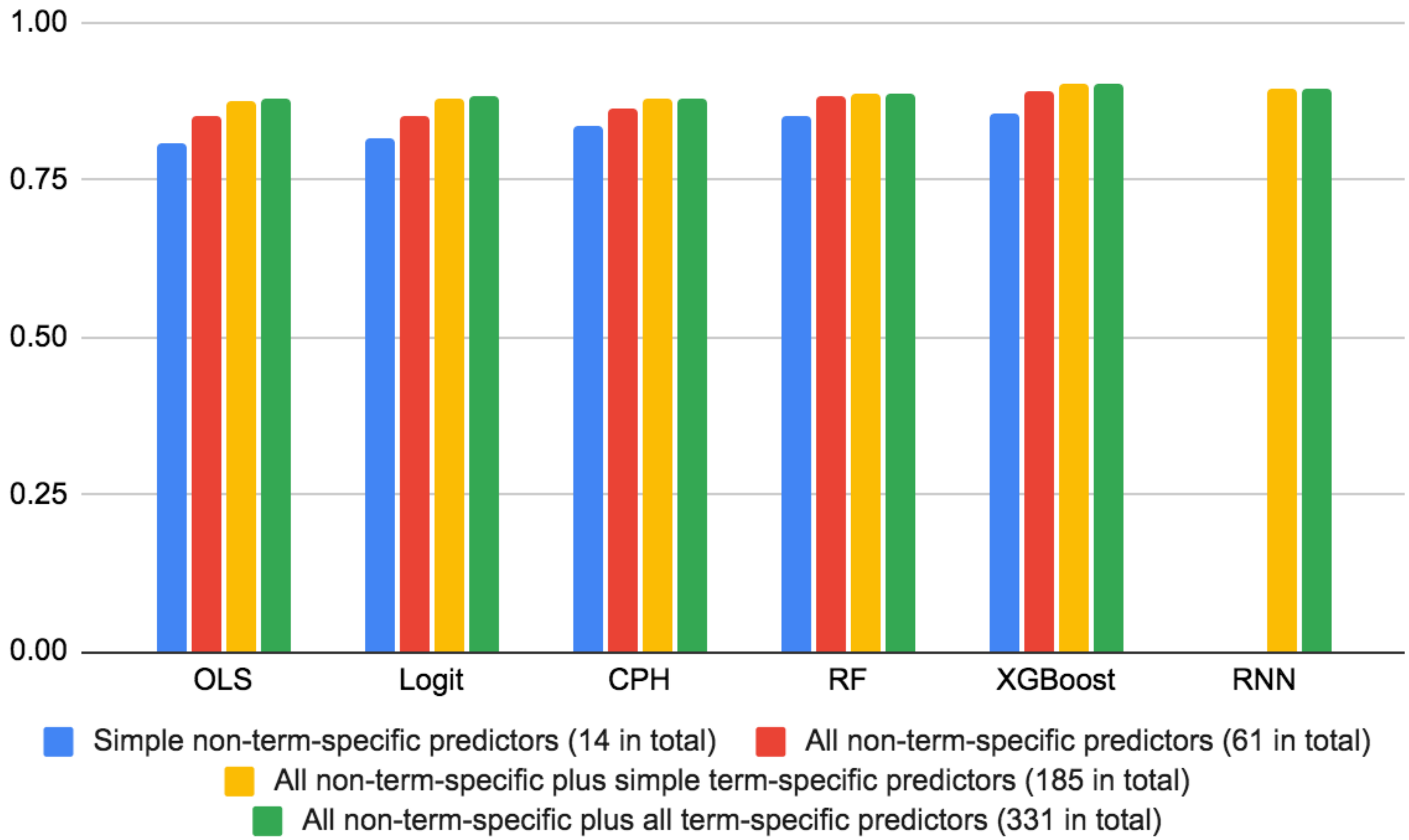


Figure 3: Evaluation statistics of the six base models

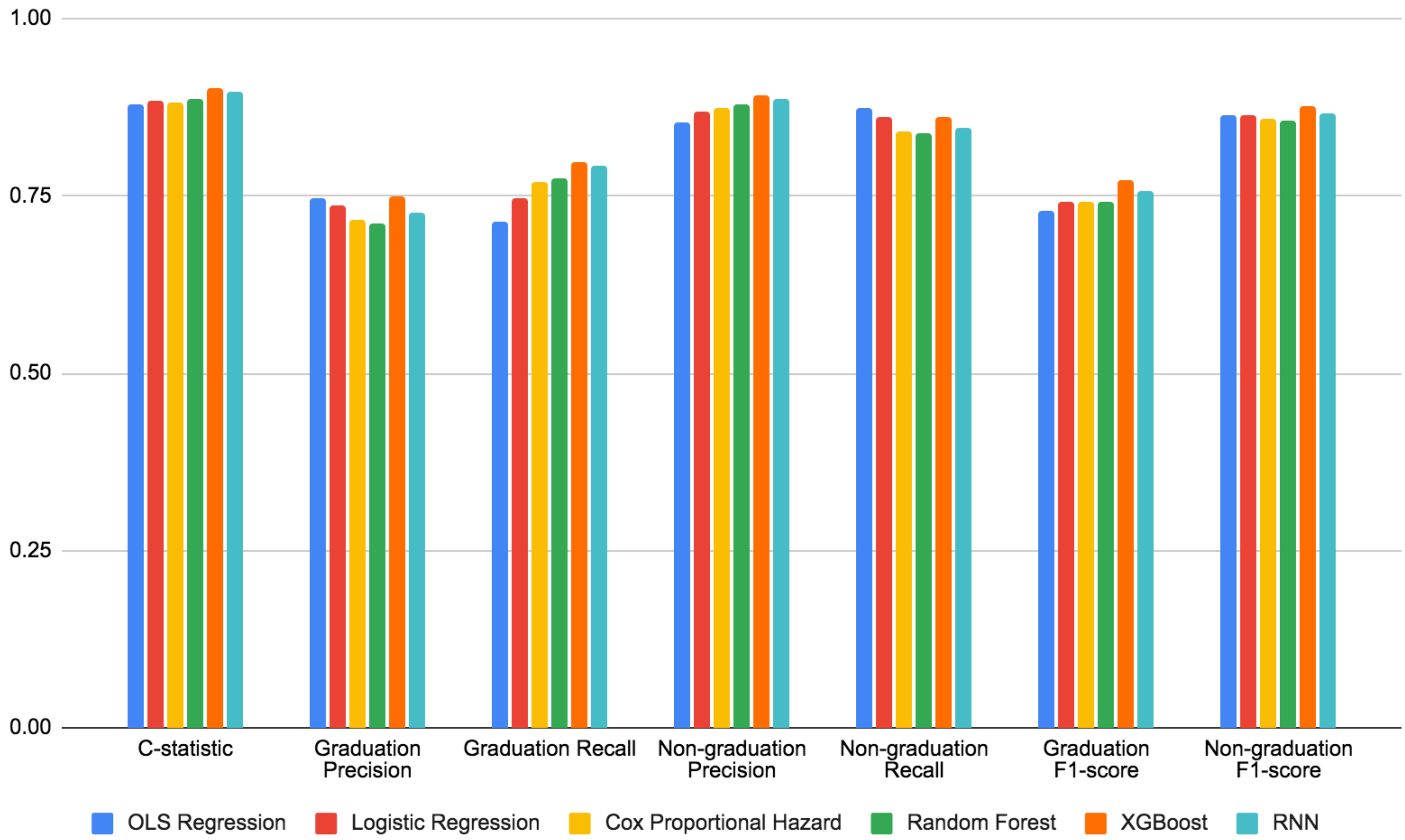
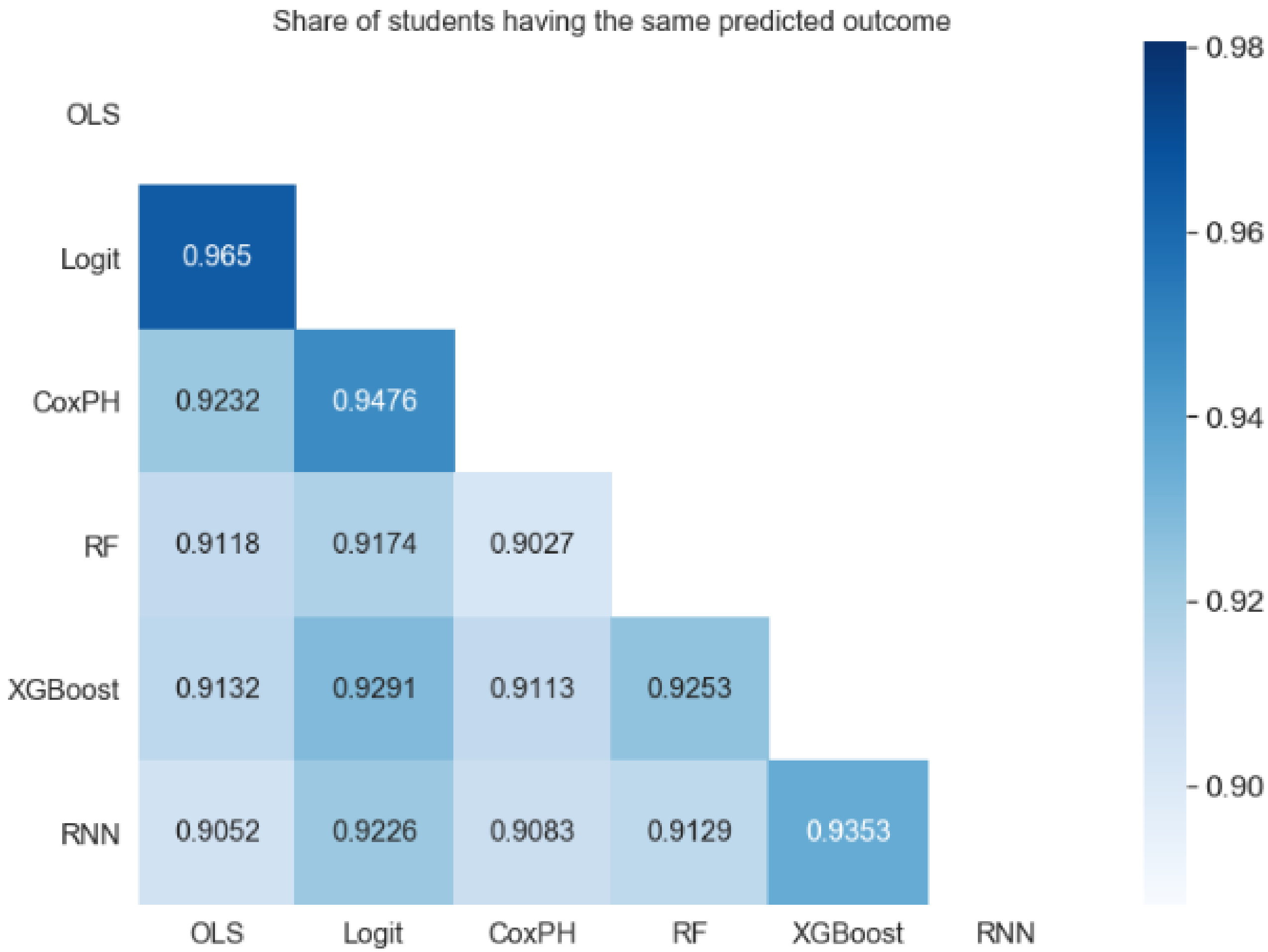


Figure 4: Consistency of students' predicted outcome across base models



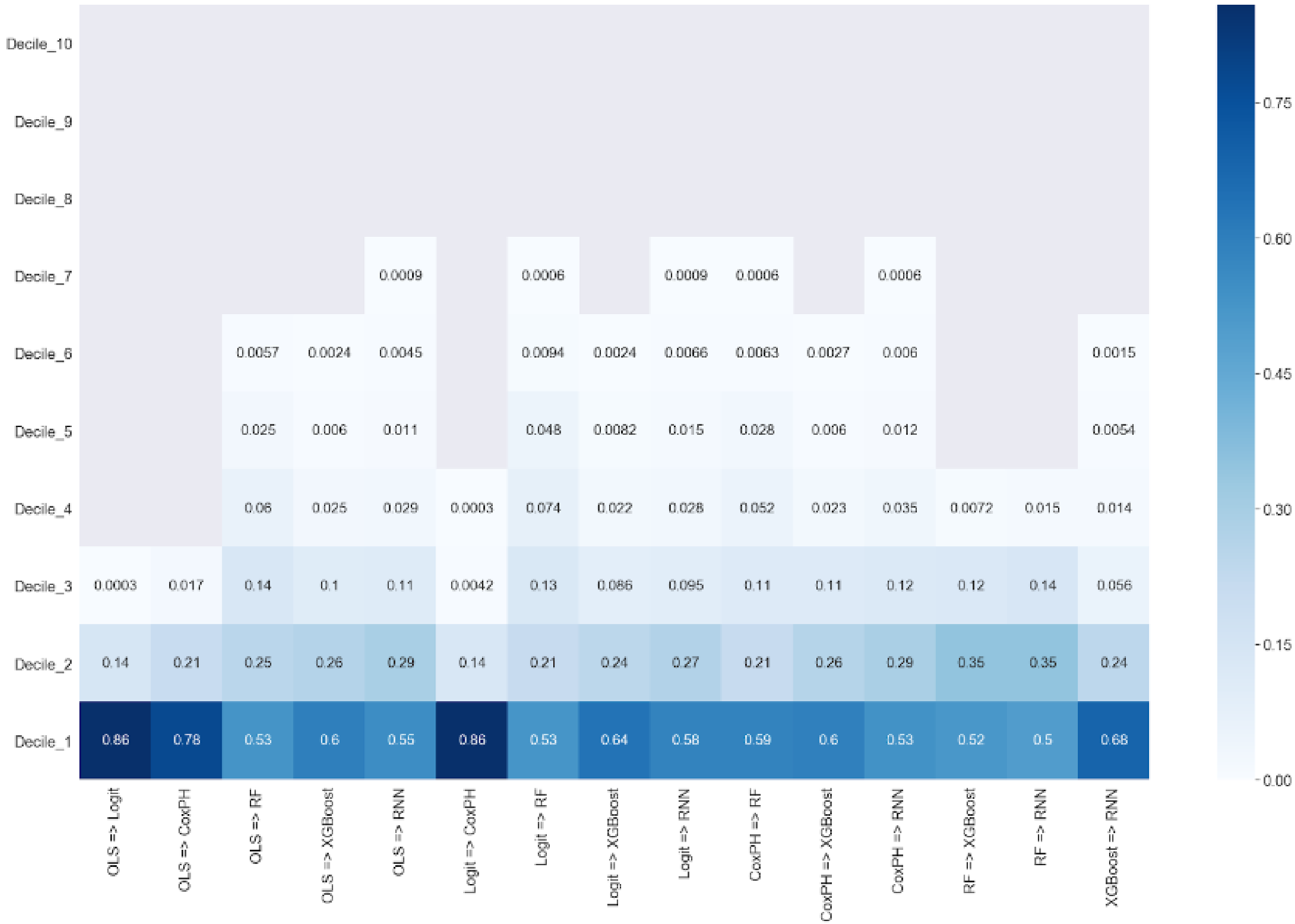
Note: this figure shows the share of students who are assigned the same predicted binary outcome (graduate or not graduate) in both Model 1 and Model 2.

Figure 5: Distribution of differences across base models in students' risk ranking percentile

Notes: these plots show the distribution of the student-level differences in percentile risk ranking between Model 1 and Model 2. For example, if a student's predicted score was in the 15th percentile in OLS but in the 10th percentile for Logistic, then that student would contribute a value equal to -5 in the upper left plot (OLS => Logit). The vertical dotted lines represent the 25th and 75th percentiles of the difference in percentile risk ranking; the solid diamonds represent the 10th and 90th percentiles.

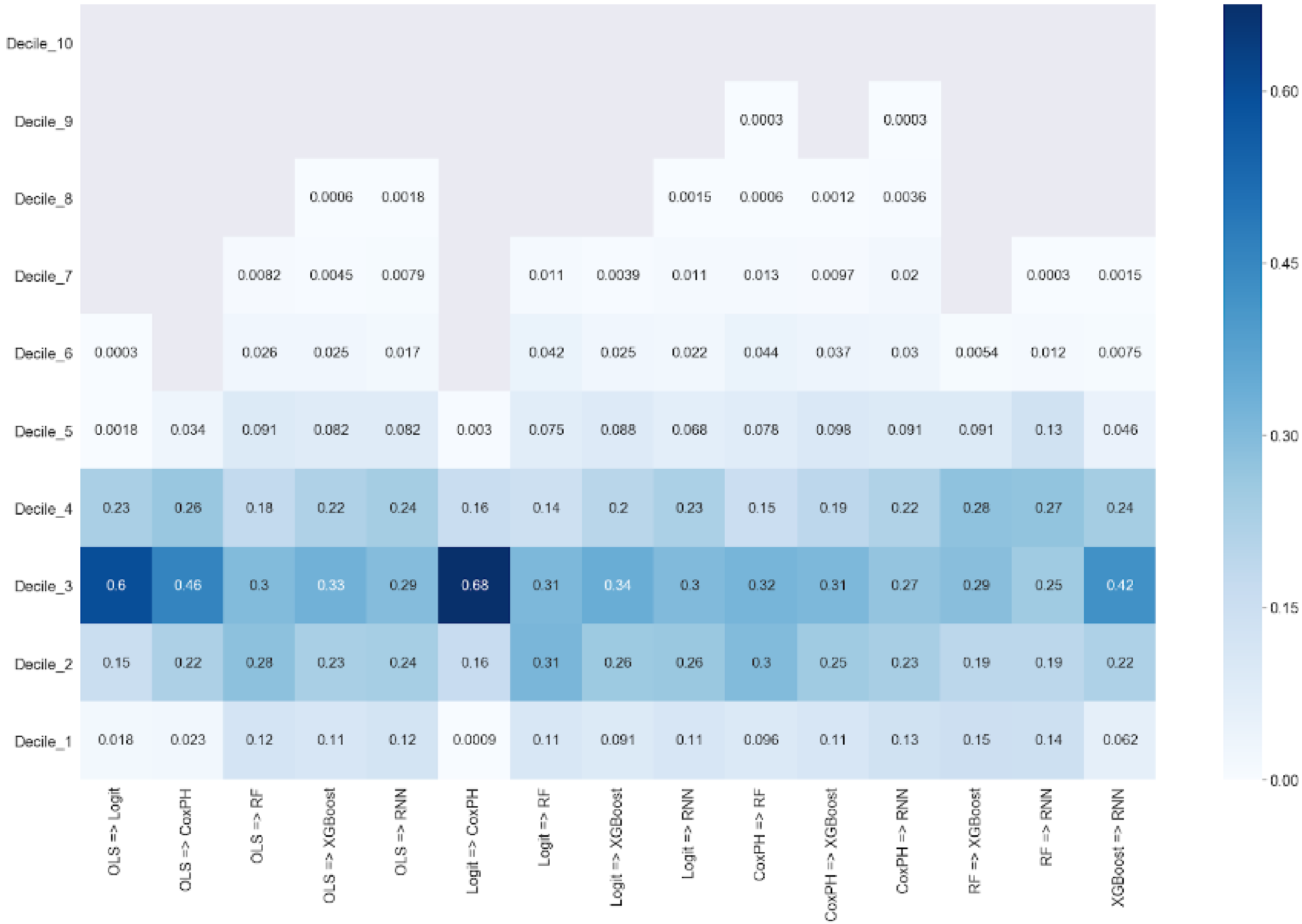
Figure 6: Consistency across models in student assignment to decile of risk rankings

Panel A: First decile of risk rankings



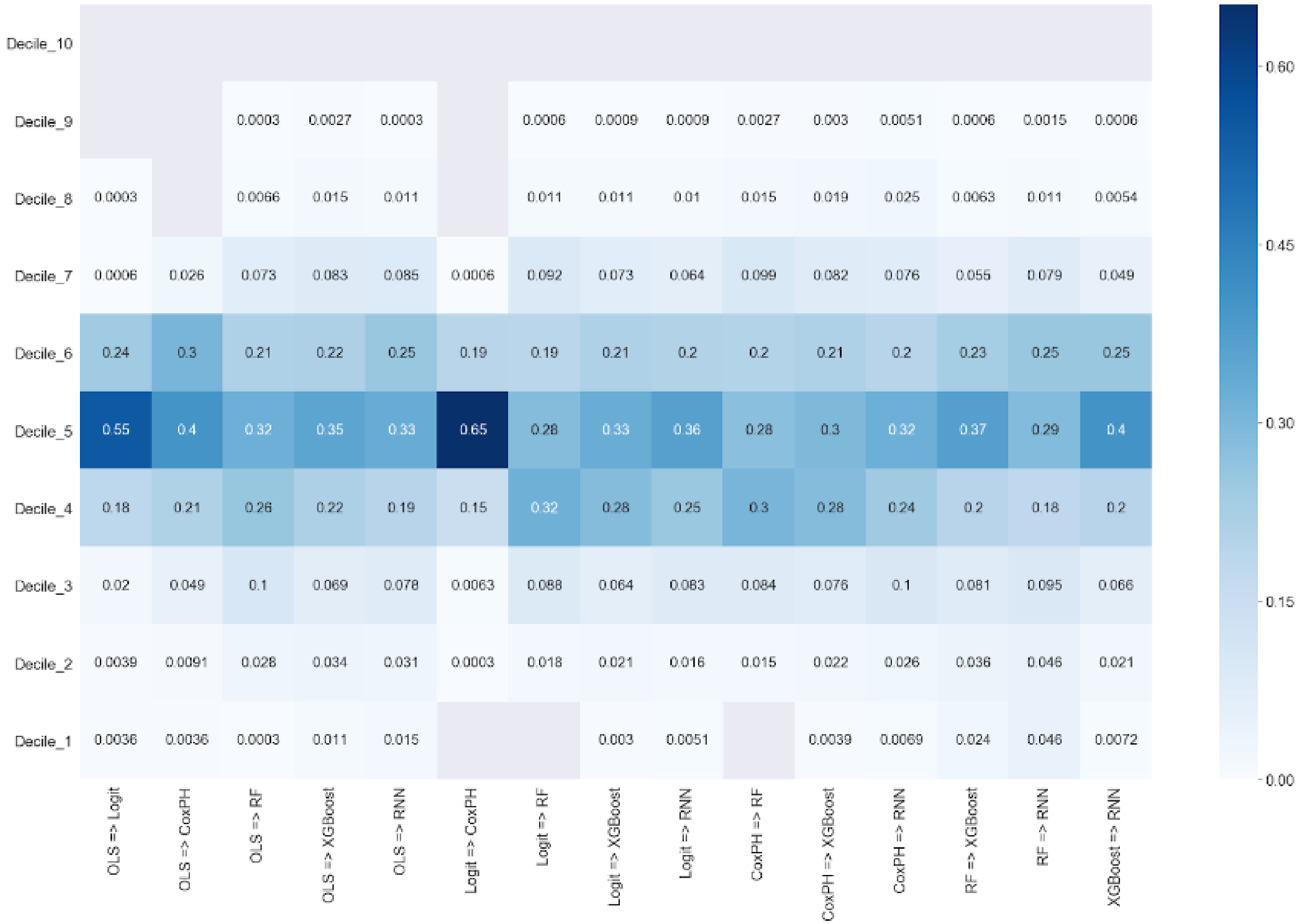
Notes: the first decile of contain the students with a risk ranking percentile between 1-10. Each column of this figure shows the share of students assigned to the first decile by Model A that are assigned to given decile by Model B.

Figure 6, Panel B: Third decile of risk rankings



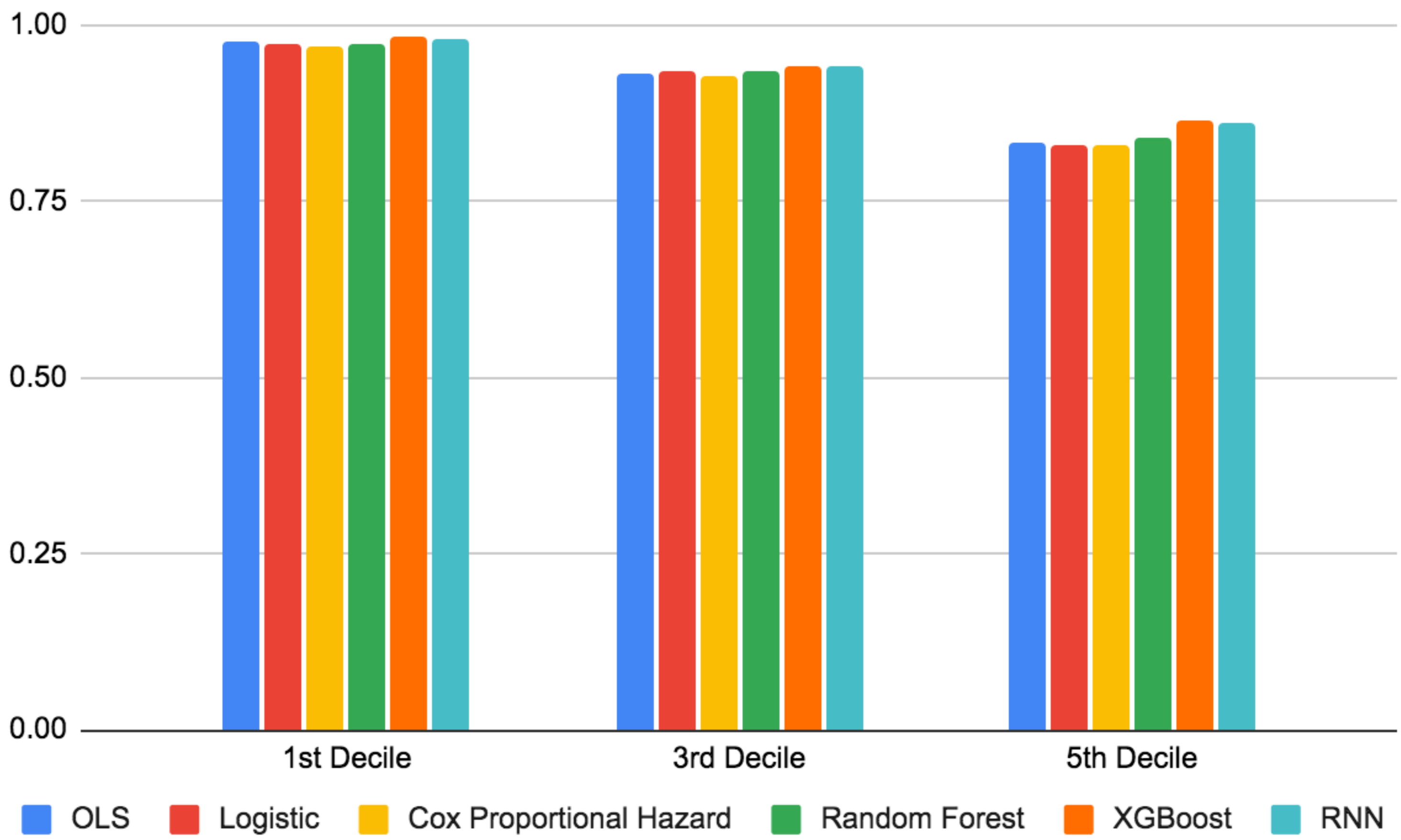
Notes: the third decile of contain the students with a risk ranking percentile between 21-30. Each column of this figure shows the share of students assigned to the third decile by Model A that are assigned to given decile by Model B.

Figure 6, Panel C: Fifth decile of risk rankings



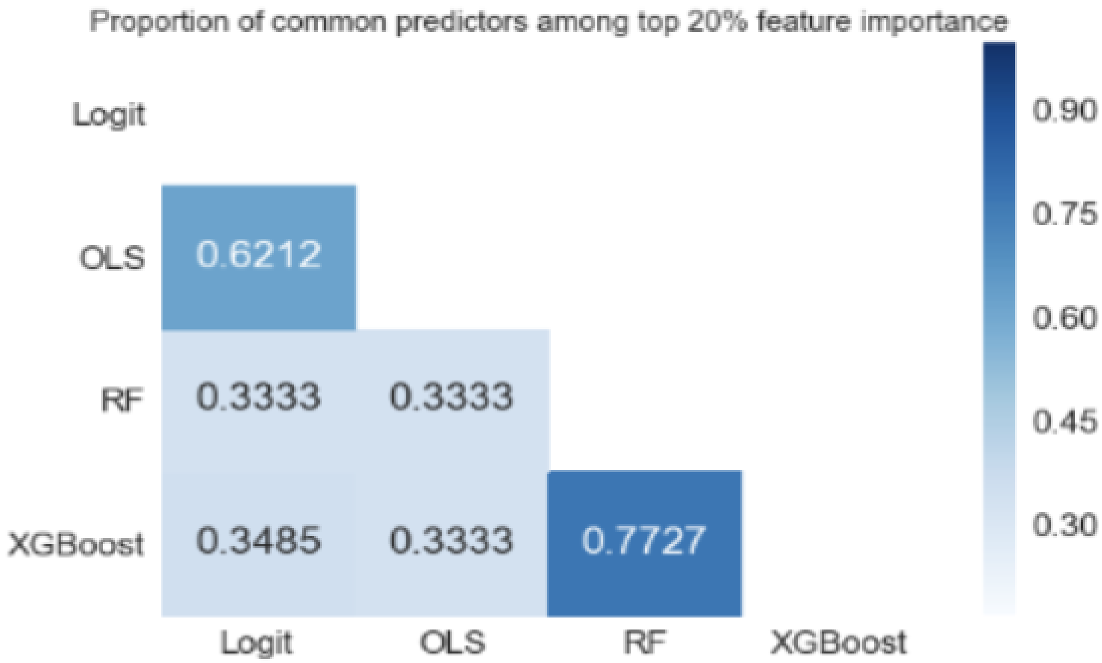
Notes: the fifth decile of contain the students with a risk ranking percentile between 41-50. Each column of this figure shows the share of students assigned to the fifth decile by Model A that are assigned to given decile by Model B.

Figure 7: Percent of non-graduates within the 1st, 3rd and 5th deciles of risk rankings



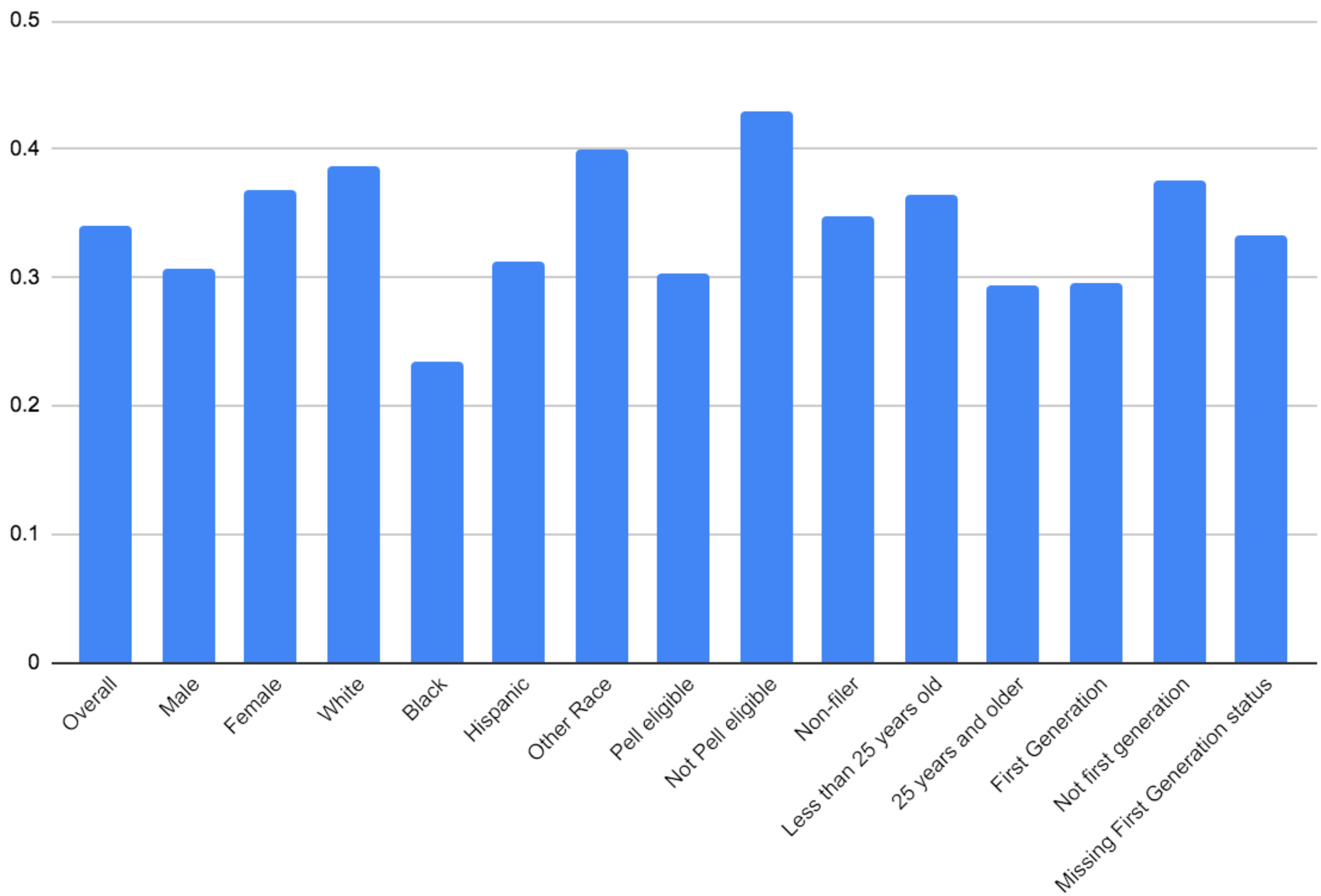
Notes: this figure shows the share of students who are assigned to either the bottom decile, the 3rd decile or the 5th decile of predicted scores (and are therefore predicted to not graduate by all base models) who actually did not graduate.

Figure 8: Commonality of top 20% of important features across base models



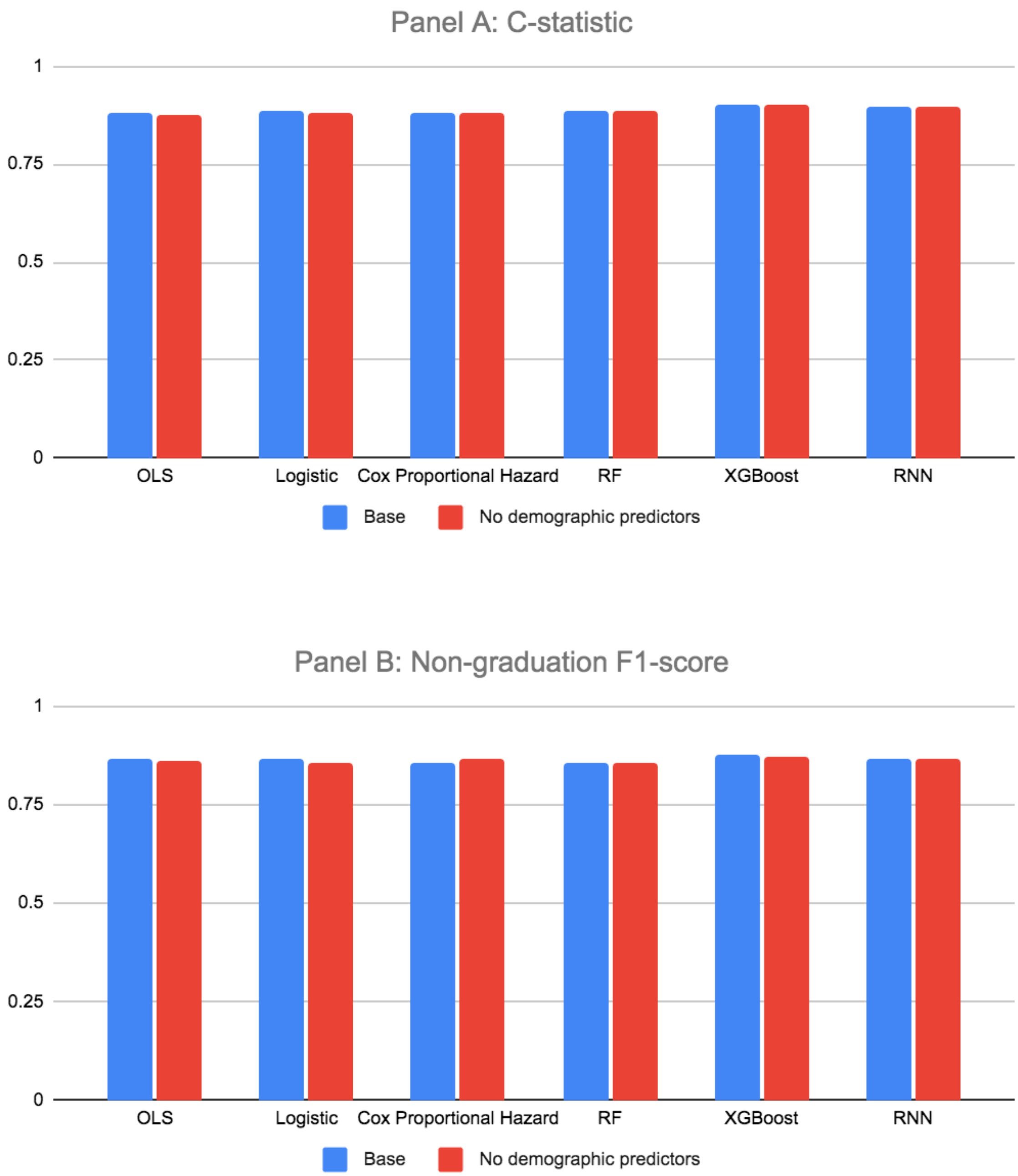
Notes: this figure shows the share of predictors that appear in the top 20% of important features in both Model A and Model B.

Figure 9: Graduation rates by subgroup



Notes: based on observed graduation (based on our outcome variable definition) within the validation sample.

Figure 10: Evaluation statistics, base models versus models excluding demographic predictors



Appendix 1: Details about Sample Construction

(1) Sample definition:

We define degree-seeking status as being enrolled in a college-level curriculum of study that would lead to a VCCS credential (including short-term and long-term certificates and Associate degrees). Note that our analysis excludes students who were only enrolled in non-credit bearing programs, as these students are not represented in our data. We also exclude students who were only ever enrolled at VCCS as a dual enrollment student; for the most part, dual enrollment students are not seeking degrees at VCCS and most enroll as a first-time college student after high school at a non-VCCS institution. In addition, we exclude students who had completed a college credential prior to their initial college-level enrollment at VCCS, as these students have already achieved the outcome we are interested in outside the VCCS context.

(2) Sample truncation method:

If we used all six years of data to construct predictors, we would expect there are certain predictors which are highly correlated with the outcome measure but are not available when applying the model to currently enrolled students. For example, the total number of credits a student has completed by their sixth year would be highly predictive of graduation, but would only be available for students currently in their sixth year. As an illustrative example, for a student enrolled in VCCS during their first academic year and for whom a college administrator would like to estimate their predicted probability of graduating, the model constructed using all six years of data will pick up the fact that this student has no records of enrollment from Year 2 to Year 6; the model would see this as a strong indication of non-graduation. As a result, this student would be assigned a low predicted score regardless of their academic performance in

Year 1. In other words, it is likely that the model is unlikely to accurately differentiate between students who will eventually graduate from the students who will not, because the model is highly dependent on the predictors in subsequent terms which are unavailable for such students. As such, a model using all six years of data to construct predictors is not likely to be generalizable to a sample of currently enrolled students with varying lengths of enrollment history. To mitigate this issue, we apply a random truncation procedure to the sample in order to obtain a new sample whose distribution of enrollment lengths is similar to the currently enrolled cohort at VCCS.

The procedure for performing random truncation for the training and validation sample is as follows: First we identify the percentage of enrollment lengths for all fall 2012 enrollees (assuming fall 2012 is an approximated representation of the currently enrolled cohort at VCCS). The starting with the non-truncated sample (this works for both training and validation samples), among all students whose last enrolled term at VCCS (during the six-year window) is the 17th term since the initial college-level enrollment term, we randomly select a certain number of students so that their truncated observation window is 17 terms, and the number is determined such that the percentage of students whose enrollment length is 17 after truncation is equal to the percentage of fall 2012 enrollees whose number of elapsed terms is 17 since initial enrollment. In other words, the students whose last enrolled term at VCCS is the 17th term that are not selected in this step will be essentially truncated in later steps. In the next steps, for all students in the sample who have not been selected, we first identify those who are enrolled in VCCS during the 16th term since their initial enrollment, and then randomly select a certain number of students so that the percentage of students whose truncated observation window is 16 terms is equal to the percentage of fall 2012 enrollees whose number of elapsed terms is 16 since initial enrollment.

We repeat this procedure for 15th, 14th, ... , until we end up with the students whose truncated observation window is one term. We perform this random truncation procedure from longer truncated observation windows to shorter ones instead of the reverse because starting from shorter truncated observation windows is more likely to cut the observation windows of students who have longer enrollment periods in the historical cohort to very short ones, resulting in insufficient number of students who end up having longer truncated observation window to match the percentage of Fall 2012 enrollees.

(3) Summary statistics of the full analytic sample:

Appendix Table A4 provides summary statistics for our full analytic sample (column 1), and then separately for the training and validation sets (columns 2 and 3, respectively). Panel A shows basic demographic baseline characteristics of the sample, and Panel B shows the academic outcomes of students in our sample. Panel A shows that the average age at initial enrollment for our sample is nearly 25 years old. A little over half of the sample are White, one-quarter are Black, and the remainder is Hispanic or other races. Female students make up 55 percent of our sample. Among students for whom we do observe parental education (57 percent), approximately one-third are classified as first-generation college goers. In Panel B, we see that the average student in our sample was enrolled at VCCS for nearly five terms, and was enrolled at a non-VCCS institution for nearly two terms -- with a little over one-third of students ever being enrolled at a non-VCCS institution. Of the 34.1 percent of the sample who graduated within six years, roughly half only earned a VCCS degree, while the other half earned a non-VCCS degree (either with or without also earning a VCCS degree). Among graduates, the average time to completion was 9.5 terms, which translates to a little over three years. However, this time to degree is highly variable. While we only use the binary outcome for whether a student graduated in our predictive models, we provide descriptive statistics on these other outcomes to better illustrate the enrollment and graduation experiences of students in our sample. Columns 2 and 3 of Appendix Table A4 show that, as expected due to the large size of both the full analytic sample, the randomly selected training and validation samples are nearly identical on these baseline characteristics and academic outcomes.

Appendix 2: List of all predictors by their type and complexity of construction

Appendix Table A5 provides the full set of predictor rankings for the OLS and Logistic models based on the RFE method, and the feature importance measure for the Random Forest and XGBoost models (see Appendix 3 for a description of the RFE method and feature importance measure). A lower value in the OLS and Logistic columns corresponds to a more important predictor; a lower value in the Random Forest and XGBoost models corresponds to a less important predictor. Appendix Table A5 is sorted based on the OLS predictor ranking. Appendix Table A6 provides the full set of coefficient estimates for the OLS and Logistic Base models.

Here is an exhaustive list of all of the predictors in our model:

<u>(1) Simple non-term-specific predictors:</u>

- Demographic predictors:
  - Age at initial enrollment term
  - Gender
  - Race/Ethnicity: four binary indicators for White, Black, Hispanic, other.
  - Parents’ highest education level (categorized)
- VCCS most recent academic predictors:
  - Percentage of terms enrolled at VCCS through the last term or the end of the observation window, whichever comes first
  - Cumulative GPA
  - Share of total credits earned (= credits passed / credits attempted, with credits attempted - credits passed = credits failed. Does not account for course withdraw and audited courses)

    - Average number of credits attempted during each enrolled term at VCCS
    - Ever received a Pell grant

<u>(2) Complexly specified non-term-specific predictors:</u>

- Predictors related to any academic experience prior to student's initial enrollment term at VCCS:
    - Ever dually enrolled at VCCS prior to initial enrollment
    - College-level credit hours accumulated prior to initial enrollment term
    - Cumulative GPA prior to initial enrollment term
    - Share of total credits earned prior to initial enrollment term -- does not account for course withdrawals.
    - Enrolled in any non-VCCS institutions in the 3 years prior to initial VCCS enrollment
    - Number of "pre-VCCS" terms enrolled at non-VCCS institutions in the 3 years prior to initial VCCS enrollment
    - Seamless enrollee indicator (a student is a seamless enrollee if his/her initial enrollment term is in the same academic year as the high school graduation date)
- Overall non-VCCS predictors:
    - Ever enrolled in non-VCCS colleges since initial enrollment term
    - Total number of enrolled terms at non-VCCS (minimum = 0, maximum = 3 * number_of_years_in_time_window)
    - Total number of non-VCCS colleges attended
    - Non-VCCS institution type ever attended -- combination of sector (public, non-profit private, for-profit); level (4-year, 2-year) and location (in-state, out-of-

state). May not be mutually exclusive if the student attended more than one non-VCCS institution

- Admission rates of non-VCCS institutions attended (averaged if attended multiple; weighted based on enrollment intensity)
- Graduation rates (150 percent of time) of non-VCCS institutions attended (averaged if attended multiple; weighted based on enrollment intensity)
- 25th and 75th percentiles of the SAT scores (separately for each subject, math and verbal) of incoming class of non-VCCS institutions attended (averaged if attended multiple non-VCCS institutions, weighted based on enrollment intensity)

- VCCS most recent academic predictors:
  - Standard deviation of term proportion of credits earned
  - Share of total credits withdrawn
  - Standard deviation of term proportion of credits withdrawn
  - Share of developmental credits attempted among total credits attempted
  - Share of 200-level credits attempted among total credits attempted
  - Trend of term enrollment intensity (term credits attempted)
  - Trend of term GPA[1]
  - Ever repeated a course
  - Average grants received by all enrolled terms at VCCS
  - Average subsidized loans received by all enrolled terms at VCCS
  - Average unsubsidized loans received by all enrolled terms at VCCS
  - Average other aids received by all enrolled terms at VCCS

<u>(3) Simple term-specific predictors</u>:[2]

- VCCS term-specific predictors:
    - Indicator for whether the data in this term are within the student's randomly truncated observation window
    - Indicator for whether the student was actively enrolled in VCCS or not
    - Credits attempted
    - Share of credits earned (does not account for course withdraw and audited courses)
    - Term GPA
    - Whether or not the student received a Pell grant

  Note: If the data during a term are not "observed" for a student, the value of all of the predictors in the above list will be zero. If the student was observed to be not actively enrolled in VCCS during a term, then the value of all of the predictors in the above list except for the first one will be zero.

<u>(4) Complexly specified term-specific predictors</u>:[3]

- VCCS term-specific predictors:
    - Proportion of credits withdrawn
    - Proportion of developmental credits attempted among credits attempted
    - Proportion of 200-level credits attempted among credits attempted
    - Repeating a previously attempted course in the current term or not
    - Amount of grants received[4]
    - Amount of subsidized loans received
    - Amount of unsubsidized loans received
    - Amount of other aids received

    - Degree-seeking or not
- Non-VCCS term-specific predictors:
    - Attended any non-VCCS institution
    - Total enrollment intensity in non-VCCS institutions. For example, consider a student who attends two non-VCCS institutions in the same term. If the student is enrolled full-time at the first institution and part-time at the second institution, then their term-specific total enrollment intensity will be equal to 1.5

Appendix 3: Details about Predictive Models

(1) Details regarding setting up, fitting and evaluating the CPH models:

In the CPH model, we define the event to be graduation, and the event time to be the number of terms elapsed since the beginning time (t = 0). For students who did not graduate within six years since initially enrolled at VCCS, the CPH model has a built-in way of handling those observations called right-censoring, which assumes that the event occurs at a future time point that is beyond the six-year observation window but the exact event time is unknown. For those observations, we define the censoring time as the last term of the six-year observation window relative to t = 0. As a result of our definition of the event, the "survival" function of the CPH model is interpreted as the likelihood of not graduating through a certain time point since the beginning time, while the "hazard" function is interpreted as instantaneous likelihood of graduating at a certain time point given the graduation event has not occurred. With this setup, the CPH model characterizes the association between the graduation time and the pre-specified covariates (predictors). CPH can be constructed either using non-time-varying covariates only, or using both non-time-varying and time-varying covariates that allows for incorporating the term-specific predictors at each time point.

The simplest and most natural way of defining the beginning time for the CPH model is to set the term in which each student initially enrolled at VCCS at the college level as t = 0. However, this definition is not feasible because it suffers from some theoretical and practical issues given our definition of the outcome of interest in this paper:

- We do not construct the CPH model that only includes non-time-varying covariates using the natural way of defining t = 0, because by doing so we are only able to include a very small number of non-time-dependent predictors (e.g., demographic predictors) which are not meaningfully predictive of the outcome of interest. For all term-specific predictors (e.g., enrollment intensity of each term) and some other non-term-specific predictors (predictors that are not anchored to a specific term but are in fact time-dependent, e.g., cumulative GPA through the end of each term) that have been demonstrated to be strongly predictive of the outcome, we're unable to include them in such a non-time-varying CPH model because the values of those predictors change over time.

- We do not construct the CPH model that includes time-varying covariates using the natural way of defining t = 0, because our goal is to predict a relatively long-term outcome, and in particular, the majority of students included in our study sample will be those who have enrolled in VCCS for less than 2 academic years. For such students, based on the mechanics of the CPH model, it will need to first predict the probability of the event occurrence for every single future time point within the six-year observation window, and then calculate the cumulative probability. But in order to generate the prediction at each future time point, the model with time-varying covariates needs to know exactly the values of those covariates at that time point, which are unknown to us. So strictly speaking, CPH model with time-varying covariates cannot generate the desired predicted score that aligns with our definition of the outcome of interest, for the currently enrolled cohort. There are two potential workarounds to tackle this issue: (1) Assume all future terms have the same covariates as the last term (the current term); (2) Use the mean values of observed covariates at time t in the training sample to "impute" all of the unobserved covariate values at time t in the validation/current sample. However, neither method is theoretically or practically sound because they have to make very strong and unrealistic assumptions about the unseen data.

To tackle the issues of using the natural way of defining the beginning time, we devised an alternative way of defining the term corresponding to t = 0. In fact, we used the same randomly truncated training sample as is used in other base models – the truncated training sample that closely resemble the currently enrolled cohort at VCCS in terms of the number of terms that elapsed since each student has initially enrolled, and set the last term in the observation window as t = 0, and finally set the event time or censoring time relative to that last term. For instance, if the observation window for a student in the truncated training sample is the 1st – 5th terms, and if the student graduated in the 8th term, then we set the 5th term as t = 0, and the event time to be t = 3. As another illustrative example, if the observation window for a student is the 1st – 3rd terms, and if the student did not graduate within 6 years, the student is right-censored at t = 15. The advantage of using the alternative definition is that all terms during the observation window will be t <= 0, so we could construct the predictors in a similar fashion as we did for fitting the other base models -- explicitly include all of the non-term-specific and term-specific predictors, and treat those predictors as if they are non-time-varying covariates in the CPH model, because they only encode the information of each student prior to last term (t = 0).

Because there is heterogeneity across students included in the training sample in terms of how many terms have elapsed since they initially enrolled in VCCS at the college level, and students who started VCCS at different times might have different likelihoods of graduating within 6 years, we applied stratification to the CPH model, which allows students with different enrollment lengths (1-17 terms) to have different baseline hazard functions.

To evaluate the performance of the CPH model, we use the same randomly truncated validation sample as used in other base models, in order to obtain an "apple-to-apple" comparison of model performance. Same as the training sample, for students in the truncated validation sample, we set the last term in their observation window as t = 0, and calculate the cumulative probability of the event (graduation) occurring during the time period from t = 1 to t = 18-n, where n is the number of terms elapsed since they initially enrolled in VCCS. And this calculated cumulative probability will be the predicted score for rank ordering students in terms of their likelihood of graduation within 6 years, which is in theory comparable to the predicted scores generated by other base models.

In order for the CPH model that uses the alternative way of defining the t = 0 term to work properly, it is necessary to apply random truncation to the training data. If we were to use the non-truncated data (with the goal of making the sample/prediction construction process as simple as possible) and set the last term in the observation window as t = 0, then the t = 0 term will be either the one in which the student graduated or the 18th term since initial enrollment at VCCS (if the student did not graduate within 6 years). In that case, all events occurred at t = 0, and there is no point of using the CPH model, which is designed to make predictions for events that occurred at different time points.

(2) <u>Additional detail and considerations about the tree-based methods:</u>

As indicated by the name, a Random Forest includes many decision trees, with each tree using a randomly selected subset of the training sample and each node splitting using a randomly selected subset of the model predictors. The results are then averaged across all trees to produce the final predictions and to evaluate overall model performance. XGBoost, which is a popular implementation of gradient boosting, is similar to the Random Forest procedure, except that

XGBoost adds decision trees sequentially to the ensemble by growing each new tree based on the residual errors of all previous trees. In this way, the predictions generated by XGBoost gradually improve as more trees are added. Due to the relatively high complexity of these tree-based methods, it takes much longer time to fit them than regression models, with XGBoost requiring particularly higher computing power because its model fitting procedure cannot be parallelized, as is the case with Random Forest. And unlike regression models, tree-based methods require fine tuning several key parameters in order to ensure the model is working under optimal conditions, which further adds to the computing power and user's skill set needed for model fitting. Because of the ensemble nature of Random Forest and XGBoost, the interpretability of those models is lower than regression models; for example, these models do not have a direct equivalent to the coefficients that OLS and Logistic estimate. But, as we describe below, tree-based methods still provide quantitative measures such as feature importance which gives insights into which predictors are playing a key role in predicting the outcome.

(3) Additional detail and considerations about the recurrent neural network model:

In each sequential "layer" of analysis, neural networks achieve a higher level of complexity by re-combining the information encoded previously. In doing so, a deep stack of layers in certain contexts can more accurately capture underlying relationships in the raw data that ultimately improve predictive performance. Recurrent Neural Networks (RNN) are a special type of neural network models that have achieved state-of-the-art performance in applications with sequential data such as natural language processing and time series predictions, by feeding information on time-dependent inputs back into "recurrent" layers sequentially (Karpathy, 2015; Che et al, 2018). The neural network model we construct in this paper is a modified RNN model that uses recurrent hidden layers to transmit the time-specific predictors and regular hidden layers to transmit the time-invariant and most-recent-term predictors, and then merges the two streams of hidden layers into a uniform output layer. The specific type of recurrent layer used to construct the RNN models in this paper is called a Long Short-Term Memory (LSTM) layer, which is a variant of the regular recurrent layer whose internal architecture is designed to overcome the technical difficulties (vanishing gradient problem) encountered in the procedure of fitting RNN models in practice.

(4) <u>Cross-validation procedure for fine tuning the Random Forest, XGBoost and RNN models:</u>

We randomly partition the training set into n equally sized folds, pick a different fold for evaluation every time and fit a model using the other n-1 folds. The evaluation result averaged over all of the n folds provides an estimation of how accurate the model is able to predict unseen data. During the model tuning phase, we first pre-specify a set of combinations of different parameter values, then for every single combination we estimate the corresponding performance measure using cross-validation, and finally select the combination of parameter values which results in the highest performance in cross-validation. For Random Forest, the tuning parameters include the maximum depth of each tree, the total number of trees in the ensemble and the number of randomly selected predictors used in node splitting. For XGBoost, the tuning parameters include the maximum depth of each tree, the step size shrinkage, the minimum number of instances required for splitting a tree node, the fraction of randomly selected predictors in node splitting and the total number of trees in the ensemble. For RNN, the tuning parameters include the quantities that specify the structure of the model architecture (i.e. the number of each type of hidden layers -- fully connected layers, LSTM layers, etc., number of neurons within each hidden layer) as well as the parameters that control the model fitting procedure of RNN, such as the dropout rate (the probability of randomly deactivating a neuron while training the RNN, used to prevent overfitting) and learning rate (the magnitude for updating coefficients values within one training iteration). The Random Forest and XGBoost models can easily underfit (poor performance on unseen data as result of failing to adequately explain the relationship between outcome variable and predictors) or overfit (poor performance on unseen data as a result of picking up the idiosyncrasies or noises in the training data) if the tuning parameters are not properly set. For RNN, overfitting is a more predominant issue than underfitting.

The performance measure used in model tuning is the c-statistic, a quantitative metric widely used for evaluating models that predict binary outcomes, which is described in detail below. Then optimal parameter values identified in this step are used in all subsequent model construction and evaluation steps.

(5) Procedure for choosing the optimal threshold value:

Because the raw output of predicted models are scores ranging from 0 to 1 which approximately represent the likelihood of graduation (the higher the predicted score is the more likely the student will graduate), a threshold is needed in order to serve as the decision boundary -- if the predicted score of a student is above the threshold, they are predicted to graduate, and vice versa. If the threshold value is too low, there is increased risk of failing to identify a substantial share of students who will actually not graduate as high-risk students. And if the threshold value is too high, it is more likely to incorrectly identify a considerable number of students who will actually graduate as non-graduates. To find a reasonable threshold value that strikes a balance between these two considerations, we use a performance measure called F1-score (described in more detail below). Specifically, we use cross-validation to identify the threshold which maximizes the F1-score for each model.

(6) Details about predictor importance measures:

For the OLS and Logistic regressions, we use recursive feature elimination (RFE), which repeatedly removes the weakest predictor (based on the coefficient values) and then refits the model until only one predictor is left. The order of predictors eliminated corresponds to the reversed feature importance. For CPH regression, as discussed in Appendix 3. (1), because we include a stratification variable into the model in order to account for the heterogeneity of the truncated training sample, there are different coefficient estimates for covariates associated with each stratum. Therefore, there are no well-defined feature importance measures based on RFE. For the Random Forest model, we use the mean decrease in importance (MDI) as the feature importance metric to rank order the predictors, which calculates the overall contribution of each predictor to improving the prediction accuracy in instances where the predictor is used to split a tree node. Similarly, the XGBoost model uses a metric calculated according to the relative frequency and depth of each predictor being used in splitting tree nodes to identify feature importance. For the RNN model, because the term-specific predictors are constructed in a different way from the other five modeling approaches (each term-specific predictor across different terms are considered as a whole predictor instead of different predictors), the quantitative feature importance measures for the term-specific predictors in RNN will not be comparable to those from other models. As a result, we do not explore the RNN feature importance in this paper.

Appendix 4: Additional details about each evaluation statistic

C-statistic: A c-statistic score of 0.80 means that when randomly selecting two students -- one who actually graduated and one who actually did not -- there is an 80 percent chance that the student who actually graduated has a higher predicted score than the student who did not. A c-statistic of 0.80 is typically considered a sufficient level of accuracy.

Precision: In other words, a higher precision value translates to fewer Type I errors (false positives). We also report the non-graduation precision values, i.e., the proportion of predicted non-graduates who actually did not graduate.

Recall: In other words, a higher recall value translates to fewer Type II errors (false negatives). We also report the non-graduation recall values, i.e., the proportion of actual non-graduates who are correctly classified as graduates.

F1: We also report the non-graduation F1-score, which is calculated using the non-graduation precision and non-graduation recall values.

Rank order of predicted scores: Specifically, we examine the distribution of all magnitudes of change and the proportion of students who fall into the same quantiles in both models versus different quantiles.

Appendix 5: Additional tests with reduced set of predictors

We explore five particular cases for reducing the set of model predictors: (1) excluding the complexly specified term-specific predictors; (2) excluding all term-specific predictors; (3) only including the simple non-term-specific predictors (4) using feature selection to reduce the set of predictors while still maintaining similar performance to the base models; (5) excluding demographic predictors. Note that RNN is excluded from all tests except for the case (1) and (5); as noted above, RNN is not compatible with our feature selection method, and relies on the time-dependency of the term-specific predictors.

(1) <u>Excluding the complexly specified term-specific predictors</u>

As we have discussed in section III.B, excluding the complexly specified time-specific predictors from each base model only results in very small reductions in model performance, with the decrease in c-statistics being less than 0.3 percent. In the case of the Random Forest models, model performance actually improves slightly when we exclude the complexly specified term-specific predictors in the model. Figure A8 also compares non-graduation F1-scores, which are also very close between each pair of base models and the model which excludes the term-specific predictors (within 1.5 percent).

In Figures A9 and A10, we explore how excluding the complexly specified term-specific predictors affects the risk rankings of students based on the model predictions. Figure A9 shows the distribution of students according to their absolute change in percentile of predicted scores between the base model and the model that excludes the complexly specified term-specific predictors. For instance, looking at the first column of Figure A9, we see that 89 percent of students' predicted scores from the OLS model changed by five percentiles or less across the two specifications. We see very similar results for the Logistic, Random Forest, and XGBoost models; however, the predicted scores vary more within CPH and RNN. Approximately 16 percent of students move six or more percentiles across the CPH derived risk distribution when using the set of predictors that exclude the complexly specified term-specific ones, and the share for RNN is 25 percent. Still, the share of students that experience extreme changes in predicted scores remains fairly small. Figure A10 focuses specifically on the proportion of students in the bottom, 3rd and 5th decile of the base models that are assigned to the same decile when using the set of predictors that exclude the complexly specified term-specific ones. For instance, 89 percent of students in the bottom decile and 73 percent of students in the 3rd decile remain in those deciles across the two OLS model specifications.

Both Figures A9 and A10 show that excluding the complexly specified term-specific predictors has a relatively larger effect on the placement of students in the distribution of predicted scores for CPH and RNN, although in absolute terms the amount of variation is minimal. We hypothesize that this pattern of results is driven by the fact that for CPH regression[5] and RNN models the relationship between predictors and risk rankings is intrinsically more complexly specified than the OLS and Logistic regression in which we do not specify any complexly specified variable interactions prior to estimation, so removing certain characteristics from CPH and RNN can lead to more dramatic movement in risk rankings. Since tree-based models routinely estimate complexly specified relationships between predictors, and predictors in decision tree classifications can also be re-used in the construction of tree nodes, removing certain predictors from the model could potentially have a significant impact on the sequence of nodes within the constructed tree-based models. However, we do not find there is a higher degree of risk ranking movement for Random Forest and XGBoost than OLS and Logit. This is probably because the tree-based methods perform feature selection by themselves -- the less important predictors will be used in a way that has less impact on the predicted risk rankings during the tree growing procedure. And the excluded more complexly specified term-specific predictors happen to be the ones of little importance when fitting those tree-based methods.

(2) Excluding all term-specific predictors

As we have discussed in section III.B, excluding all term-specific predictors from each base model will result in slight but noticeable reductions in model performance. Figure A11 compares the c-statistics and non-graduation F1-scores between the base model and the model without term-specific predictors. Across all five models, we see a reduction in performance when excluding the term-specific predictors, and the reductions in c-statistic values are at least twice as large for the regression models compared to the tree-based models. In other words, when excluding term-specific predictors, the added gain from a more complexly specified model is much more pronounced. Figures A12 and A13 illustrate how removing term-specific predictors affects the movement of students within the distribution of predicted scores. Unlike the results from the model which excludes the complexly specified term-specific predictors presented above, this modeling change results in greater movement with respect to the regression models than the tree-based models, although across all models we observe considerable fluctuation in the relative risk rankings of students. The share of students that move more than 10 percentiles in the distribution of predicted scores with the removal of term-specific predictors ranges from 14 percent (XGBoost) to 36 percent (OLS) across the five model types. This degree of instability is not surprising, given that many of the predictors with highest feature importance in the base models are the simple term-specific ones. Models that exclude those variables necessarily rely more heavily on other predictors that can generate divergent predictions. The pronounced instability of the OLS- and Logistic-derived risk rankings is also likely explained by the degree of performance degradation observed in the regression models. After excluding the term-specific predictors, the OLS and Logistic models become much worse at correctly rank ordering students compared to the Random Forest and XGBoost models, which translates into larger changes in the distribution of predicted scores generated by OLS and Logistic regression. In summary, the results of this analysis indicate that the benefits to using tree-based models are more pronounced when institutions are constrained in their ability to leverage term-level data to identify at-risk students.

(3) Only including the simple non-term-specific predictors

As we have discussed in section III.B, only including the simple non-term-specific predictors from each base model will result in considerable reductions in model performance. Figure A14 compares the c-statistics and non-graduation F1-scores between the base model and the model with only the simple non-term-specific predictors. Across all five models, we see a reduction in performance when excluding the complexly specified non-term-specific predictors and all term-specific predictors, and the reductions in c-statistic values are more pronounced for the regression models compared to the tree-based models. Figures A15 and A16 illustrate how removing term-specific predictors and the complexly specified non-term-specific predictors affects the movement of students within the distribution of predicted. Compared with Figure A12 and A13 where risk ranking movement results for models excluding term-specific predictors are displayed, we find even more pronounced movement of students within the distribution of predicted scores. The share of students that move more than 10 percentiles in the distribution of predicted scores with the removal of term-specific predictors ranges from 38 percent (XGBoost) to 52 percent (OLS) across the five model types. This is as expected, since removal of certain complexly specified non-term-specific predictors that contribute substantially to model performance will force the models to rely more heavily on the simple non-term-specific predictors that can generate more divergent predictions. Similar to the results from the model which excludes the term-specific predictors presented above, the instability of the OLS- and Logistic-derived risk rankings are pronounced than other ones, which can also be explained by the fact that the OLS and Logistic models that only include the simple non-term-specific predictors have significantly poorer performance in terms of correctly rank ordering at-risk students than CPH, Random Forest and XGBoost. Therefore, using a simple variable construction approach can substantially influence which students are targeted for intervention.

(4) Automated feature Selection to maintain model performance

In addition to manually checking the different variable construction methods that have reduced set of predictors, in order to achieve a more parsimonious model, we implement an automated feature selection method using penalized logistic regression with a 2-SE rule, which reduces the number of predictors from 331 to 147.[6] We then estimate each of the five regression and tree-based models using the same set of 147 selected predictors. Figure A17 compares overall model performance between the full model and the model with 147 selected predictors. We observe very small reductions in c-statistics and non-graduation F1-scores (all less than 0.3 percent) as a result of limiting the predictor set. In the case of the Random Forest models, model performance actually improves slightly when we include only the selected predictors in the model.[7]

We extend the penalized logistic regression method of feature selection to show the relationship between the number of predictors in the model and model performance.[8] Figure A18 reports the results from this exploration, with the number of predictors displayed on the x-axis and the corresponding c-statistic from the penalized Logistic model displayed on the y-axis.[9] The upper dashed horizontal line denotes the c-statistic value for the model using the 2-SE selection rule, which crosses the curve at 147 predictors. The lower dotted horizontal line is positioned on a c-statistic value of 0.80, which is a common lower-bound benchmark of acceptable model performance.

In Figures A19 and A20, we explore how restricting the model to the 147 selected predictors affects the risk rankings of students based on the model predictions. Figure A19 shows the distribution of students according to their absolute change in percentile of predicted scores between the base model and the model with 147 selected predictors. Looking at the first column of Figure A19, we see that 96 percent of students' predicted scores from the OLS model changed by five percentiles or less across the two specifications. We see very similar results for the Logistic model; however, the predicted scores vary more within CPH, Random Forest and XGBoost. Approximately 12 percent of students move six or more percentiles across the Random Forest derived risk distribution when using the 147 selected predictors. Still, the share of students that experience extreme changes in predicted scores remains fairly small. Figure A20 focuses specifically on the proportion of students in the bottom, 3rd and 5th decile of the base models that are assigned to the same decile when using the 147 selected predictors. For instance, 95 percent of students in the bottom decile and 88 percent of students in the 3rd decile remain in those deciles across the two OLS model specifications.

Both Figures A19 and A20 show that using the 147 selected features has a relatively larger effect on the placement of students in the distribution of predicted scores for CPH, Random Forest and XGBoost, although in absolute terms the amount of variation is minimal. Comparing those two figures with Figure A9 and A10 for the risk ranking movement results for the models that exclude the complexly specified term-specific predictors, we find that for OLS, Logit and CPH, the variation in risk rankings upon removing the complexly specified term-specific predictors is significantly less than removing the predictors which are dropped in the feature selection. One possible explanation is that the penalized logistic method of feature selection is linear by nature, so it is fairly effective at identifying the predictors to be dropped that have the least impact on both the prediction accuracy and the relative risk rankings for generalized linear models including OLS, Logit and CPH, while those dropped predictors might not necessarily be the least impactful ones for risk rankings generated by tree-based methods.

(5) Excluding all demographic predictors

In Figures A21 and A22, we explore how excluding demographic characteristics affects the risk level rankings of students based on the model predictions. Similar to the pattern of results from the model with 147 selected features, we observe that excluding the demographic predictors from the tree-based models (in this case, particularly XGBoost) and especially RNN has a significantly larger effect on the placement of students in the distribution of predicted scores compared to the regression-based models. Nearly one-quarter of students (XGBoost) and nearly one-third of students (RNN) in the bottom decile of scores derived from the XGBoost base model fall out of the bottom decile when demographic characteristics are excluded from the prediction model. This level of instability arises, despite the fact that only one demographic predictor -- age -- is listed among the top-20 predictors with respect to feature importance in the XGBoost base model. Similar to our explanation above, we hypothesize that this result is driven by tree-based and neural network models' ability to estimate more complex relationships and interactions between predictors. In other words, even though we do not observe meaningful changes in overall model performance when demographic characteristics are excluded from the predictor set, removing them has the potential to more significantly alter which students are targeted for intervention when predictions are generated with more complex machine learning algorithms.

Appendix 6: Additional tests for models using smaller samples

Because we have access to many years of data for the entire system of 23 community colleges in Virginia, the training and validation samples we use to construct the base models are large. However, it is likely the case that individual colleges with a much smaller number of students would want to use predictive modeling to identify at-risk students. We therefore also estimated college-specific models for Piedmont Virginia Community College (PVCC), a medium-sized community college within the VCCS system. Specifically, we limit both the training and validation samples to students who enrolled at PVCC during their last available term within their truncated enrollment window. This restriction reduces the training and validation samples to 2.4 percent and 2.5 percent of their original size (n = 7,132 for the training sample; n = 837 for the validation sample). Note that the graduation rate for PVCC-only validation set is slightly higher compared to the base model, 35.7 percent compared to 34.1 percent, respectively. In order to compare base model performance to the PVCC only models[10], we apply the base models built using the full training sample to a validation sample consisting only of PVCC students. This provides a more "apples-to-apples" comparison between the two models, because each model is being evaluated with the same set of out-of-sample students. Similarly, we perform a second analysis using a 10 percent random sample of the training set to build the model (n = 29,813), which we then compare to the base model using a 10 percent random sample of the validation set (n = 3,312).[11]

We report the c-statistic and non-graduation F1-scores for the base models and the PVCC-only models in Figure A23. The PVCC-only models perform slightly worse (decrease in c-statistic of 1-2 percent; decrease in F1-score of 1-3 percent) but are still very high, ranging from 0.898 (Random Forest) to 0.911 (XGBoost); in fact, the c-statistics from the PVCC-only trained models are higher than the overall base models reported in Figure 3. This performance for the PVCC-only model is beyond what we expected given a significant reduction in the training sample, which is typically accompanied by significantly worse model performance (Abu-Mostafa et al, 2012; Banko et al, 2001). One hypothesis to explain this high level of performance is that the PVCC-only model has a more may be due to a more narrowed focus on a specific context. In other words, because PVCC students may be more alike than VCCS students as a whole, this similarity may improve the models' ability to predict graduation, which partly compensates for the significant reduction in training sample size. However, when we build a training sample the same size as the PVCC-only set using a random sample from all 23 colleges, and apply it to the PVCC-only validation model, we find similarly high levels of performance. Therefore, while we are able to achieve higher accuracy for PVCC students across all training sample configurations, this better performance is likely not driven by a more focused context. In Figure A24, we show that model performance is similarly high when we rely on the 10 percent random sample to build and validate the prediction models. Taken together, the results in Figures A23 and A24 indicate that the value of prediction models are not limited to "Big Data" contexts. They can also be leveraged effectively in smaller contexts to target additional outreach and support to at-risk students.

In Figures A25 and A26, we further examine the extent to which the rank ordering of predicted scores in the PVCC validation sample changes when predictions are generated using the base model versus the PVCC-specific model. Figure A25 reports the distribution of the change in predicted score percentiles and Figure A26 reports the share of students in the bottom, 3rd and 5th decile of base model-derived predicted scores who remain in the same decile when scores are based on the PVCC-specific model. We observe significant movement of students along the distribution of predicted scores, most notably with respect to predictions generated from XGBoost models. In Figure A25, more than half of students move more than five percentiles, and 10 percent of students move more than 15 percentiles between the base and PVCC-specific XGBoost models. In addition, Figure A26 shows that nearly half of students move out of the bottom decile across the two XGBoost models. Figures A27 and A28 show the same results for the 10 percent random sample model. While there is still significant movement in predicted scores between the base model and the 10 percent random sample model, particularly for the tree-based methods, there is less movement compared to the PVCC-only model. These results indicate that relying on a relatively small sample (particularly by focusing on a specific college within a system) can significantly alter which students are targeted. We hypothesize that this pattern exists due to the sensitivity of predictive models to changes in the training data. This is particularly true of tree-based models, which make more complex assumptions about the relationship between the outcome variable and predictors. By virtue of having more model fitting flexibility, Random Forest and XGBoost models are more sensitive to changes in training sample construction than regression models, which rely on simpler assumptions about the data.

In conclusion, the results in Figures A23-A28 show that it is possible for predictive models to have similarly high performance in smaller settings. What’s more, we do see that the choice of system-wide or institution-level model can significantly impact which students are targeted for intervention. However, we caution against too strong interpretation of these results, as we only tested one college-specific model; the same pattern of results may not hold for the other 22 VCCS colleges, or within a separate state system.

Appendix 7: Additional tests for NSC data

The context of our study is within the community college sector. Many community college students attend community colleges with the goal of transferring to a four-year school and earning a bachelor's degree, sometimes without earning a degree from the community college before doing so. Indeed, we see that 37.1 percent of our sample transfers to a four-year school, 72.8 percent of whom do not earn a degree from VCCS before transferring. For this reason, we include information from the NSC data both in constructing our graduation outcome and predictors which provide information on student academic experiences outside the VCCS. In this section, we test how excluding the NSC data may change the models' performance and behavior. We perform this test because, while NSC data is a fairly commonly used source, not all institutions currently have access to this data. NSC data can be costly to obtain and maintain; what is more, an institution may only be interested in the outcome of whether a student graduates from that institution. Therefore, we investigate how our results differ when removing the NSC data entirely from our modeling process.

The exclusion of the NSC data reduces the number of model predictors from 331 to 274; the exclusion also reduces the graduation rate of the sample from 34.1 percent to 23.7 percent. This reduction is due to no longer observing true graduation for students who earn a degree outside VCCS. As a result of this change in outcome variable specification, we expect three opposing forces in terms of the effect on model performance:

1. Increased model performance because the outcome is measuring a more specific and narrowly focused occurrence -- i.e., whether a student earns a degree from a VCCS college is likely to be less related to whether a student earns a degree from a non-VCCS college.

2. Decreased model performance due to additional noise in the model because we can't fully observe graduation. For example, suppose that the students with the best GPAs are the most likely to go on to earn degrees at non-VCCS institutions, without necessarily earning a VCCS degree before transferring. By re-assigning the outcome for these students from "graduated" to "did not graduate", the relationship between graduation and GPA would become distorted.
3. Decreased model performance due to removing the NSC predictors, which decreases the amount of information at the models' disposal.

Figure A29 shows that, in terms of c-statistics, all models that exclude all NSC data actually perform better than the base model. Specifically, the c-statistic increases by 2-3 percent across models, with the largest increase occurring for XGBoost; in fact, the XGBoost model that excludes all NSC data has an impressive c-statistic of 0.933. A practical concern regarding interpreting these c-statistic results is whether it's an "apple-to-apple" comparison, given that there are two different outcome definitions between the base model and the model with no NSC data. Therefore, we also compute the c-statistics for both model versions excluding students who earn a non-VCCS degree (but no VCCS degree) from the validation sample (results not displayed). The c-statistics are still higher for the no NSC data compared to the base model, although the differences are smaller than what we see in Figure A29.

In Figure A29, we also see increases in the non-graduation F1-scores for all models, with the increases ranging from 4-5 percent, with the largest increases for the CPH and tree-based models. These increases in F1-scores are expected, due to the reduced observed graduation rate with the exclusion of the NSC data. Put another way, if there are two models with a similar c-statistic but one has a lower outcome mean, then that model will mechanically have a higher non-graduation F1-score. Therefore, we conduct a separate test where we compute the F1-scores for many random subsamples of the validation set that have the same proportion of graduates and non-graduates, separately for the base model and the model excluding NSC data (results not displayed). Other than a very small decrease in F1-score for OLS, we find increases from the base model to the model with NSC predictors.

Figures A30 and A31 show the movement of students across the distribution of predicted scores from the base models to the no NSC data models. Figure A30 shows that between 32-41 percent of students move 11+ percentiles when NSC data is excluded from the model. Similarly, Figure A31 shows that 40-66 percent of students in the bottom decile in the base model move out of the bottom decile when we exclude NSC data. Similar to other results we discuss above, we see the greatest movement of students for the XGBoost model. In the context of this particular test, it is not very surprising to see these dramatic shifts in the predicted scores percentiles because the outcome of interest changed significantly. When we recreate Figures A30 and A31 using the alternative validation sample tests mentioned above for the c-statistic and non-graduation F1-score, we find very similar results (not displayed).

Next, we conduct two additional tests regarding the use of the NSC data. Our first test is to remove all of the predictors we constructed using the NSC data. While the information available in the NSC data is far more limited than what we observe in the VCCS administrative files, we still observe enrollment periods, enrollment intensity, and quality metrics of the non-VCCS institutions that students attend. Second, we test how the model performance differs when we drop all students who attended a non-VCCS college during their randomly truncated observation window. The purpose of this second test is to show how the models may differ when exclusively focused on students enrolled at VCCS. This exclusion results in a training sample size of n =

281,317 (94 percent of base model) and a validation sample size of n = 31,200 (94 percent of base model). Note that this does not exclude students who attended a non-VCCS institution outside of their observation window. Because none of the students in these samples attended a non-VCCS institution during their observation window, we exclude all non-VCCS predictors from this model as well. Figure A32 compares the c-statistics and non-graduation F1-scores between the base model and these two additional model variants. We observe very slight reductions (1 percent or less across all five models) in the c-statistics for both tests. For the No NSC Predictors test, the F1-score decreases slightly (1 percent or less); For the No NSC Enrollees test, the F1-score increases slightly (less than 1 percent).

Finally, Figures A33 through A36 show how the distribution of students' predicted scores changes between the base models and these two model variants. Despite the similarity in the model variants, we see that there is still non-trivial movement across predicted score deciles, particularly for Random Forest models. For the Random Forest models, nearly one quarter of students in the bottom decile in the base model are not in the bottom decile for the test variants.

Appendix 8: Details about back-of-the-envelope benefit-cost calculation

We find that using XGBoost instead of OLS to predict the probability of graduation correctly identifies an additional 2 percent of at-risk students who would not graduate in the absence of intervention. In order to make these calculations, we set the threshold delineating predicted graduates from non-graduates by forcing the graduation rate of the validation sample to be the same as the graduation rate of the training sample (34.1%). This alternate method for setting the threshold allows for a better comparison of non-graduation recall across models, since as we discussed above the recall values can be quite sensitive to the threshold set by optimizing the F1-score.

This estimate differs from the marginal increase of 681 students reported in Section III.C. for two reasons: 1) it is based on the size of the average VCCS college, while the latter is derived from all 33,000 students in the validation sample, and 2) it only includes the additional number of non-graduates correctly identified, while the latter captures both the marginal increase in correctly identified graduates and non-graduates.

Notes from Appendices:

1. If there are less than 2 prior terms this value will be missing. If there are exactly 2 prior terms then the trend in GPA will be the slope of the difference in term GPA between the two terms. If there're n prior terms (n>=3), we fit a linear regression (the x variable are simply t=1,2,3,…) using the most recent 3,4,5,…,n terms respectively, and use the slope of whichever linear regression line that has the highest R^2 as the value for trend in GPA. For instance, if there are 5 prior terms, we fit the slope of term GPA using terms 3-5, terms 2-5 and terms 1-5, and if terms 2-5 has the highest linear regression R^2, the slope of terms 2-5 will be the value of trend in GPA.

2. When constructing RNN models, for both VCCS and non-VCCS term-specific predictors, we organize the values of each predictor in the chronological order (each time stamp corresponds to the term in which a student was actively enrolled in VCCS or non-VCCS institutions). Therefore, unlike other models where a certain term-specific predictor in different terms will be considered as distinct predictors (e.g., term GPA in the 1st Sprint term is a different predictor from term GPA in the 2nd fall term), in RNN each term-specific predictor is treated as an object which consists of a sequence of its values at all of the different time steps.

3. We do not create term-specific for the non-VCCS institutional type and quality measures, as including these measures would significantly increase the total number of predictors within the model, and because few students have relevant data to contribute toward these term-specific predictors that differ meaningfully from time-invariant predictors. Specifically, only 6.9 percent of students in the training sample and 5.6 percent of students in the validation sample have ever attended non-VCCS institutions during their randomly truncated observation windows. Furthermore, among such students, only 7.0 percent (for training sample) and 5.9 percent (for

validation sample) of them have attended more than one non-VCCS institution during their randomly truncated observation windows, we do not create term-specific non-VCCS institutional type and quality measure predictors.

4. Grants, subsidized loans, unsubsidized loans and other aid predictors are constructed based on the academic year, as opposed to term. Also, logarithm transformations have been applied to those financial aid variables in order to prevent the extremely large values from making the model predictions unstable.

5. For CPH, the predicted outcome needs to be further processed in order to generate risk rankings, and the inclusion of the stratification variable increases the complexity of the model by one order of magnitude.

6. Although a 1-SE rule is more frequently used for feature selection, we employ a 2-SE rule to push the degree of model parsimony as high as possible without meaningfully losing predictive performance. Specifically, from an algorithmic perspective the penalized logistic regression with L1 regularization is a variant of the standard logistic regression that appends an additional term (a function of model coefficients) to the objective function of the standard model. By increasing the magnitude of this additional term (i.e., from 1-SE to 2-SE), it forces the optimal model solution to become increasingly sparse. Optimization is achieved by choosing the fewest number of predictors that maintain model performance within 2 standard errors of the best-performing model specification when no constraints are imposed on the predictor set. Model performance did not decrease with the 2-SE rule, which suggests that using the more aggressive rule is justified in this case.

7. We also tested a second feature selection method called cost complexity pruning of decision trees. However, we found that the cost complexity pruning did not perform as well as the penalized logistic regression with the 2-SE rule.

8. Specifically, we gradually increased the tuning parameter at each step and then recorded the corresponding number of selected predictors as well as the out-of-sample model performance.

9. We report the average c-statistic value from a 10-fold cross-validation process.

10. We do not build the RNN model for the PVCC-only training set primarily because RNN is very computationally intensive in terms of modeling fitting and fine tuning, and we did not find the performance of other models trained on smaller training samples have any significant difference from the base models. And we do not build the CPH model for the PVCC-only training set either, as the model fitting procedure runs into a convergence error with such a small training sample size and large number of predictors, which is likely caused by the variance of certain predictors being too small.

11. We do not build the RNN model for the randomly selected 10% of the training sample primarily because RNN is very computationally intensive in terms of modeling fitting and fine tuning, and we did not find the performance of other models trained on smaller training samples have any significant difference from the base models.

Appendix Table A1: Model Performance Under Different Sample & Predictor Construction Methods

	Simple non-term-specific predictors (14 total)		All non-term-specific predictors (61 total)		All non-term-specific predictors and simple term-specific predictors (185 total)	All term-specific and non-term-specific predictors (331 total)
	Truncated (1)	c-statistic (2)	Truncated (3)	c-statistic (4)	c-statistic (5)	c-statistic (6)
OLS	Yes	0.8105	Yes	0.8513	0.8773	0.8795
	No	0.7611	No	0.7597		
Logit	Yes	0.8183	Yes	0.8539	0.8816	0.8837
	No	0.7837	No	0.8038		
CPH	Yes	0.8368	Yes	0.8643	0.8794	0.881
	No	N/A	No	N/A		
RF	Yes	0.8516	Yes	0.8841	0.8892	0.8859
	No	0.8163	No	0.8474		
XGBoost	Yes	0.855	Yes	0.8906	0.9024	0.9032
	No	0.8162	No	0.8504		
RNN	Yes	N/A	Yes	N/A	0.8956	0.8959
	No	N/A	No	N/A		

Appendix Table A2: Cross-model correlation of risk rankings

Panel A: Pearson correlation coefficient

	OLS	Logit	CoxPH	RF	XGBoost
OLS					
Logit	0.9621				
CoxPH	0.9362	0.9736			
RF	0.9157	0.9386	0.9231		
XGBoost	0.899	0.946	0.9255	0.9518	
RNN	0.8922	0.9408	0.9193	0.9322	0.9607

Panel B: Spearman's coefficient of rank correlation

	OLS	Logit	CoxPH	RF	XGBoost
OLS					
Logit	0.9864				
CoxPH	0.957	0.9787			
RF	0.9289	0.9244	0.9151		
XGBoost	0.9322	0.9436	0.9238	0.9432	
RNN	0.9256	0.9365	0.9123	0.92	0.9576

Appendix Table A3: Share of top 20% important features in common, base models versus models excluding demographic predictors

Model	Share of predictors in common
OLS	0.939
Logit	0.924
RF	0.939
XGBoost	0.924

Appendix Table A4: Summary statistics of analytic sample

	Full sample (1)	Training sample (2)	Validation sample (3)
Panel A: Baseline Characteristics			
Age at first enrollment	24.64	24.64	24.64
White	55.56%	55.58%	55.38%
Black	25.71%	25.67%	26.16%
Hispanic	7.85%	7.85%	7.85%
Other	8.73%	8.75%	8.52%
Male	44.61%	44.57%	44.95%
Female	55.39%	55.43%	55.05%
First Generation	19.52%	19.53%	19.47%
Not First Generation	37.52%	37.48%	37.81%
Missing Parental education	42.96%	42.99%	42.72%
Panel B: Academic outcomes, within six years of initial VCCS enrollment			
Ever enrolled at non-VCCS?	37.09%	37.10%	36.97%
Total terms enrolled			
VCCS only	4.7	4.71	4.69
VCCS or non-VCCS	6.56	6.56	6.54
Non-VCCS only	1.94	1.94	1.94
Earned credential?			
VCCS only	17.75%	17.75%	17.73%
Non-VCCS only	10.44%	10.45%	10.37%
VCCS and non-VCCS	6.01%	6.01%	6.04%
Time to credential (# terms)			
Mean	9.5	9.5	9.52
Standard Deviation	4.11	4.12	4.1
N	331,254	298,139	33,115

Notes: total terms enrolled can include up to three terms in a calendar year: Spring, Summer, and Fall. For non-VCCS enrollment, we use the enrollment beginning and end dates for each enrollment records to determine whether the student was enrolled in a given Spring, Summer, or Fall term. We consider all levels of postsecondary credentials, including diplomas, short- and long-term certificates, Associate degrees, Bachelor's degrees, and graduate degrees when determining the outcome of credential completion.

Appendix Table A5: Predictor importance measures for each base model

Notes: The ranking values in columns (1) and (2) are based on recursive feature elimination (RFE), described in more detail in the text. A lower value in columns (1) and (2) indicate a more important feature, with the feature ranked as "1" being the most important. The feature importance values in columns (3) and (4) are also described in more detail in the text. A higher value in columns (3) and (4) indicate a more important feature.

Predictor	OLS (ranking)	Logistic (ranking)	Random Forest (feature importance)	XGBoost (feature importance)
Weighted average of the 1st quartiles of SAT verbal scores of all non-VCCS institutions attended	1	200	0.002037	0.001634
Weighted average of the 3rd quartiles of SAT verbal scores of all non-VCCS institutions attended	2	37	0.0023394	0.0013617
Number of cumulative college-level credit hours earned prior to initial enrollment at VCCS	3	1	0.0053223	0.008382
Number of credit hours attempted in the 2nd summer term	4	2	0.0065079	0.0065135
Weighted average of the 3rd quartiles of SAT math scores of all non-VCCS institutions attended	5	97	0.0023282	0.0011423
Number of credit hours attempted in the 2nd fall term	6	6	0.0197854	0.0137607
Standard deviation of term-level proportion of earned credits among attempted credits since initial enrollment term	7	13	0.0543757	0.0199186
Number of credit hours attempted in the 1st summer term	8	15	0.0118768	0.0120132
Number of terms in which student was enrolled in non-VCCS institutions prior to initial enrollment term	9	7	0.0169086	0.0100085
Number of credit hours attempted in the 6th fall term	10	5	0.0007055	0.0015584
Slope of term-level number of credits attempted through the end of observation window	11	14	0.0082708	0.0359185
Number of credit hours attempted in the 1st fall term	12	9	0.0182714	0.0200245
Cumulative GPA through the end of observation window	13	4	0.0719979	0.0393455

Number of credit hours attempted in the 6th summer term	14	54	8.64E-05	0.0004615
Slope of term GPA through the end of observation window	15	12	0.0201959	0.040518
Age at initial enrollment at VCCS	16	16	0.0107476	0.0293975
Indicator for whether student repeated a previously taken course in the 2nd fall term	17	31	0.0009596	0.0013239
Indicator for whether student repeated a previously taken course in the 1st spring term	18	25	0.0006474	0.0007716
Standard deviation of term-level proportion of withdrawn credits among attempted credits since initial enrollment term	19	27	0.0265366	0.0210836
Indicator for whether student is actively enrolled in VCCS in the 6th summer term	20	92	4.89E-05	0
Total enrollment intensity in non-VCCS institutions in the 1st spring term	21	8	0.0017681	0.0023678
Indicator for whether student is actively enrolled in VCCS in the 6th fall term	22	45	0.0004651	6.81E-05
Number of credit hours attempted in the 1st spring term	23	26	0.021539	0.0169153
Total enrollment intensity in non-VCCS institutions in the 4th fall term	24	52	0.0001913	0.0010894
Indicator for whether student is actively enrolled in non-VCCS institutions in the 4th fall term	25	89	0.0001277	0.0001286
Indicator for whether student repeated a previously taken course in the 2nd summer term	26	68	0.0004559	0.0007035
Number of credit hours attempted in the 3rd fall term	27	28	0.0063712	0.0103413
Indicator for whether student is actively enrolled in non-VCCS institutions in the 5th spring term	28	86	4.26E-05	4.54E-05
Total enrollment intensity in non-VCCS institutions in the 5th spring term	29	85	6.76E-05	0.000295
Indicator for whether student repeated a previously taken course in the 1st fall term	30	72	0.0004031	0.0004463
Indicator for whether student is actively enrolled in VCCS in the 3rd fall term	31	34	0.000881	0.0002118

Term GPA in the 3rd fall term	32	112	0.0065446	0.0089645
Overall proportion of attempted credits of developmental courses	33	19	0.0223806	0.0248056
Logarithm of other aids received in year 6	34	79	7.38E-06	1.51E-05
Indicator for data availability in the 6th spring term	35	11	0.0004567	0.0010364
Total enrollment intensity in non-VCCS institutions in the 1st fall term	36	38	0.0012021	0.0021258
Proportion of withdrawn credits among attempted credits the 4th fall term	37	32	0.001078	0.0030109
Number of credit hours attempted in the 4th fall term	38	33	0.0028362	0.0071716
Indicator for whether student is actively enrolled in VCCS in the 2nd fall term	39	61	0.0052024	0.000522
Term GPA in the 2nd fall term	40	62	0.0242318	0.0122099
Overall proportion of withdrawn credits among attempted credits since initial enrollment term	41	10	0.0276241	0.0278921
Indicator for whether student is in degree-seeking status in the 6th summer term	42	109	5.00E-05	0
Proportion of attempted credits of developmental courses in the 6th spring term	43	41	2.35E-05	6.81E-05
Number of credit hours attempted in the 4th spring term	44	64	0.0021387	0.005825
Indicator for whether student is actively enrolled in VCCS in the 4th spring term	45	60	0.0006149	0.0002799
Proportion of earned credits among attempted credits in the 4th spring term	46	59	0.0010921	0.0023678
Number of credit hours attempted in the 3rd spring term	47	46	0.0040111	0.008261
Indicator for whether student is actively enrolled in VCCS in the 3rd spring term	48	47	0.000758	0.0002723
Proportion of earned credits among attempted credits in the 3rd spring term	49	48	0.0023129	0.0032
Total enrollment intensity in non-VCCS institutions in the 6th spring term	50	70	4.48E-06	2.27E-05

Indicator for whether student is actively enrolled in non-VCCS institutions in the 1st summer term	51	84	0.0001978	8.32E-05
Indicator for whether student is actively enrolled in non-VCCS institutions in the 6th summer term	52	230	3.89E-06	0
Total enrollment intensity in non-VCCS institutions in the 6th summer term	53	231	3.35E-06	0
Total enrollment intensity in non-VCCS institutions in the 3rd spring term	54	93	0.0003308	0.0018837
Number of terms in which student was enrolled in non-VCCS institutions since initial enrollment term	55	291	0.0024128	0.0030865
Total enrollment intensity in non-VCCS institutions in the 2nd spring term	56	81	0.0005321	0.0015735
Total enrollment intensity in non-VCCS institutions in the 3rd summer term	57	154	0.0001143	0.0002723
Total enrollment intensity in non-VCCS institutions in the 5th summer term	58	56	3.05E-05	0.000174
Indicator for data availability in the 4th fall term	59	51	0.0021805	0.0027915
Number of credit hours attempted in the 5th summer term	60	75	0.0005323	0.0016492
Term GPA in the 5th summer term	61	74	0.0004642	0.0023905
Indicator for whether student is actively enrolled in VCCS in the 5th summer term	62	76	0.0002217	0.0001135
Indicator for data availability in the 6th summer term	63	55	6.25E-05	0.0001059
Cumulative GPA prior to initial enrollment term at VCCS	64	98	0.0080178	0.0072246
Indicator for whether student was ever enrolled in VCCS prior to initial enrollment term	65	99	0.0020839	0.0002194
Proportion of earned credits among attempted credits in the 4th fall term	66	77	0.001509	0.0032076
Indicator for whether student is actively enrolled in VCCS in the 4th fall term	67	78	0.000694	0.0002269
Proportion of withdrawn credits among attempted credits the 3rd fall term	68	73	0.0017939	0.0038808

Indicator for two-year, private, in-state	69	42	4.22E-05	0.0001664
Term GPA in the 4th summer term	70	175	0.0012517	0.0033664
Proportion of withdrawn credits among attempted credits the 2nd fall term	71	88	0.0030399	0.0038354
Proportion of withdrawn credits among attempted credits the 3rd spring term	72	80	0.0013158	0.0027385
Indicator for whether student is actively enrolled in VCCS in the 1st fall term	73	115	0.0013982	0.0005144
Term GPA in the 1st fall term	74	114	0.027933	0.0161739
Number of credit hours attempted in the 5th fall term	75	21	0.0013478	0.0043272
Indicator for whether student is actively enrolled in VCCS in the 5th fall term	76	22	0.0005746	0.0001513
Proportion of earned credits among attempted credits in the 5th fall term	77	23	0.0007349	0.0019518
Term GPA in the 1st summer term	78	171	0.0156963	0.0081097
Number of credit hours attempted in the 2nd spring term	79	82	0.0100978	0.0119602
Total enrollment intensity in non-VCCS institutions in the 5th fall term	80	110	6.34E-05	0.0003253
Indicator for whether student is actively enrolled in non-VCCS institutions in the 5th fall term	81	111	5.68E-05	3.03E-05
Indicator for whether student repeated a previously taken course in the 6th summer term	82	63	1.01E-05	0
Term GPA in the 6th summer term	83	66	7.46E-05	0.0001664
Proportion of withdrawn credits among attempted credits the 4th spring term	84	69	0.0006341	0.0022317
Indicator for whether student is actively enrolled in VCCS in the 2nd spring term	85	129	0.0022544	0.0005295
Term GPA in the 2nd spring term	86	225	0.0101114	0.0101068
Proportion of earned credits among attempted credits in the 3rd fall term	87	35	0.003778	0.0041456

Indicator for whether student repeated a previously taken course in the 3rd summer term	88	210	0.0003216	0.0007716
Indicator for whether student is actively enrolled in VCCS in the 1st spring term	89	116	0.0018176	0.000643
Term GPA in the 1st spring term	90	117	0.0233063	0.0146836
Indicator for whether student repeated a previously taken course in the 5th fall term	91	39	0.0002442	0.0007338
Proportion of withdrawn credits among attempted credits the 5th fall term	92	44	0.000363	0.0013995
Indicator for whether student is actively enrolled in non-VCCS institutions in the 6th spring term	93	321	4.09E-06	0
Number of credit hours attempted in the 5th spring term	94	29	0.0011696	0.0036766
Indicator for whether student is actively enrolled in VCCS in the 5th spring term	95	50	0.0005534	0.0001664
Proportion of earned credits among attempted credits in the 5th spring term	96	49	0.0006023	0.0013239
Indicator for whether student is actively enrolled in non-VCCS institutions in the 4th summer term	97	90	6.82E-05	0.0001135
Indicator for dual enrollment prior to initial enrollment term	98	100	0.0035016	0.0007111
Term GPA in the 4th fall term	99	118	0.0025024	0.006748
Proportion of withdrawn credits among attempted credits the 3rd summer term	100	155	0.000343	0.0008321
Proportion of earned credits among attempted credits in the 3rd summer term	101	135	0.0018354	0.0022922
Indicator for whether student repeated a previously taken course in the 6th fall term	102	162	6.60E-05	0.0002118
Proportion of withdrawn credits among attempted credits the 5th spring term	103	65	0.0002271	0.0011045
Proportion of attempted credits of developmental courses in the 2nd fall term	104	157	0.0023684	0.0030714

Indicator for whether student is actively enrolled in non-VCCS institutions in the 3rd summer term	105	153	8.76E-05	6.05E-05
Total enrollment intensity in non-VCCS institutions in the 2nd fall term	106	130	0.0005256	0.001861
Indicator for data availability in the 1st fall term	107	67	0.0017427	0.002678
Term GPA in the 2nd summer term	108	259	0.0081105	0.0049324
Weighted average of the 1st quartiles of SAT math scores of all non-VCCS institutions attended	109	36	0.0024611	0.0015811
Term GPA in the 6th spring term	110	147	0.0003894	0.0005674
Number of credit hours attempted in the 6th spring term	111	3	0.0003717	0.0009078
Number of credit hours attempted in the 3rd summer term	112	133	0.0021506	0.0048189
Indicator for whether student is actively enrolled in VCCS in the 3rd summer term	113	134	0.0009306	0.0003782
Indicator for Pell-eligible the 3rd summer term	114	120	0.0003491	0.0011726
Indicator for whether student repeated a previously taken course in the 2nd spring term	115	172	0.0007519	0.0013466
Proportion of attempted credits of developmental courses in the 1st summer term	116	185	0.0020621	0.002852
Proportion of withdrawn credits among attempted credits the 2nd spring term	117	113	0.0021369	0.0033589
Indicator for not Pell-eligible the 6th summer term	118	149	1.79E-05	6.81E-05
Number of credit hours attempted in the 4th summer term	119	20	0.0011096	0.0030789
Indicator for whether student is actively enrolled in VCCS in the 4th summer term	120	192	0.0005391	0.0003177
Indicator for Pell-eligible the 4th summer term	121	121	0.0001843	0.0007262
Indicator for Pell-eligible the 4th spring term	122	122	0.0002807	0.0004842
Proportion of withdrawn credits among attempted credits the 2nd summer term	123	184	0.000612	0.0012331
Proportion of withdrawn credits among attempted credits the 5th summer term	124	102	5.50E-05	0.0003177

Total enrollment intensity in non-VCCS institutions in the 3rd fall term	125	176	0.0003064	0.001218
Indicator for data availability in the 3rd fall term	126	94	0.0016701	0.0032908
Proportion of earned credits among attempted credits in the 4th summer term	127	139	0.0007974	0.0016113
Term GPA in the 3rd spring term	128	170	0.003931	0.0079281
Indicator for data availability in the 5th fall term	129	91	0.0022776	0.0026704
Proportion of withdrawn credits among attempted credits the 4th summer term	130	151	0.0001683	0.0004993
Overall proportion of attempted credits of 2XX level courses	131	146	0.0221612	0.0276424
Indicator for whether student repeated a previously taken course in the 5th spring term	132	106	0.0001527	0.0003102
Indicator for not Pell-eligible the 5th summer term	133	108	0.0001338	0.0007262
Indicator for whether student repeated a previously taken course in the 4th fall term	134	220	0.0004302	0.0008548
Number of non-VCCS institutions in which student was enrolled prior to initial enrollment term	135	96	0.0155781	0.0054014
Indicator for not a seamless enrollee	136	216	0.0039405	0.0015357
Total enrollment intensity in non-VCCS institutions in the 2nd summer term	137	255	0.0001601	0.0002572
Indicator for data availability in the 1st summer term	138	132	0.005825	0.0019896
Indicator for data availability in the 1st spring term	139	83	0.001968	0.0017702
Term GPA in the 3rd summer term	140	95	0.0025241	0.004017
Indicator for whether student is actively enrolled in non-VCCS institutions in the 2nd summer term	141	256	0.0001229	7.56E-05
Indicator for whether student is actively enrolled in non-VCCS institutions in the 1st fall term	142	199	0.0010606	0.0003026
Total enrollment intensity in non-VCCS institutions in the 4th summer term	143	143	8.60E-05	0.0004009
Indicator for whether student is in degree-seeking status in the 5th spring term	144	103	0.0005411	7.56E-05

Overall proportion of earned credits among attempted credits since initial enrollment term	145	43	0.0733521	0.0239205
Term GPA in the 5th fall term	146	123	0.0011111	0.0041154
Proportion of earned credits among attempted credits in the 1st spring term	147	211	0.0144905	0.0059612
Proportion of earned credits among attempted credits in the 1st fall term	148	223	0.0156395	0.006695
Indicator for whether student repeated a previously taken course in the 6th spring term	149	71	2.86E-05	0.0001135
Indicator for whether student is in degree-seeking status in the 6th fall term	150	87	0.0004677	2.27E-05
Proportion of attempted credits of 2XX level courses in the 5th spring term	151	159	0.0004565	0.0014449
Proportion of attempted credits of developmental courses in the 2nd spring term	152	101	0.001675	0.002973
Proportion of withdrawn credits among attempted credits the 6th spring term	153	53	1.89E-05	0.0003177
Proportion of attempted credits of 2XX level courses in the 6th spring term	154	310	0.00014	0.0002875
Proportion of attempted credits of 2XX level courses in the 2nd spring term	155	173	0.0038408	0.0067858
Indicator for whether student repeated a previously taken course in the 3rd spring term	156	145	0.0005605	0.0009456
Indicator for two-year, public, out-of-state	157	183	0.0001493	0.0002496
Total enrollment intensity in non-VCCS institutions in the 1st summer term	158	275	0.0003255	0.0004917
Overall proportion of earned credits among attempted credits prior to initial enrollment term	159	314	0.0034647	0.0011801
Negative of logarithm of the maximum proportion of cumulative credits attempted at one VCCS institution	160	150	0.0017449	0.0064832

Proportion of attempted credits of 2XX level courses in the 1st fall term	161	186	0.004562	0.009078
Proportion of attempted credits of developmental courses in the 1st fall term	162	187	0.0052849	0.0082307
Total enrollment intensity in non-VCCS institutions in the 4th spring term	163	57	0.0001532	0.0008624
Indicator for whether student is actively enrolled in non-VCCS institutions in the 4th spring term	164	58	0.0001089	8.32E-05
Weighted average of the 3rd quartiles of SAT writing scores of all non-VCCS institutions attended	165	160	0.0022397	0.0007716
Indicator for two-year, public, in-state	166	189	9.36E-06	0
Weighted average of the 1st quartiles of SAT writing scores of all non-VCCS institutions attended	167	161	0.0022795	0.0011121
Indicator for four-year, public, out-of-state	168	202	0.000166	8.32E-05
Indicator for four-year, public, in-state	169	201	0.0007089	0.0003253
Number of non-VCCS institutions in which student was enrolled since initial enrollment term	170	203	0.0018355	0.0003102
Indicator for four-year, private, out-of-state	171	204	0.0002095	0.0002799
Indicator for Pell-eligible the 5th spring term	172	166	0.0001631	0.0001513
Logarithm of total grants received in year 5	173	167	0.0007315	0.0027007
Indicator for Pell-eligible the 5th fall term	174	168	0.0001931	0.0001816
Proportion of attempted credits of developmental courses in the 6th fall term	175	224	5.36E-05	0.0004615
Proportion of attempted credits of 2XX level courses in the 1st spring term	176	198	0.004517	0.0087905
Indicator for whether student is actively enrolled in VCCS in the 1st summer term	177	297	0.0073172	0.0010742
Indicator for data availability in the 2nd fall term	178	131	0.0018638	0.0032227
Logarithm of subsidized loans received in year 6	179	104	7.71E-05	0.0004615
Indicator for whether student repeated a previously taken course in the 3rd fall term	180	222	0.0006736	0.0013012

Proportion of earned credits among attempted credits in the 6th fall term	181	24	0.0003878	0.000643
Term GPA in the 6th fall term	182	163	0.0005399	0.0011953
Proportion of attempted credits of developmental courses in the 2nd summer term	183	251	0.0006896	0.0014298
Indicator for whether student is in degree-seeking status in the 5th summer term	184	254	0.0002473	8.32E-05
Logarithm of other aids received in year 5	185	174	2.35E-05	7.56E-06
Indicator for not Pell-eligible the 4th summer term	186	191	0.0002688	0.0005901
Indicator for whether student is in degree-seeking status in the 4th summer term	187	142	0.000497	0.0002269
Proportion of attempted credits of 2XX level courses in the 4th fall term	188	197	0.0011552	0.0035707
Indicator for male	189	180	0.0015069	0.0052879
Indicator for whether student is in degree-seeking status in the 3rd spring term	190	188	0.0007852	0.0003177
Indicator for Pell-eligible the 2nd summer term	191	193	0.0005147	0.0014071
Indicator for Pell-eligible the 1st spring term	192	194	0.0011306	0.0011121
Proportion of attempted credits of 2XX level courses in the 3rd spring term	193	219	0.0018173	0.0045844
Proportion of attempted credits of developmental courses in the 1st spring term	194	212	0.0042696	0.0067631
Indicator for two-year, private, out-of-state	195	266	1.06E-05	0
Indicator for not Pell-eligible the 6th fall term	196	306	3.44E-05	1.51E-05
Proportion of attempted credits of developmental courses in the 4th summer term	197	141	0.0001668	0.0005295
Indicator for whether student repeated a previously taken course in the 4th summer term	198	190	0.0002039	0.0003177
Indicator for not Pell-eligible the 2nd summer term	199	215	0.0005181	0.0011499
Indicator for Pell-eligible the 3rd spring term	200	214	0.0004002	0.000469
Indicator for Pell-eligible the 5th summer term	201	127	6.00E-05	0.0002043

Indicator for highest parental education being having earned Post-Bachelor's degree	202	217	0.0009557	0.0014525
Proportion of attempted credits of developmental courses in the 5th summer term	203	312	6.57E-05	0.0001891
Proportion of attempted credits of 2XX level courses in the 5th summer term	204	181	0.0002573	0.0006279
Proportion of earned credits among attempted credits in the 2nd summer term	205	137	0.0057481	0.0021258
Indicator for whether student changed degree/major program pursued	206	209	0.001304	0.0032908
Indicator for whether student was ever enrolled in any non-VCCS institutions since initial enrollment term	207	234	0.0016352	0
Weighted average of admission rates of all non-VCCS institutions attended	208	105	0.0029963	0.0041229
Proportion of attempted credits of 2XX level courses in the 4th spring term	209	243	0.0009676	0.0033135
Proportion of attempted credits of 2XX level courses in the 6th summer term	210	233	3.19E-05	6.05E-05
Indicator for data availability in the 5th summer term	211	182	0.0017042	0.0016038
Indicator for whether student is in degree-seeking status in the 5th fall term	212	169	0.0006306	0.0001362
Term GPA in the 4th spring term	213	318	0.0017909	0.0048567
Logarithm of unsubsidized loans received in year 1	214	232	0.0015781	0.0053182
Proportion of withdrawn credits among attempted credits the 6th fall term	215	18	8.66E-05	0.0006203
Indicator for seamless enrollee	216	119	0.0033386	0.001634
Indicator for whether student is actively enrolled in non-VCCS institutions in the 2nd fall term	217	156	0.0003168	0.0002345
Indicator for whether student is actively enrolled in non-VCCS institutions in the 3rd fall term	218	177	0.0002051	0.0002269
Indicator for data availability in the 3rd spring term	219	179	0.0019378	0.0036539

Indicator for whether student is actively enrolled in VCCS in the 6th spring term	220	138	0.0003464	0.0001286
Proportion of attempted credits of 2XX level courses in the 2nd fall term	221	229	0.0100912	0.0068766
Indicator for not Pell-eligible the 5th spring term	222	206	0.0001008	0.0001589
Logarithm of total grants received in year 4	223	213	0.0014898	0.0050005
Proportion of withdrawn credits among attempted credits the 1st summer term	224	221	0.0010582	0.00233
Indicator for Pell-eligible the 6th summer term	225	126	6.04E-06	0
Indicator for data availability in the 4th spring term	226	148	0.0024147	0.003321
Proportion of attempted credits of developmental courses in the 3rd spring term	227	196	0.0008274	0.0016567
Logarithm of other aids received in year 1	228	195	0.0003256	0.0010818
Proportion of earned credits among attempted credits in the 6th summer term	229	152	5.21E-05	1.51E-05
Logarithm of unsubsidized loans received in year 6	230	205	6.96E-05	0.0001589
Indicator for not Pell-eligible the 6th spring term	231	311	1.59E-05	4.54E-05
Indicator for Pell-eligible the 6th spring term	232	125	6.30E-05	0.0001362
Indicator for whether student repeated a previously taken course in the 4th spring term	233	305	0.0003466	0.0007414
Indicator for not Pell-eligible the 3rd spring term	234	227	0.0002985	0.0002043
Indicator for never Pell-eligible	235	236	0.0009911	0.0012785
Logarithm of unsubsidized loans received in year 2	236	237	0.0011712	0.0034799
Logarithm of subsidized loans received in year 2	237	238	0.0012819	0.0041456
Indicator for data availability in the 5th spring term	238	30	0.0020158	0.0026856
Indicator for data availability in the 2nd spring term	239	207	0.0013975	0.0029882
Indicator for highest parental education being having earned Bachelor's degree	240	245	0.0008788	0.0015735
Logarithm of total grants received in year 3	241	226	0.0024607	0.0074288
Indicator for Pell-eligible the 3rd fall term	242	228	0.0004519	0.0005069

Indicator for whether student is in degree-seeking status in the 3rd fall term	243	262	0.001308	0.0003858
Indicator for whether student is in degree-seeking status in the 1st fall term	244	241	0.0019667	0.00146
Logarithm of subsidized loans received in year 5	245	252	0.0002841	0.0012255
Logarithm of unsubsidized loans received in year 5	246	270	0.0002444	0.0008548
Indicator for data availability in the 6th fall term	247	17	0.0007152	0.0009532
Indicator for whether student is actively enrolled in VCCS in the 2nd summer term	248	296	0.0054352	0.0005749
Indicator for data availability in the 2nd summer term	249	264	0.0016355	0.0023527
Indicator for Pell-eligible the 1st summer term	250	235	0.0007488	0.0013239
Indicator for whether student is in degree-seeking status in the 6th spring term	251	165	0.0002707	6.81E-05
Proportion of attempted credits of 2XX level courses in the 5th fall term	252	317	0.000583	0.001861
Proportion of withdrawn credits among attempted credits the 1st fall term	253	286	0.0043747	0.0046752
Indicator for whether student is actively enrolled in non-VCCS institutions in the 3rd spring term	254	208	0.0001982	0.0002875
Indicator for whether student is actively enrolled in non-VCCS institutions in the 1st spring term	255	246	0.0010923	0.0004842
Indicator for whether student is actively enrolled in non-VCCS institutions in the 6th fall term	256	292	6.92E-06	0
Indicator for whether student is actively enrolled in non-VCCS institutions in the 2nd spring term	257	324	0.0003221	0.0003556
Proportion of attempted credits of developmental courses in the 3rd summer term	258	307	0.0003479	0.0006884
Proportion of earned credits among attempted credits in the 5th summer term	259	124	0.0003026	0.00087
Proportion of attempted credits of developmental courses in the 5th fall term	260	178	0.0002392	0.0006808

Proportion of attempted credits of developmental courses in the 3rd fall term	261	218	0.0010236	0.0016189
Proportion of attempted credits of 2XX level courses in the 3rd fall term	262	290	0.0025481	0.0048643
Indicator for not Pell-eligible the 2nd fall term	263	263	0.0005337	0.0004463
Indicator for not Pell-eligible the 2nd spring term	264	250	0.0004433	0.0004615
Indicator for whether student is in degree-seeking status in the 2nd spring term	265	282	0.0012761	0.0005295
Indicator for Pell-eligible the 2nd spring term	266	249	0.0005859	0.0003707
Logarithm of total grants received in year 2	267	248	0.0039298	0.0094941
Proportion of earned credits among attempted credits in the 2nd fall term	268	144	0.0206284	0.00466
Indicator for data availability in the 3rd summer term	269	277	0.0013881	0.0019896
Indicator for not Pell-eligible the 3rd summer term	270	300	0.000352	0.0006203
Indicator for highest parental education being having attended college	271	268	0.0006914	0.0013012
Indicator for highest parental education being having graduated from high school	272	267	0.0008182	0.0016113
Proportion of withdrawn credits among attempted credits the 6th summer term	273	107	5.20E-06	7.56E-06
Logarithm of total grants received in year 6	274	164	0.0003101	0.0011801
Indicator for Pell-eligible the 1st fall term	275	273	0.0011669	0.000817
Indicator for whether student is in degree-seeking status in the 1st spring term	276	272	0.001664	0.0011423
Indicator for whether student is in degree-seeking status in the 2nd summer term	277	271	0.0035506	0.0009759
Proportion of earned credits among attempted credits in the 2nd spring term	278	128	0.0086272	0.0036161
Logarithm of subsidized loans received in year 4	279	239	0.0006203	0.0024208
Logarithm of unsubsidized loans received in year 4	280	240	0.0005179	0.0014903
Indicator for not Pell-eligible the 4th spring term	281	260	0.0001829	0.0002269

Proportion of attempted credits of developmental courses in the 6th summer term	282	319	3.72E-06	0
Indicator for whether student is in degree-seeking status in the 3rd summer term	283	258	0.0008817	0.000469
Indicator for White	284	274	0.0012777	0.0046525
Indicator for whether student has ever repeated a previously taken course	285	302	0.001666	0.0020425
Indicator for not Pell-eligible the 3rd fall term	286	284	0.0003328	0.0003102
Indicator for Pell-eligible the 6th fall term	287	265	7.74E-05	3.03E-05
Indicator for not Pell-eligible the 5th fall term	288	247	0.0001185	0.0002345
Indicator for African American	289	280	0.0017113	0.0038657
Indicator for other race/ethnicity	290	279	0.0007103	0.0015206
Indicator for whether student is in degree-seeking status in the 4th spring term	291	244	0.0005467	0.000174
Proportion of attempted credits of 2XX level courses in the 6th fall term	292	158	0.0002506	0.0006128
Logarithm of other aids received in year 4	293	278	3.86E-05	0.0001664
Logarithm of other aids received in year 3	294	285	0.0001018	0.0003404
Proportion of attempted credits of developmental courses in the 4th fall term	295	299	0.0005448	0.0010591
Proportion of attempted credits of developmental courses in the 5th spring term	296	276	0.0001535	0.0004161
Indicator for whether student is in degree-seeking status in the 4th fall term	297	253	0.0006222	0.0002194
Proportion of withdrawn credits among attempted credits the 1st spring term	298	327	0.0035524	0.0046222
Proportion of attempted credits of developmental courses in the 4th spring term	299	242	0.0003484	0.0009381
Indicator for not Pell-eligible the 1st spring term	300	261	0.0007035	0.0006355
Indicator for not Pell-eligible the 1st fall term	301	294	0.000816	0.0007338

Indicator for whether student is actively enrolled in non-VCCS institutions in the 5th summer term	302	288	2.98E-05	1.51E-05
Term GPA in the 5th spring term	303	257	0.0008146	0.0027083
Logarithm of other aids received in year 2	304	323	0.0002382	0.000469
Indicator for Pell-eligible the 4th fall term	305	281	0.0003272	0.0002799
Indicator for Pell-eligible the 2nd fall term	306	295	0.0007671	0.0004009
Logarithm of total grants received in year 1	307	293	0.0047828	0.0186099
Indicator for whether student repeated a previously taken course in the 1st summer term	308	301	0.0004583	0.0006052
Indicator for not Pell-eligible the 1st summer term	309	289	0.0008587	0.0011574
Logarithm of subsidized loans received in year 1	310	298	0.0017186	0.0056662
Indicator for whether student is in degree-seeking status in the 2nd fall term	311	322	0.0047448	0.0007641
Proportion of attempted credits of 2XX level courses in the 2nd summer term	312	303	0.0020093	0.0026629
Proportion of attempted credits of 2XX level courses in the 3rd summer term	313	308	0.0011554	0.0021182
Total enrollment intensity in non-VCCS institutions in the 6th fall term	314	140	9.76E-06	7.56E-06
Indicator for whether student is in degree-seeking status in the 1st summer term	315	325	0.0057032	0.0019291
Logarithm of unsubsidized loans received in year 3	316	329	0.0008147	0.0021712
Logarithm of subsidized loans received in year 3	317	315	0.0008846	0.0024435
Proportion of attempted credits of 2XX level courses in the 1st summer term	318	313	0.0048933	0.0048567
Proportion of attempted credits of 2XX level courses in the 4th summer term	319	287	0.0005797	0.0011877
Indicator for highest parental education being having earned Associate's degree	320	326	0.0005131	0.0006279
Proportion of earned credits among attempted credits in the 1st summer term	321	269	0.0103658	0.0036766

Proportion of earned credits among attempted credits in the 6th spring term	322	40	0.0002688	0.0002648
Indicator for highest parental education being having attended high school	323	316	0.0003556	0.0003102
Indicator for highest parental education being less than high school	324	309	0.0002726	0.0001816
Indicator for ever Pell-eligible	325	304	0.0016752	0.0027915
Weighted average of graduation rates of all non-VCCS institutions attended	326	331	0.0037802	0.0085106
Indicator for whether student repeated a previously taken course in the 5th summer term	327	136	8.61E-05	0.000174
Indicator for four-year, private, in-state	328	283	0.0002466	0.0005674
Indicator for data availability in the 4th summer term	329	330	0.001463	0.0015357
Indicator for Hispanic	330	320	0.000576	0.000991
Indicator for not Pell-eligible the 4th fall term	331	328	0.0002346	0.0002269

Appendix Table A6: Coefficient estimates from base OLS and Logistic models

Predictor	OLS	Logit
Weighted average of admission rates of all non-VCCS institutions attended	-0.022	-0.2076
	(0.0249)	(0.1849)
Indicator for African American	-.0080*	-0.0399
	(0.0048)	(0.0359)
Age at initial enrollment at VCCS	-.0021***	-.0128***
	0.0000	(0.0006)
Indicator for data availability in the 1st fall term	-.0613***	-.5191***
	(0.0041)	(0.0336)
Indicator for data availability in the 2nd fall term	-.0345***	-.3099***
	(0.0042)	(0.0322)
Indicator for data availability in the 3rd fall term	-.0487***	-.4269***
	(0.0052)	(0.0402)
Indicator for data availability in the 4th fall term	-.0422***	-.3992***
	(0.0067)	(0.0552)
Indicator for data availability in the 5th fall term	-.0400***	-.5005***
	(0.0094)	(0.0919)
Indicator for data availability in the 6th fall term	-0.0167	-1.539***
	(0.0243)	(0.5592)
Indicator for data availability in the 1st spring term	-.0511***	-.4925***
	(0.0038)	(0.0312)
Indicator for data availability in the 2nd spring term	-.0241***	-.2309***
	(0.0043)	(0.0330)
Indicator for data availability in the 3rd spring term	-.0261***	-.2614***
	(0.0055)	(0.0425)
Indicator for data availability in the 4th spring term	-.0193***	-.3230***
	(0.0072)	(0.0590)
Indicator for data availability in the 5th spring term	-0.0164	-.3829***
	(0.0101)	(0.0989)
Indicator for data availability in the 6th spring term	-.0827**	-3.307***
	(0.0416)	(1.1050)
Indicator for data availability in the 1st summer term	.0662***	.2687***
	(0.0027)	(0.0205)
Indicator for data availability in the 2nd summer term	.0174***	.0649**
	(0.0037)	(0.0270)
Indicator for data availability in the 3rd summer term	.0108**	0.05
	(0.0049)	(0.0363)
Indicator for data availability in the 4th summer term	-0.001	0.0041
	(0.0064)	(0.0496)
Indicator for data availability in the 5th summer term	-.0159*	-.1890**

	(0.0090)	(0.0787)
Indicator for data availability in the 6th summer term	-0.0457	-.9738**
	(0.0342)	(0.4448)
Number of cumulative college-level credit hours earned prior to initial enrollment at VCCS	.1104***	.6374***
	(0.0046)	(0.0317)
Negative of logarithm of the maximum proportion of cumulative credits attempted at one VCCS institution	-.0455***	-.3823***
	(0.0069)	(0.0542)
Cumulative GPA through the end of observation window	.0282***	.3257***
	(0.0012)	(0.0111)
Cumulative GPA prior to initial enrollment term at VCCS	.0437***	.2082***
	(0.0023)	(0.0199)
Indicator for whether student is in degree-seeking status in the 1st fall term	-.0119***	-.0828**
	(0.0044)	(0.0321)
Indicator for whether student is in degree-seeking status in the 2nd fall term	0.0041	-0.011
	(0.0084)	(0.0613)
Indicator for whether student is in degree-seeking status in the 3rd fall term	0.0124	0.0528
	(0.0124)	(0.0919)
Indicator for whether student is in degree-seeking status in the 4th fall term	-0.0058	-0.1046
	(0.0176)	(0.1366)
Indicator for whether student is in degree-seeking status in the 5th fall term	0.0181	0.1996
	(0.0240)	(0.2164)
Indicator for whether student is in degree-seeking status in the 6th fall term	0.0302	.9076*
	(0.0369)	(0.4849)
Indicator for whether student is in degree-seeking status in the 1st spring term	-.0095**	-.0637*
	(0.0048)	(0.0352)
Indicator for whether student is in degree-seeking status in the 2nd spring term	0.0123	0.0505
	(0.0091)	(0.0670)
Indicator for whether student is in degree-seeking status in the 3rd spring term	.0369***	.2710***
	(0.0140)	(0.1051)
Indicator for whether student is in degree-seeking status in the 4th spring term	0.0093	0.1462
	(0.0197)	(0.1570)

Indicator for whether student is in degree-seeking status in the 5th spring term	0.0405	.4375*
	(0.0272)	(0.2609)
Indicator for whether student is in degree-seeking status in the 6th spring term	-0.0144	-0.3784
	(0.0491)	(0.6327)
Indicator for whether student is in degree-seeking status in the 1st summer term	-0.0037	-0.0084
	(0.0046)	(0.0318)
Indicator for whether student is in degree-seeking status in the 2nd summer term	0.01	0.0359
	(0.0097)	(0.0707)
Indicator for whether student is in degree-seeking status in the 3rd summer term	0.009	0.0555
	(0.0145)	(0.1076)
Indicator for whether student is in degree-seeking status in the 4th summer term	-.0474**	-.3663**
	(0.0203)	(0.1548)
Indicator for whether student is in degree-seeking status in the 5th summer term	0.0238	0.0869
	(0.0303)	(0.2546)
Indicator for whether student is in degree-seeking status in the 6th summer term	0.0742	0.7256
	(0.0641)	(0.7766)
Overall proportion of attempted credits of developmental courses	-.0887***	-.7823***
	(0.0047)	(0.0428)
Proportion of attempted credits of developmental courses in the 1st fall term	-.0471***	-.2581***
	(0.0035)	(0.0286)
Proportion of attempted credits of developmental courses in the 2nd fall term	-.0555***	-.2859***
	(0.0055)	(0.0443)
Proportion of attempted credits of developmental courses in the 3rd fall term	-.0169**	-.1558**
	(0.0085)	(0.0728)
Proportion of attempted credits of developmental courses in the 4th fall term	0.0075	-0.0267
	(0.0120)	(0.1070)
Proportion of attempted credits of developmental courses in the 5th fall term	0.0123	-0.2537
	(0.0174)	(0.1890)
Proportion of attempted credits of developmental courses in the 6th fall term	.0470*	-0.0486

	(0.0284)	(0.4395)
Proportion of attempted credits of developmental courses in the 1st spring term	-.0310***	-.1864***
	(0.0036)	(0.0293)
Proportion of attempted credits of developmental courses in the 2nd spring term	-.0488***	-.3491***
	(0.0061)	(0.0504)
Proportion of attempted credits of developmental courses in the 3rd spring term	-.0172*	-.2234***
	(0.0094)	(0.0798)
Proportion of attempted credits of developmental courses in the 4th spring term	-0.0061	-0.1507
	(0.0137)	(0.1243)
Proportion of attempted credits of developmental courses in the 5th spring term	0.0055	-0.0327
	(0.0202)	(0.2169)
Proportion of attempted credits of developmental courses in the 6th spring term	.1630***	1.541***
	(0.0367)	(0.5184)
Proportion of attempted credits of developmental courses in the 1st summer term	-.0602***	-.2618***
	(0.0038)	(0.0294)
Proportion of attempted credits of developmental courses in the 2nd summer term	-.0314***	-.1173**
	(0.0074)	(0.0549)
Proportion of attempted credits of developmental courses in the 3rd summer term	-0.0169	-0.0228
	(0.0111)	(0.0841)
Proportion of attempted credits of developmental courses in the 4th summer term	0.0265	.3921***
	(0.0166)	(0.1299)
Proportion of attempted credits of developmental courses in the 5th summer term	-0.0292	0.0495
	(0.0245)	(0.2288)
Proportion of attempted credits of developmental courses in the 6th summer term	0.0094	0.2306
	(0.0607)	(0.7378)
Indicator for dual enrollment prior to initial enrollment term	.0819***	.5289***
	(0.0062)	(0.0471)
Total enrollment intensity in non-VCCS institutions in the 1st fall term	.0883***	.5544***
	(0.0152)	(0.1089)

Total enrollment intensity in non-VCCS institutions in the 2nd fall term	.0675***	.4535***
	(0.0192)	(0.1440)
Total enrollment intensity in non-VCCS institutions in the 3rd fall term	.0470**	.3389*
	(0.0233)	(0.1791)
Total enrollment intensity in non-VCCS institutions in the 4th fall term	.1134***	.9896***
	(0.0317)	(0.2526)
Total enrollment intensity in non-VCCS institutions in the 5th fall term	0.0494	0.6455
	(0.0509)	(0.4439)
Total enrollment intensity in non-VCCS institutions in the 6th fall term	-0.0049	0.9195
	(0.1286)	(1.1850)
Total enrollment intensity in non-VCCS institutions in the 1st spring term	.1522***	.9757***
	(0.0162)	(0.1167)
Total enrollment intensity in non-VCCS institutions in the 2nd spring term	.0666***	.4103***
	(0.0207)	(0.1566)
Total enrollment intensity in non-VCCS institutions in the 3rd spring term	.0783***	.6435***
	(0.0236)	(0.1853)
Total enrollment intensity in non-VCCS institutions in the 4th spring term	.0561*	.5784**
	(0.0338)	(0.2767)
Total enrollment intensity in non-VCCS institutions in the 5th spring term	.1019*	0.8043
	(0.0573)	(0.5385)
Total enrollment intensity in non-VCCS institutions in the 6th spring term	0.1897	1.937
	(0.1591)	(1.6110)
Total enrollment intensity in non-VCCS institutions in the 1st summer term	-0.0248	-0.1018
	(0.0311)	(0.2363)
Total enrollment intensity in non-VCCS institutions in the 2nd summer term	-0.0319	-0.1996
	(0.0327)	(0.2616)
Total enrollment intensity in non-VCCS institutions in the 3rd summer term	-0.0441	-0.3922
	(0.0369)	(0.3004)
Total enrollment intensity in non-VCCS institutions in the 4th summer term	0.0343	0.2032

	(0.0514)	(0.4177)
Total enrollment intensity in non-VCCS institutions in the 5th summer term	0.102	1.601**
	(0.0855)	(0.7664)
Total enrollment intensity in non-VCCS institutions in the 6th summer term	-0.3935	-3.154
	(0.4372)	(3.7080)
Slope of term-level number of credits attempted through the end of observation window	.0040***	.0295***
	(0.0002)	(0.0019)
Indicator for whether student is actively enrolled in VCCS in the 1st fall term	-.0475***	-.2494***
	(0.0068)	(0.0574)
Indicator for whether student is actively enrolled in VCCS in the 2nd fall term	-.1229***	-.6140***
	(0.0097)	(0.0762)
Indicator for whether student is actively enrolled in VCCS in the 3rd fall term	-.1510***	-1.047***
	(0.0139)	(0.1085)
Indicator for whether student is actively enrolled in VCCS in the 4th fall term	-.1218***	-1.012***
	(0.0194)	(0.1586)
Indicator for whether student is actively enrolled in VCCS in the 5th fall term	-.1245***	-1.571***
	(0.0268)	(0.2599)
Indicator for whether student is actively enrolled in VCCS in the 6th fall term	-.1343***	-2.922***
	(0.0463)	(0.8083)
Indicator for whether student was ever enrolled in any non-VCCS institutions since initial enrollment term	0.0361	0.1207
	(0.0633)	(0.4560)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 1st fall term	0.026	0.2457
	(0.0184)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 2nd fall term	-.0370*	-0.1723
	(0.0209)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 3rd fall term	-0.0343	-0.1431
	(0.0235)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 4th fall term	-.1140***	-0.7583
	(0.0287)	(33169.0000)

Indicator for whether student is actively enrolled in non-VCCS institutions in the 5th fall term	-.0751*	-0.5053
	(0.0422)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 6th fall term	-0.0032	-0.2205
	(0.0937)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 1st spring term	-0.0138	-0.0606
	(0.0190)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 2nd spring term	-0.0066	0.0622
	(0.0219)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 3rd spring term	-0.0191	-0.1537
	(0.0238)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 4th spring term	-.0540*	-0.4024
	(0.0305)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 5th spring term	-.0927*	-0.6526
	(0.0477)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 6th spring term	-0.0757	-0.7234
	(0.1170)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 1st summer term	.1071***	0.6111
	(0.0236)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 2nd summer term	0.0277	0.1924
	(0.0248)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 3rd summer term	.0679***	0.5505
	(0.0263)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 4th summer term	.0728**	0.5676
	(0.0342)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 5th summer term	0.0062	-0.2665
	(0.0508)	(33169.0000)
Indicator for whether student is actively enrolled in non-VCCS institutions in the 6th summer term	0.2056	1.761
	(0.2091)	(33169.0000)
Indicator for whether student was ever enrolled in VCCS prior to initial enrollment term	-.1350***	-.9144***

	(0.0094)	(0.0887)
Indicator for whether student is actively enrolled in VCCS in the 1st spring term	-.0306***	-.1482***
	(0.0067)	(0.0558)
Indicator for whether student is actively enrolled in VCCS in the 2nd spring term	-.0551***	-.2043**
	(0.0105)	(0.0819)
Indicator for whether student is actively enrolled in VCCS in the 3rd spring term	-.1260***	-.8852***
	(0.0155)	(0.1221)
Indicator for whether student is actively enrolled in VCCS in the 4th spring term	-.0954***	-.8105***
	(0.0217)	(0.1832)
Indicator for whether student is actively enrolled in VCCS in the 5th spring term	-.1282***	-1.661***
	(0.0305)	(0.3123)
Indicator for whether student is actively enrolled in VCCS in the 6th spring term	0.0406	0.5
	(0.0678)	(1.3430)
Indicator for whether student is actively enrolled in VCCS in the 1st summer term	-.0342***	0.0325
	(0.0067)	(0.0527)
Indicator for whether student is actively enrolled in VCCS in the 2nd summer term	-.0279**	0.0383
	(0.0119)	(0.0888)
Indicator for whether student is actively enrolled in VCCS in the 3rd summer term	-.0670***	-.4387***
	(0.0172)	(0.1309)
Indicator for whether student is actively enrolled in VCCS in the 4th summer term	-0.0284	-0.1063
	(0.0240)	(0.1900)
Indicator for whether student is actively enrolled in VCCS in the 5th summer term	-.1221***	-1.203***
	(0.0352)	(0.3248)
Indicator for whether student is actively enrolled in VCCS in the 6th summer term	-.2012**	-2.094*
	(0.0856)	(1.1070)
Slope of term GPA through the end of observation window	.0419***	.3188***
	(0.0009)	(0.0076)
Weighted average of graduation rates of all non-VCCS institutions attended	-0.0012	-0.0924
	(0.0251)	(0.1916)
Logarithm of total grants received in year 1	-0.0005	-0.0029

	(0.0004)	(0.0031)
Logarithm of total grants received in year 2	.0017***	.0159***
	(0.0005)	(0.0037)
Logarithm of total grants received in year 3	.0023***	.0236***
	(0.0007)	(0.0052)
Logarithm of total grants received in year 4	.0026***	.0246***
	(0.0009)	(0.0073)
Logarithm of total grants received in year 5	.0051***	.0310***
	(0.0013)	(0.0115)
Logarithm of total grants received in year 6	0.0019	0.0455
	(0.0027)	(0.0283)
Indicator for Hispanic	0.0002	0.012
	(0.0052)	(0.0387)
Overall proportion of attempted credits of 2XX level courses	.1170***	.6265***
	(0.0053)	(0.0376)
Proportion of attempted credits of 2XX level courses in the 1st fall term	-.0570***	-.2892***
	(0.0040)	(0.0287)
Proportion of attempted credits of 2XX level courses in the 2nd fall term	-.0209***	-.1402***
	(0.0040)	(0.0286)
Proportion of attempted credits of 2XX level courses in the 3rd fall term	-.0120**	-0.0344
	(0.0049)	(0.0355)
Proportion of attempted credits of 2XX level courses in the 4th fall term	-.0305***	-.2075***
	(0.0064)	(0.0479)
Proportion of attempted credits of 2XX level courses in the 5th fall term	-0.0135	-0.0058
	(0.0090)	(0.0734)
Proportion of attempted credits of 2XX level courses in the 6th fall term	-0.0072	.2614*
	(0.0151)	(0.1471)
Proportion of attempted credits of 2XX level courses in the 1st spring term	-.0499***	-.2491***
	(0.0038)	(0.0270)
Proportion of attempted credits of 2XX level courses in the 2nd spring term	-.0463***	-.2702***
	(0.0042)	(0.0301)
Proportion of attempted credits of 2XX level courses in the 3rd spring term	-.0286***	-.1470***
	(0.0054)	(0.0390)

Proportion of attempted credits of 2XX level courses in the 4th spring term	-.0236***	-.1073**
	(0.0073)	(0.0547)
Proportion of attempted credits of 2XX level courses in the 5th spring term	-.0427***	-.2639***
	(0.0105)	(0.0876)
Proportion of attempted credits of 2XX level courses in the 6th spring term	-.0496**	0.0276
	(0.0202)	(0.2280)
Proportion of attempted credits of 2XX level courses in the 1st summer term	0.0032	-0.0127
	(0.0038)	(0.0268)
Proportion of attempted credits of 2XX level courses in the 2nd summer term	-0.0035	-0.0229
	(0.0047)	(0.0338)
Proportion of attempted credits of 2XX level courses in the 3rd summer term	-0.0035	-0.0187
	(0.0063)	(0.0460)
Proportion of attempted credits of 2XX level courses in the 4th summer term	-0.0028	-0.0344
	(0.0088)	(0.0650)
Proportion of attempted credits of 2XX level courses in the 5th summer term	-.0261**	-.2371**
	(0.0129)	(0.1026)
Proportion of attempted credits of 2XX level courses in the 6th summer term	-0.0265	-0.1248
	(0.0334)	(0.3157)
Indicator for male	-.0347***	-.2515***
	(0.0014)	(0.0105)
Indicator for two-year, private, out-of-state	-0.03	0.1082
	(0.0554)	(0.4727)
Indicator for two-year, private, in-state	.0858**	1.068***
	(0.0361)	(0.2711)
Indicator for two-year, public, out-of-state	-.0683***	-.3718*
	(0.0249)	(0.1942)
Indicator for two-year, public, in-state	-0.0544	-0.3501
	(0.0501)	(0.3675)
Indicator for four-year, private, out-of-state	-0.0354	-0.1734
	(0.0234)	(0.1835)
Indicator for four-year, private, in-state	-0.001	0.081
	(0.0235)	(0.1844)
Indicator for four-year, public, out-of-state	-.0435*	-0.1942
	(0.0237)	(0.1850)
Indicator for four-year, public, in-state	-.0419*	-0.2149

	(0.0236)	(0.1849)
Number of terms in which student was enrolled in non-VCCS institutions since initial enrollment term	-0.0123	-0.0675
	(0.0138)	(33169.0000)
Number of non-VCCS institutions in which student was enrolled since initial enrollment term	0.0093	0.0468
	(0.0178)	(0.1414)
Indicator for other race/ethnicity	-0.0076	-.0744*
	(0.0051)	(0.0380)
Logarithm of other aids received in year 1	0.0017	.0166*
	(0.0012)	(0.0087)
Logarithm of other aids received in year 2	0.0005	0.001
	(0.0015)	(0.0113)
Logarithm of other aids received in year 3	-0.001	-0.0055
	(0.0021)	(0.0157)
Logarithm of other aids received in year 4	0.001	0.011
	(0.0033)	(0.0244)
Logarithm of other aids received in year 5	-0.0049	-0.0446
	(0.0050)	(0.0388)
Logarithm of other aids received in year 6	0.0142	0.1148
	(0.0089)	(0.0768)
Indicator for not Pell-eligible the 1st fall term	0.0059	0.0382
	(0.0037)	(0.0283)
Indicator for not Pell-eligible the 2nd fall term	.0133***	.0589*
	(0.0046)	(0.0333)
Indicator for not Pell-eligible the 3rd fall term	0.0089	0.0411
	(0.0063)	(0.0455)
Indicator for not Pell-eligible the 4th fall term	0	0.0048
	(0.0084)	(0.0619)
Indicator for not Pell-eligible the 5th fall term	0.0081	0.1163
	(0.0120)	(0.0942)
Indicator for not Pell-eligible the 6th fall term	-.0335*	-0.0414
	(0.0197)	(0.1909)
Indicator for never Pell-eligible	.0152***	.1146***
	(0.0033)	(0.0243)
Indicator for not Pell-eligible the 1st spring term	-.0063*	-.0693**
	(0.0036)	(0.0269)
Indicator for not Pell-eligible the 2nd spring term	-.0211***	-.1352***
	(0.0049)	(0.0358)
Indicator for not Pell-eligible the 3rd spring term	-.0275***	-.1604***
	(0.0068)	(0.0494)
Indicator for not Pell-eligible the 4th spring term	-0.0125	-0.0898
	(0.0094)	(0.0695)
Indicator for not Pell-eligible the 5th spring term	-.0239*	-.2028*

	(0.0133)	(0.1070)
Indicator for not Pell-eligible the 6th spring term	-0.0331	-0.055
	(0.0246)	(0.2594)
Indicator for not Pell-eligible the 1st summer term	0.0041	.0368*
	(0.0030)	(0.0219)
Indicator for not Pell-eligible the 2nd summer term	.0270***	.1473***
	(0.0045)	(0.0322)
Indicator for not Pell-eligible the 3rd summer term	.0104*	0.0309
	(0.0061)	(0.0448)
Indicator for not Pell-eligible the 4th summer term	.0306***	.1885***
	(0.0086)	(0.0638)
Indicator for not Pell-eligible the 5th summer term	.0626***	.4629***
	(0.0123)	(0.0995)
Indicator for not Pell-eligible the 6th summer term	0.0488	0.2634
	(0.0298)	(0.2758)
Indicator for Pell-eligible the 1st fall term	-.0083**	-0.0228
	(0.0034)	(0.0272)
Indicator for Pell-eligible the 2nd fall term	-0.0055	-0.029
	(0.0043)	(0.0321)
Indicator for Pell-eligible the 3rd fall term	-.0107*	-.0944**
	(0.0058)	(0.0432)
Indicator for Pell-eligible the 4th fall term	-0.0045	-0.0443
	(0.0077)	(0.0602)
Indicator for Pell-eligible the 5th fall term	-.0311***	-0.1517
	(0.0111)	(0.0942)
Indicator for Pell-eligible the 6th fall term	-0.0088	-0.1521
	(0.0219)	(0.2319)
Indicator for ever Pell-eligible	0.0021	-0.0226
	(0.0033)	(0.0262)
Indicator for Pell-eligible the 1st spring term	-.0191***	-.1448***
	(0.0031)	(0.0245)
Indicator for Pell-eligible the 2nd spring term	-.0162***	-.1194***
	(0.0043)	(0.0317)
Indicator for Pell-eligible the 3rd spring term	-.0388***	-.2671***
	(0.0058)	(0.0430)
Indicator for Pell-eligible the 4th spring term	-.0503***	-.4082***
	(0.0078)	(0.0602)
Indicator for Pell-eligible the 5th spring term	-.0492***	-.3123***
	(0.0113)	(0.0965)
Indicator for Pell-eligible the 6th spring term	-0.0317	-.5164**
	(0.0233)	(0.2618)
Indicator for Pell-eligible the 1st summer term	.0209***	.1542***
	(0.0042)	(0.0302)
Indicator for Pell-eligible the 2nd summer term	.0506***	.2820***

	(0.0062)	(0.0439)
Indicator for Pell-eligible the 3rd summer term	.0686***	.3992***
	(0.0085)	(0.0613)
Indicator for Pell-eligible the 4th summer term	.0724***	.4434***
	(0.0116)	(0.0858)
Indicator for Pell-eligible the 5th summer term	0.0234	.2728**
	(0.0170)	(0.1376)
Indicator for Pell-eligible the 6th summer term	-0.034	-0.656
	(0.0462)	(0.5180)
Indicator for highest parental education being less than high school	-0.0022	-0.0191
	(0.0050)	(0.0376)
Indicator for highest parental education being having attended high school	-0.0022	-0.0139
	(0.0039)	(0.0308)
Indicator for highest parental education being having graduated from high school	-.0103***	-.0661***
	(0.0020)	(0.0158)
Indicator for highest parental education being having attended college	-.0114***	-.0623***
	(0.0022)	(0.0174)
Indicator for highest parental education being having earned Associate's degree	-0.0027	-0.0076
	(0.0030)	(0.0225)
Indicator for highest parental education being having earned Bachelor's degree	.0126***	.0776***
	(0.0022)	(0.0161)
Indicator for highest parental education being having earned Post-Bachelor's degree	.0262***	.1568***
	(0.0027)	(0.0194)
Number of terms in which student was enrolled in non-VCCS institutions prior to initial enrollment term	.0435***	.2494***
	(0.0010)	(0.0071)
Number of non-VCCS institutions in which student was enrolled prior to initial enrollment term	.0121***	.1549***
	(0.0027)	(0.0195)
Indicator for whether student changed degree/major program pursued	.0230***	.1491***
	(0.0019)	(0.0140)
Overall proportion of earned credits among attempted credits since initial enrollment term	.0872***	1.025***
	(0.0055)	(0.0541)
Proportion of earned credits among attempted credits in the 1st fall term	-.0469***	-.1903***

	(0.0048)	(0.0443)
Proportion of earned credits among attempted credits in the 2nd fall term	.0109*	.2357***
	(0.0060)	(0.0510)
Proportion of earned credits among attempted credits in the 3rd fall term	.0684***	.7291***
	(0.0082)	(0.0687)
Proportion of earned credits among attempted credits in the 4th fall term	.0672***	.8746***
	(0.0111)	(0.0973)
Proportion of earned credits among attempted credits in the 5th fall term	.0541***	1.086***
	(0.0158)	(0.1591)
Proportion of earned credits among attempted credits in the 6th fall term	.0437*	2.325***
	(0.0263)	(0.4055)
Overall proportion of earned credits among attempted credits prior to initial enrollment term	-.0355***	0.0048
	(0.0099)	(0.0846)
Standard deviation of term-level proportion of earned credits among attempted credits since initial enrollment term	-.1980***	-1.079***
	(0.0061)	(0.0616)
Proportion of earned credits among attempted credits in the 1st spring term	-.0564***	-.2389***
	(0.0047)	(0.0433)
Proportion of earned credits among attempted credits in the 2nd spring term	0.01	.2599***
	(0.0067)	(0.0559)
Proportion of earned credits among attempted credits in the 3rd spring term	.0491***	.5801***
	(0.0093)	(0.0771)
Proportion of earned credits among attempted credits in the 4th spring term	.0669***	.8999***
	(0.0129)	(0.1147)
Proportion of earned credits among attempted credits in the 5th spring term	.0688***	1.273***
	(0.0189)	(0.2018)
Proportion of earned credits among attempted credits in the 6th spring term	-0.0023	1.244**
	(0.0353)	(0.5664)
Proportion of earned credits among attempted credits in the 1st summer term	-0.0024	-.0942**
	(0.0056)	(0.0466)

Proportion of earned credits among attempted credits in the 2nd summer term	.0351***	.1550**
	(0.0087)	(0.0655)
Proportion of earned credits among attempted credits in the 3rd summer term	.0500***	.3745***
	(0.0120)	(0.0920)
Proportion of earned credits among attempted credits in the 4th summer term	.0568***	.4836***
	(0.0168)	(0.1351)
Proportion of earned credits among attempted credits in the 5th summer term	0.0137	.4875**
	(0.0248)	(0.2380)
Proportion of earned credits among attempted credits in the 6th summer term	-0.0119	0.5392
	(0.0637)	(0.7711)
Indicator for whether student repeated a previously taken course in the 1st fall term	-.0191***	-.0857***
	(0.0036)	(0.0284)
Indicator for whether student repeated a previously taken course in the 2nd fall term	-.0230***	-.1154***
	(0.0021)	(0.0158)
Indicator for whether student repeated a previously taken course in the 3rd fall term	-.0060**	-.0304*
	(0.0025)	(0.0185)
Indicator for whether student repeated a previously taken course in the 4th fall term	-.0084**	-0.0311
	(0.0034)	(0.0259)
Indicator for whether student repeated a previously taken course in the 5th fall term	-.0142***	-.0998**
	(0.0049)	(0.0421)
Indicator for whether student repeated a previously taken course in the 6th fall term	0.0119	0.06
	(0.0083)	(0.0894)
Indicator for whether student has ever repeated a previously taken course	-.0082***	-0.0238
	(0.0026)	(0.0201)
Indicator for whether student repeated a previously taken course in the 1st spring term	-.0182***	-.0764***
	(0.0026)	(0.0205)
Indicator for whether student repeated a previously taken course in the 2nd spring term	-.0113***	-.0583***
	(0.0024)	(0.0179)
Indicator for whether student repeated a previously taken course in the 3rd spring term	.0057*	.0595***

	(0.0030)	(0.0223)
Indicator for whether student repeated a previously taken course in the 4th spring term	-0.0026	0.0038
	(0.0041)	(0.0318)
Indicator for whether student repeated a previously taken course in the 5th spring term	0.0101	.1368**
	(0.0062)	(0.0543)
Indicator for whether student repeated a previously taken course in the 6th spring term	0.0135	.4157***
	(0.0125)	(0.1452)
Indicator for whether student repeated a previously taken course in the 1st summer term	0.0012	0.0135
	(0.0038)	(0.0278)
Indicator for whether student repeated a previously taken course in the 2nd summer term	-.0121***	-.0796***
	(0.0038)	(0.0275)
Indicator for whether student repeated a previously taken course in the 3rd summer term	-0.0066	-0.0307
	(0.0049)	(0.0360)
Indicator for whether student repeated a previously taken course in the 4th summer term	-0.0012	0.0254
	(0.0068)	(0.0516)
Indicator for whether student repeated a previously taken course in the 5th summer term	0.0002	0.1105
	(0.0098)	(0.0833)
Indicator for whether student repeated a previously taken course in the 6th summer term	0.0393	0.3929
	(0.0257)	(0.2474)
Weighted average of the 1st quartiles of SAT math scores of all non-VCCS institutions attended	0	-0.0003
	(0.0003)	(0.0026)
Weighted average of the 3rd quartiles of SAT math scores of all non-VCCS institutions attended	.0007***	.0050***
	(0.0002)	(0.0019)
Weighted average of the 1st quartiles of SAT verbal scores of all non-VCCS institutions attended	0.0005	0.0024
	(0.0004)	(0.0033)
Weighted average of the 3rd quartiles of SAT verbal scores of all non-VCCS institutions attended	-.0010***	-.0052**
	(0.0003)	(0.0025)
Weighted average of the 1st quartiles of SAT writing scores of all non-VCCS institutions attended	.0006*	.0058*
	(0.0004)	(0.0030)

Weighted average of the 3rd quartiles of SAT writing scores of all non-VCCS institutions attended	-0.0006	-.0054*
	(0.0004)	(0.0031)
Indicator for not a seamless enrollee	-.0244***	-.1367**
	(0.0072)	(0.0577)
Indicator for seamless enrollee	.0233***	.2009***
	(0.0072)	(0.0582)
Logarithm of subsidized loans received in year 1	-0.0004	-0.0032
	(0.0004)	(0.0031)
Logarithm of subsidized loans received in year 2	.0019***	.0127***
	(0.0006)	(0.0044)
Logarithm of subsidized loans received in year 3	0.0003	0.0015
	(0.0008)	(0.0057)
Logarithm of subsidized loans received in year 4	-0.0017	-.0153*
	(0.0010)	(0.0079)
Logarithm of subsidized loans received in year 5	-0.0012	-0.0129
	(0.0015)	(0.0119)
Logarithm of subsidized loans received in year 6	-0.0035	-.0750***
	(0.0024)	(0.0247)
Number of credit hours attempted in the 1st fall term	.0115***	.0756***
	(0.0002)	(0.0017)
Number of credit hours attempted in the 2nd fall term	.0113***	.0638***
	(0.0003)	(0.0023)
Number of credit hours attempted in the 3rd fall term	.0100***	.0678***
	(0.0004)	(0.0033)
Number of credit hours attempted in the 4th fall term	.0098***	.0772***
	(0.0006)	(0.0047)
Number of credit hours attempted in the 5th fall term	.0091***	.0893***
	(0.0009)	(0.0075)
Number of credit hours attempted in the 6th fall term	.0072***	.1097***
	(0.0014)	(0.0134)
Number of credit hours attempted in the 1st spring term	.0109***	.0689***
	(0.0002)	(0.0018)
Number of credit hours attempted in the 2nd spring term	.0065***	.0367***
	(0.0003)	(0.0026)
Number of credit hours attempted in the 3rd spring term	.0058***	.0386***
	(0.0005)	(0.0037)
Number of credit hours attempted in the 4th spring term	.0078***	.0557***
	(0.0007)	(0.0055)
Number of credit hours attempted in the 5th spring term	.0070***	.0747***
	(0.0010)	(0.0089)
Number of credit hours attempted in the 6th spring term	.0036*	.0997***
	(0.0020)	(0.0237)
Number of credit hours attempted in the 1st summer term	.0059***	.0343***

	(0.0003)	(0.0024)
Number of credit hours attempted in the 2nd summer term	.0039***	.0219***
	(0.0005)	(0.0040)
Number of credit hours attempted in the 3rd summer term	.0040***	.0289***
	(0.0008)	(0.0059)
Number of credit hours attempted in the 4th summer term	.0045***	.0302***
	(0.0011)	(0.0086)
Number of credit hours attempted in the 5th summer term	.0079***	.0697***
	(0.0017)	(0.0139)
Number of credit hours attempted in the 6th summer term	.0178***	.1642***
	(0.0045)	(0.0419)
Term GPA in the 1st fall term	.0349***	.1716***
	(0.0011)	(0.0090)
Term GPA in the 2nd fall term	.0344***	.1510***
	(0.0015)	(0.0118)
Term GPA in the 3rd fall term	.0203***	.0928***
	(0.0021)	(0.0163)
Term GPA in the 4th fall term	.0184***	.0991***
	(0.0029)	(0.0227)
Term GPA in the 5th fall term	.0108***	.0950***
	(0.0042)	(0.0355)
Term GPA in the 6th fall term	-0.0085	0.0244
	(0.0070)	(0.0726)
Term GPA in the 1st spring term	.0287***	.1423***
	(0.0011)	(0.0094)
Term GPA in the 2nd spring term	.0152***	.0334**
	(0.0017)	(0.0131)
Term GPA in the 3rd spring term	.0152***	.0605***
	(0.0024)	(0.0181)
Term GPA in the 4th spring term	.0057*	0.0017
	(0.0033)	(0.0261)
Term GPA in the 5th spring term	0.0013	0.0194
	(0.0049)	(0.0425)
Term GPA in the 6th spring term	-0.0143	0.0887
	(0.0092)	(0.1093)
Term GPA in the 1st summer term	.0242***	.0739***
	(0.0014)	(0.0104)
Term GPA in the 2nd summer term	.0085***	0.0158
	(0.0021)	(0.0156)
Term GPA in the 3rd summer term	.0146***	.0833***
	(0.0030)	(0.0221)

Term GPA in the 4th summer term	.0151***	.0676**
	(0.0042)	(0.0318)
Term GPA in the 5th summer term	.0230***	.2098***
	(0.0063)	(0.0527)
Term GPA in the 6th summer term	0.0221	.2957*
	(0.0161)	(0.1620)
Logarithm of unsubsidized loans received in year 1	-.0020***	-.0117***
	(0.0004)	(0.0033)
Logarithm of unsubsidized loans received in year 2	-.0022***	-.0130***
	(0.0006)	(0.0046)
Logarithm of unsubsidized loans received in year 3	-0.0003	-0.0001
	(0.0008)	(0.0061)
Logarithm of unsubsidized loans received in year 4	0.0014	0.0118
	(0.0011)	(0.0084)
Logarithm of unsubsidized loans received in year 5	0.0012	0.0062
	(0.0015)	(0.0126)
Logarithm of unsubsidized loans received in year 6	0.002	0.0202
	(0.0026)	(0.0259)
Indicator for White	-.0144***	-.0924***
	(0.0047)	(0.0347)
Overall proportion of withdrawn credits among attempted credits since initial enrollment term	-.1284***	-1.239***
	(0.0062)	(0.0585)
Proportion of withdrawn credits among attempted credits the 1st fall term	-.0167***	-0.05
	(0.0054)	(0.0496)
Proportion of withdrawn credits among attempted credits the 2nd fall term	-.1033***	-.5981***
	(0.0056)	(0.0488)
Proportion of withdrawn credits among attempted credits the 3rd fall term	-.1035***	-.6981***
	(0.0067)	(0.0574)
Proportion of withdrawn credits among attempted credits the 4th fall term	-.1125***	-1.046***
	(0.0086)	(0.0806)
Proportion of withdrawn credits among attempted credits the 5th fall term	-.0814***	-1.007***
	(0.0122)	(0.1324)
Proportion of withdrawn credits among attempted credits the 6th fall term	-0.0262	-1.611***
	(0.0198)	(0.3322)
Standard deviation of term-level proportion of withdrawn credits among attempted credits since initial enrollment term	-.1455***	-1.185***

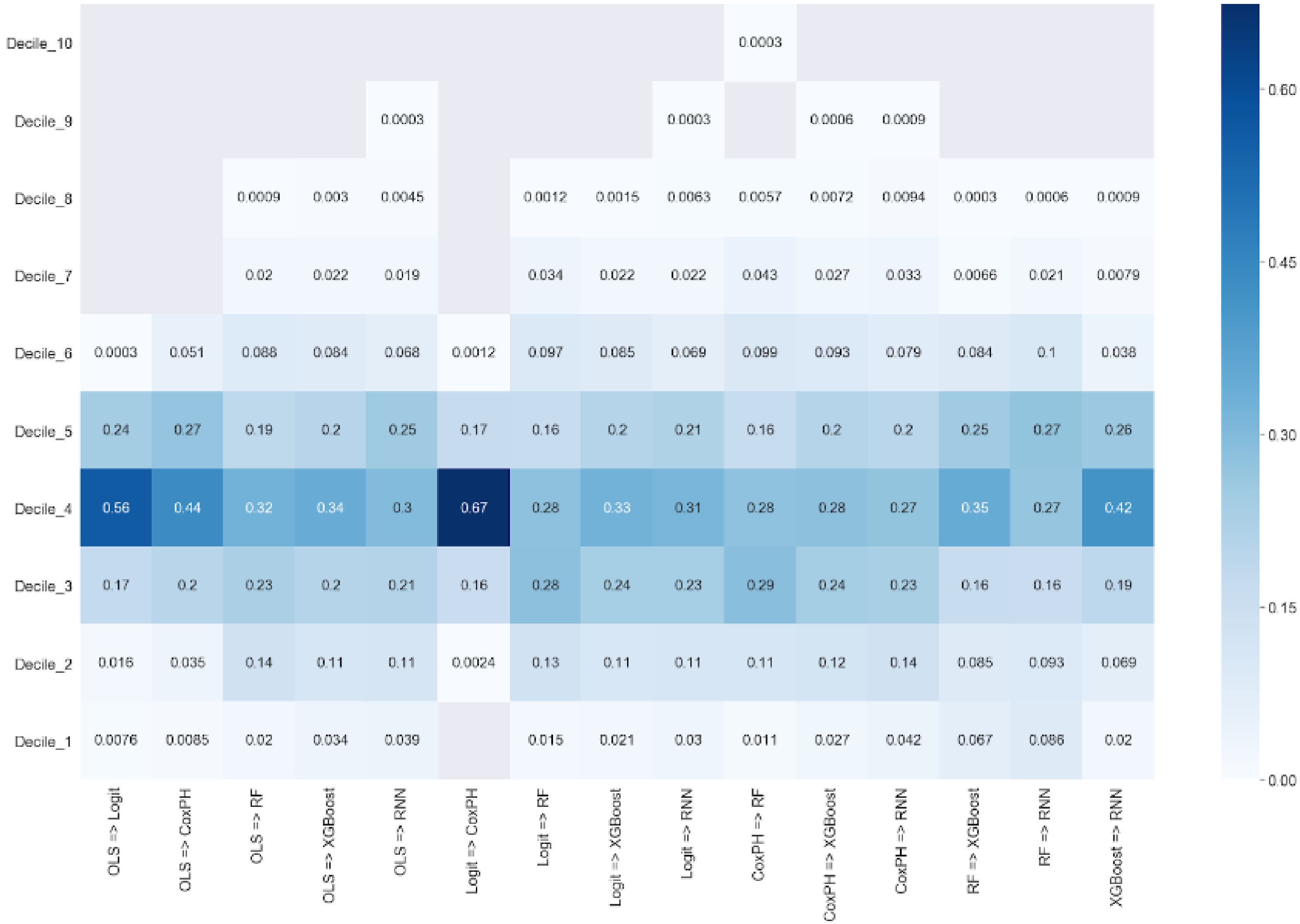
	(0.0077)	(0.0730)
Proportion of withdrawn credits among attempted credits the 1st spring term	-0.0058	-0.0097
	(0.0052)	(0.0475)
Proportion of withdrawn credits among attempted credits the 2nd spring term	-.0640***	-.3777***
	(0.0059)	(0.0500)
Proportion of withdrawn credits among attempted credits the 3rd spring term	-.0844***	-.6183***
	(0.0076)	(0.0651)
Proportion of withdrawn credits among attempted credits the 4th spring term	-.0815***	-.7604***
	(0.0103)	(0.0967)
Proportion of withdrawn credits among attempted credits the 5th spring term	-.0635***	-.8166***
	(0.0151)	(0.1685)
Proportion of withdrawn credits among attempted credits the 6th spring term	0.0421	-0.7067
	(0.0268)	(0.5016)
Proportion of withdrawn credits among attempted credits the 1st summer term	-.0253***	.1406***
	(0.0059)	(0.0490)
Proportion of withdrawn credits among attempted credits the 2nd summer term	-.0598***	-.2129***
	(0.0077)	(0.0583)
Proportion of withdrawn credits among attempted credits the 3rd summer term	-.0683***	-.3082***
	(0.0102)	(0.0770)
Proportion of withdrawn credits among attempted credits the 4th summer term	-.0517***	-.2524**
	(0.0139)	(0.1107)
Proportion of withdrawn credits among attempted credits the 5th summer term	-.0518**	-.4122**
	(0.0210)	(0.1987)
Proportion of withdrawn credits among attempted credits the 6th summer term	0.009	-0.6603
	(0.0475)	(0.5415)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure A1: Consistency across models in student assignment to the second decile of risk rankings

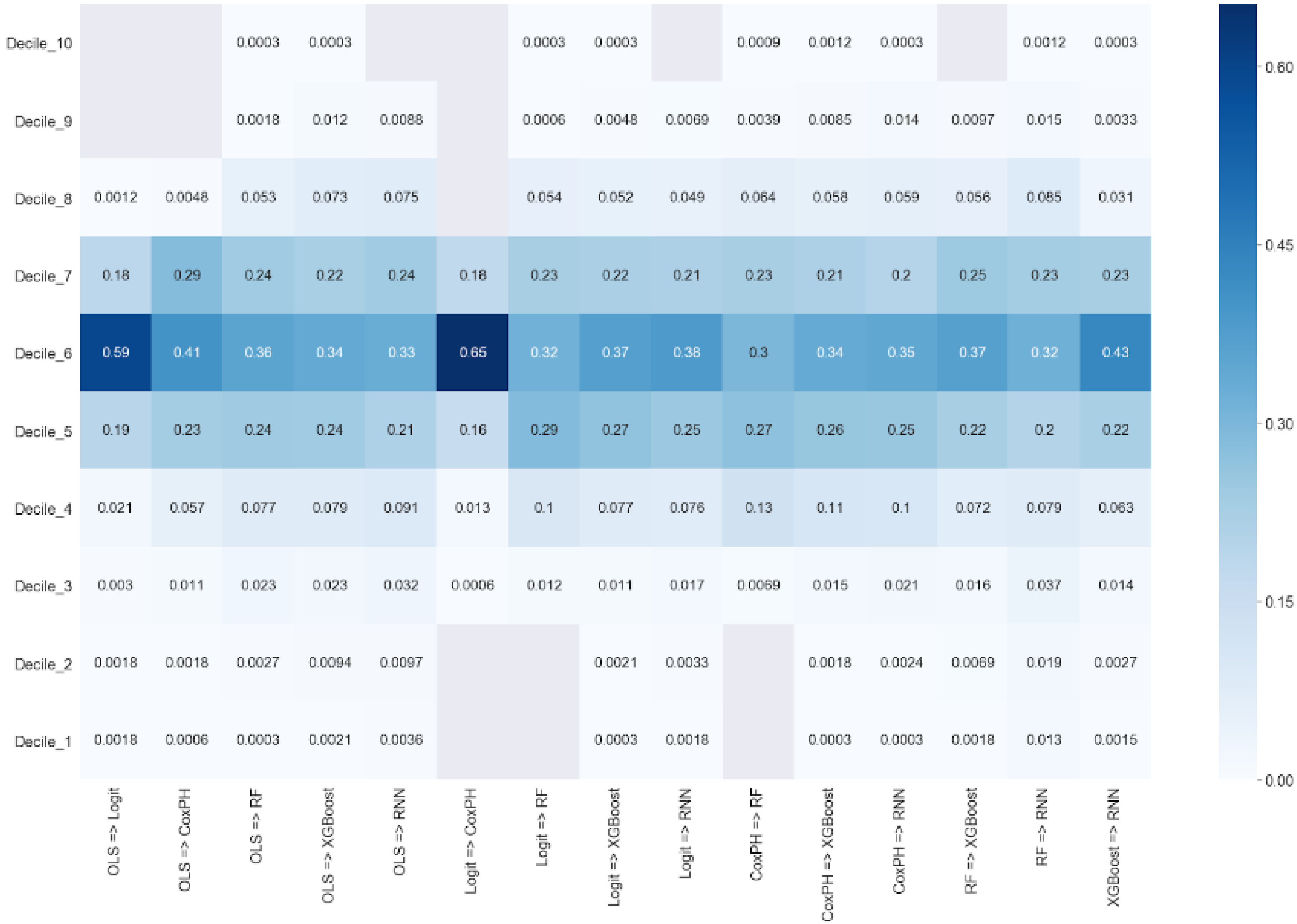
Notes: the second decile of contain the students with a risk ranking percentile between 11-20. Each column of this figure shows the share of students assigned to the second decile by Model A that are assigned to given decile by Model B.

Figure A2: Consistency across models in student assignment to the fourth decile of risk rankings



Notes: the fourth decile of contain the students with a risk ranking percentile between 31-40. Each column of this figure shows the share of students assigned to the fourth decile by Model A that are assigned to given decile by Model B.

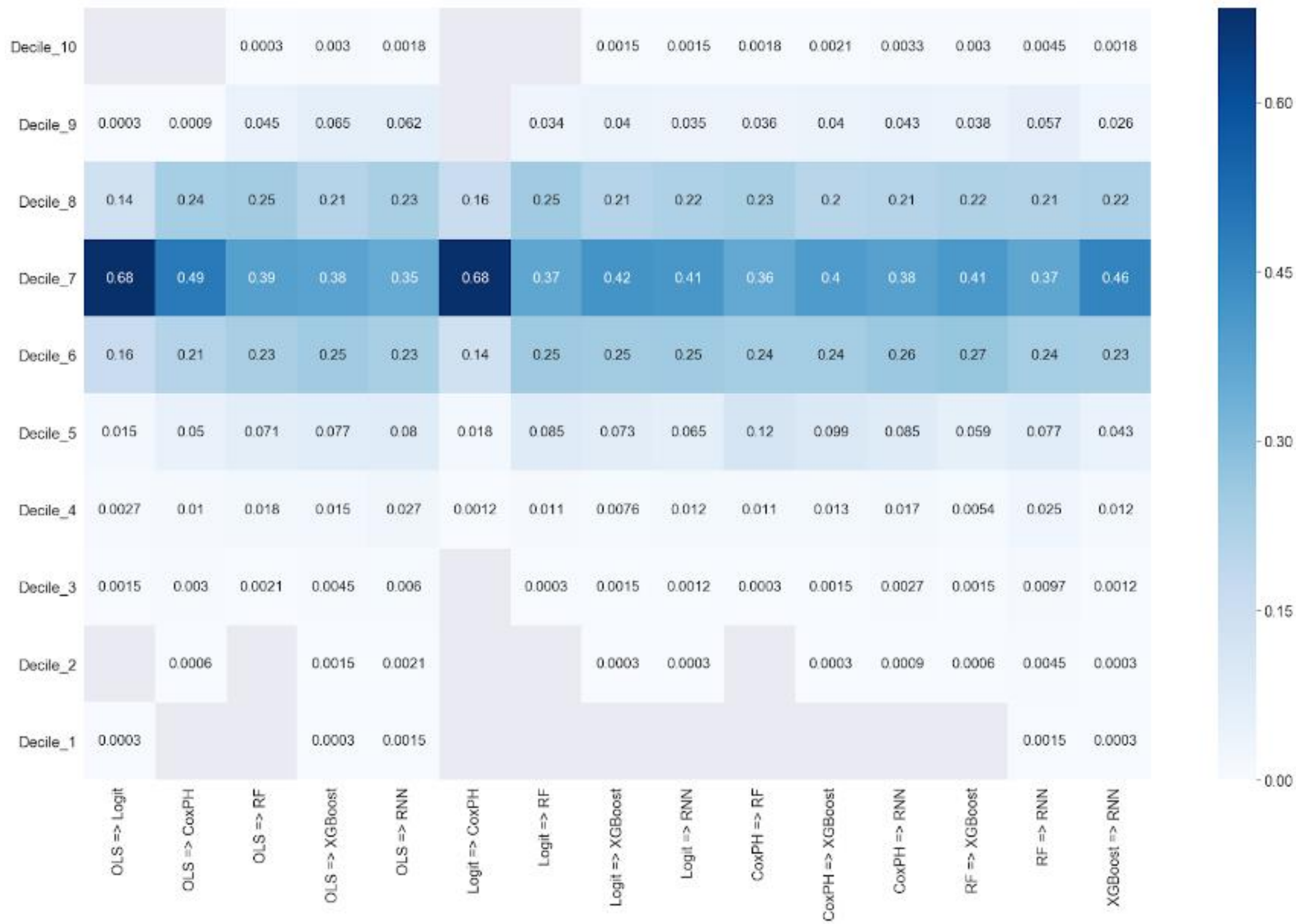
Figure A3: Consistency across models in student assignment to the sixth decile of risk rankings



	OLS => Logit	OLS => CoxPH	OLS => RF	OLS => XGBoost	OLS => RNN	Logit => CoxPH	Logit => RF	Logit => XGBoost	Logit => RNN	CoxPH => RF	CoxPH => XGBoost	CoxPH => RNN	RF => XGBoost	RF => RNN	XGBoost => RNN
Decile_10			0.0003	0.0003			0.0003	0.0003		0.0009	0.0012	0.0003		0.0012	0.0003
Decile_9			0.0018	0.012	0.0088		0.0006	0.0048	0.0069	0.0039	0.0085	0.014	0.0097	0.015	0.0033
Decile_8	0.0012	0.0048	0.053	0.073	0.075		0.054	0.052	0.049	0.064	0.058	0.059	0.056	0.085	0.031
Decile_7	0.18	0.29	0.24	0.22	0.24	0.18	0.23	0.22	0.21	0.23	0.21	0.2	0.25	0.23	0.23
Decile_6	0.59	0.41	0.36	0.34	0.33	0.65	0.32	0.37	0.38	0.3	0.34	0.35	0.37	0.32	0.43
Decile_5	0.19	0.23	0.24	0.24	0.21	0.16	0.29	0.27	0.25	0.27	0.26	0.25	0.22	0.2	0.22
Decile_4	0.021	0.057	0.077	0.079	0.091	0.013	0.1	0.077	0.076	0.13	0.11	0.1	0.072	0.079	0.063
Decile_3	0.003	0.011	0.023	0.023	0.032	0.0006	0.012	0.011	0.017	0.0069	0.015	0.021	0.016	0.037	0.014
Decile_2	0.0018	0.0018	0.0027	0.0094	0.0097			0.0021	0.0033		0.0018	0.0024	0.0069	0.019	0.0027
Decile_1	0.0018	0.0006	0.0003	0.0021	0.0036			0.0003	0.0018		0.0003	0.0003	0.0018	0.013	0.0015

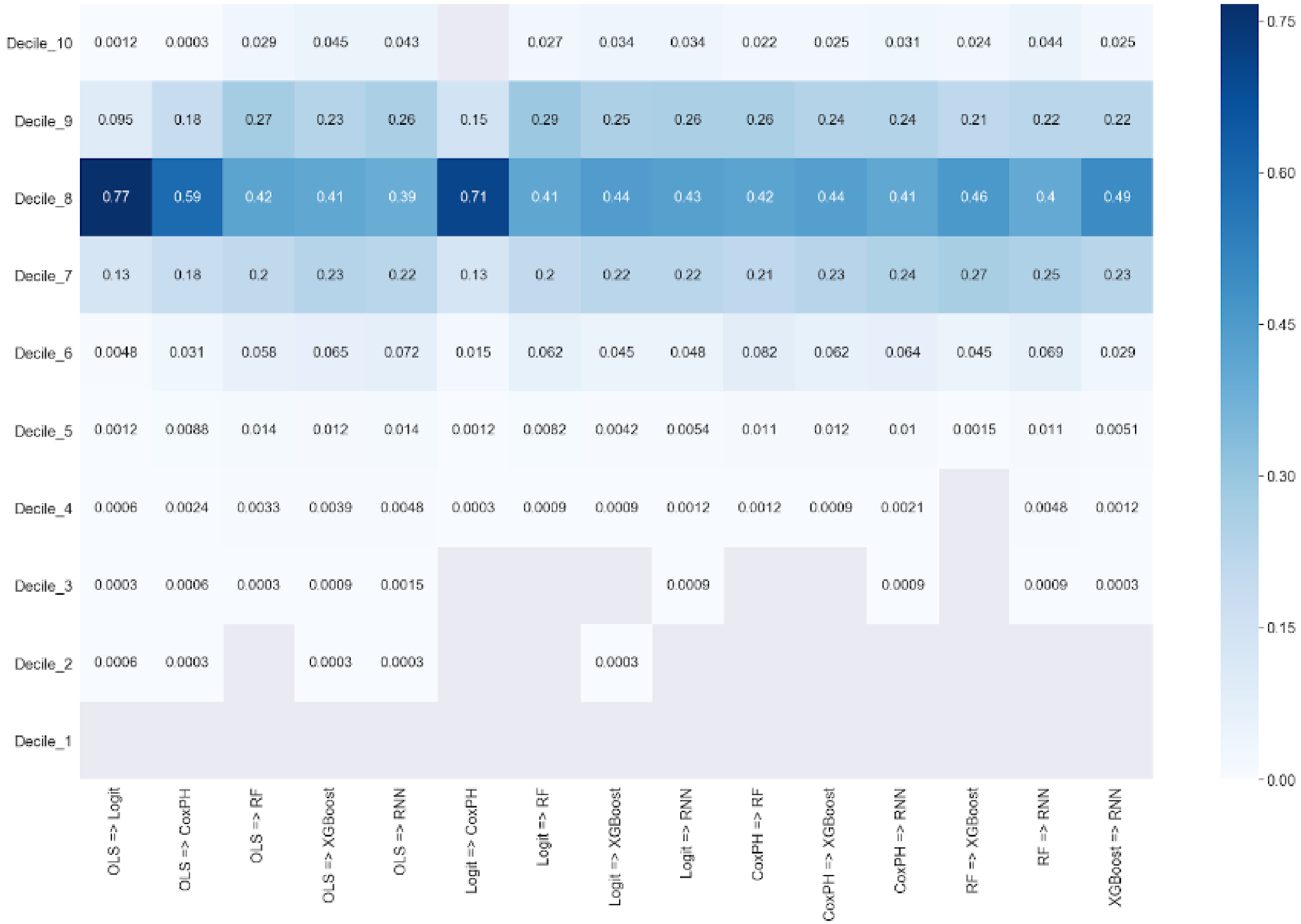
Notes: the sixth decile of contain the students with a risk ranking percentile between 51-60. Each column of this figure shows the share of students assigned to the sixth decile by Model A that are assigned to given decile by Model B.

Figure A4: Consistency across models in student assignment to the seventh decile of risk rankings



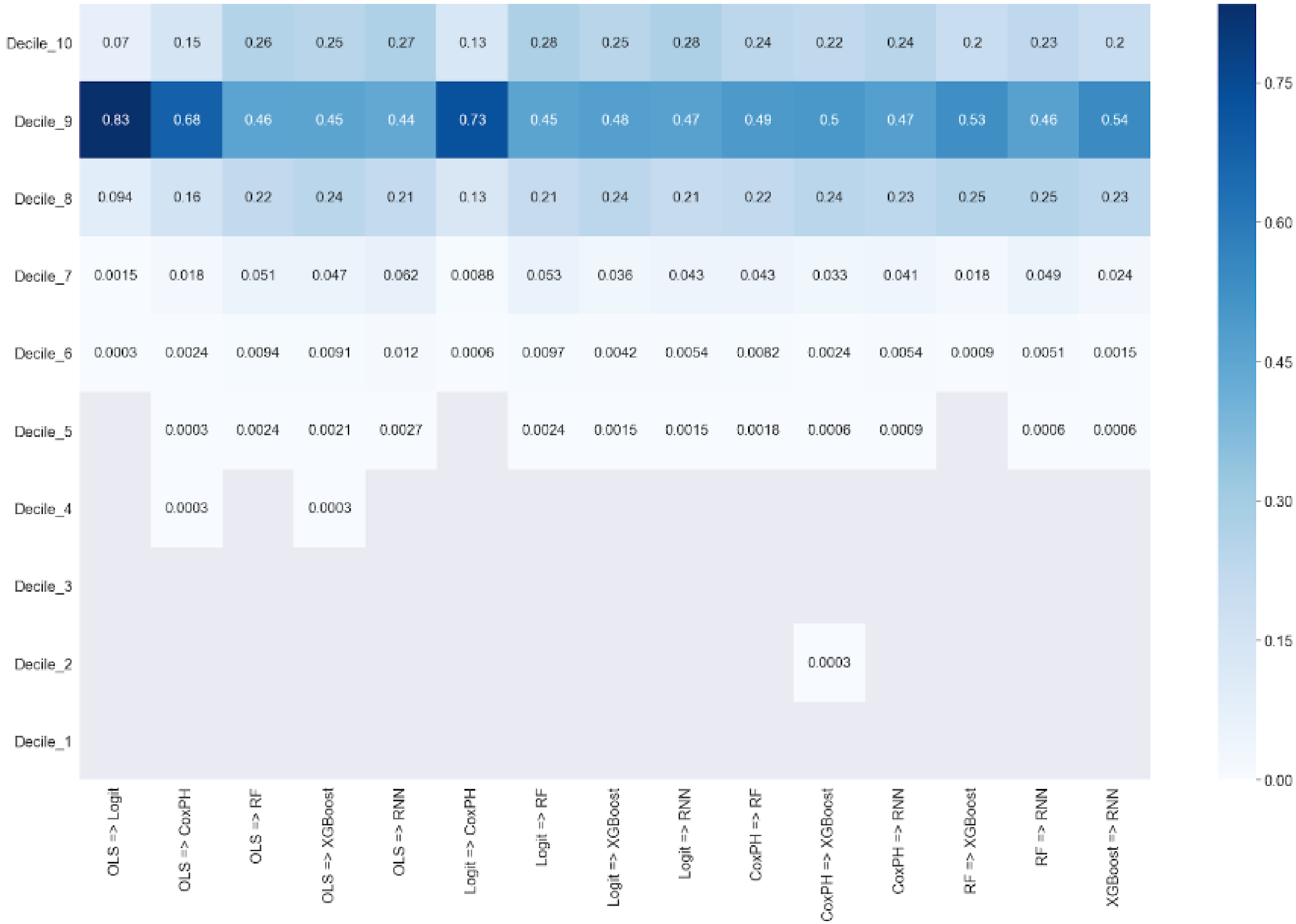
Notes: the seventh decile of contain the students with a risk ranking percentile between 61-70. Each column of this figure shows the share of students assigned to the seventh decile by Model A that are assigned to given decile by Model B.

Figure A5: Consistency across models in student assignment to the eighth decile of risk rankings



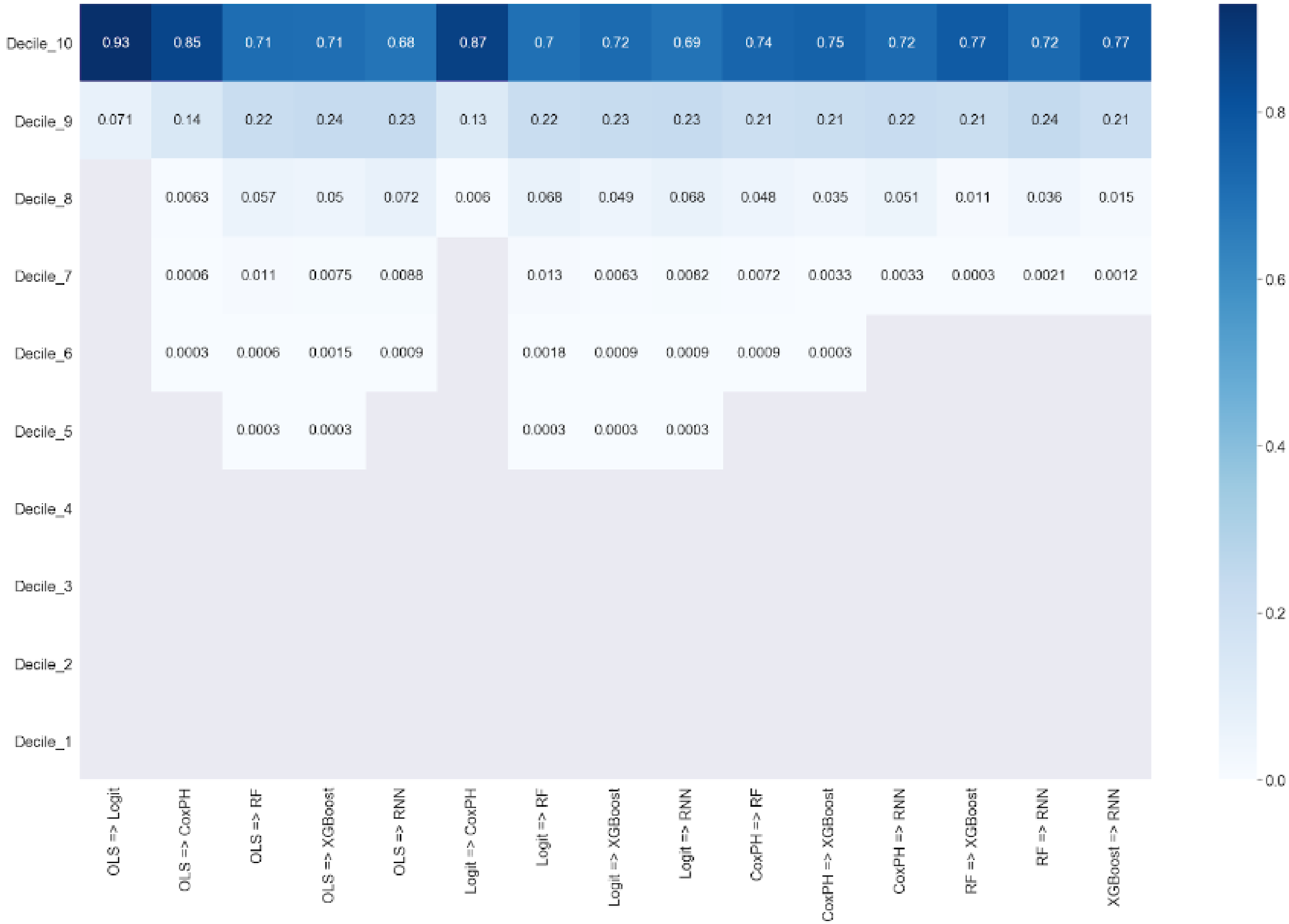
Notes: the eighth decile of contain the students with a risk ranking percentile between 71-80. Each column of this figure shows the share of students assigned to the eighth decile by Model A that are assigned to given decile by Model B.

Figure A6: Consistency across models in student assignment to the ninth decile of risk rankings



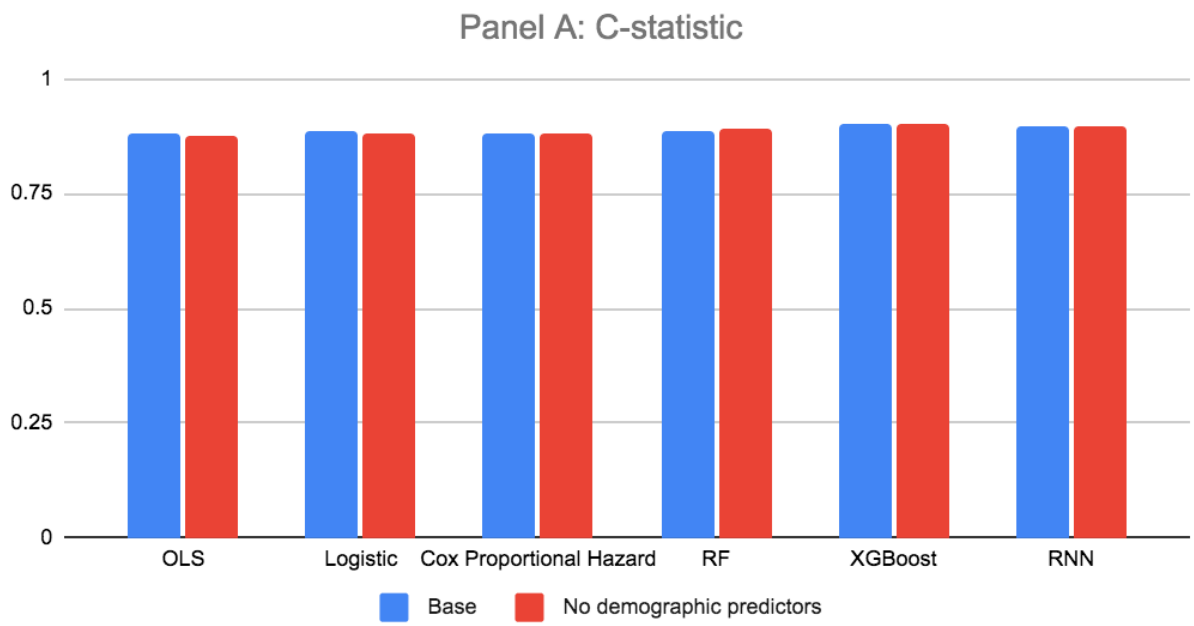
Notes: the ninth decile of contain the students with a risk ranking percentile between 81-90. Each column of this figure shows the share of students assigned to the ninth decile by Model A that are assigned to given decile by Model B.

Figure A7: Consistency across models in student assignment to the tenth decile of risk rankings



Notes: the tenth decile of contain the students with a risk ranking percentile between 91-100. Each column of this figure shows the share of students assigned to the tenth decile by Model A that are assigned to given decile by Model B.

Figure A8: Evaluation statistics, base models versus models that exclude the complexly specified term-specific predictors



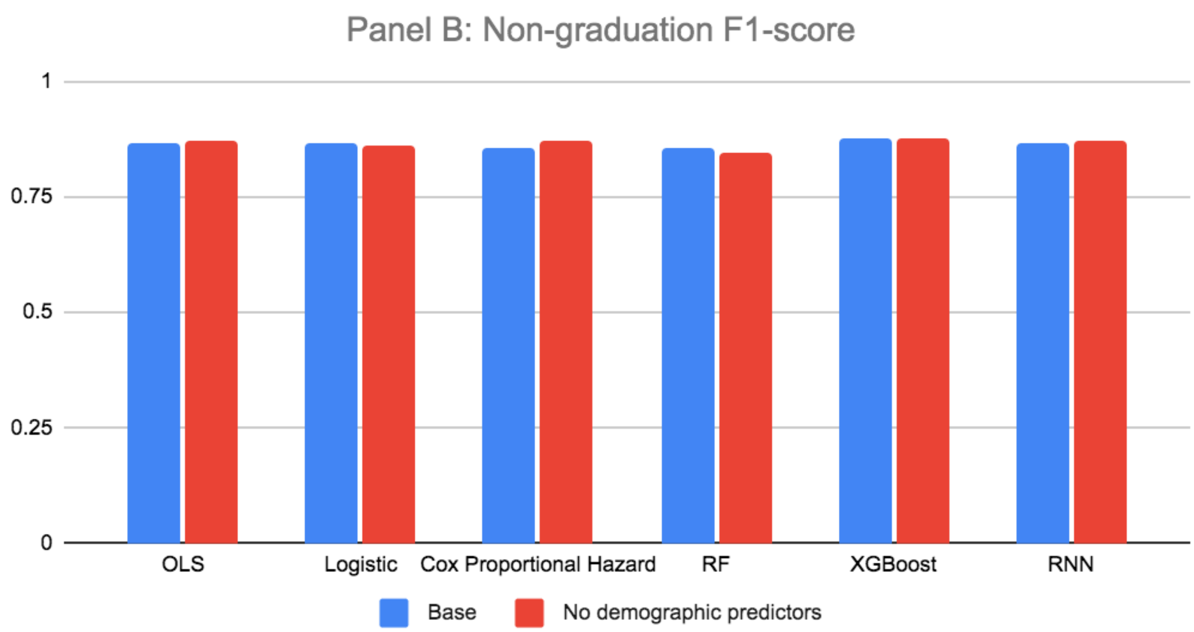
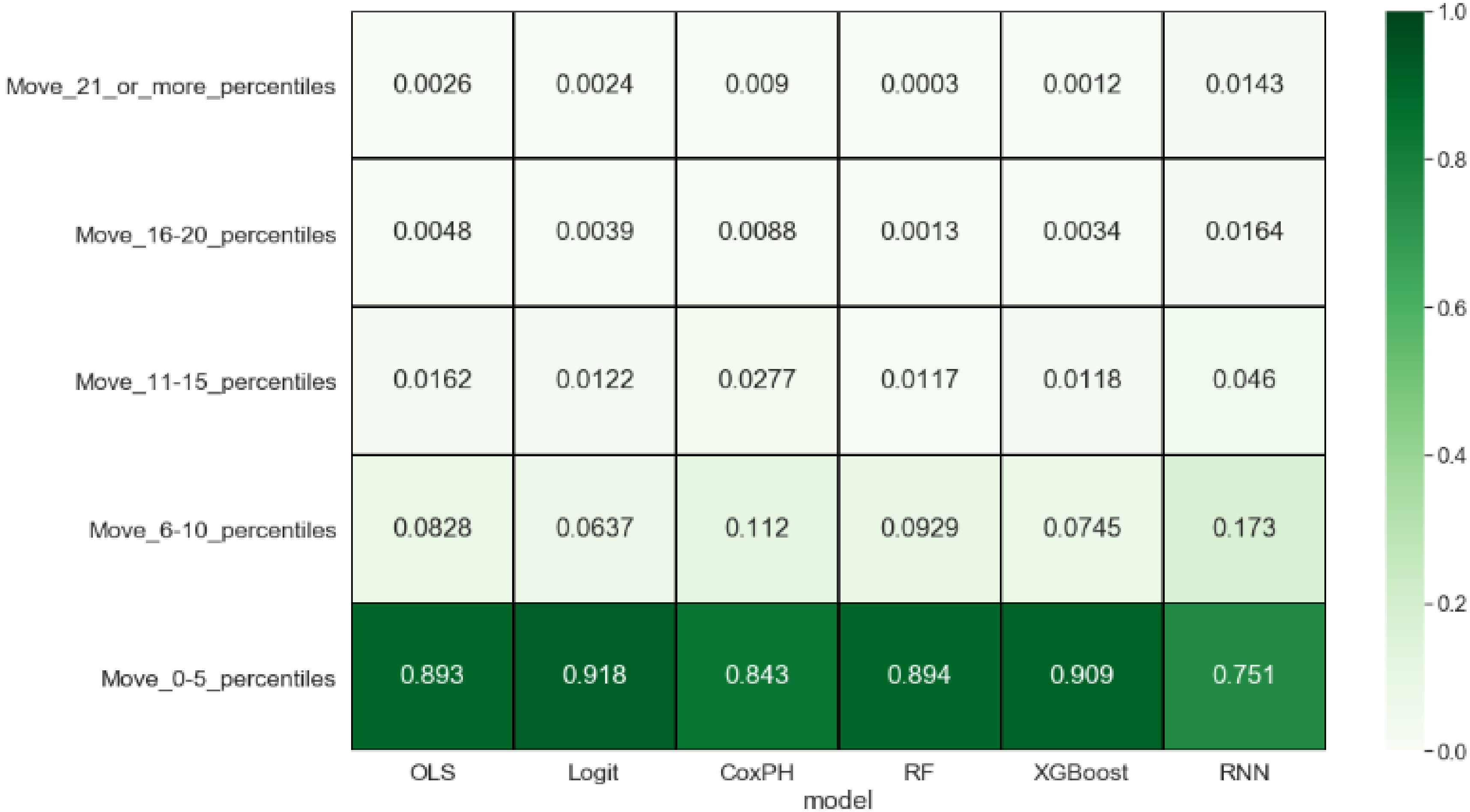
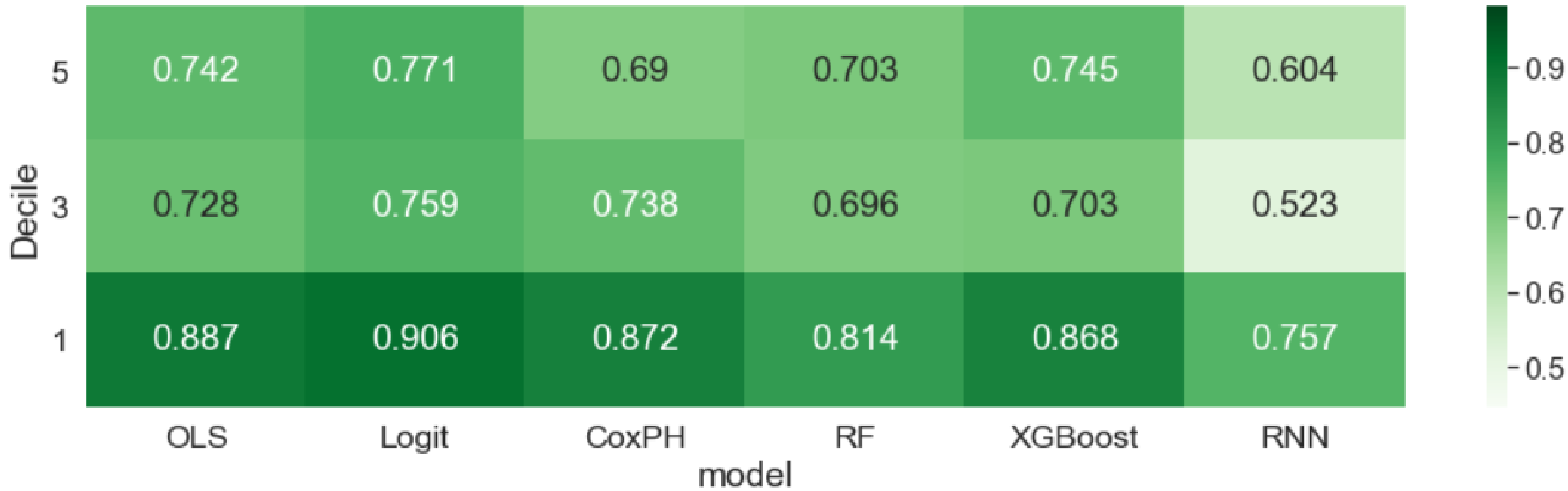


Figure A9: Student-level differences in risk ranking percentile, base models versus models exclude the complexly specified term-specific predictors



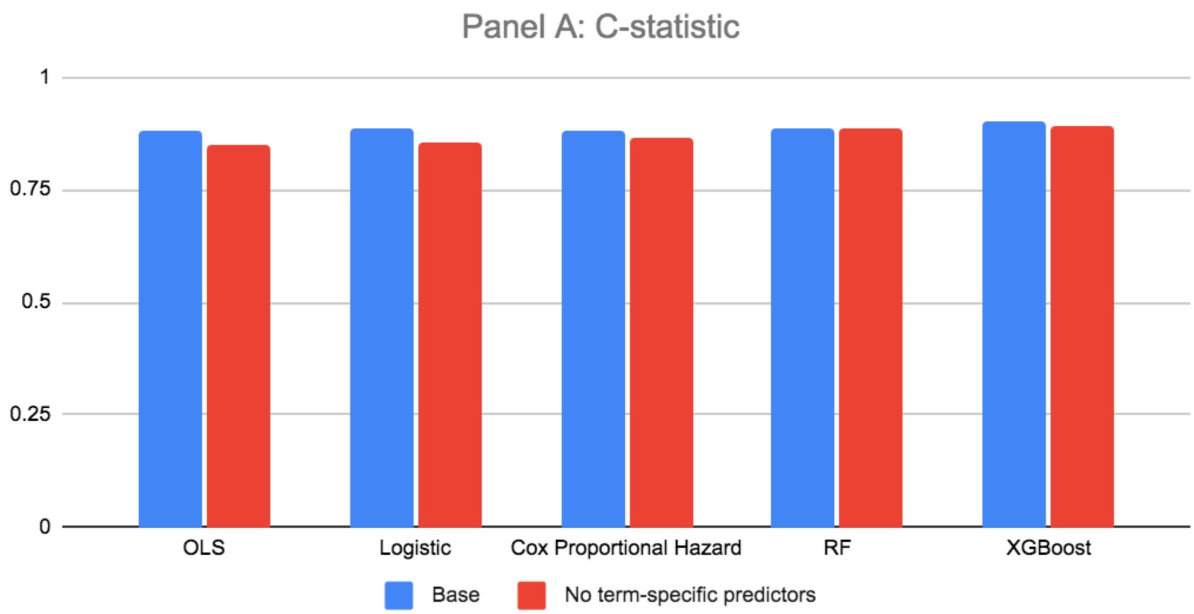
Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, XGBoost, or RNN), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model excluding the complexly specified term-specific predictors. These changes in risk ranking percentiles are measured in absolute value.

Figure A10: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models that exclude the complexly specified term-specific predictors



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, XGBoost, or RNN), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in the model excluding the complexly specified term-specific predictors.

Figure A11: Evaluation metrics, base models versus models excluding all term-specific predictors



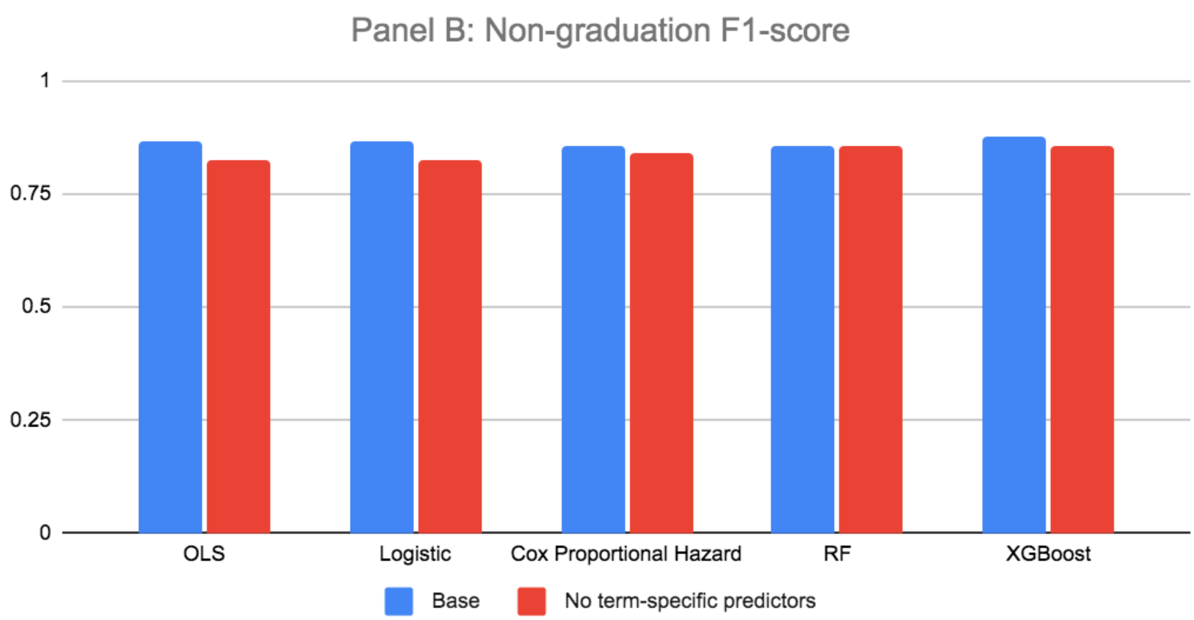
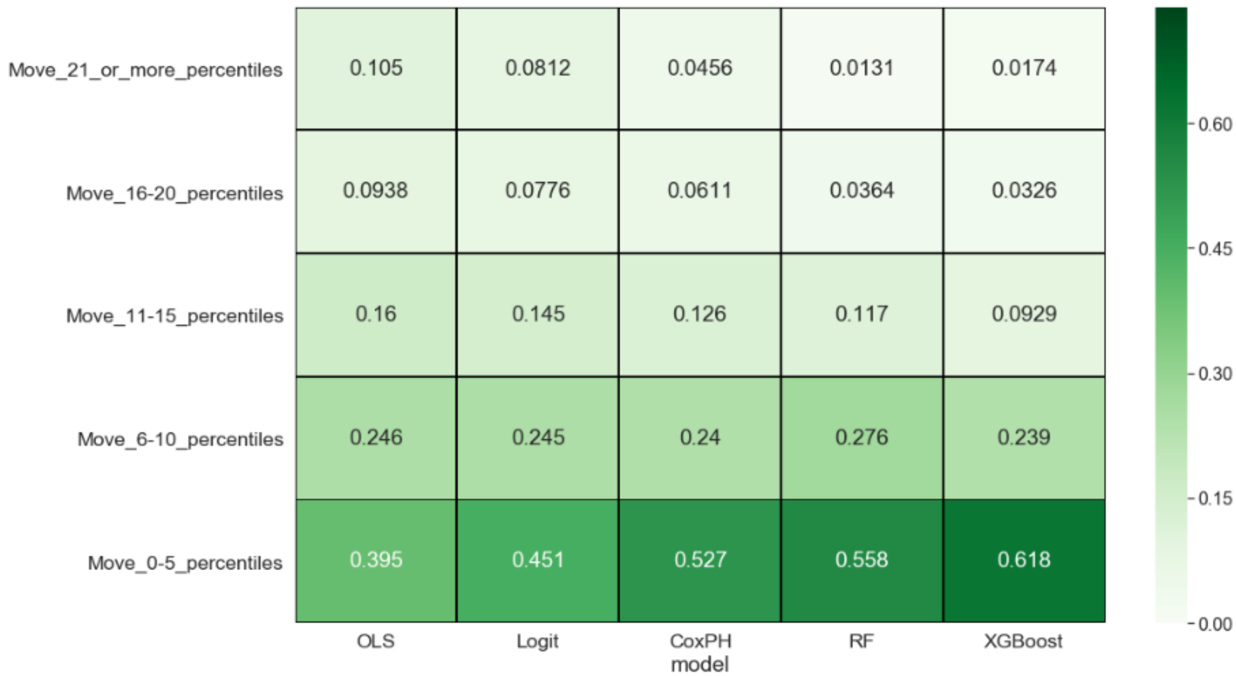


Figure A12: Student-level differences in risk ranking percentile, base models versus models excluding all term-specific predictors



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model excluding term-specific predictors. These changes in risk ranking percentiles are measured in absolute value.

Figure A13: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models excluding all term-specific predictors

Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in the model excluding term-specific predictors.

Figure A14: Evaluation Statistics, base models versus models that only include the simple non-term-specific predictors

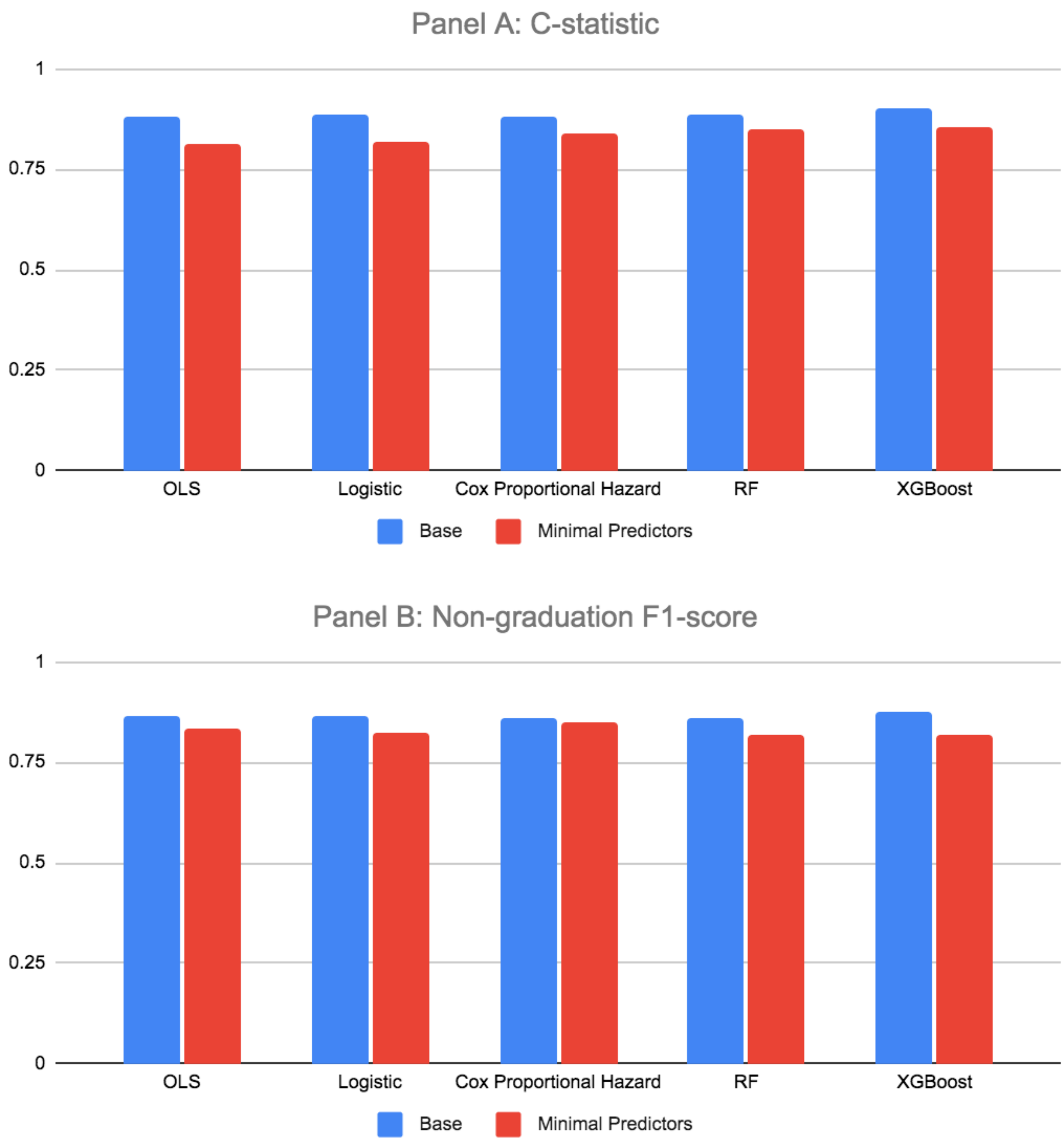
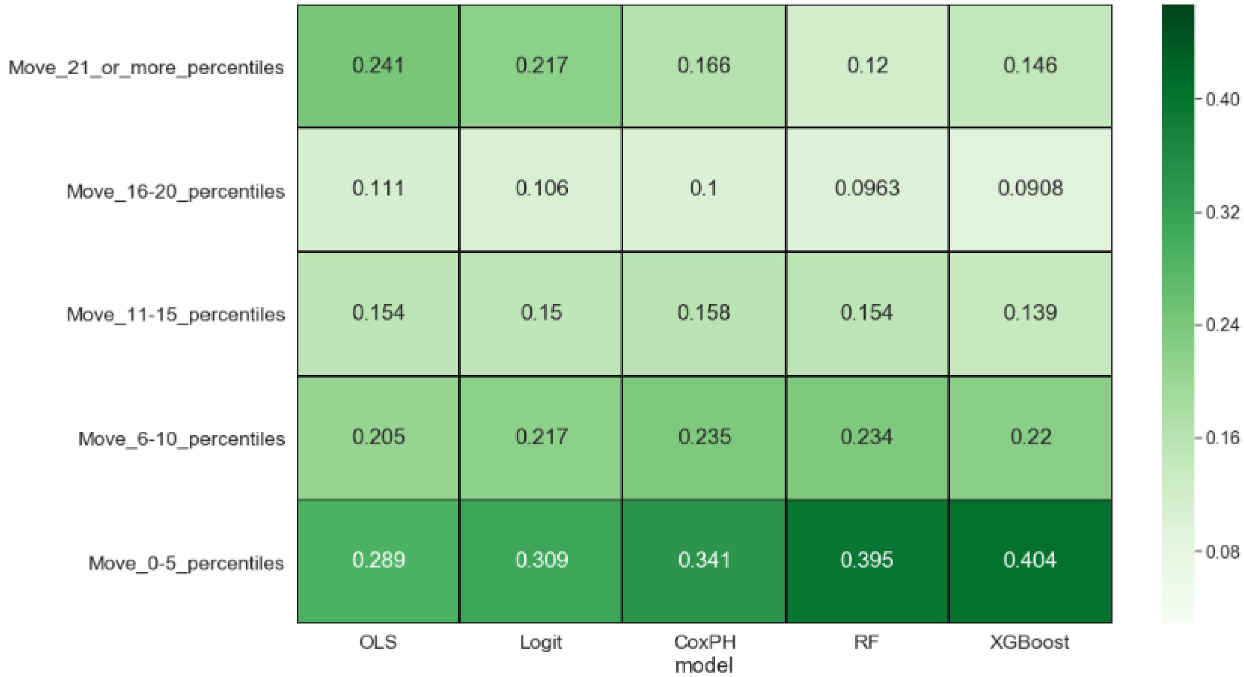
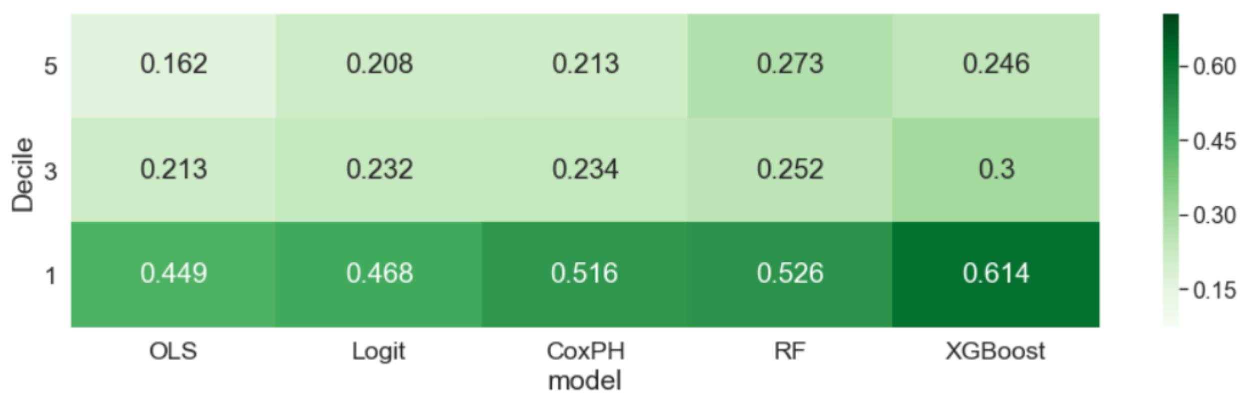


Figure A15: Student-level differences in risk ranking percentile, base models versus models that only include the simple non-term-specific predictors



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model that only includes the simple non-term-specific predictors. These changes in risk ranking percentiles are measured in absolute value.

Figure A16: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models that only include the simple non-term-specific predictors



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in the model that only includes the simple non-term-specific predictors.

Figure A17: Evaluation statistics, base models versus models with 147 selected predictors

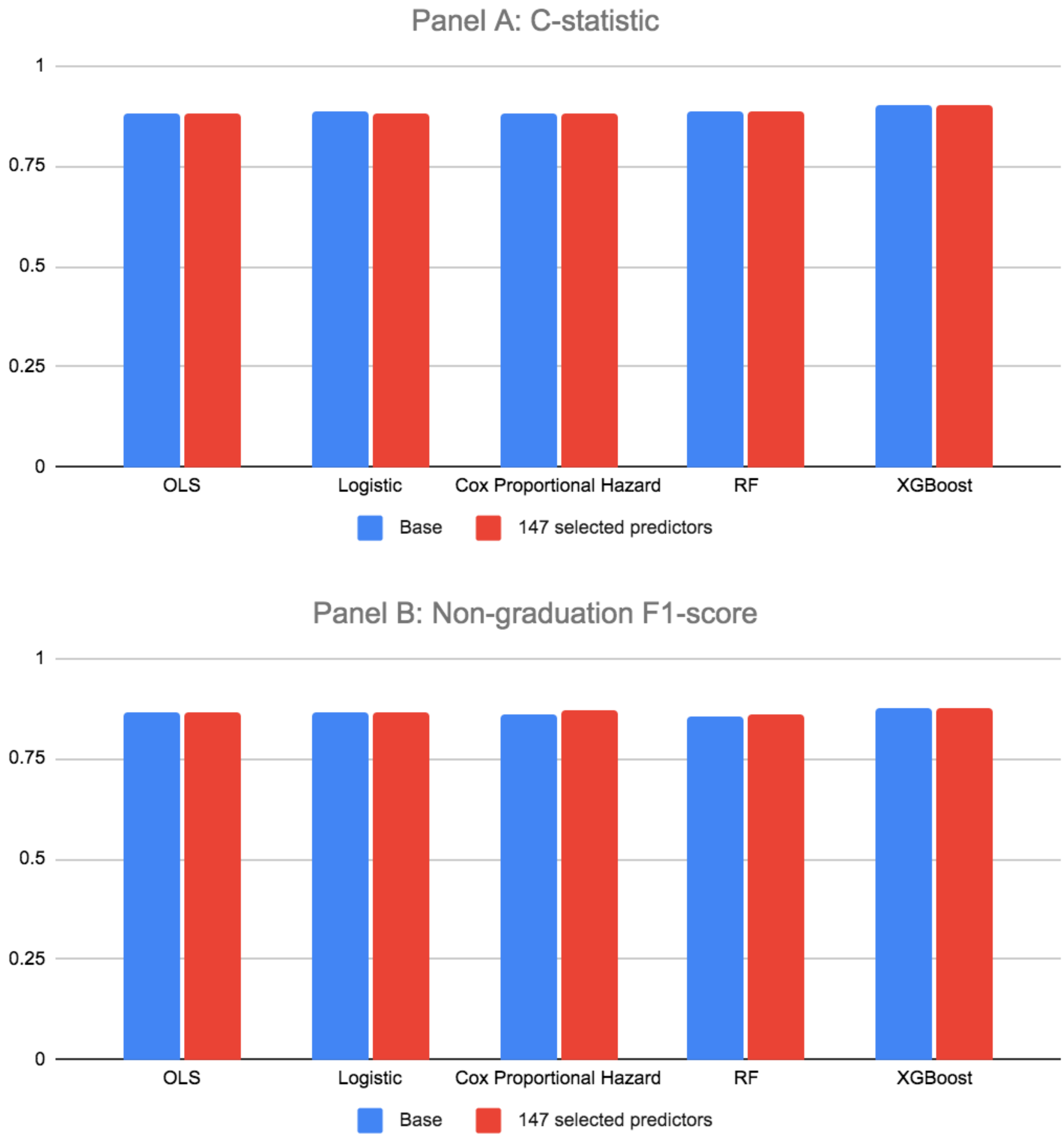
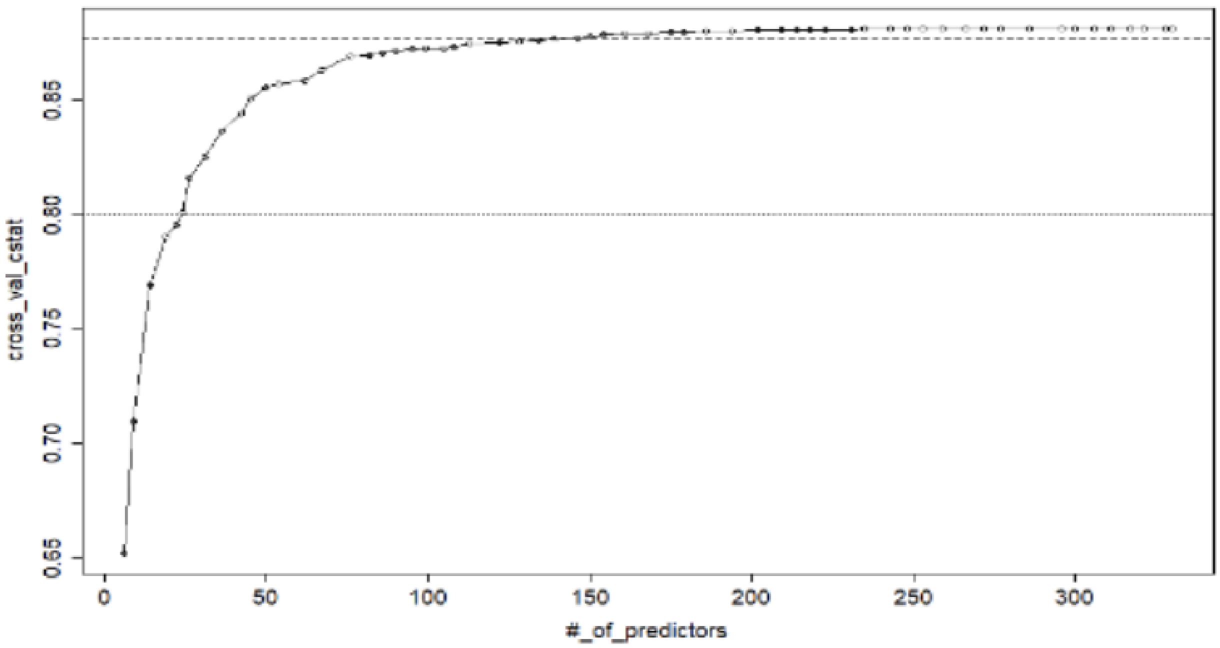
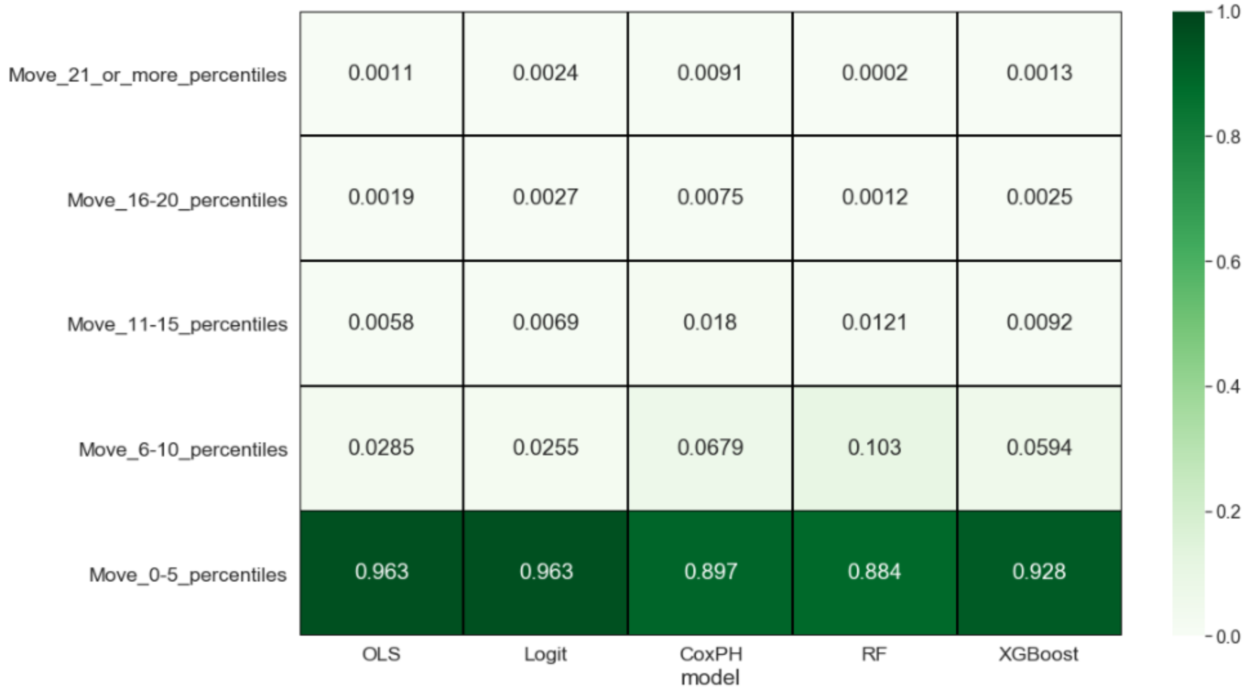


Figure A18: Relationship between c-statistic and number of predictors, using penalized Logistic feature selection



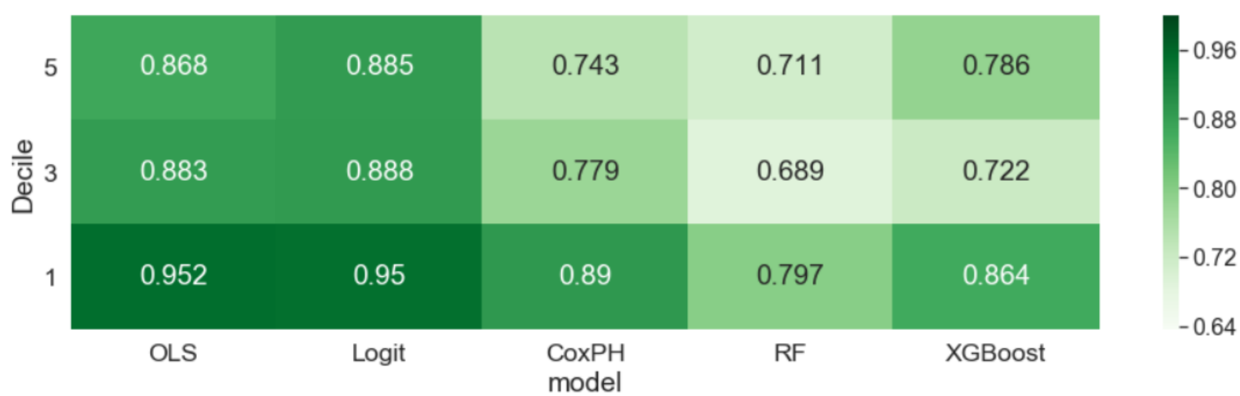
Notes: this figure shows the relationship between the 10-fold cross-validation c-statistic (y-axis) and the number of predictors left in the model, as a result of a stepwise increase in the tuning parameter of the penalized logistic feature selection process. Specifically, we slightly increased the tuning parameter so that the model becomes gradually more selective of which predictors to keep in the model. The upper dashed horizontal line denotes the c-statistic values for the model using the 2-SE selection rule, which crosses the curve at 147 predictors. The lower dotted horizontal line is positioned on a c-statistic value of 0.80, which is a common lower-bound benchmark of acceptable performance.

Figure A19: Student-level differences in risk ranking percentile, base models versus models with 147 selected predictors



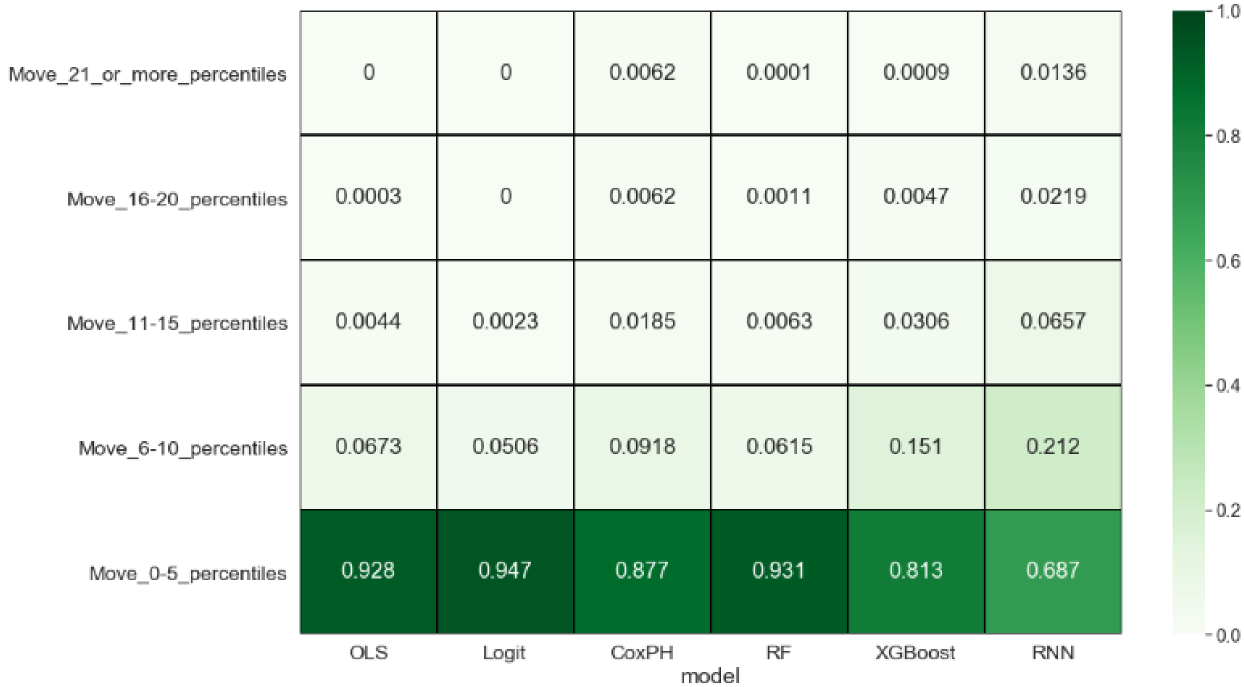
Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model with 147 selected predictors. These changes in risk ranking percentiles are measured in absolute value.

Figure A20: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models with 147 selected predictors



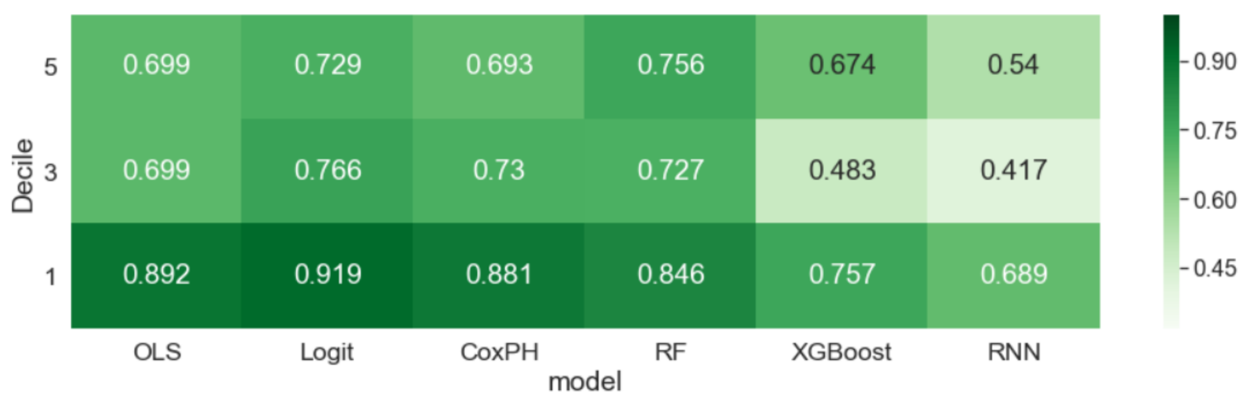
Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the bottom quartile or decile in the model with 147 selected predictors.

Figure A21: Student-level differences in risk ranking percentile, base models versus models excluding demographic predictors



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, XGBoost, or RNN), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model excluding demographic predictors. These changes in risk ranking percentiles are measured in absolute value.

Figure A22: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models excluding demographic predictors



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, XGBoost, or RNN), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in the model excluding demographic predictors.

Figure A23: Evaluation statistics, base models versus PVCC-only models

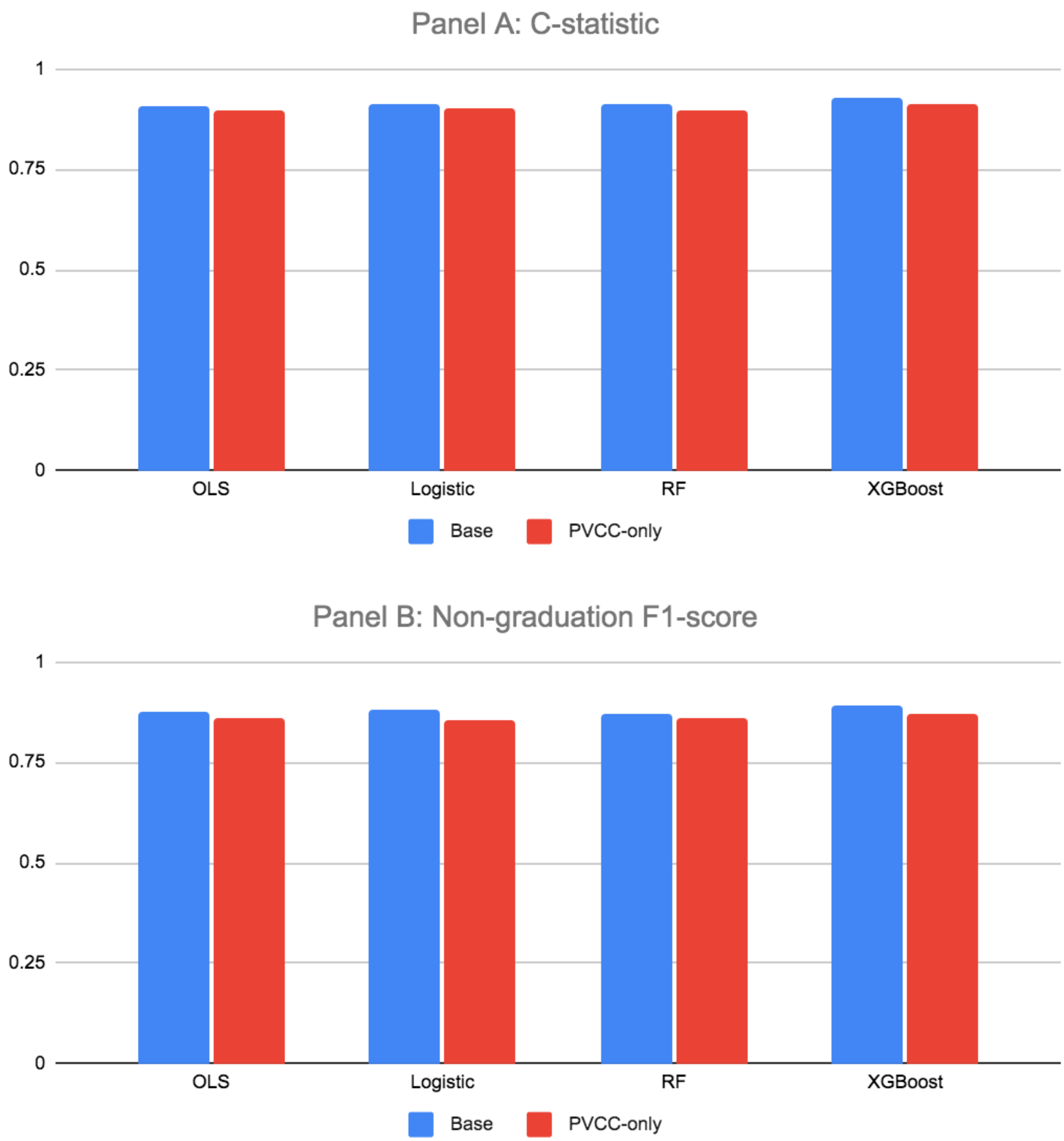
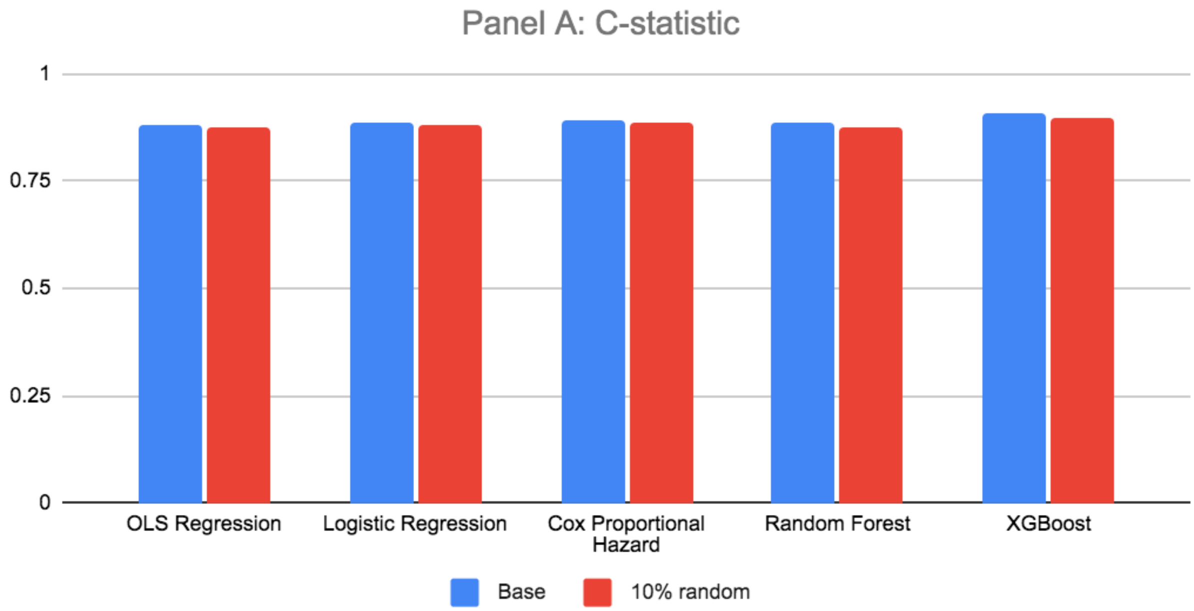
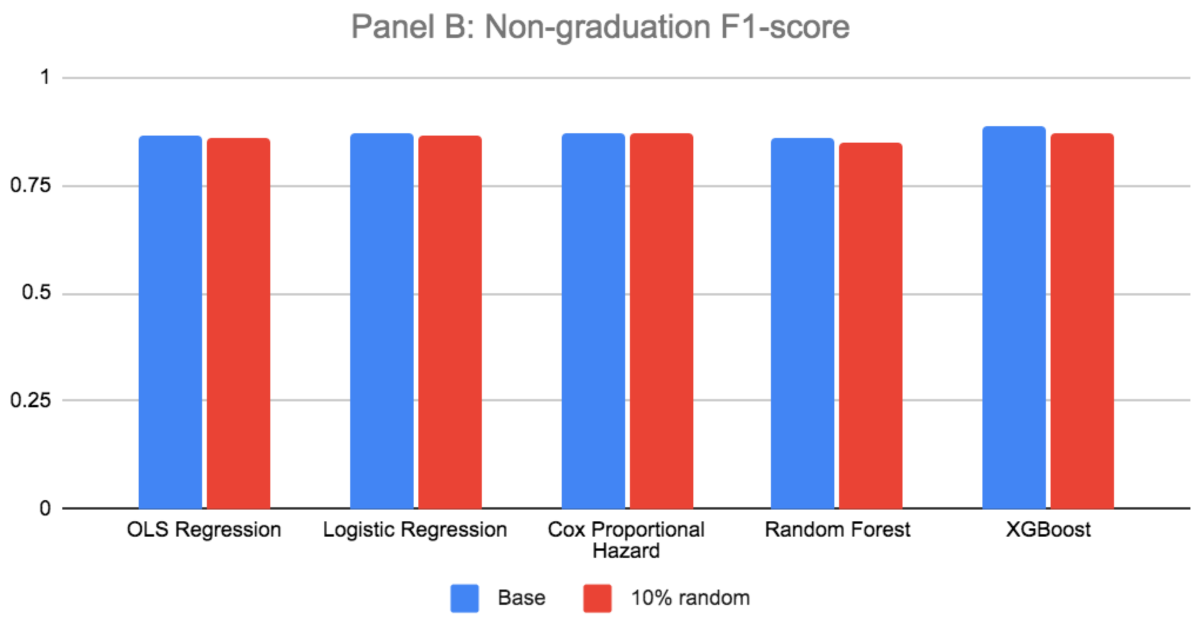


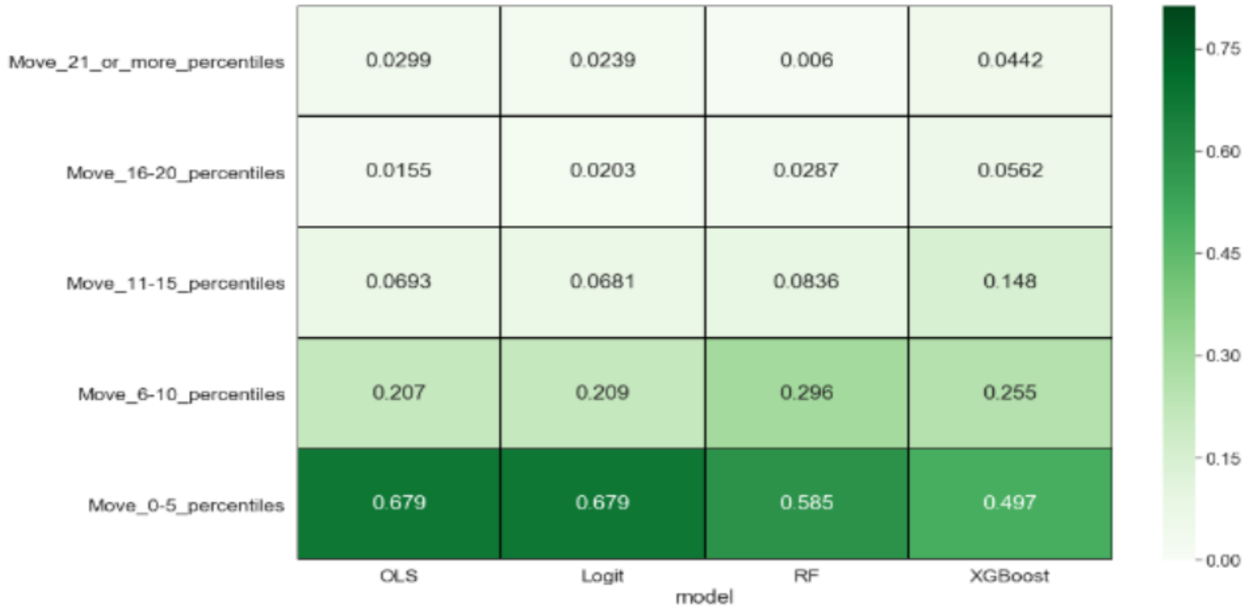
Figure A24: Evaluation statistics, base models versus 10% random sample models





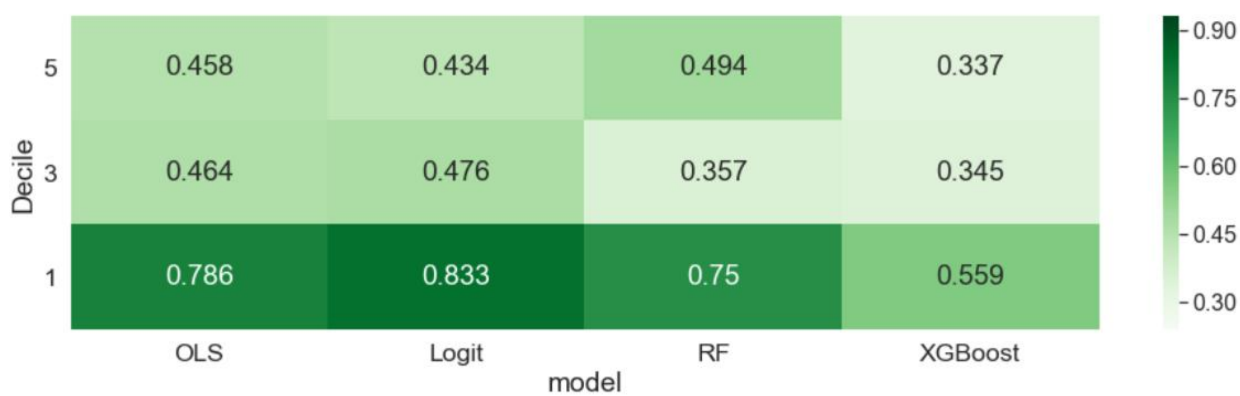
Notes: we use the 10% random validation sample to compute the evaluation statistics for both the base models (using the full training sample) and the 10% random sample models.

Figure A25: Student-level differences in risk ranking percentile, base models versus PVCC-only models



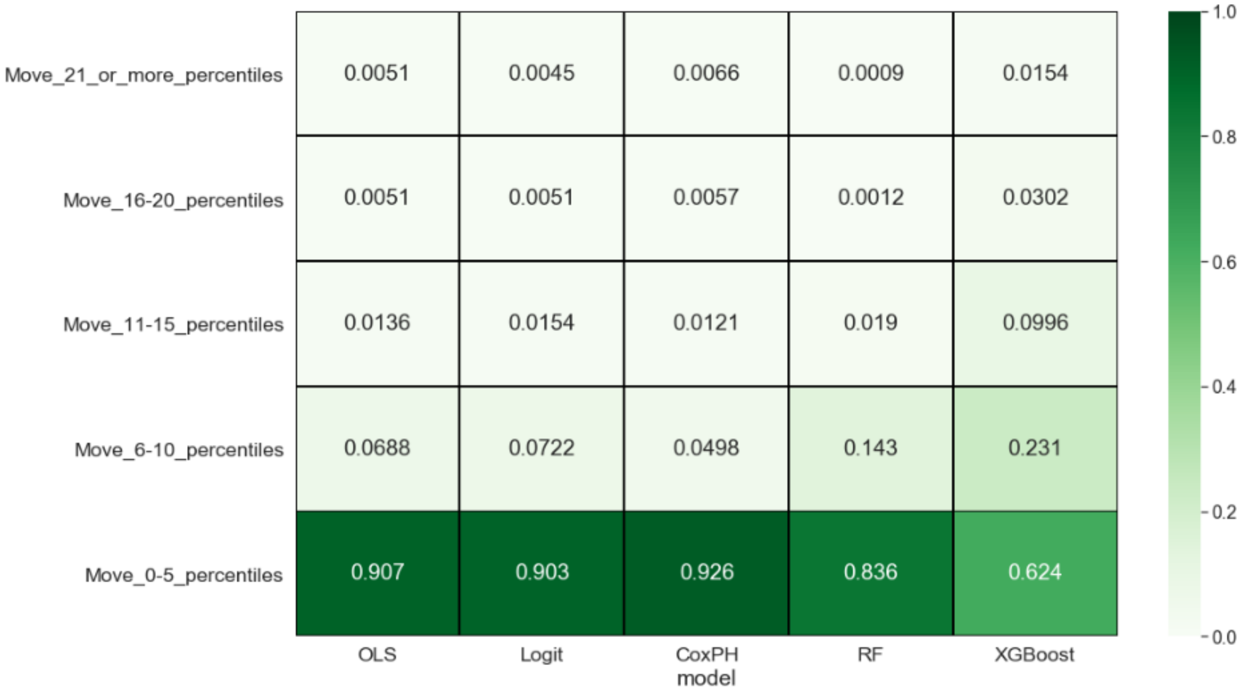
Notes: this figure shows, within a given model type (OLS, Logistic, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the PVCC-only model. These changes in risk ranking percentiles are measured in absolute value. In calculating these differences, we use the PVCC-only validation sample for the base models as well as the PVCC-only model.

Figure A26: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus PVCC-only models



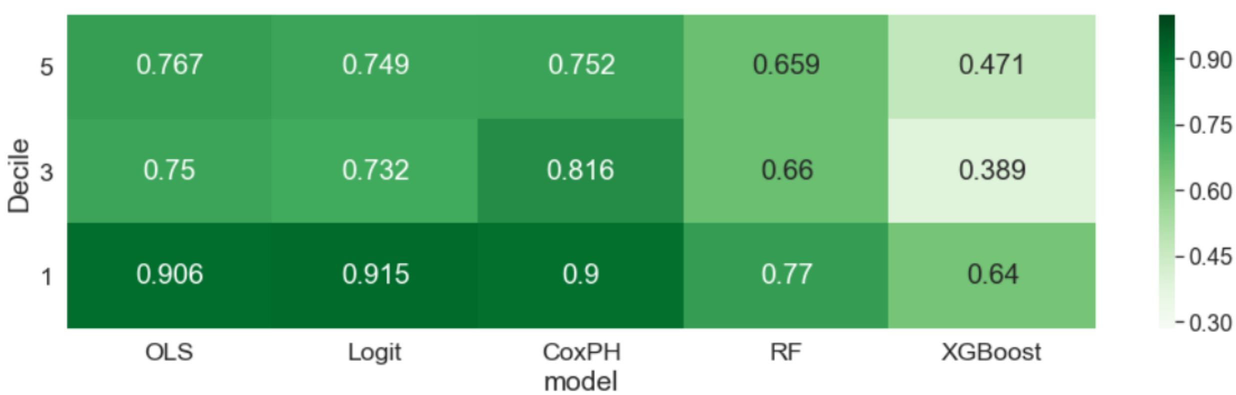
Notes: this figure shows, within a given model type (OLS, Logistic, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in PVCC-only model. In calculating these differences, we use the PVCC-only validation sample for the base models as well as the PVCC-only model.

Figure A27: Student-level differences in risk ranking percentile, base models versus 10% random sample models



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the 10% random sample model. These changes in risk ranking percentiles are measured in absolute value. In calculating these differences, we use the 10% random validation sample for the base models as well as the 10% random sample model.

Figure A28: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus 10% random sample models



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in 10% random sample model. In calculating these differences, we use the 10% random validation sample for the base models as well as the 10% random sample model.

Figure A29: Evaluation statistics, base models versus models excluding NSC data

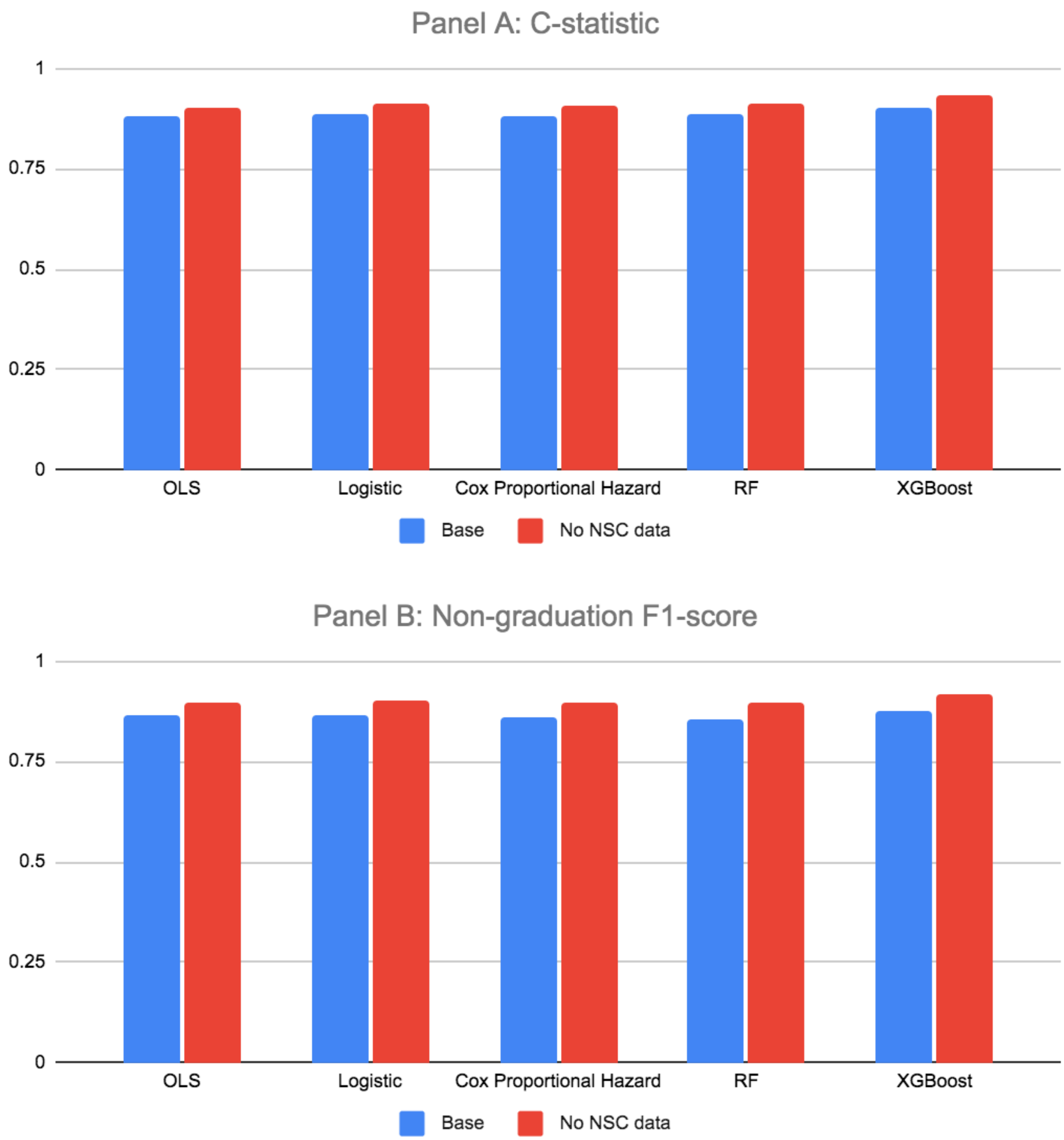
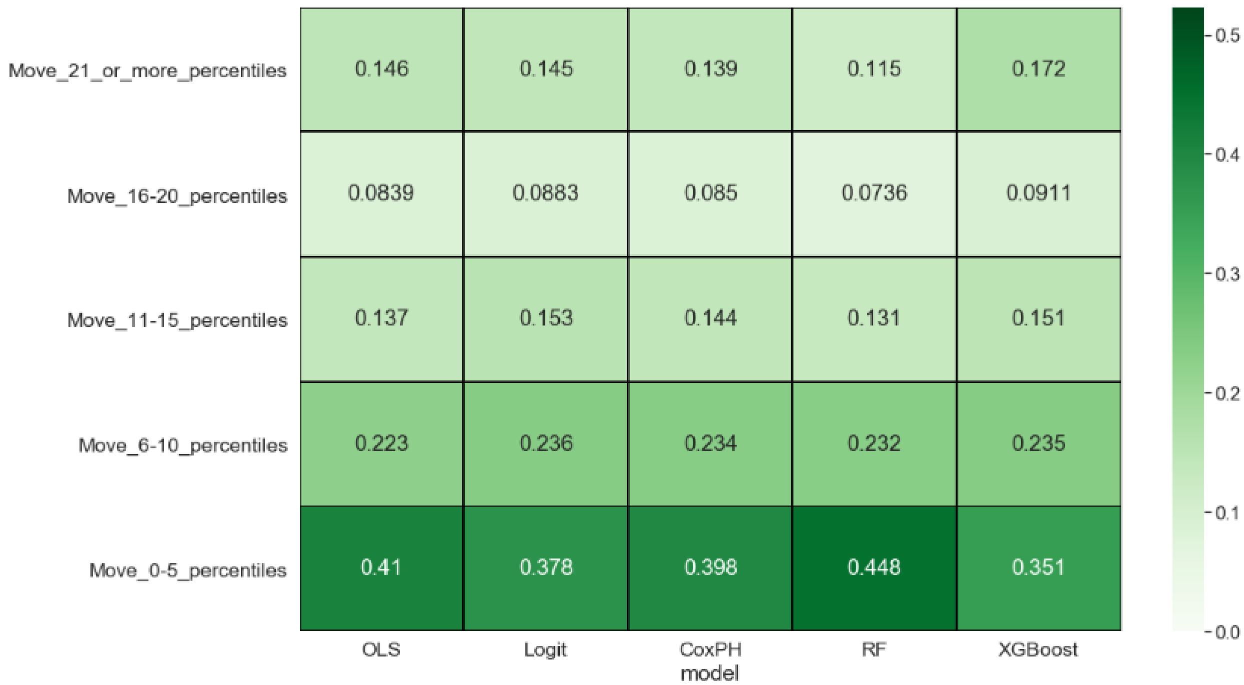


Figure A30: Student-level differences in risk ranking percentile, base models versus models excluding NSC data

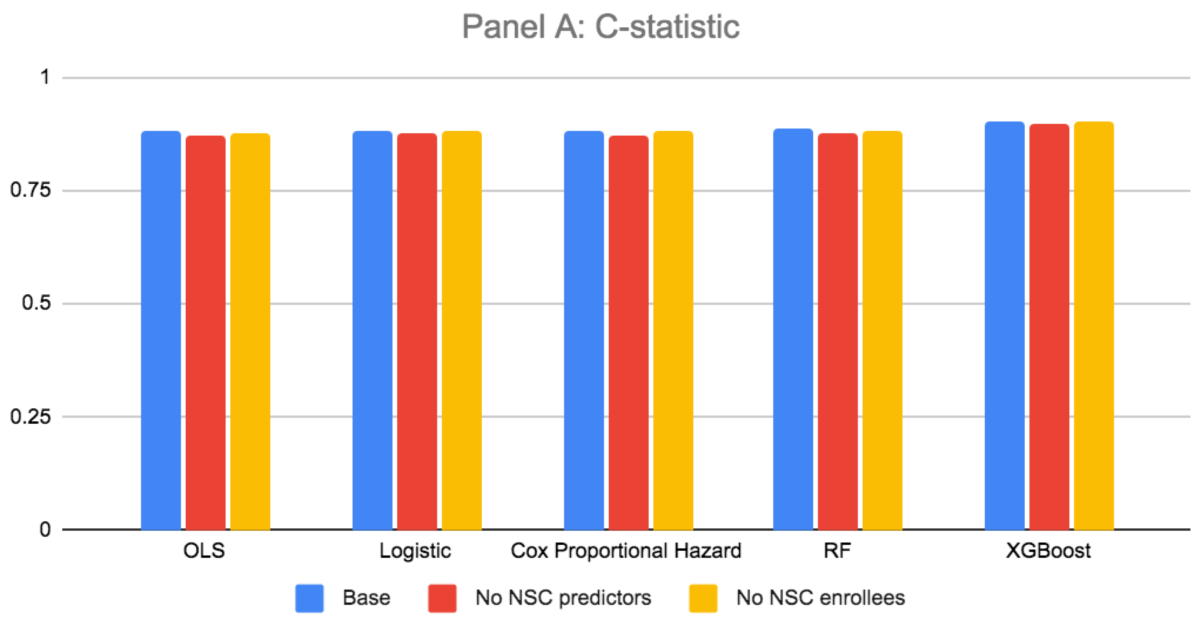


Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model excluding all NSC data (both in the construction of predictors and the outcome of interest). These changes in risk ranking percentiles are measured in absolute value.

Figure A31: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models excluding NSC data

Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in the model excluding NSC data (from predictor construction and outcome definition).

Figure A32: Evaluation statistics for base models, models excluding NSC predictors, and models excluding NSC enrollees



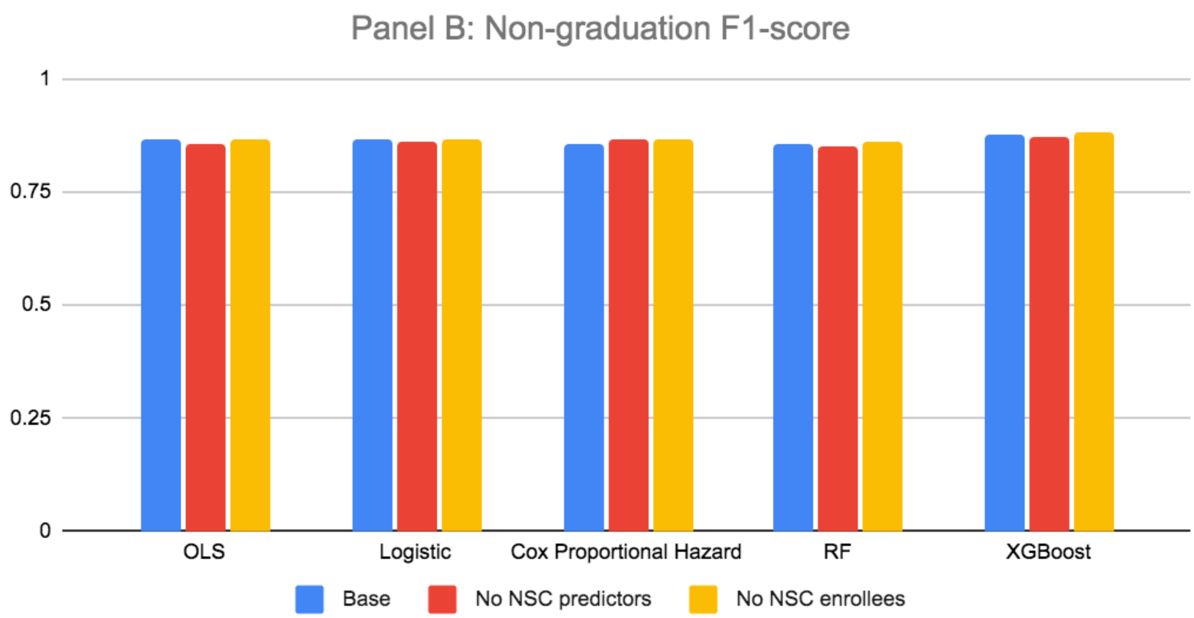
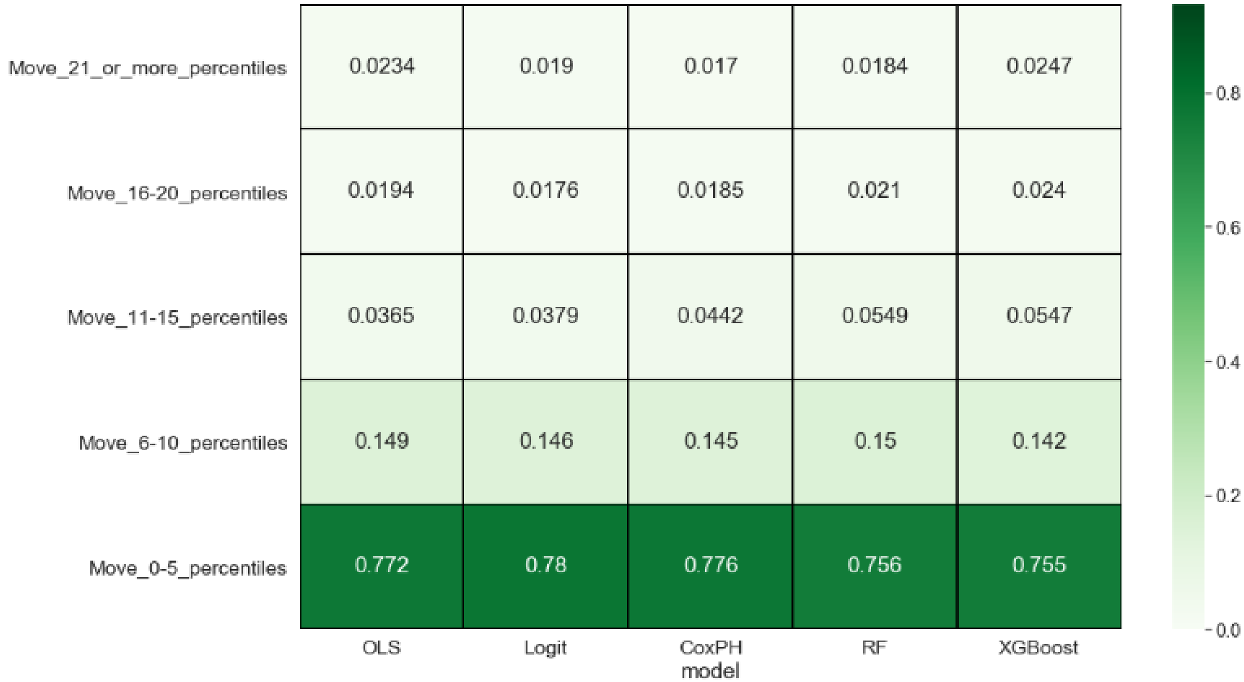
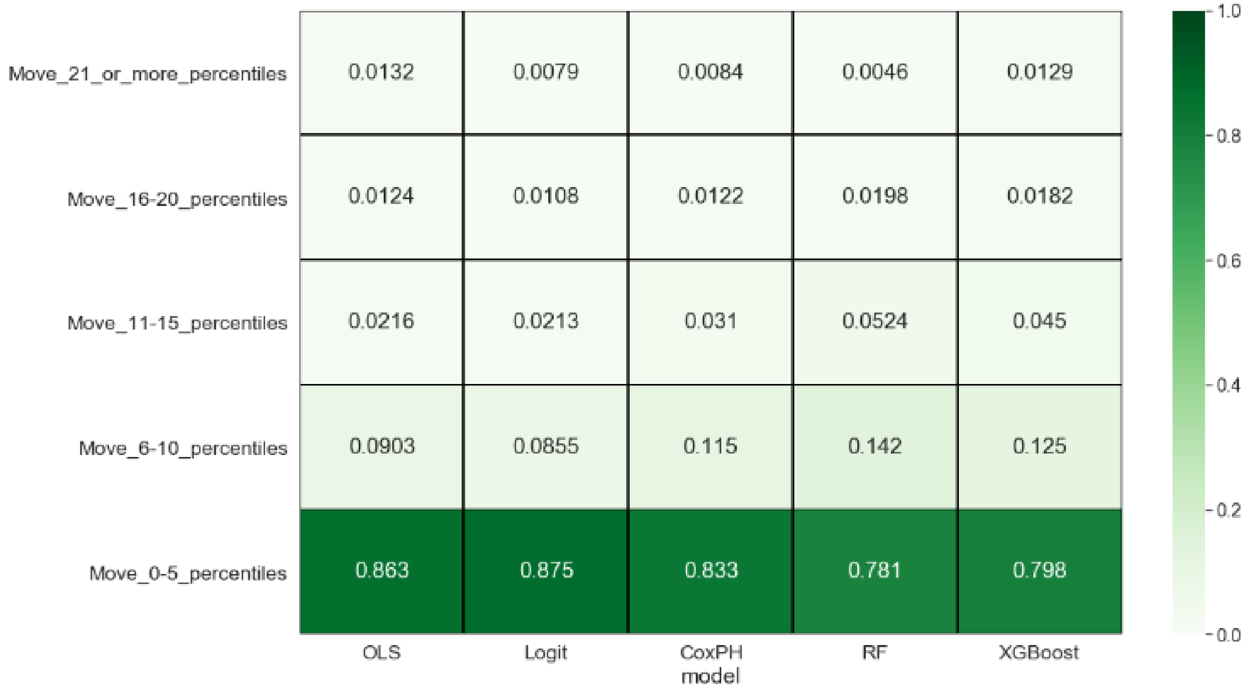


Figure A33: Student-level differences in risk ranking percentile, base models versus models without NSC predictors



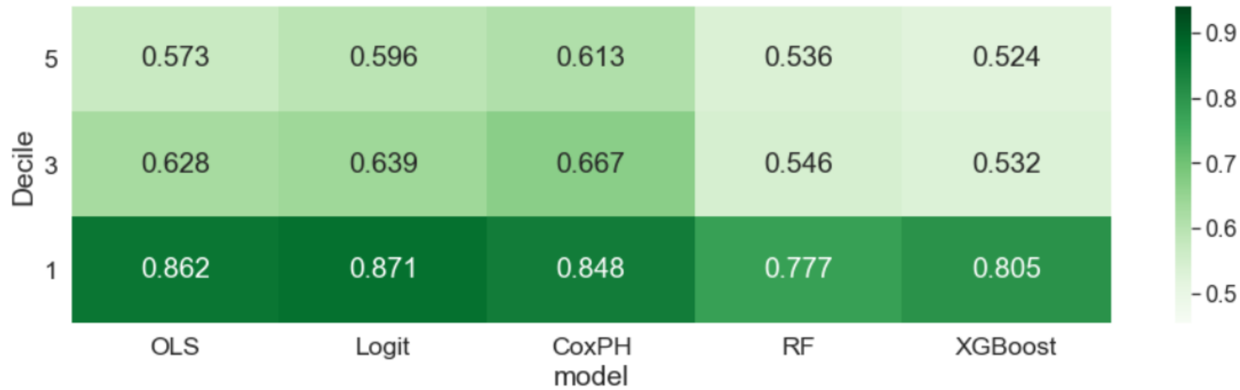
Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model excluding NSC predictors. These changes in risk ranking percentiles are measured in absolute value.

Figure A34: Student-level differences in risk ranking percentile, base models versus models without NSC enrollees



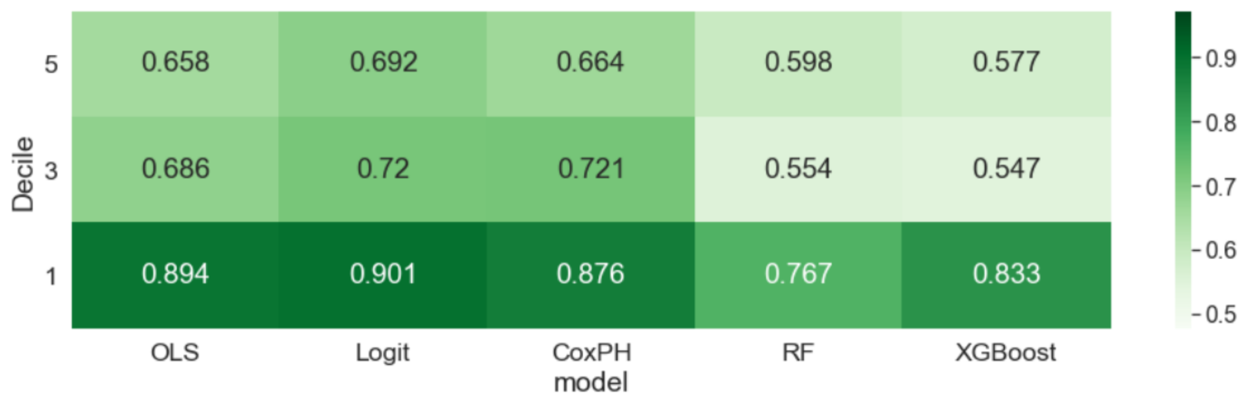
Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students whose risk ranking percentile changes by a certain amount between the base model and the model excluding NSC enrollees. These changes in risk ranking percentiles are measured in absolute value.

Figure A35: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models excluding NSC predictors



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in the model excluding NSC predictors.

Figure A36: Consistency of 1st, 3rd & 5th deciles of risk rankings, base models versus models excluding NSC enrollees



Notes: this figure shows, within a given model type (OLS, Logistic, CPH, Random Forest, or XGBoost), the share of students assigned to the 5th decile (top row), 3rd decile (middle row) and bottom decile (bottom row) by the base model who are also assigned to the same deciles in the model excluding NSC enrollees.