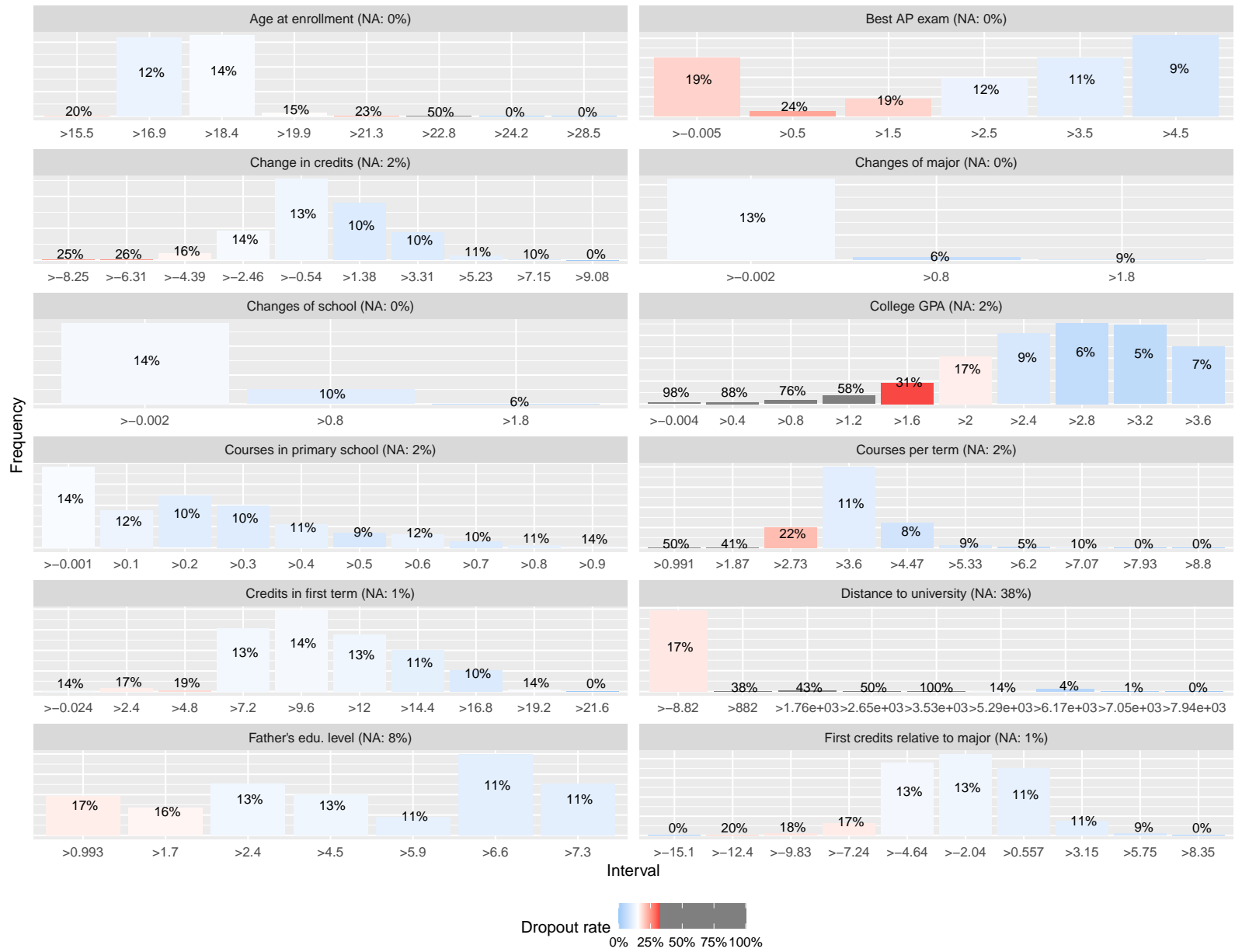# Supplementary material: Temporal and Between-Group Variability in College Dropout Prediction

Glandorf et al.
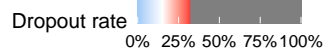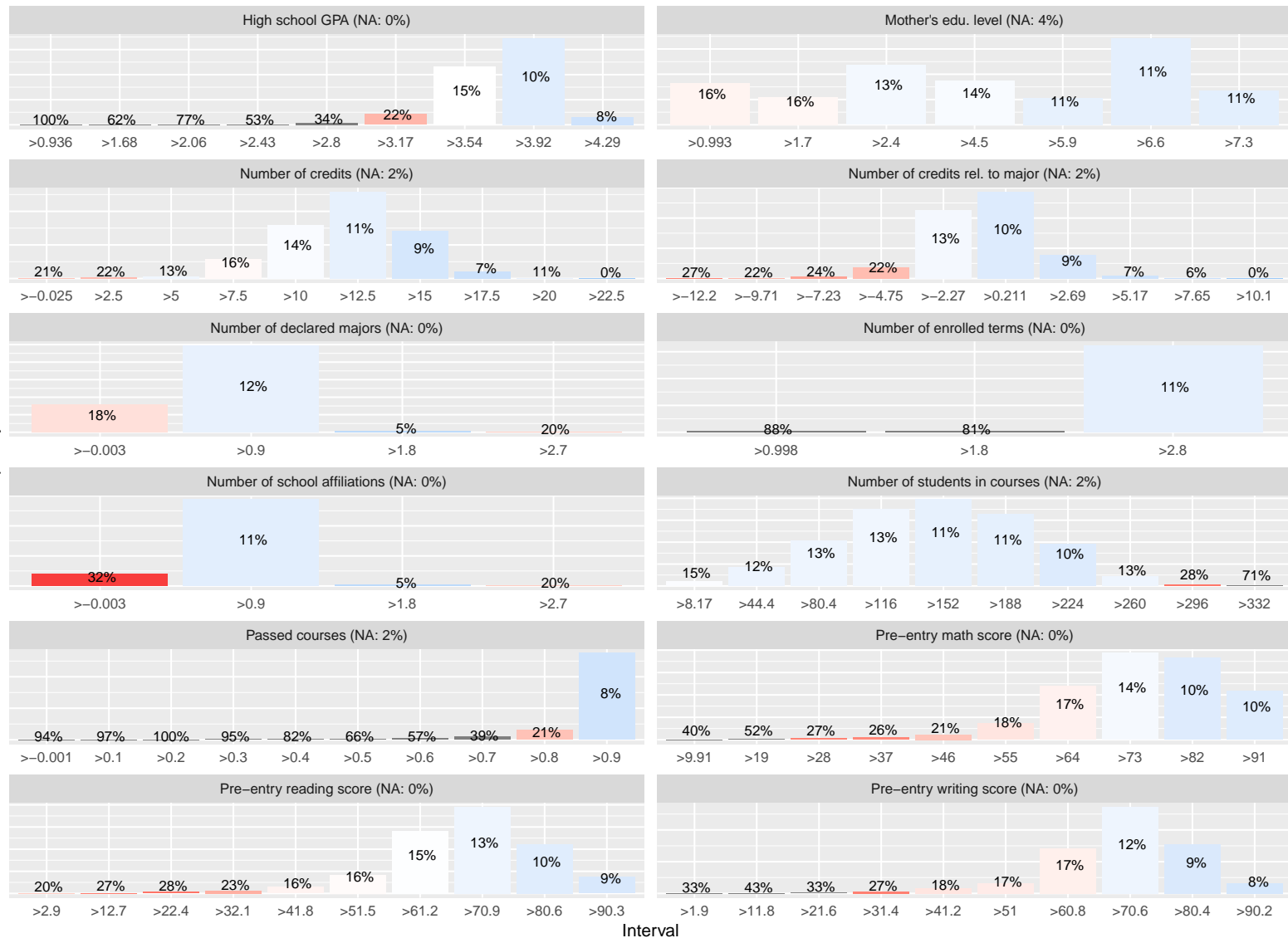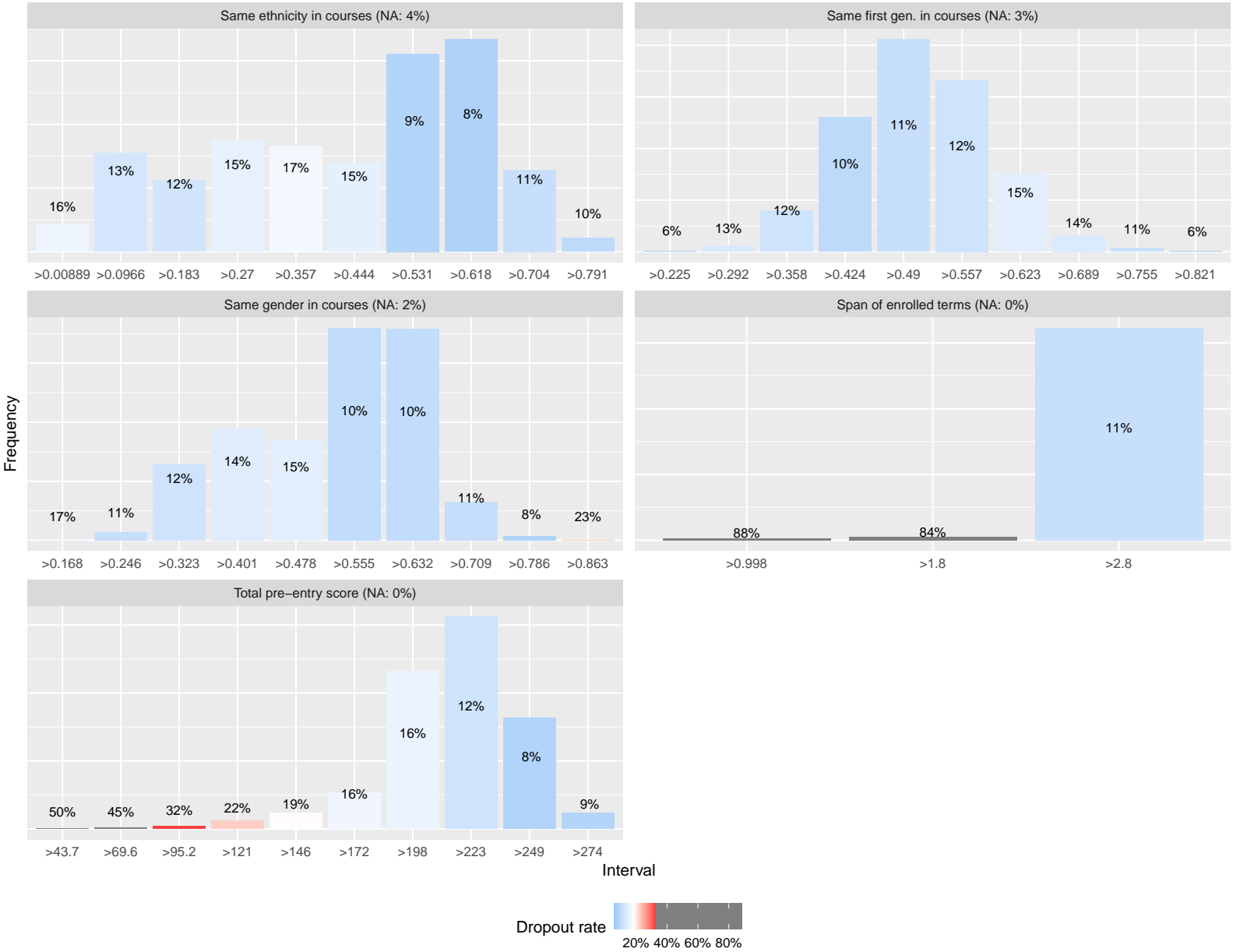
December 12, 2023

1 Descriptive statistics

The following plots depict the histograms for numerical and non-numerical predictors after the first year. For each bar, the corresponding dropout rate is indicated by color and the label.

High school GPA (NA: 0%)

100%	62%	77%	53%	34%	22%	15%	10%	8%
>0.936	>1.68	>2.06	>2.43	>2.8	>3.17	>3.54	>3.92	>4.29

Mother's edu. level (NA: 4%)

16%	16%	13%	14%	11%	11%	11%
>0.993	>1.7	>2.4	>4.5	>5.9	>6.6	>7.3

Number of credits (NA: 2%)

21%	22%	13%	16%	14%	11%	9%	7%	11%	0%
>−0.025	>2.5	>5	>7.5	>10	>12.5	>15	>17.5	>20	>22.5

Number of credits rel. to major (NA: 2%)

27%	22%	24%	22%	13%	10%	9%	7%	6%	0%
>−12.2	>−9.71	>−7.23	>−4.75	>−2.27	>0.211	>2.69	>5.17	>7.65	>10.1

Number of declared majors (NA: 0%)

18%	12%	5%	20%
>−0.003	>0.9	>1.8	>2.7

Number of enrolled terms (NA: 0%)

88%	81%	11%
>0.998	>1.8	>2.8

Number of school affiliations (NA: 0%)

32%	11%	5%	20%
>−0.003	>0.9	>1.8	>2.7

Number of students in courses (NA: 2%)

15%	12%	13%	13%	11%	11%	10%	13%	28%	71%
>8.17	>44.4	>80.4	>116	>152	>188	>224	>260	>296	>332

Passed courses (NA: 2%)

94%	97%	100%	95%	82%	66%	57%	39%	21%	8%
>−0.001	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9

Pre−entry math score (NA: 0%)

40%	52%	27%	26%	21%	18%	17%	14%	10%	10%
>9.91	>19	>28	>37	>46	>55	>64	>73	>82	>91

Pre−entry reading score (NA: 0%)

20%	27%	28%	23%	16%	16%	15%	13%	10%	9%
>2.9	>12.7	>22.4	>32.1	>41.8	>51.5	>61.2	>70.9	>80.6	>90.3

Pre−entry writing score (NA: 0%)

33%	43%	33%	27%	18%	17%	17%	12%	9%	8%
>1.9	>11.8	>21.6	>31.4	>41.2	>51	>60.8	>70.6	>80.4	>90.2

Frequency — Interval

Dropout rate: 0% 25% 50% 75% 100%

Same ethnicity in courses (NA: 4%)

Interval	Value
>0.00889	16%
>0.0966	13%
>0.183	12%
>0.27	15%
>0.357	17%
>0.444	15%
>0.531	9%
>0.618	8%
>0.704	11%
>0.791	10%

Same first gen. in courses (NA: 3%)

Interval	Value
>0.225	6%
>0.292	13%
>0.358	12%
>0.424	10%
>0.49	11%
>0.557	12%
>0.623	15%
>0.689	14%
>0.755	11%
>0.821	6%

Same gender in courses (NA: 2%)

Interval	Value
>0.168	17%
>0.246	11%
>0.323	12%
>0.401	14%
>0.478	15%
>0.555	10%
>0.632	10%
>0.709	11%
>0.786	8%
>0.863	23%

Span of enrolled terms (NA: 0%)

Interval	Value
>0.998	88%
>1.8	84%
>2.8	11%

Total pre-entry score (NA: 0%)

Interval	Value
>43.7	50%
>69.6	45%
>95.2	32%
>121	22%
>146	19%
>172	16%
>198	16%
>223	12%
>249	8%
>274	9%

Frequency

Interval

Dropout rate 20% 40% 60% 80%

STEM major (NA: 0%)	Residency at application (NA: 0%)	Citizenship (NA: 1%)	English language learner (NA: 0%)
FALSE 15%, TRUE 11%	Bona... 15%, Cali... 12%, Non-... 16%	Not ... 16%, Perm... 3%, US C... 13%, NA 7%	FALSE 14%, TRUE 5%

Ethnicity (NA: 2%)	Female (NA: 0%)	First generation (NA: 1%)	1st term year of study (NA: 0%)
Asia... 10%, Black 19%, Hisp... 19%, Indi... 17%, Whit... 14%, NA 12%	FALSE 15%, TRUE 11%, NA 24%	FALSE 11%, TRUE 15%, NA 26%	Fres... 15%, Junior 11%, Soph... 7%

Geographic category (NA: 0%)	International student (NA: 0%)	Low-income family (NA: 0%)	Honors (NA: 0%)
Fore... 16%, Non-... 14%, Oran... 11%, Othe... 11%, Out ... 22%	no 13%, yesN... 20%, yesW... 13%, NA 100%	FALSE 12%, TRUE 14%, NA 95%	FALSE 13%, TRUE 5%

Current year of study (NA: 0%)
Fres... 34%, Junior 5%, Senior 5%, Soph... 9%

Frequency

Value

Dropout rate   0.25 0.50 0.75 1.00

Primary major name (NA: 0%)

Frequency

Major	Rate
AEROSPACE ENGINEERING	14%
ANTHROPOLOGY	15%
ART	18%
BIOLOGICAL SCIENCES	11%
BIOMEDICAL ENGINEERING	9%
BUSINESS ADMINISTRATION	7%
BUSINESS ECONOMICS	10%
BUSINESS INFORMATION M...	13%
CHEMICAL ENGINEERING	11%
CHEMISTRY	14%
CIVIL ENGINEERING	12%
COMPUTER ENGINEERING	13%
COMPUTER GAME SCIENCE	15%
COMPUTER SCIENCE	9%
COMPUTER SCIENCE AND E...	10%
CRIMINOLOGY, LAW AND S...	13%
DRAMA	12%
ECONOMICS	14%
EDUCATION SCIENCES	8%
ELECTRICAL ENGINEERING	15%
ENGLISH	12%
FILM AND MEDIA STUDIES	13%
HISTORY	10%
INFORMATICS	10%
INTERNATIONAL STUDIES	11%
MATHEMATICS	13%
MECHANICAL ENGINEERING	12%
NURSING SCIENCE	6%
OTHER	13%
PHARMACEUTICAL SCIENCES	9%
PHYSICS	21%
POLITICAL SCIENCE	10%
PSYCHOLOGY AND SOCIAL ...	11%
PSYCHOLOGY_BA	13%
PUBLIC HEALTH POLICY	7%
PUBLIC HEALTH SCIENCES	12%
QUANTITATIVE ECONOMICS	4%
SOCIAL ECOLOGY	11%
SOCIOLOGY	13%
SOFTWARE ENGINEERING	10%
UNDECL–UNAFF	18%

Primarily affiliated school (NA: 2%)

School	Rate
Div of Undergraduate E...	23%
Pharmacy	8%
Program in Public Health	9%
School of Biological S...	10%
School of Business	6%
School of Education	7%
School of Engineering	11%
School of Humanities	12%
School of Info & Compu...	11%
School of Nursing	6%
School of Physical Sci...	14%
School of Social Ecology	10%
School of Social Sciences	10%
School of the Arts	14%
NA	78%

Value

Dropout rate
0.2 0.4 0.6

2 RQ1

The following table depicts the same content as Table 2 in the paper for the model comparison but includes the standard deviations induced by the missing data imputation.

Time of prediction	After first year			After second year		
Model	AUPRC	AUROC	Accuracy	AUPRC	AUROC	Accuracy
No predictors	0.131 (0.0000)	0.500 (0.0000)	0.869 (0.0000)	0.112 (0.0000)	0.500 (0.0000)	0.888 (0.0000)
Logistic Regression	0.553 (0.0022)	0.816 (0.0008)	0.898 (0.0004)	0.712 (0.0014)	0.899 (0.0003)	0.938 (0.0002)
k-Nearest Neighbors	0.493 (0.0018)	0.785 (0.0006)	0.890 (0.0005)	0.620 (0.0022)	0.863 (0.0020)	0.922 (0.0004)
Naive Bayes	0.501 (0.0011)	0.789 (0.0008)	0.890 (0.0004)	0.680 (0.0007)	0.885 (0.0004)	0.928 (0.0003)
Neural Network	0.561 (0.0016)	0.823 (0.0010)	0.899 (0.0004)	0.722 (0.0027)	0.903 (0.0012)	0.939 (0.0003)
Random Forest	0.577 (0.0015)	0.827 (0.0006)	0.902 (0.0003)	0.735 (0.0013)	0.908 (0.0005)	0.940 (0.0003)
Support Vector Machine	0.556 (0.0021)	0.821 (0.0012)	0.898 (0.0004)	0.718 (0.0019)	0.903 (0.0006)	0.938 (0.0002)

3 RQ3
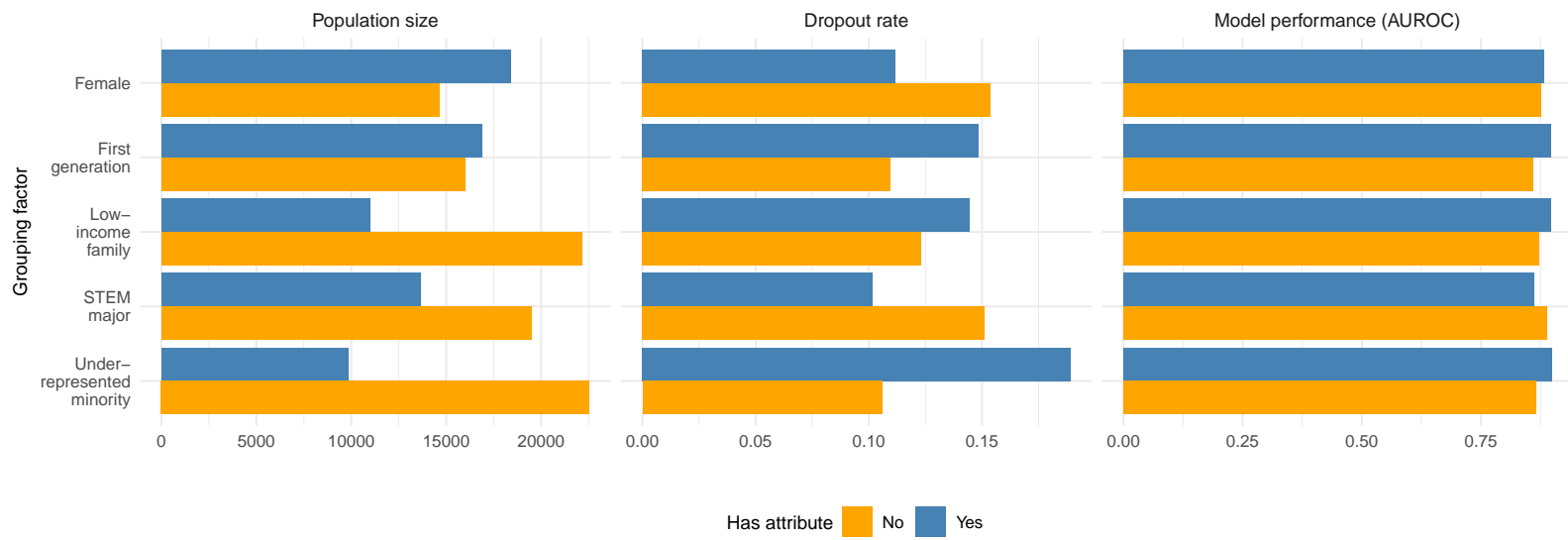
The following figures include the same analysis as RQ3 but for an observation span of 2 and 4 spans.

3.1 After two terms

Population and performance by groups

Predictor importance by groups

3.2   After four terms

Population and performance by groups

Predictor importance by groups