CPB Netherlands Bureau for
Economic Policy Analysis

The Costs of Affirmative Action. Evidence from a Medical School Lottery

We find that the efficiency costs of data driven affirmative action are small. The average graduation rate decreases only slightly, from 65.0% to 64.4%, when the share of selected students with a minority background is almost doubled to match that of a lottery. Moreover, data-driven selection consistently outperforms lotteries, enhancing outcome efficiency substantially.

The costs and merits of affirmative action are source of much debate. We quantify how affirmative action in data driven admission selections affects education and labor market performance of students.

CPB Discussion Paper

Bas Scheer, Brinn Hekkelman, Mark Kattenberg

mei 2024

Doi: https://doi.org/10.34932/wofv-7977

The Costs of Affirmative Action. Evidence from a Medical School Lottery *

Bas J. Scheer[a,b], Brinn Hekkelman[a], and Mark A.C. Kattenberg[a]

[a] *Netherlands Institute for Economic Policy Analysis (CPB)*
[b] *Corresponding author: b.j.scheer@cpb.nl*

May 14, 2024

Abstract

The perceived costs and merits of affirmative action are source of much debate. We quantify how affirmative action in data driven admission selections affects education and labor market performance of students. Our approach exploits lottery-based variation in medical school admissions in the Netherlands. First, we show that the efficiency costs of data driven affirmative action are small. The average graduation rate decreases only slightly from 65.0% to 64.4% when the share of selected students with a minority background is almost doubled to match that of a lottery. Second, we find that data-driven selection consistently outperforms lotteries, enhancing outcome efficiency substantially for any minority group share.

Keywords: affirmative action, lottery, college admission, medical school, algorithmic fairness, supervised machine learning
JEL codes: J15; D63; I24; J24

*For this research we used microdata of Statistics Netherlands. These data cannot be shared publicly, because it is highly confidential personal data. We will support getting access to the data-project of Statistics Netherlands. We are grateful for feedback from seminar participants at CPB. We thank *Robert Dur* in particular for their helpful comments. All remaining errors are ours.
Declarations of interest: none.

1 Introduction

Affirmative action in college admissions is both controversial and widespread. College admission that favors minority groups could improve minority outcomes, as it allows people from minority groups to get into higher quality colleges. This improves the academic and labor market opportunities of individuals in these groups. However, too strong affirmative action can lead to lower student academic performance if not enough suitable candidates are available from the minority group Arcidiacono and Lovenheim (2016).

In order to get the full picture, one needs to investigate the overall effect of affirmative action plans on student academic and labor market performance to judge the effectiveness of such plans. Yet, when evaluating economic effects of affirmative action plans, scholars have mainly focused on academic outcomes of the minority group, such as college completion rates, academic achievements and pre-college investment decisions (see the review by Arcidiacono and Lovenheim (2016) and references therein). Little is known about how affirmative action plans are best designed given the goals universities want to achieve as a whole (Arcidiacono and Lovenheim, 2016). Universities generally care about study completion, as funding is often tied to the number of graduates (the Netherlands, 2017). At the same time, many universities also have explicit equal opportunity goals.

The aim of this paper is to show how data driven selection procedures can be used to design an affirmative action plan that produces desirable selections in terms of both efficiency and equity. In order to do that we connect statistics of student population diversity, such as ethnic background, gender or age, to group-wide outcomes, such as graduation and employment rates. Our empirical approach is inspired by the framework of Kleinberg et al. (2018b).

Central to this approach is the imposing of a certain diversity restriction on the selection procedure. For example: 20% of admitted students should be from a certain ethnic background. The selection procedure then selects the most suitable candidates, as determined by its selection method, while assuring that this restriction is met. In essence, we allow the selection method to find the most efficient selection with a constraint on the composition of the selection group.

We extend the framework to selection with restrictions on any number of diversity characteristics. Often, a single characteristic such as gender is discussed, with the implication that this generalizes to more characteristics. However, characteristics may very well correlate with each other, meaning that restrictions on one affect selection of the other. We devise a method that generalizes selection restrictions to an arbitrary number and combination of characteristics. With this addition to the framework, over- or underselection of characteristics due to restrictions on correlated characteristics can be completely mitigated.

Once the selection is made, we record resulting group-wide outcomes, which consist of the graduation rate and employment in the healthcare sector. Similar to Kleinberg et al. (2018b), this allows us to create a possibility frontier for each selection method that we test, which shows the highest possible outcome efficiency reached for any restriction on selection group composition. We also run each selection method without restrictions on group composition, which yields the point on the possibility frontier that a social optimizer with no equity preference would find.

We apply the framework using administrative data on the universe of admissions to medical schools in the Netherlands between 2000 and 2004. We use these data because admittance to medical schools in the Netherlands at that time was determined at random using a weighted lottery. In essence, this selection procedure constituted a separate lottery for six groups based on Grade Point Average (GPA). The lottery was run such that groups with a higher GPA had a higher rate of admission.

After correcting for the differences in admission rates between GPA groups using inverted probability weights, this data allows for estimation of unbiased and consistent predictions for students who were not enrolled in medical school. Students that were selected and those that were not have the same observed and unobserved characteristics, because random chance (conditional on GPA) determined who would be selected. Therefore, the estimated models on selected students can be used to generate unbiased predictions for students that were not selected. They can also be used to generate unbiased predictions for a new wave of medical school applicants under the assumption that characteristics of different waves of applicants are equal.

Our analysis consists of a simulation exercise in which we test four selection methods: a simple lottery, a weighted lottery, selection using a prediction model based on high school GPA, and finally selection using a prediction model based on GPA with additional predictors. These additional predictors are: age, gender, minority background and social economic background as reflected by income of the parents. We run each of the selection methods on the same set of applicants from the data set described above.

The results of our case show that affirmative action brings a cost in terms of efficiency. For a given selection method, this cost is small. For example, we find that increasing the share of students with a migration background[1] by 1%-point decreases the graduation rate by less than 0.1%-point.[2] Since the difference in suitability between majority and minority groups for the marginal applicant is small, increasing the share of minority group does not have an economically significant impact on selection efficiency. However, the differences in efficiency between selection methods are significant. For example, switching from a data driven selection method to a lottery reduces the graduation rate by roughly 8.8%-points, selecting an equal share with a migration background. This is equivalent to around 260 extra yearly graduates. As such, we can conclude that in our case the cost of affirmative action is relatively small compared to the choice for selection method.

Moreover, we find that a data driven selection model achieves higher efficiency than a lottery for any share of minority group in the selection. Using this method, no matter the preferred share of minority group in admission, the data driven selection method always outperforms the lottery in terms of efficiency. This efficiency gain is attained by selecting the most suitable students, as determined by a prediction model, from majority *and* minority groups.

This study contributes to the existing literature in several ways. First, it presents results for a real-world use case where selection is essential and prediction models can improve the process. Second, we assess multiple long-term outcomes to evaluate the performance of selection methods, encompassing not only obtaining a degree but also the subsequent career path of working as a doctor. Last, we measure various facets of diversity, including gender, migrant status, and parental income. In addition to setting a restriction on any one of these dimensions at a time (e.g. Kleinberg et al. (2018b); Rambachan et al. (2020)), we present an algorithm that allows for any combination of diversity restrictions in the selection process[3]. This can be used to set a joint restriction on, for example, gender and migrant status.

Our results are policy relevant given the current public and political debate on affirmative action. In the US, the Supreme Court decided that the

[1]In accordance with the Statistics Netherlands definition, in this paper we categorize an individual as having a migration background if at least one of their parents was born outside the Netherlands, or the individual themselves was born outside the Netherlands.

[2]Example calculated for the data driven selection method using the full prediction model: if the selection is unrestricted, it selects 11.7% students with a migration background, 65.0% of the selected students graduate. If the selection method is set to select 20.0% students with a migration background, which reflects the share among applicants, 64.4% of the selected students graduate. Hence the slope is equal to -0.07: $(0.644 - 0.650)/(0.200 - 0.116)$.

[3]Provided that the selected restrictions are attainable in the candidate group.

race-based admission plans adopted by Harvard University and University of North Carolina are unconstitutional, because they 'lack sufficiently focused and measurable objectives warranting the use of race'.[4] However, Bleemer (2022) shows evidence that suggests that the benefits of affirmative action for underrepresented minority groups exceed the costs for on-the-margin white and Asian applicants. As such, affirmative action may be defendable as a measure to improve overall social welfare. In the Netherlands, the source of our data set, implementing a lottery for medical school admission is one of the options considered to reduce differences in admission chance between ethnic groups. Starting in study year 2024-2025 (weighted) lotteries will once again be permitted as a selection mechanism. Consequently, future selection procedures may benefit from incorporating the insights from this paper.

While our research offers valuable insights, there are also limitations. The main limitation is that it is impossible to set an optimal level of equity. Equity preferences are subjective and therefore cannot be optimized for. What constitutes an equitable outcome – which minorities should be targeted and how much extra representation should be allowed to cost in terms of efficiency – are debatable subjects. To strike an appropriate equity-efficiency balance, clear equity preferences must be established through discussions among those responsible for the study program. However, for a given equity preference in terms of representation of minority groups our framework gives the highest possible efficiency for a given outcome, such as graduation rate.

The remainder of the paper is structured as follows. The next two sections provide institutional context for the US and the Netherlands. Section 2 presents a summary of affirmative action in higher education in the US. Section 3 presents the history of medical school admission in the Netherlands, which gives the context for the data that our case is based on. The subsequent sections concern the empirical findings of our paper: section 4 introduces our empirical approach, section 5 presents the data, and section 6 shows the results. Finally, section 7 discusses our findings and section 8 concludes.

[4]Students for Fair Admissions, Inc. v. Harvard College, 600, US, 181 (2023).

2 Affirmative action in higher education

[5]The US Supreme Court's decision in *Regents of the University of California v. Bakke* (1978) has had a major influence on the development of affirmative action. This ruling has allowed institutions to use affirmative action policies to achieve and maintain a diverse student population, as long as they are subject to strict scrutiny. Between 1997 and 2013, eight states have passed laws that restrict the use of affirmative action in college admissions.[6] In *Students for Fair Admissions, Inc. v. Harvard College* (2023) the Supreme Court decided that the race-based admission plans adopted by Harvard University and University of North Carolina are unconstitutional, because they did not justify the use of race well enough.

The effect of affirmative action plans on selected minorities depends on the quality-fit trade-off. Affirmative action benefits underrepresented minority groups (UMGs) if more selective colleges provide better quality schooling that yields higher payoffs. Yet, if the fit between selective colleges and students enrolled based on affirmative action rules is worse than the fit between these students and less-selective schools, affirmative action might harm students. This is because it generates a mismatch between the college and qualities of the student. This mismatch might occur even if more selective colleges provide larger benefits than less-selective colleges (Arcidiacono and Lovenheim, 2016). Moreover, Loury and Garman (1993) indicate that affirmative action can lead to a mismatch effect, making it more difficult for those admitted to achieve high grades, graduate, and pursue more profitable majors. However, more recent empirical investigations suggests that students admitted through affirmative action may perform as well or better than their peers (Fischer and Massey, 2007).

Structural empirical analysis suggests that eliminating or reducing affirmative action plans will significantly alter university racial compositions, minority educational attainment, and the distribution of benefits. In particular, race-neutral admission procedures moderately decrease minority enrollment, but drastically affect the most selective institutions (Howell, 2010; Epple et al., 2008; Arcidiacono, 2005). One explanation for this is that only 20 to 30 percent of four-year colleges practice racial preferences in admissions, as most schools simply are not selective.

[5]This section is by-and-large a concise and selective summary of the literature on affirmative action reviewed in Arcidiacono et al. (2015) and Arcidiacono and Lovenheim (2016).

[6]These states are Texas (1997), California (1998), Washington (1999), Florida (2001), Michigan (2006), Arizona (2010), New Hampshire (2012) and Oklahoma (2013).

Recent empirical work on affirmative action exploits quasi-experimental variation in affirmative action that results from its banning by the eight states listed before. Using a subset of these states, Hinrichs (2012) concludes that affirmative action bans reduce the quality of schools attended by UMG students without affecting graduation rates. As explained in Arcidiacono et al. (2015) one explanation is that under affirmative action, the effect of an increase in college quality might offset the decrease in academic match quality. A second explanation could be that universities help minority students more to succeed after the ban on affirmative action (Arcidiacono et al., 2014). For California, Bleemer (2022) concludes that banning affirmative action not only reduced the quality of colleges that UMG students are admitted to, but also negatively affected degree attainment rates.[7]

Affirmative action favors UMG applicants, which by definition harms students from other groups. Bleemer (2022) studies the 1998 ban of affirmative action at California public universities and find it negatively affected average wages of UMG students when they are in their twenties or thirties. It suggests that costs to other students are such that the net educational and wage benefits of affirmative action for UMG applicants exceed the net costs for on-the-margin white and Asian applicants.

3 Medical school admission in the Netherlands

One of our main contributions is quantifying the difference in efficiency and diversity between selection methods. For this we require a data set that lets us draw random samples with unbiased counterfactuals. Such a data set is provided by the Dutch system for student admissions to medical school. Its historical weighted lottery selection provides an ideal basis for our research.

Medical schools in the Netherlands have a quota on the yearly inflow of first-year students; each year the number of applicants exceeds the number of available medical school places. This quota was introduced in 1976. At that time the number of medical students was increasing rapidly, and it was feared that the large influx of students would harm education quality (Raad voor Volksgezondheid en Samenleving, 2010). Starting in 1976 a lottery was held to determine admission. This lottery was weighted such that students with a higher high school exam grade point average (GPA) had a higher probability of being admitted.

However, there were doubts about the effectiveness of this lottery in

[7] Arcidiacono et al. (2015) discuss other ways in which affirmative action bans could change the behavior of universities and students.

terms of selecting the best prospective students. In 1999, a reform was implemented that entailed that applicants with a GPA above eight are automatically admitted.[8] This reform was implemented as a response to a large public discussion about a candidate who finished high school with an exceptionally high GPA of 9.6, who lost the lottery three times in a row (Van Walsum, 1998). Moreover, from 1999 onward medical schools were allowed to select up to half of the available places themselves, the so-called "decentralized selection" (the Netherlands, 1999). Instead of a centrally run lottery, each medical school could now, in part, select the best students using motivational letters and entry exams. These decentralized selections were first implemented in the study year 2000-2001.[9]

From study year 2017-2018 onward, selection using a lottery was disallowed entirely (the Netherlands, 2015), as it was deemed to be too ineffective. Policy makers feared that too many motivated and qualified applicants would not be selected in a (weighted) lottery. Before 2017, medical schools could still opt for selection using the central lottery for the full 100%. Thus since 2017/2018, Dutch universities select students based on merit, as expressed in interviews, tests and high-school grades. This is similar to admission to college in the United States, when universities would not use affirmative action plans.

However the admission lottery may make a come-back in the Netherlands. Recently, the public debate has focused more on the inequality in opportunity between applicants of different socio-economic backgrounds. Under the current selection method applicants that are female, have a native Dutch background, and wealthy parents, are more likely to be admitted compared to applicants that are male, have a migration background, and poorer parents (Mulder et al., 2022). This was also the case in the weighted lottery, as there are, on average, more women and fewer persons with non-Western migration backgrounds in the lottery categories with a higher GPA (Ketel et al., 2016), which receive a higher weight in the lottery. However, Mulder et al. (2022) conclude that the switch from lottery-based to selection-based admission has led to a stronger link between applicants' background characteristics and the odds of admission.

From 2024 onward, (weighted) lotteries will again be allowed as a selection procedure. Policy makers want to re-introduce this option for medical schools as a method that can reduce the inequality of opportunity. This

[8] with a GPA above eight (out of ten)

[9] Unfortunately, data on these decentralized selections are not available, and for that reason they are out of scope of this paper.

re-ignited political and public debate about the efficiency and fairness of lottery and other selection procedures. Although there is much discussion on the effectiveness of the current selection procedure and the benefits of a lottery in terms of equality of opportunity, the debate is severely lacking empirical evidence on the effectiveness of data-driven selection procedures and the interplay between effectiveness and inequality in opportunity. This paper aims to present a framework for practitioners to set up a selection procedure with equity restrictions.

4 Empirical approach

This section demonstrates how universities can employ data-oriented selection processes to design an affirmative action plan that yields efficient and equitable selections. Specifically, we create a possibility frontier for various selection methods that signifies the maximum efficiency outcome achievable for any given restriction on the composition of the selection group. Section 4.1 details how we exploit lottery-based variation in medical school admissions to estimate counterfactual outcomes for students who were selected out. Section 4.2 outlines the selection algorithms that we examine, while the enforcement of equity restrictions on single and multiple attributes are described in sections 4.3 and 4.4.

4.1 Exogeneous variation in selection into medical school

We run each of the selection methods on the same set of applicants that were admitted through the weighted lottery selection procedure. Since applicants were selected using a lottery, we can generalize the predictive performance of the selection models to the full set of applicants without fear of selective labels bias (Kleinberg et al., 2018a). After all, conditional on lottery weight, the difference between admission and rejection is only the luck of the draw. As such, we can ascertain that the results of selection models trained on these data can be generalized to the full set of applicants, and new batches of applicants. In order to do that, we need to use Inverse Probability Weights (IPW) when sampling applicants from the data set:

$$w = p(Admission|G,T)^{-1}, \tag{1}$$

with $p(Admission|G,T)$ representing the probability of admission conditional on grade category G and year of application T.

First, we draw applicants from the data set of medical school admissions. Then, a selection method is used to determine who is admitted and who is rejected. Note that selection mechanisms based on prediction models are trained on the remainder of the data set that was not drawn as applicants. After the admitted applicants are selected, we collect diversity statistics, such as migration background and gender, and the outcome variables of interest, such as study completion and employment in the medical sector. In order to rule out that our results are affected by any specific draw, we run this procedure a 1000 times and present the average result of these runs.

4.2 Selection algorithms

For each of the four selection methods we created a selection algorithm to determine which applicants are admitted and which are rejected. The algorithm gets a set of applicants and number of available places as input and needs to provide a set of admitted applicants as output. Algorithm 1 formally describes the selection decision for a simple lottery. This is the simplest selection method: randomly order the applicants and admit the top N, with N being the number of available places. The weighted lottery algorithm is more involved. We reproduce the procedure as it was implemented between 1976 and 2017, see section 3. The idea behind the weighted lottery is to give applicants with a higher grade (high school exam GPA) a larger admission probability using a weighting scheme. Inclusion of this selection method serves as a baseline reference. The weighted lottery algorithm is formally described in Algorithm 2.

For data-driven selection the procedure is altered slightly. After drawing a random subset as applicants, the remainder is used to train a prediction model. This prediction model is then used to rank the applicants and the top N are selected. In our analysis we will test two prediction models: a simple model that uses only grade category to predict the outcome variable, and an expanded model with added background characteristics. These data-driven selection models are formally described in Algorithm 3.

Note that the simple selection model uses the same information as the weighted lottery. However, they differ in the way this information is used to determine which applicants to select. The simple model checks which grade categories score best on the outcome – one would generally expect higher grades means a better outcome – and fills up all available places from best to worse grade category. This means that applicants from lower grade categories are never selected if there are still applicants from higher categories available. In contrast, in the weighted lottery applicants from lower grade categories do have a chance to be admitted; this chance is just lower than for applicants from higher grade categories.

4.3 Equity restrictions

Using a data-driven selection model that was trained on a specific outcome will lead to a selection that scores high, in expectation, on that outcome. Such a selection may, however, end up selecting applicants with certain background characteristics to greater or lesser extent. There are various reasons why it could be desirable to predetermine a certain share of candidates with specific background characteristics. For example, one may suspect that certain groups have a lower GPA on average while not being less qualified, one may suspect that a minimal share size of a specific group will benefit their study results through peer networks, or one may find representation of a certain group in doctors to-be desirable for other reasons. For instance, ethnically diverse doctors provide better access to healthcare for patients with a minority background (Mulder et al., 2022).

Therefore, we add controls over the shares of characteristics in the selection procedure. Let N be the selection capacity, and let f be the share of characteristic c that we would like our selection to include. In other words, we want our selection to consist of $N \cdot f$ (rounded to an integer) candidates with characteristic c and the rest, $N \cdot (1 - f)$ candidates, without characteristic c. At the same time, we want the selection to score as high as possible (in expectation) on the chosen outcome. We can achieve this by simply having the algorithm select the best candidates as determined by the model from both groups separately. Algorithm 4 implements this technique.

This algorithm allows us to show the expected outcome for the selected applicants for any share f, thus creating a line that plots the relationship between the outcomes of the selection (efficiency) and the diversity measure (equity). We can also plot the point on that line that a social optimizer with no equity preference would choose; the point with maximal efficiency. This creates figures comparable to those of Kleinberg et al. (2018b), which can be found in the results in Section 6.

For the weighted lottery setting this restriction is more involved, as the selection is split up over multiple grade categories. Therefore, we need to impose this restriction for each grade category, and if it is not possible for a certain grade category this needs to be compensated for in the selection of the other grade categories.

4.4 Equity restrictions on multiple characteristics

Similar to how a focus on the chosen outcome may result in a skewed share of a characteristic c, determining the share of a characteristic c_1 may affect the share of another characteristic c_2. For example, determining a certain share of male candidates in the selection may affect the share of candidates with a migration background in the selection.

If these possible effects are undesirable then the shares of multiple characteristics may be determined simultaneously. The principle of the restricted selection algorithm can be extended to any number of characteristic restrictions by rewriting the problem as a linear program. A column of binary dummy variables indicating presence of characteristic c is interpreted as a restriction by equating its selected sum to the desired share $N \cdot f_c$. The object of the linear program is then to find a binary solution vector that maximizes the expected outcome as determined by the model while satisfying all characteristic share constraints. Computational complexity of finding such a solution, if it exists, is limited since the constraint matrix is binary and the right-hand sides of the constraints are integers. Algorithm 5 formally describes this generalization to multiple characteristic constraints.

Note that the more constraints are added, the more likely it is that a solution does not exist. Even with one constraint a solution may not exist if for example we demand the share $f = 1$ for a characteristic that is not common enough to fill the entire selection with. Additional constraints further increase the complexity of the composition of the desired selection and are thus more likely to not be possible to realize with the available pool of applicants.

5 Data

For our analysis, we use the same data set on lottery selection used by Ketel et al. (2016). This data contains the registration of the centrally led selection procedures for all studies with a quota on admissions. In this paper we focus on admission to medical school.[10] We have access to all lottery applications and results between 1987 and 2004. Since our empirical analysis relies on the weighted lottery data (see Section 4), we only include applications through the weighted lottery.[11] After these restrictions, the data set contains 63,792

[10]We select applications for ISAT-code 6551: Medicine.

[11]This means we exclude people that were applied for a "ministersplaats" (258 applications) and applications of people that did not have a Dutch high school GPA (6,070 applications).

applications, of which 32,753 were admitted.

Figure 1a shows the number of applications and admissions between 1997 and 2004. When the quota was first introduced, the number of applications did not exceed the number of places by much, allowing for an admission rate of around 70%. However, over time the number of applicants increased, whereas the number of places remained stable. This caused a decrease in the admission rate, which reached its lowest point in 1997 at 34.6%. After 1997 the number of applicants dropped somewhat, and in 1999 the capacity of the medical schools increased, which meant that the admission rate could increase again. In the last year of our data set, 2004, the number of available places had doubled compared to 1987, and the admission rate was 62.9%.

Figure 1b shows the admission rate for each GPA category of the weighted lottery. By design, the admission rate increases with grade category. The admission rates shown here are used to calculate the IPW of equation 1. Between 1987 and 1990 all of the applicants in the two highest GPA categories were admitted. However, as the number of applicants started to increasingly outnumber the number of available places, the admission rate dropped below 100% for these categories too. As discussed in section 3, in 1999 the weighted lottery admission system was altered such that the highest grade category A was automatically granted admission. So between 1999 and 2004 the admission rate for this group was 100%. Moreover, the lowest grade categories, E & F were combined, which meant that from 2000 onward E became the lowest grade category.

We add background characteristics based on registry provided by Statistics Netherlands (CBS). Migration background, gender and birth date is collected from the municipal population register.[12] Additionally, we find the parents of the applicants using children-parents registration data made available by CBS, which allows us to construct an indicator for low parental income, which we define as having an income that is in the bottom three deciles[13].

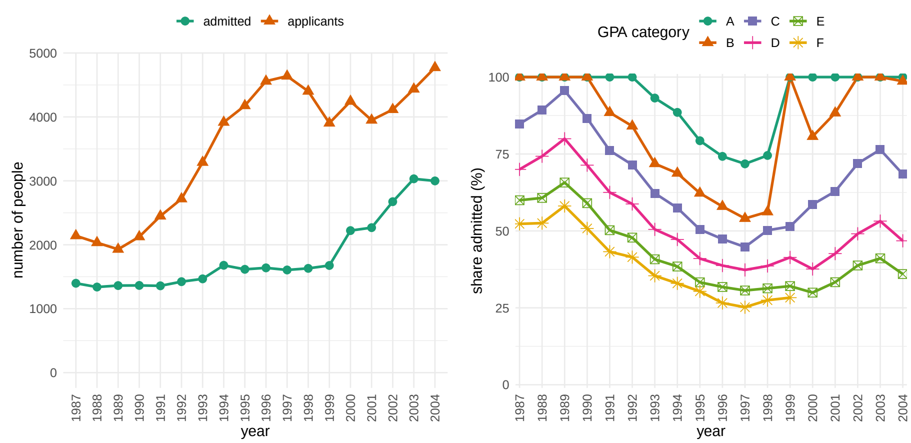
Figure 2 shows the admission rate of different groups in the data set. We can see that the admission rate of men and women was comparable between 1987 and 1998. However, after 1998 the admission rate for women gradually become higher than for men. Figure 3a shows the log odds ratio, which allows us to conclude that the weighted lottery statistically significantly under-sampled men between 1998 and 2004. This must be due to the fact

[12] For 500 applications there is no valid person identification code available, so for those applications no information can be collected from the register.

[13] The income deciles are based on the parental income of all applicants in our data sample

Figure 1: Applications and admission rates over time

(a) Applications (b) Admission rate



Number of applications and admission rates by grade category, based on actual results of the weighted lottery for medical school admission conducted by DUO. The weighted lottery gives applicants with a higher high school exam GPA a larger admission probability. The weighted lottery algorithm is formally described in Algorithm 2.

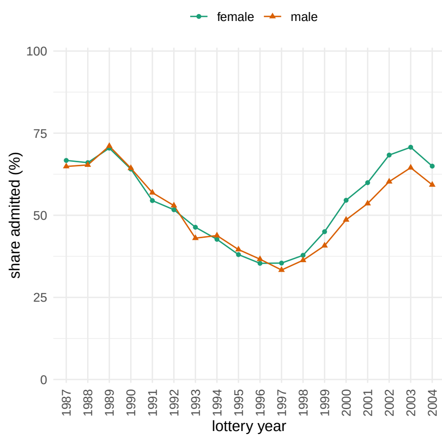
that male applicants attained lower grades than female applicants, as this is the only factor that affected the chance of admission. We can see a similar pattern for migrant compared to native Dutch applicants, although the differences in admission rates are smaller. Figure 3b shows that the admission rate for native Dutch applicants was significantly higher than for persons with a migration background in 2001, 2002, and 2004. Again, this implies that migrant applicants attained lower grades in these years than native Dutch applicants. The current selection procedure probably under-samples these groups even more, as shown by Mulder et al. (2022).

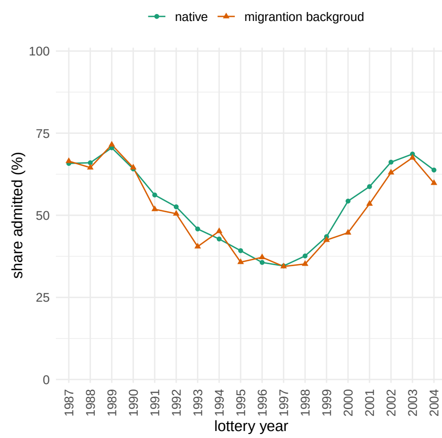
In our results we will show the relationship between the outcomes of the selection (efficiency) and the diversity measure (equity). In order to do that, we add study completion and employment in the medical sector as outcome variables. To track study completion we use degree data from DUO registrations. We use Dutch tax registration data to see if people are employed in the medical sector. Since this registration only covers employees, we extend this data set with registration on self-employment from the CBS individual income statistics. Only the self-employed persons that are registered as active in the medical sector are included. As such, our employment outcome definition includes both employees and self-employed individuals.

Figure 2: Admission rate over time

(a) Women compared to men

(b) Native Dutch compared to persons with a migration background





Admission rates by gender and migration background over time, based on actual results of the weighted lottery for medical school admission conducted by DUO. The weighted lottery gives applicants with a higher high school exam GPA a larger admission probability. The weighted lottery algorithm is formally described in Algorithm 2.

The fact that our sample covers a long period of time can distort the results of our simulation analysis. For example, assume that both the graduation rate and share of students with a migration background increase over time. As we explain outcomes using characteristics of persons with a migration background only (and do not include year effects) the parameter for being a migrant would be biased upwards in the early years of our sample and it would be biased downwards in the later years. As a result, the estimated model might wrongly rank candidates. In order to limit such biases, we can reduce the length of the sample period. The problem is non-existent if we sample from one point in time, i.e. within one application year.[14] In the main analysis we limit ourselves to the application data of 2000-2004, the results of which can be found in Section 6 below. Results for the full sample period (1987-2004) can be found in Appendix B.

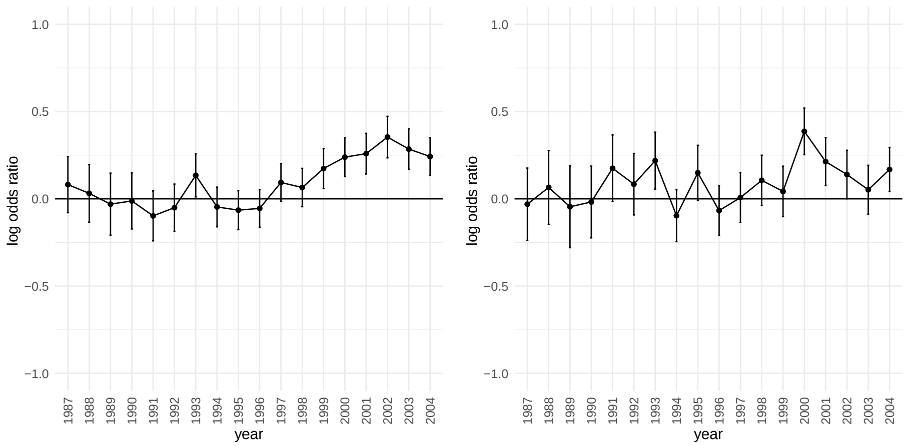
The main analysis data set that forms the basis of our simulation exercise consists of 11,119 applicants that were admitted between 2000 and 2004.

[14]For real-life application, this means that a selection should be re-estimated periodically. Ideally using data that was randomly sampled, for example by admitting a small percentage of students randomly each year.

Figure 3: Log odds ratio over time

(a) Women compared to men

(b) Native Dutch compared to persons with a migration background



Log odds ratios calculated as $log((p_{01}/p_{00})/(p_{11}/p_{10}))$, where p_{01} and p_{00} are the admission and non-admission probabilities for one group and p_{11} and p_{10} for the other group. Bars represent 0.95 confidence intervals, based on the standard error calculated as: $se = \sqrt{1/n_{01} + 1/n_{00} + 1/n_{11} + 1/n_{10}}$.

The estimation data set contains only applicants for whom all background information is known: GPA category, age at time of application, prior education before application, migration background, gender, and parental income. For 2076 applications at least one element of this information set is missing, which is 15.7% of the total number of admitted applicants. We set the number of applicants equal to 5000 and the number of students to select equal to 3000 per run. This roughly equals the number of students entering medical schools each year in 2003 and 2004 (see figure 1a).

6 Results

This section shows the results of our simulation set-up for study completion rate as outcome of interest and share of admitted students with a migration background as diversity measure of interest.

We find that the simple lottery gives a graduation rate of 55.7%, with a student population that consists of 20.2% with a migration background. This share with a migration background reflects the share in the applicant group. The weighted lottery attains a slightly higher graduation rate of

58.7%, with a share of 19.1% with a migration background in the selected group. Data driven selection using prediction models performs much better than the lotteries in terms of study completion rates. The prediction model based on grade only – using the same information as the weighted lottery – attains a graduation rate of 63.9% and the extended model attains an even higher graduation rate of 65.0%. A 9.3%-point improvement over the simple lottery selection. However, the share of admitted students with a migration background is much smaller in the data driven selection: 17.5% for the grade-only model and only 11.7% for the full model.

By placing restrictions on the share with a migration background in the selection procedure we can set the share with a migration background in the student population. With these restrictions the data-driven selection methods outperform the lottery selections in terms of graduation rate. If the share with a migration background is set to 20%, the full model attains a graduation rate of 64.4% and the grade-only model 63.5%. Much higher than the simple lottery's 55.7% and weighted lottery's 58.6%. As such, using data-driven selection nets up to 8.8%-points higher graduation rate without loss of representation of persons with a migration background.

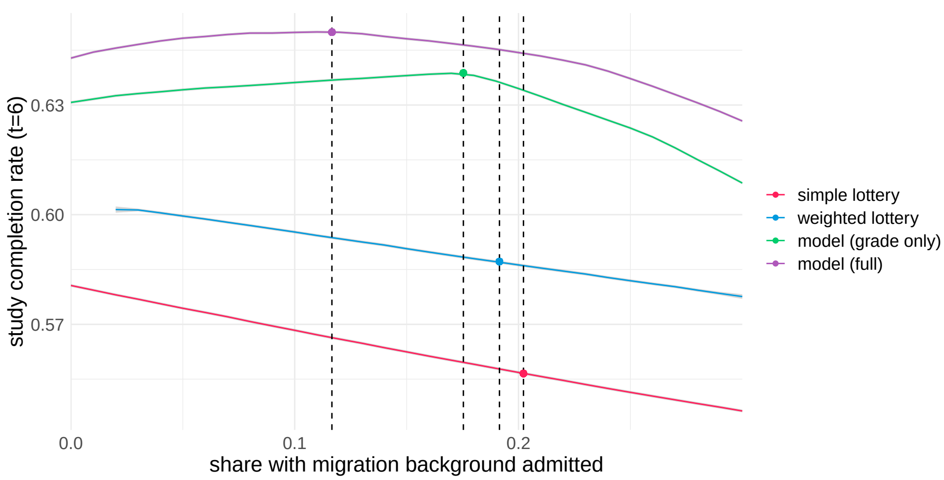
6.1 Equity restriction on one group at a time

Since we can set the share with a migration background, we can show the graduation rate for a sequence of shares with a migration background for each selection method. This creates a line that shows the trade-off between graduation rate and the share with a migration background in the student population. The resulting line shows that costs of affirmative action are small. For example for the data driven selection with the full model, when the share with a migration background is increased from 11.7% to 20.0%, the graduation rate decreases by 0.6%-points.

Figure 4 shows the study completion rate of the selected groups for each selection method. Note that the results for the prediction models are out-of-sample: the models were estimated on the data excluding the applicants that they had to select from. In other words, the applicants that the prediction model had to predict the outcome for were not part of the estimation data. The points indicate the realized study completion rate and share with a migration background in the selection when the selection procedure is unrestricted. In line with Kleinberg et al. (2018b), this point can be interpreted as the points an efficient planner without equity concern would choose, i.e. the highest expected value for study completion rate given the information available in the selection process. The lines indicate the completion rate and migrant share combination that one can reach by setting a restriction on the share with a migration background to select (see algorithm 4).

These results clearly show that the data-driven selection processes outperform the lotteries for any share with a migration background admitted. The lines for the selection processes using the prediction models are higher than the weighted and simple lotteries, for any restriction on the share with a migration background admitted. Given that one opts for a share with a migration background that mirrors the application group (20%), a linear model that only considers high school grade yields a 4.8%-point increase in graduation rate compared to the weighted lottery selection. This equals an 8.2% increase, or 145 yearly graduates (based on 3000 yearly admissions), an economically significant increase in potential labor supply of medical professionals.

Figure 4: selection models outperform weighted lottery for any share with a migration background



Simulation results for selection with restriction on share with a migration background. Results are based on the average outcome from 1000 simulations. Outcome is share of selected students that graduate medical school in 6 school years. The solid points represent the social optimizer with no equity preference, i.e. the outcome attained without diversity restrictions.

Moreover, we find that the efficiency costs associated with affirmative action are small compared to the differences between selection methods. Increasing the share of students with a migration background does result in lower overall study completion rates, but the decrease is relatively small compared to the differences between selection methods. That is to say, the changes in study completion rate due to switching between selection methods (switching lines) is much larger than increasing the share with a migration background within a selection method (moving along the line).

Take for example the results for the full model: introducing a restriction to set the share with a migration background equal to the share among applicants – 20% – would reduce the study completion rate by 0.6%-points, which translates to 0.07%-point reduction per 1%-point increase in the share with a migration background. The study completion rate does not fall below that attained by the efficient planner using the simple model – GPA category only – until a share of 25% with a migration background is reached. When comparing to the weighted lottery, this point is not reached for any reasonable share with a migration background. Even at 30% share with a migration background, which implies 10%-point oversampling compared to the share in the applicants, the expanded model attains a far higher study completion rate than the weighted lottery.

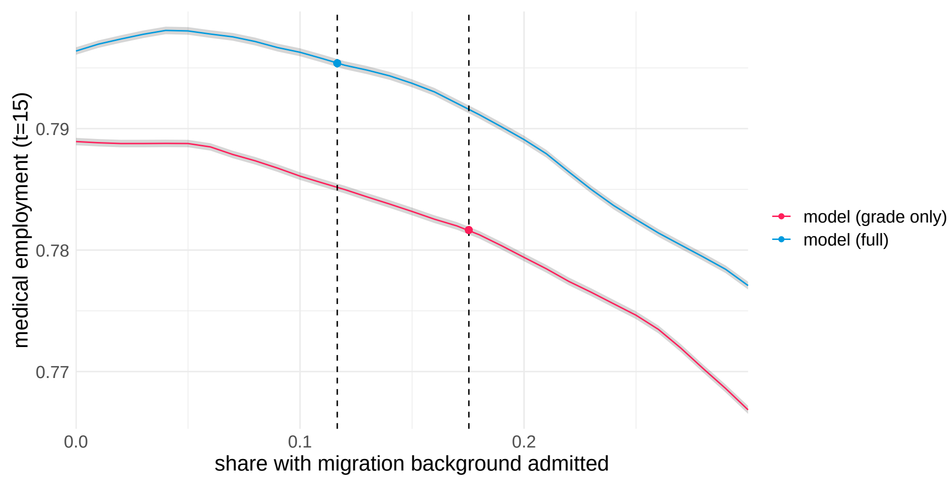
6.2 Results for different outcomes

Of key importance to the evaluation of a selection in terms of efficiency is the chosen outcome. Since a data-driven model is trained to maximize this particular outcome, the final selection is to a large extent tailored towards this outcome. Which outcome to chose depends on multiple factors. First, an outcome can be more or less representative of the underlying quality that we are trying to select for. For example, do students that graduate on time make good doctors? Second, different parties may have different goals for the selection. For example, a university may want students to graduate on time while a hospital may want good doctors.

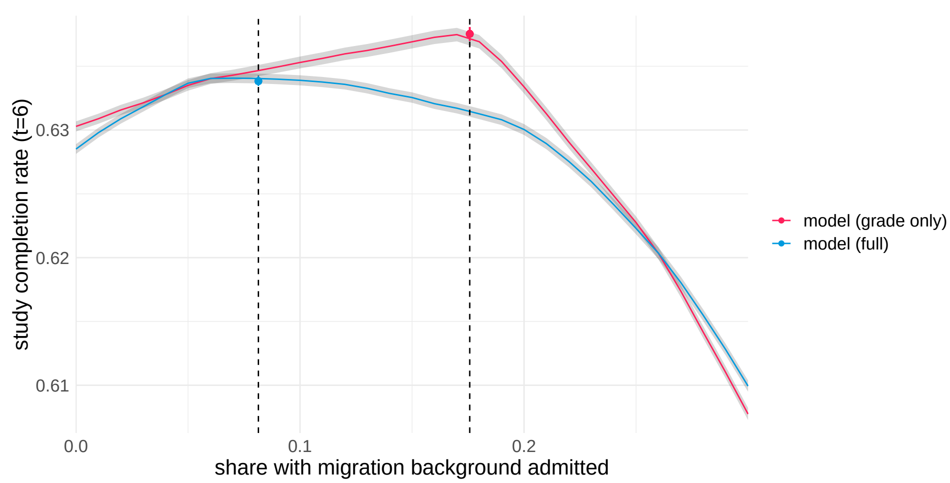
When applying outcome-focused data-driven selection techniques it is therefore important to analyse how our selections score on different outcomes. Figure 5 shows how the selection scores on medical employment when the selection was made by maximizing the expected graduation rate, and vice versa. The full model trained on the medical employment outcome attains an expected study completion rate of 63.4%. Compare this to the attained expected study completion rate of 65.0% when trained for that outcome (see fig. 4). The relatively small difference suggests that these two outcomes are correlated to some extent. As we can see, however, the optimal share with a migration background associated with study completion does not coincide with that of medical employment. About 8 percent of selected students has a minority background when the full model trained on medical employment is used, whereas this is about 12 percent when the full model is

Figure 5: Outcome efficiency when trained for another outcome

(a) Medical employment for model trained on graduation



(b) Graduation for model trained on medical employment



Simulation results for selection with restriction on share with a migration background. Results are based on the average outcome from 1000 simulations. Panel (a) shows the share of selected students that are employed in the medical sector 15 school years after admission to medical school, with models trained to optimize the share of students that graduate medical school 6 years after admission. Panel (b) shows the share of selected students that graduate medical school 6 years after admission, with models trained to optimize the share of students that are employed in the medical sector 15 school years after admission. The solid points represent the social optimizer with no equity preference, i.e. the outcome attained without diversity restrictions.

trained on graduation. This difference could be explained by the fact that students with a minority background are less likely to continue working in the medical after attaining a medical degree.

6.3 Results for other diversity measures

Besides migration background we can compute the efficiency frontier for other characteristics as well. In fig. 6 we study efficiency for the share of men and share of low parental income, defined as having parents with an income in the bottom 30% of parent incomes among applicants.[15] As we would expect, the optimal study completion rate of the full model is around 65% in both cases. This is because it is the optimal rate regardless of which characteristic share we are studying, and adjusting the share just means moving away from the optimal rate along a different axis (see section 6.4 for a representation of moving along different axes simultaneously).

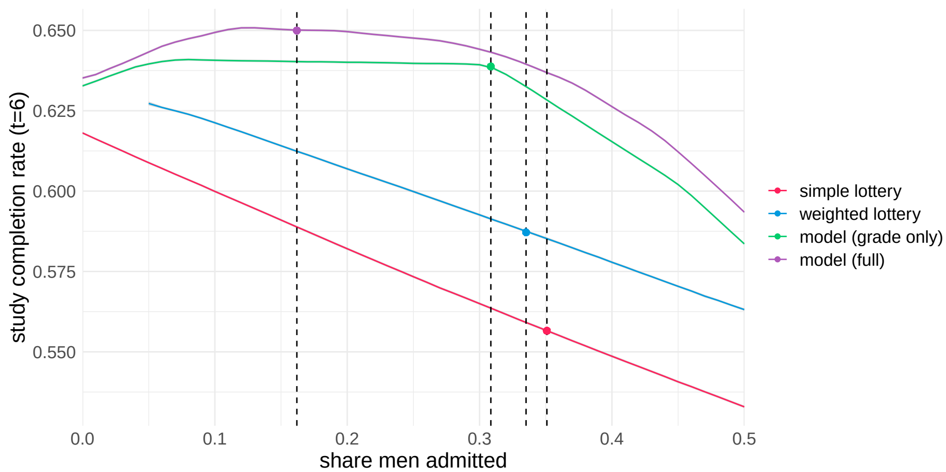
Figure 6a shows that for the full model, a more steep decline in efficiency starts around 30% share. This is lower than the share of men among applicants, which is about 35%. We see that adjusting the share to an equal representation of men and women, i.e. a share of men of 50%, drops the efficiency by almost 6%-points to 59.3% for the full model. A more modest increase of the share of men to reflect the share in applicants (35%), leads to a drop of efficiency of 1.3%-points to 63.7%.

In fig. 6b we see a contrasting scenario for the low parental income characteristic. There is almost no difference in efficiency between the full model optimal point and the full model restricted to the share among applicants (the share for the simple lottery). This means that one can select from a wide range of shares of low parental income without affecting the outcome much. In other words, the trade-off is small.

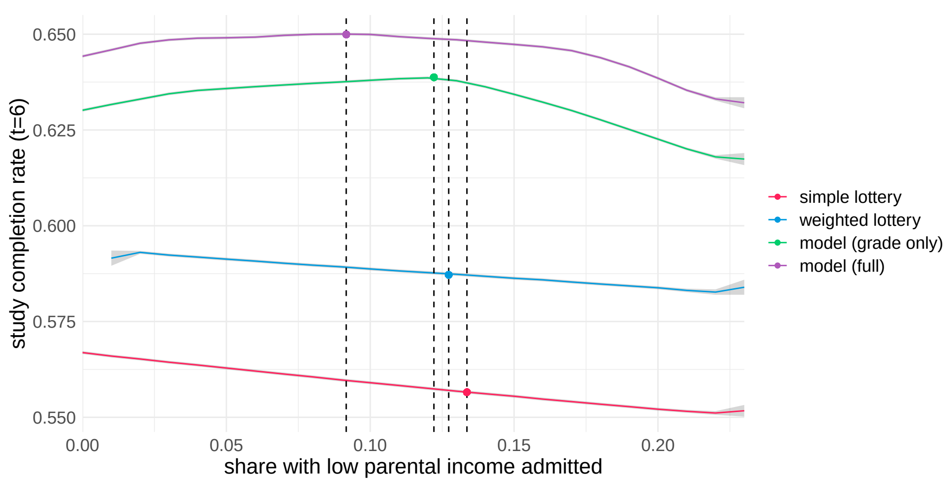
[15] This definition of low income puts those parents in the bottom 50% of incomes across the entire Dutch population, which shows the strong the selection of children from higher income families into application for medical school.

Figure 6: Graduation rate for different diversity measures

(a) Gender



(b) Low parental income



Simulation results for selection with restriction on share of men in panel (a) and share of students with low income parents in panel (b). Results are based on the average outcome from 1000 simulations. Outcome is share of selected students that graduate medical school in 6 school years. The solid points represent the social optimizer with no equity preference, i.e. the outcome attained without diversity restrictions.

6.4 Equity restriction on multiple diversity measures simultaneously

When multiple restrictions are imposed across characteristics, the efficiency frontiers become more complicated to plot. For two restrictions we can draw contour plots of the outcome across different shares of both restricted characteristics along the axes, as shown in fig. 7. On the left we see the outcomes for the extended data-driven model. As expected, we see a decrease in the outcome along both axes when the shares move away from the optimal unrestricted shares. We also see that the two shares affect each other: for male shares far away from the optimal share we see a smaller effect of changing migrant shares, and vice versa.

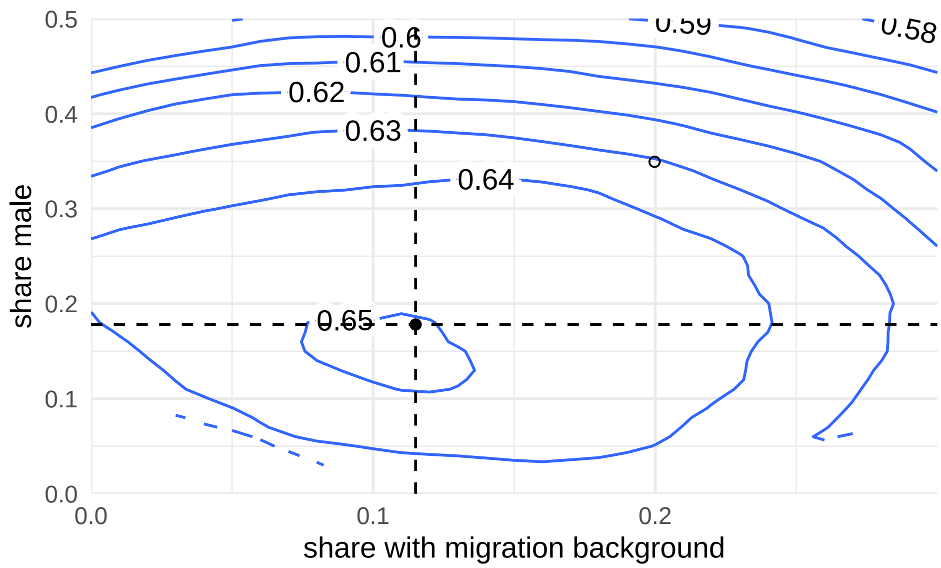
On the right we see the outcomes for the lottery model. Here we again see a decrease in the outcome along both axes when shares move away from the optimal unrestricted shares, but this time the effect is purely based on an average group effect and we do not see an interaction between the two shares, as expected from the lottery.

Finally, the left panel illustrates that the efficiency loss of two restrictions are very low as well. For instance, suppose that we want to restrict the data driven model such that it selects the same share of males and same share with a migration background as the lottery. Thus, whereas about 18 percent of the candidates selected by the data driven model are male, and about 12 percent has a migration background, these shares are restricted to be equal to 35 percent and 20 percent respectively. The left plot shows that this causes the graduation rate to fall from 65 percent (in the point (12, 18)) to 63 percent (in the point (20, 35)). This suggests that it is possible to have additional 150 graduates from medical school each year[16], with an entry composition similar to that of a lottery in terms of gender and migration background.

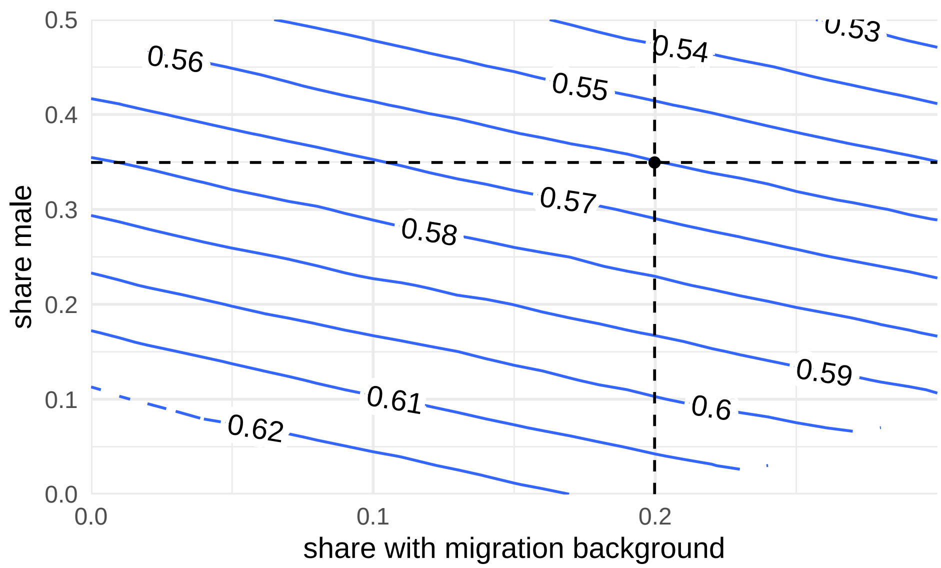
[16] Each year about 3000 students are admitted for Medical School. The graduation rate of the weighted lottery is 58 percent, that of the restricted data driven model is 63 percent. $(0.63 - 0.58) * 3000 = 150$.

Figure 7: Results for simultaneous restriction on share with a migration background and men

(a) outcome = study completion rate, selection method = model (full)



(b) outcome = study completion rate, selection method = lottery



Simulation results for selection with restrictions on share with a migration background and men. Results are based on the average outcome from 500 simulations. Outcome is share of selected students that graduate medical school in 6 school years. The solid points represent the social optimizer with no equity preference, i.e. the outcome attained without diversity restrictions. The unfilled dot in panel (a) shows the lottery outcome without diversity restrictions.

7 Discussion

Our results show the potential of using data driven selection methods. However, for effective use of selection methods in practice there are additional challenges. What are the best practices for constructing a predictive model to inform the selection decision? And what are the alternatives to using a prediction model for data driven selection?

For an effective prediction model, good predictors are vital. These predictors should contain information that helps predict whether someone would be a good fit for the selection. Ideally, a predictor is both predictive and motivating. An example of this is the final exam grade. The final exam grade gives a good indication of study completion chances later. Moreover, if this grade is an important determinant for selection into medical school, then prospective students have an incentive to achieve higher grades in high school. A predictor should also not be easily adjusted by the candidate in order to have a better chance of selection (Cowgill and Tucker, 2019). If candidates do that, then that predictor will quickly lose its predictive value. For example, if candidates are able to illegally purchase a certain certification that is used as a predictor, then this certification loses its actual value as a predictor of study completion.

Although it is not necessary to exclude any predictors from the prediction model to combat under-selection of disadvantaged groups (Rambachan et al., 2020; Kleinberg et al., 2018c), keeping the model simple keeps the resulting ranking rule simple and explainable as well. Moreover, certain predictors may inherently be viewed as unfair to use in important selection decisions, even if using these predictors do not lead to under-selection of disadvantaged groups. For example, giving prospective students a higher chance of admission based on the income of their parents may be viewed as unfair, even if a certain share is reserved for low-income students. One can remove such sensitive predictors from the prediction model, possibly at the cost of some effectiveness. In our case a simple prediction model, which only uses high school grade as a predictor, provides outcomes that are not far off of the extended model. As such, the cost of removing the additional predictors is not that high: one good predictor was enough to attain adequate results. Note that even if a sensitive predictor is removed from the model, one can still use its information to adjust the final selection as described in the paper.

In order to ensure the quality of the prediction model and selection process as a whole over time, the outcome of the selection procedure needs to be monitored. This means checking the actual performance of the selected students compared to the expected performance and periodically re-estimating the prediction model to update its parameters. The analysis we present gives valuable insights, but we cannot simulate second order effects of a change in the selection process using historical data. After all, we can only observe the outcome for students conditional on the selection process and resulting student population at the time. It is possible that expected student outcomes change if the selection process is changed, for example due to changing dynamics among student groups (e.g. peer effects). As such, monitoring the actual student outcomes after adjusting the selection process is still necessary.

Finally, it is important to note that the framework we introduce in this paper does not necessarily rely on model predictions to rank prospective students. Any ranking can be used. For example, universities could use exams and personality tests to create a ranking of applicants that they deem a good fit. If such a procedure is applied then a prediction model is no longer needed in the selection procedure. The university needs to formulate a clear view on the requirements for attending university, which can be measured using (a combination of) aptitude tests and examinations for substantive knowledge. Of course, the students' outcomes under this alternative selection method can still be tracked; the same way the results of the lottery selection were collected in our analysis.

In practice, one Dutch university has proposed to use this procedure to select students under diversity constraints. However, in the weeks following the announcement of the proposal, the Dutch education authority disallowed it based on current law. The Delft University of Technology announced a plan to increase the share of women studying Aerospace Engineering. An academic aptitude test and entrance exam is used to rank prospective students. The proposal involved reserving 30% of available places for the highest ranking women. The remaining places would be assigned to the highest ranked remaining applicants (TU Delft, 2024). The proposed set-up reflects the framework presented in this paper, which shows that the framework does have potential for real-world application.

8 Conclusions

The purpose of this paper is to show the breadth of possibilities that exist to assure diversity in selection procedures. The debate on affirmative action does not have to be limited to a dichotomous choice between pure efficiency maximization and diversity measures. Instead, empirical evidence can be collected on the possibilities which lie in between the two, the results of which can inform and reinforce an affirmative action plan that considers both diversity and efficiency in conjunction.

Within the framework presented in this paper it is still possible to opt for a selection procedure that either only considers efficiency or diversity in the selection group. At the same time, the framework also allows for a mixture of the two: trading off efficiency for equity and vice versa. Measurable insights into the effects of affirmative action on outcome efficiency may positively affect willingness to adopt such policies if the effects on outcome efficiency are small (Kleinberg et al., 2018c). In addition, this may also positively affect legal implementation, as previously their 'lack of sufficiently focused and measurable objectives warranting the use of race' proved problematic.[17]

The framework entails computing a ranking of applicants and subsequently applying a diversity selection. The ranking is defined on a chosen outcome metric, and can be computed by any means ranging from random assignment to data driven prediction models. The diversity selection is then made by selecting the highest ranked applicants from designated subgroups in a specified ratio. The advantages of this framework are threefold. First, any ranking method can be used without affecting the share of subgroups selected. Second, the entire range of outcome efficiencies for all shares of subgroups can be computed. Third, there is absolute transparency and control over the share of subgroups in the final selection.

Our case on selection for Dutch medical schools shows how this could work in practice. We show that the cost of affirmative action is relatively small compared to the differences in efficiency between selection methods. Increasing the share with a migration background by 1%-point decreases the graduation rate by less than 0.1%-point. This is due to the fact that the difference in suitability between majority and minority groups for the marginal applicant is small. Switching from a data driven selection method to a lottery reduces the graduation rate by roughly 8.8%-points, with an equal share with a migration background selected. In other words, replacing lottery selection with a data driven selection method yields around 260 additional yearly graduates at no diversity cost.

Turning to the entire range of diversity, we find that a data driven selection model achieves higher outcome efficiency than a lottery for any share of minority groups in the selection. This holds for all outcomes considered in this paper: graduation rate and employment in the healthcare sector. These results are especially policy relevant in The Netherlands given the

[17]Students for Fair Admissions, Inc. v. Harvard College, 600, US, 181 (2023).

current public and political debate on affirmative action. Implementing a lottery for medical school admission is one of the options considered to promote diversity in the student population. Starting in study year 2024-2025 (weighted) lotteries will once again be permitted as a selection mechanism in The Netherlands.

Even though an objectively optimal equity cannot be defined, any selection diversity can be manifested and evaluated in terms of outcome efficiency. Equity preferences are subjective and therefore cannot be optimized for. Rather, an informed equity-efficiency balance can be struck once clear equity preferences are established through discussions among those responsible for the medicine studies. Such discussions can be informed by the empirical analyses presented in this paper.

We would like to encourage practitioners to employ techniques similar to ours that collect empirical evidence on both efficiency and diversity outcomes, such that more insight can be obtained in the selection procedures that are currently in use at private and public organizations.

References

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, 73(5):1477–1524.

Arcidiacono, P., Aucejo, E., Coate, P., and Hotz, V. J. (2014). Affirmative action and university fit: Evidence from proposition 209. *IZA Journal of Labor Economics*, 3:1–29.

Arcidiacono, P. and Lovenheim, M. (2016). Affirmative action and the quality–fit trade-off. *Journal of Economic Literature*, 54(1):3–51.

Arcidiacono, P., Lovenheim, M., and Zhu, M. (2015). Affirmative action in undergraduate education. *Annu. Rev. Econ.*, 7(1):487–518.

Bleemer, Z. (2022). Affirmative action, mismatch, and economic mobility after california's proposition 209. *The Quarterly Journal of Economics*, 137(1):115–160.

Cowgill, B. and Tucker, C. E. (2019). Economics, fairness and algorithmic bias. *preparation for: Journal of Economic Perspectives*.

Epple, D., Romano, R., and Sieg, H. (2008). Diversity and affirmative action in higher education. *Journal of Public Economic Theory*, 10(4):475–501.

Fischer, M. J. and Massey, D. S. (2007). The effects of affirmative action in higher education. *Social Science Research*, 36(2):531–549.

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics*, 94(3):712–722.

Howell, J. S. (2010). Assessing the impact of eliminating affirmative action in higher education. *Journal of Labor Economics*, 28(1):113–166.

Ketel, N., Leuven, E., Oosterbeek, H., and van der Klaauw, B. (2016). The returns to medical school: Evidence from admission lotteries. *American Economic Journal: Applied Economics*, 8(2):225–54.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2018a). Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1):237–293.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Rambachan, A. (2018b). Algorithmic fairness. In *AEA papers and proceedings*, volume 108, pages 22–27.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Sunstein, C. R. (2018c). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10:113–174.

Loury, L. D. and Garman, D. (1993). Affirmative action in higher education. *The American Economic Review*, 83(2):99–103.

Mulder, L., Wouters, A., Twisk, J. W., Koster, A. S., Akwiwu, E. U., Ravesloot, J. H., Croiset, G., and Kusurkar, R. A. (2022). Selection for health professions education leads to increased inequality of opportunity and decreased student diversity in The Netherlands, but lottery is no solution: A retrospective multi-cohort study. *Medical Teacher*, 44(7):790–799.

Raad voor Volksgezondheid en Samenleving (2010). Numerus fixus geneeskunde: Loslaten of vasthouden.

Rambachan, A., Kleinberg, J., Mullainathan, S., and Ludwig, J. (2020). An economic approach to regulating algorithms. Technical report, National Bureau of Economic Research.

the Netherlands (1999). Wet van 3 april 1999 tot wijziging van de wet op het hoger onderwijs en wetenschappelijk onderzoek, houdende aanpassingen in het systeem van selectie voor opleidingen waarvoor een toelatingsbeperking is vastgesteld.

the Netherlands (2015). Regeling van de minister van onderwijs, cultuur en wetenschap van 8 november 2015, nr. 759279, houdende wijziging van de regeling aanmelding en toelating hoger onderwijs in verband met de afschaffing van de loting.

the Netherlands (2017). Beschrijving bekostigingsmodel hoger onderwijs. https://www.eerstekamer.nl/overig/20170524/beschrijving_bekostigingsmodel/meta.

TU Delft (2024). TU Delft Faculty of Aerospace Engineering takes measure for 30% female intake in bachelor's programme. https://www.tudelft.nl/en/2024/tu-delft/tu-delft-faculty-of-aerospace-engineering-takes-measure-for-30-female-intake-in-bachelors-programme. Accessed: 03-01-2024.

Van Walsum, S. (1998). Meike Vernooy weer uitgeloot voor geneeskunde. *Volkskrant*.

Appendices

A Selection algorithms

We reproduce the weighted lottery algorithm as it was implemented between 1976 and 2017. The idea behind the weighted lottery is to give applicants with a higher grade a larger admission probability, according to a given weighting scheme. The applicants are divided into categories based on their high school exam GPA, with higher grade categories being assigned relatively more places then lower grade categories. The number of places is determined by the weights: $n_g = a_g * w_g$, where n_g represents the share of places for grade category g, a_g represents the share of applicants in grade category g, and w_g is the weight of category g. In other words, if the weight of a grade category is equal to 2, the share of places available for that category should be equal to twice its share in the applicants.

Algorithm 2 shows the weighted lottery selection procedure. The inputs consist of a data set, containing the applicants with their grade category and corresponding lottery weight, and the number of applicants to select. The algorithm gives the set of selected applicants as output. For each grade category c the number of applicants to admit is determined using the category weight: $n_c = a_c * w_c$. Note that if the average weight for the grade categories that still have to be selected is not equal to 1, we will not end up with the correct number of admitted applicants N. As such, we need to adjust the share to admit to correct for that: the difference in admission rate is added according to the share in applicants a_c. Finally we check if the share to admit n_c is within 0 and 1 and that the number of admitted applicants N_c does not exceed the number of applicants A_c. The top N_c are then collected from the randomly ordered set of applicants in the corresponding grade category. This is done for each grade category, in order from highest to lowest, and the complete set of admitted applicants is returned.[18]

[18] This algorithm has been tested, and produces results comparable to Table A1 in the Appendix of Ketel et al. (2016) for the years before 1999. After 1999 applicants with a GPA above 8 were automatically admitted, which could be implemented in our algorithm by adjusting the weights for those categories.

Algorithm 1: Lottery selection (A, N)

Input: *A set of applicants A and an integer selection capacity N*

Output: *A selection applicants S of length N*

$A' \leftarrow \texttt{Randomize}(A)$

$S \leftarrow A'[1:N]$

return S

Algorithm 2: Weighted lottery selection (A, G, W, N)

Input: *A set of applicants A with associated grade categories G and lottery weights W, and an integer selection capacity N*

Output: *A set of selected applicants S of length N*

$S \leftarrow \emptyset$

$D \leftarrow \texttt{Randomize}(D)$

$A_{remaining} \leftarrow \texttt{Count\_Elements}(A)$

$N_{remaining} \leftarrow N$

$C \leftarrow \texttt{Order}(g \in G)$

for $c \in C$ **do**

- $A_c \leftarrow \texttt{Count\_Elements}(G = c)$
- $w \leftarrow W[G = c][1]$
- $a_c \leftarrow A_c / A_{remaining}$
- $n_c \leftarrow w * a_c$
- **if** $\texttt{Average}(W[G\ in\ C]) \neq 1$ **then**
  - $n_c \leftarrow n_c + a_c * (1 - \texttt{Average}(W[G\, in\, C]))$
- $n_c \leftarrow \texttt{Min}(n_c, 1)$
- $n_c \leftarrow \texttt{Max}(n_c, 0)$
- $N_c \leftarrow \texttt{Round}(\texttt{Min}(n_c * N_{remaining}, A_c))$
- $A_{remaining} \leftarrow A_{remaining} - A_c$
- $N_{remaining} \leftarrow N_{remaining} - N_c$
- $C \leftarrow C[-c]$
- $S \leftarrow S + A[G = c][1 : N_c]$

return S

Algorithm 3: Selection without equity restriction (A, N, M)

Input: *A set of applicants A, an integer selection capacity N and a prediction model M*

Output: *The selection of applicants S of length N which maximizes the expected outcome predicted by the model M*

`// Compute the predicted outcome for all applicants`

$P \leftarrow M(A)$

`// Select the best applicants`

$S \leftarrow \texttt{Sort(}A \textit{ by } P\texttt{)}[1:N]$

return S

Algorithm 4: Selection with equity restriction (A, N, M, c, f)

Input: *A set of applicants A, an integer selection capacity N, a prediction model M, a protected characteristic c and an inclusion share f*

Output: *The selection of applicants S of length N that contains $N \cdot f$ applicants with characteristic c which maximizes the expected outcome predicted by the model M*

`// Compute the predicted outcome for all applicants`

$P \leftarrow M(A)$

`// Split applicants by characteristic` c

$X \leftarrow \{a \in A \,|\, a \text{ has characteristic } c\}$

$Y \leftarrow \{a \in A \,|\, a \text{ does not have characteristic } c\}$

`// Select the best applicants from both sets`

$X' \leftarrow \texttt{Sort(}X \textit{ by } P\texttt{)}[1:\texttt{Round}(N \cdot f)]$

$Y' \leftarrow \texttt{Sort(}Y \textit{ by } P\texttt{)}[1:\texttt{Round}(N \cdot (1-f))]$

return $S = X' \cup Y'$

Algorithm 5: Selection with multi-restriction (A, N, M, C, F)

Input: *A set of applicants A, an integer selection capacity N, a prediction model M, a set of protected characteristics C and an associated set of inclusion shares F*

Output: *The selection of applicants S of length N that for each characteristic c in C contains $N \cdot f_c$ applicants, which maximizes the expected outcome predicted by the model M*

`// Compute the predicted outcome for all applicants`

$P \leftarrow M(A)$

`// Create dummy of characteristics` c `in` C

for $c \in C$, $a \in A$ **do**

$$\mathbf{D}_{c,a} \leftarrow \begin{cases} 1 & \text{if } a \text{ has characteristic } c \\ 0 & \text{if } a \text{ does not have characteristic } c \end{cases}$$

`// Solve a binary linear program`

$$\begin{array}{ll} \text{Find a binary vector} & X \\ \text{that maximizes} & P^{\intercal} X \\ \text{subject to} & \mathbf{D}X = \texttt{Round}(N \cdot F) \end{array}$$

return $S = A[X]$

B Results for full time sample

Figure 8: Results for full time sample

(a) Graduation rate