# NERC ‘Pushing the Frontiers’

**Analysis of the influence of the partial randomisation intervention**

## Summary

Analysis of NERC’s ‘Pushing the Frontiers’ (PtF) scheme included partial randomisation in the selection of applications for funding. The analysis of the outcomes is complicated by the existence of a comparison group of predecessor schemes and the availability of a counterfactual of randomised outcomes. The information which might tell us whether PtF is inherently more/less biased than the schemes that preceded it is hard to interpret.

It seems safe to say that both PtF and its predecessors displayed biases in award outcomes against some groups who were likely to be minorities in the pool of potential applicants: those with disabilities, those with an ethnicity other than a White ethnicity and those who were female.

This analysis provides some evidence that PtF was less biased than were the preceding schemes – a combination of NERC ‘Standard Grants’ and ‘New Investigators’ schemes – but the difference is small and was strongly affected by the specifics of the actual randomisation that occurred.

The power of partial randomisation to reduce bias is limited by its partial nature: it cannot reduce bias associated with the decisions made when placing applications in bands. Therefore, we would expect the impact on bias to be relatively small. However, randomisation is a low-cost, low-effort intervention which may also bring benefits in reducing the burden on funding panels (panel members do not have to spend time carefully ranking the applications that fall around the funding cut off point) and over a large portfolio of grants, the impact could be significant. As we might expect, this analysis showed a small effect, that was often not statistically significant. It also demonstrates some of the challenges associated with experimenting in peer review.

In this context, it is important to point out that information that we collect to support equalities monitoring is not made available to assessors. However, assessors may be able to infer some of these characteristics from an applicant’s name or CV.

## Introduction

In 2023 NERC replaced its existing responsive mode Standard Grants and New Investigators schemes with ‘Pushing the Frontiers’ (PtF). Among other process changes PtF may now allocate funding based on a partial randomisation scheme. In this, proposals rated below the very highest tier but still considered to be highly competitive are randomly selected for funding.

PtF and its predecessors have demand management quotas, meaning that most applications are likely go through a pre-sifting process at the applicants’ host organisations prior to being submitted to NERC. Applications received may not reflect the diversity of the population eligible to submit applications.

This analysis looks at these questions:

- Did the introduction of partial randomisation lead to a change of diversity in the applicant and awardee populations?
- How did the outcomes of PtF differ from those seen in its predecessor schemes, in terms of award rates of applicants based on their age, disability status, ethnicity and sex?
- What was the ability of the randomisation scheme used, or other possible randomisation schemes, to influence those same outcomes?

Data from the predecessor NERC schemes is used as a ‘Comparison’ group, with the PtF applicants being described as a ‘Treatment’ group (although the two schemes did not run in parallel at any point). The question of whether/how age-, disability status-, ethnicity- and sex-related outcomes might have varied across Treatment and Comparison groups, and of how the process changes as a whole may have influenced outcomes, is of particular interest.

The Comparison group comprises ‘Standard Grants’ and ‘New Investigator’ funding decisions from January 2020 to January 2022, while the Treatment group comprises PtF decisions in January and July of both 2023 and 2024. Direct comparisons of award rates across Comparison and Treatment groups are of little interest and the focus will be on differences in award rates between the demographic groups of interest and derived measures, including differences-in-differences.

The outcome of interest is the group award rate, calculated as the number of applications where the lead applicant shared a characteristic and which were funded, divided by the total number of applications with a lead applicant sharing that characteristic. It is very unlikely that different groups of applicants will have exactly the same award rate. Chance variation will mean that there are always differences in award rates, even if there is no bias in decision processes that is related to applicants’ characteristics. To help distinguish between meaningful variation in groups’ outcomes and what is best thought of as ‘noise’, this analysis applies statistical tests to the data.

The general idea of these tests is to determine the range of outcomes that we might have seen if there had been no bias relating to applicants’ personal characteristics, and to compare the actual outcome with this. If the actual outcome is sufficiently extreme relative to the likely range of outcomes, it can be described as being statistically significant. Details of the method used are in the annex.

## Findings

### Was the introduction of partial randomisation associated with a change of composition of the applicant or awardee population?

The distribution of the proportion of applicants and awardees for the comparison group (Standard Grants rounds) and the treatment group (PtF grant rounds) for the measured characteristics are shown in the box plots Figures 1 to 4 and summarised Table 1.


**Figure 1** Proportion of applicants and awardees by sex for funding rounds in the comparison group (n=4) and the treatment group (n=4)

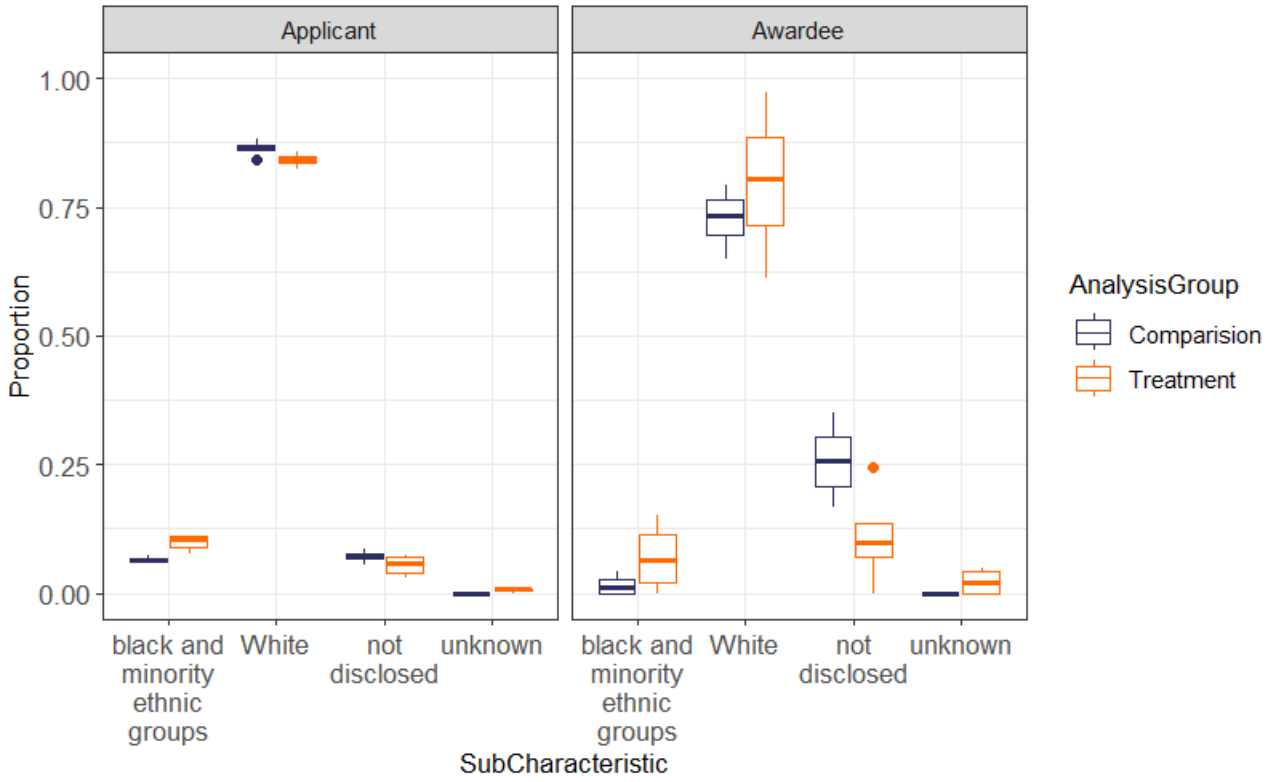


**Figure 2** Proportion of applicants and awardees by ethnicity for funding rounds in the comparison group (n=4) and treatment group (n=4)

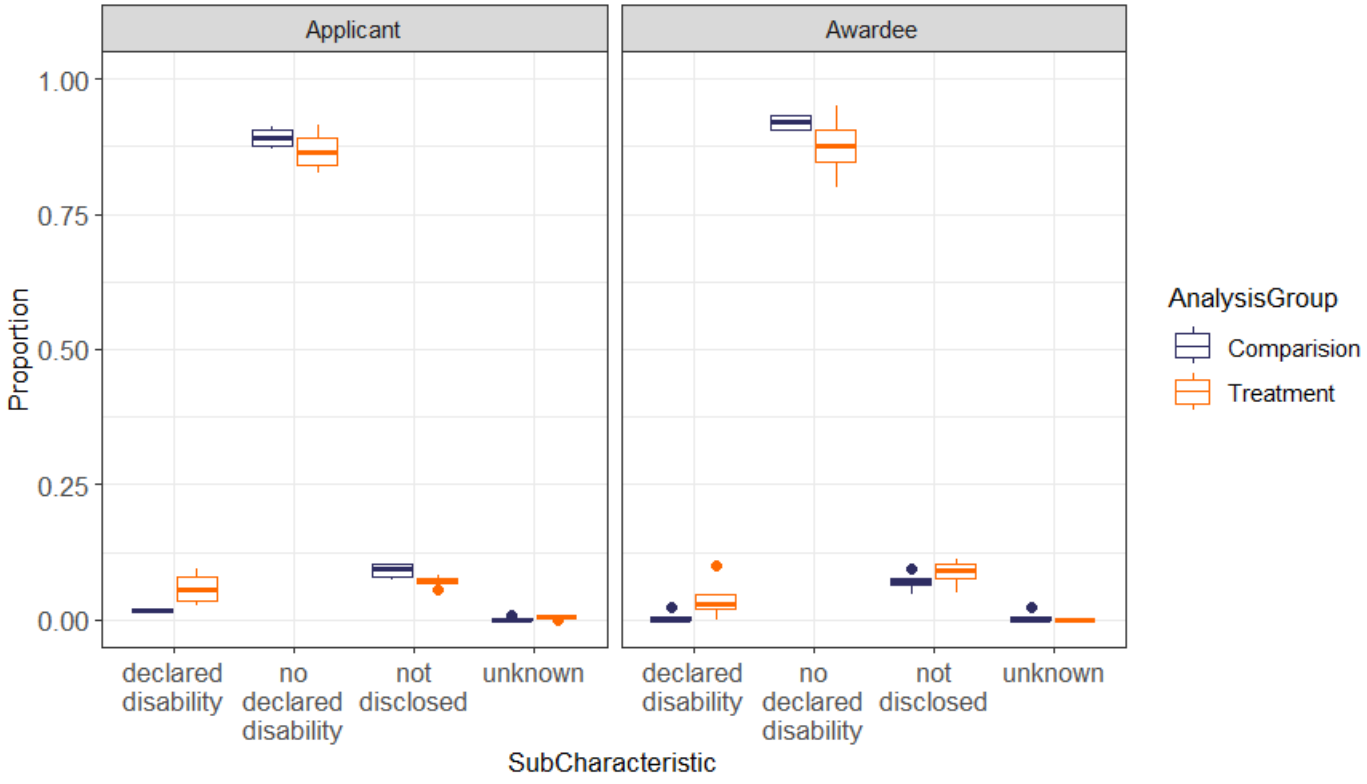


**Figure 3** Proportion of applicants and awardees by ethnicity for funding rounds

in comparison group (n=4) and treatment group (n=4)

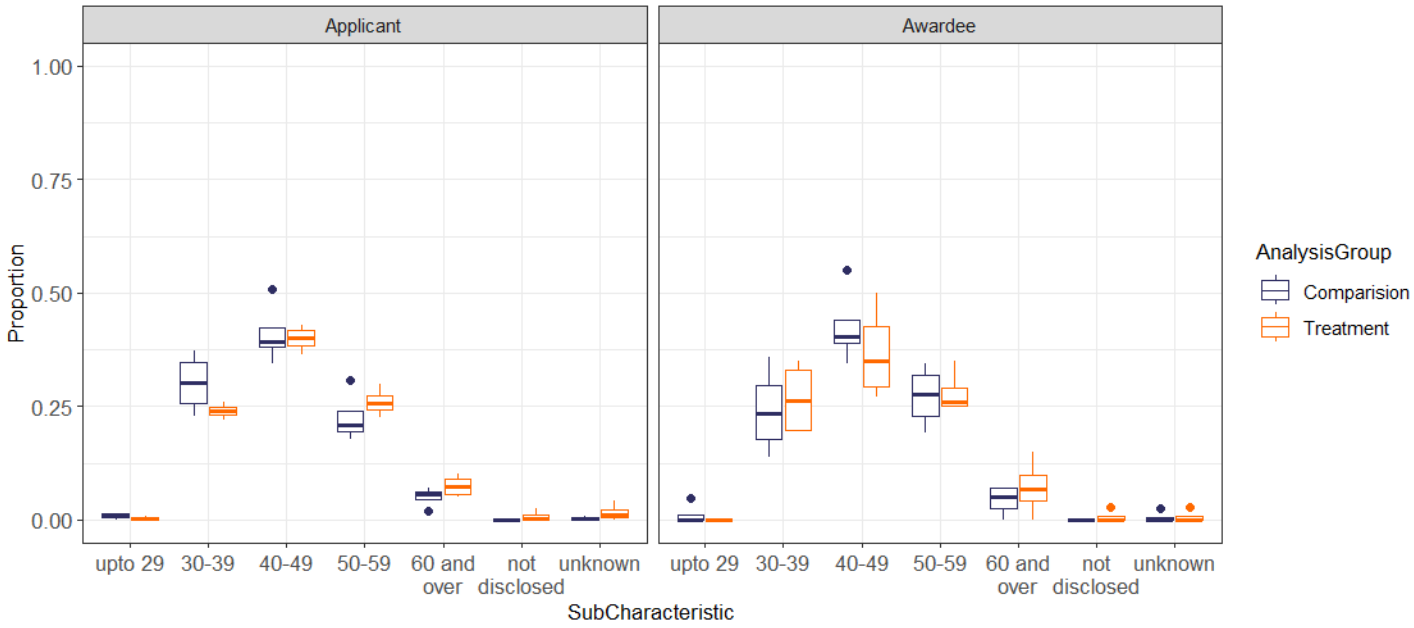


**Figure 4** Proportion of applicants and awardees by age for funding rounds in the comparison group (n=4) and treatment group (n=4)

**Table 1.** Numbers and proportions of applicants and awardees by characteristic and analysis group. Central tendency (mean, median) values are calculated over all funding rounds in each analysis group

| | | | Comparison Group | | | | | | Treatment Group | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Rounds | Total (n) | Mean (n) | Mean (%) | Median (n) | Median (%) | Rounds | Total | Mean (n) | Mean (%) | Median (n) | Median (%) |
| Applicants | | **Total** | 4 | 605 | 150 | | 150 | | 4 | 605 | 150 | | 150 | |
| | Sex | female | | 170 | 45 | 28 | 45 | 28 | | 165 | 40 | 27 | 40 | 27 |
| | | Male | | 425 | 105 | 70 | 105 | 71 | | 415 | 105 | 68 | 100 | 68 |
| | | not disclosed | | 10 | 0 | 1 | 0 | 1 | | 5 | 5 | 2 | 5 | 2 |
| | | unknown | | 0 | 0 | 0 | 0 | 0 | | 20 | 5 | 5 | 10 | 6 |
| | Ethnicity | ethnicities other than white | | 40 | 10 | 6 | 10 | 6 | | 60 | 15 | 10 | 15 | 9 |
| | | white | | 520 | 130 | 86 | 130 | 87 | | 510 | 130 | 84 | 125 | 84 |
| | | not disclosed | | 45 | 10 | 7 | 10 | 7 | | 35 | 10 | 5 | 10 | 6 |
| | | unknown | | 0 | 0 | 0 | 0 | 0 | | 5 | 0 | 1 | 0 | 1 |
| | Disability | declared disability | | 10 | 5 | 2 | 5 | 2 | | 35 | 10 | 6 | 10 | 5 |
| | | no declared disability | | 540 | 135 | 89 | 135 | 89 | | 525 | 130 | 87 | 130 | 87 |
| | | not disclosed | | 55 | 15 | 9 | 15 | 9 | | 45 | 10 | 7 | 10 | 7 |
| | | unknown | | 0 | 0 | 1 | 0 | 1 | | 5 | 0 | 1 | 0 | 1 |
| | Age | Up to 29 | | 5 | 0 | 1 | 0 | 1 | | 0 | 0 | 1 | 0 | 1 |
| | | 30-39 | | 180 | 45 | 30 | 45 | 30 | | 145 | 35 | 24 | 35 | 25 |
| | | 40-49 | | 250 | 60 | 41 | 60 | 40 | | 240 | 60 | 40 | 65 | 42 |
| | | 50-59 | | 135 | 35 | 22 | 35 | 22 | | 155 | 40 | 26 | 40 | 26 |
| | | 60 and over | | 30 | 10 | 5 | 10 | 6 | | 45 | 10 | 7 | 10 | 7 |
| | | not disclosed | | 0 | 0 | 0 | 0 | 0 | | 5 | 5 | 2 | 5 | 2 |
| | | unknown | | 0 | 0 | 1 | 0 | 1 | | 10 | 5 | 2 | 0 | 1 |
| Awardees | | **Total** | | 140 | 35 | | 35 | | | 110 | 30 | | 30 | |
| | Sex | female | | 35 | 10 | 25 | 10 | 27 | | 30 | 5 | 25 | 5 | 21 |
| | | male | | 105 | 25 | 73 | 25 | 72 | | 80 | 20 | 71 | 20 | 70 |
| | | not disclosed | | 5 | 0 | 2 | 0 | 1 | | 0 | 0 | 2 | 0 | 2 |
| | | unknown | | 0 | 0 | 0 | 0 | 0 | | 5 | 0 | 5 | 0 | 7 |
| | Ethnicity | ethnicities other than white | | 5 | 0 | 2 | 0 | 1 | | 5 | 0 | 5 | 0 | 5 |
| | | white | | 125 | 30 | 88 | 30 | 86 | | 100 | 25 | 88 | 25 | 86 |
| | | not disclosed | | 15 | 5 | 10 | 5 | 7 | | 10 | 0 | 7 | 0 | 7 |
| | | unknown | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 |
| | Disability | declared disability | | 0 | 0 | 1 | 0 | 0 | | 5 | 0 | 4 | 0 | 4 |
| | | no declared disability | | 130 | 35 | 92 | 35 | 92 | | 100 | 25 | 88 | 25 | 89 |
| | | not disclosed | | 10 | 5 | 7 | 0 | 6 | | 10 | 5 | 9 | 5 | 9 |
| | | unknown | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 |
| | Age | Up to 29 | | 0 | 0 | 2 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 |
| | | 30-39 | | 35 | 10 | 25 | 10 | 23 | | 30 | 10 | 27 | 5 | 25 |
| | | 40-49 | | 60 | 15 | 42 | 15 | 46 | | 40 | 10 | 38 | 10 | 32 |
| | | 50-59 | | 40 | 10 | 27 | 10 | 25 | | 30 | 10 | 28 | 10 | 29 |
| | | 60 and over | | 5 | 0 | 4 | 0 | 4 | | 10 | 0 | 7 | 5 | 9 |
| | | not disclosed | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 |
| | | unknown | | 0 | 0 | 1 | 0 | 1 | | 0 | 0 | 1 | 0 | 0 |

To reduce the risk of identifying individuals from the data published rounding, in accordance with HESA rounding methodology, has been applied to the figures in Table 1. These rules are applied after any calculations (sums, averages, percentages etc.) have been done so that changes to the data don't compound each other to give less accurate results.

Results from statistical tests for proportionality, summarised in Table 2, show there are significant differences in the applicant population between the treatment group and comparison group for applicants from ethnicities other than white ethnicities, those with declared disabilities and those in the 30-39 age group.

For ethnicities other than white ethnicities the proportion of applicants in the treatment group (PtF schemes) is 9.7% compared to 6.5% in the comparison group (Standard Grant schemes). For individuals with a declared disability the proportion of applicants in the treatment group is 5.8% and 1.7% in the comparison group. The proportion of applicants aged between 30-39 in the treatment group is 11.2%, significantly less than the 30.1% of applicants in the comparison group. The differences observed in the 30-39 age groups are likely to due to application processes in the comparison group having specific provision for ‘new investigators’ which was not a feature of the PtF scheme.

It is not possible to say conclusively that significant differences observed between the treatment group and the comparison are down to the introduction of partial randomisation in the award process. There are other plausible reasons for these differences, such as changes in the underlying population eligible to apply for NERC responsive mode awards and the impact of wider activities to promote diversity, equity & inclusion in research and innovation.

Other formal statistical tests of proportionality for the awardee populations do not indicate any significant differences between the treatment group and comparison group across the characteristics measured. This suggests that any variation seen is within the range we might expect to see if there had in fact been no underlying change.

Of note, the treatment group has a greater proportion of ‘unknown’ data points, across several characteristics. The number of ‘unknowns’ is unlikely to affect the interpretation of results because the absolute numbers are small in comparison to the wider ‘known’ data points, particularly if it is assumed the distribution of the ‘unknowns’ follows the same distribution of the known populations, although this is unknowable.

**Table 2.**
Results from 2-sample test for equality of proportions with continuity correction

| Population | | Sub-characteristic | $X^2$ | DF | p-value | | prop 1 | prop 2 |
|---|---|---|---|---|---|---|---|---|
| Sex | Applicant | Female | 0.18 | 1 | 0.67 | | 0.28 | 0.27 |
| | | Male | 0.60 | 1 | 0.44 | | 0.70 | 0.68 |
| | | not disclosed | 0.00 | 1 | 0.99 | | 0.01 | 0.01 |
| | | Unknown | 20.35 | 1 | 0.00 | * | 0.00 | 0.04 |
| | Awardee | Female | 0.00 | 1 | 1.00 | | 0.25 | 0.25 |
| | | Male | 0.04 | 1 | 0.83 | | 0.73 | 0.71 |
| | | not disclosed | 0.07 | 1 | 0.79 | | 0.02 | 0.01 |
| | | Unknown | 3.11 | 1 | 0.08 | | 0.00 | 0.04 |
| Ethnicity (binary) | Applicant | White | 1.06 | 1 | 0.30 | | 0.86 | 0.84 |
| | | other than white | 3.94 | 1 | 0.05 | * | 0.06 | 0.10 |
| | | not disclosed | 1.17 | 1 | 0.28 | | 0.07 | 0.05 |
| | | Unknown | 2.25 | 1 | 0.13 | | 0.00 | 0.01 |
| | Awardee | White | 0.00 | 1 | 1.00 | | 0.88 | 0.88 |
| | | other than white | 1.10 | 1 | 0.30 | | 0.02 | 0.05 |
| | | not disclosed | 0.29 | 1 | 0.59 | | 0.10 | 0.07 |
| | | Unknown | NA | 1 | NA | | 0.00 | 0.00 |
| Disability | Applicant | Declared disability | 13.21 | 1 | 0.00 | * | 0.02 | 0.06 |
| | | No declared disability | 1.47 | 1 | 0.23 | | 0.89 | 0.87 |
| | | not disclosed | 1.38 | 1 | 0.24 | | 0.09 | 0.07 |
| | | Unknown | 0.25 | 1 | 0.62 | | 0.00 | 0.00 |
| | Awardee | Declared disability | 1.39 | 1 | 0.24 | | 0.01 | 0.04 |
| | | No declared disability | 1.10 | 1 | 0.29 | | 0.92 | 0.88 |
| | | not disclosed | 0.10 | 1 | 0.75 | | 0.07 | 0.09 |
| | | Unknown | NA | 1 | NA | | 0.00 | 0.00 |
| Age | Applicant | up to 29 | 0.58 | 1 | 0.45 | | 0.01 | 0.00 |
| | | 30-39 | 5.28 | 1 | 0.02 | * | 0.30 | 0.24 |
| | | 40-49 | 0.16 | 1 | 0.69 | | 0.41 | 0.40 |
| | | 50-59 | 1.72 | 1 | 0.19 | | 0.22 | 0.26 |
| | | 60 and over | 0.13 | 1 | 0.71 | | 0.05 | 0.06 |
| | | not disclosed | 3.20 | 1 | 0.07 | | 0.00 | 0.01 |
| | | Unknown | 4.10 | 1 | 0.04 | | 0.00 | 0.02 |
| | Awardee | up to 29 | 0.30 | 1 | 0.59 | | 0.01 | 0.00 |
| | | 30-39 | 0.06 | 1 | 0.81 | | 0.25 | 0.27 |
| | | 40-49 | 0.41 | 1 | 0.52 | | 0.42 | 0.38 |
| | | 50-59 | 0.00 | 1 | 0.98 | | 0.27 | 0.28 |
| | | 60 and over | 0.54 | 1 | 0.46 | | 0.04 | 0.07 |
| | | not disclosed | Nan | 1 | NA | | 0.00 | 0.00 |
| | | Unknown | 0.00 | 1 | 1.00 | | 0.01 | 0.01 |

## How did the outcomes of PtF differ from those seen in its predecessor schemes, in terms of award rates of applicants based on their age, disability status, ethnicity and sex?

Figure 5 shows award rates for applicants in a lead role in the Comparison and Treatment groups, grouped by age, disability status, ethnicity and sex. Two ethnicity groupings are used, one which gathers more detailed categories into a five-way grouping, and a second which further aggregates into a binary 'White' and 'Ethnicities other than white ethnicities' scheme.

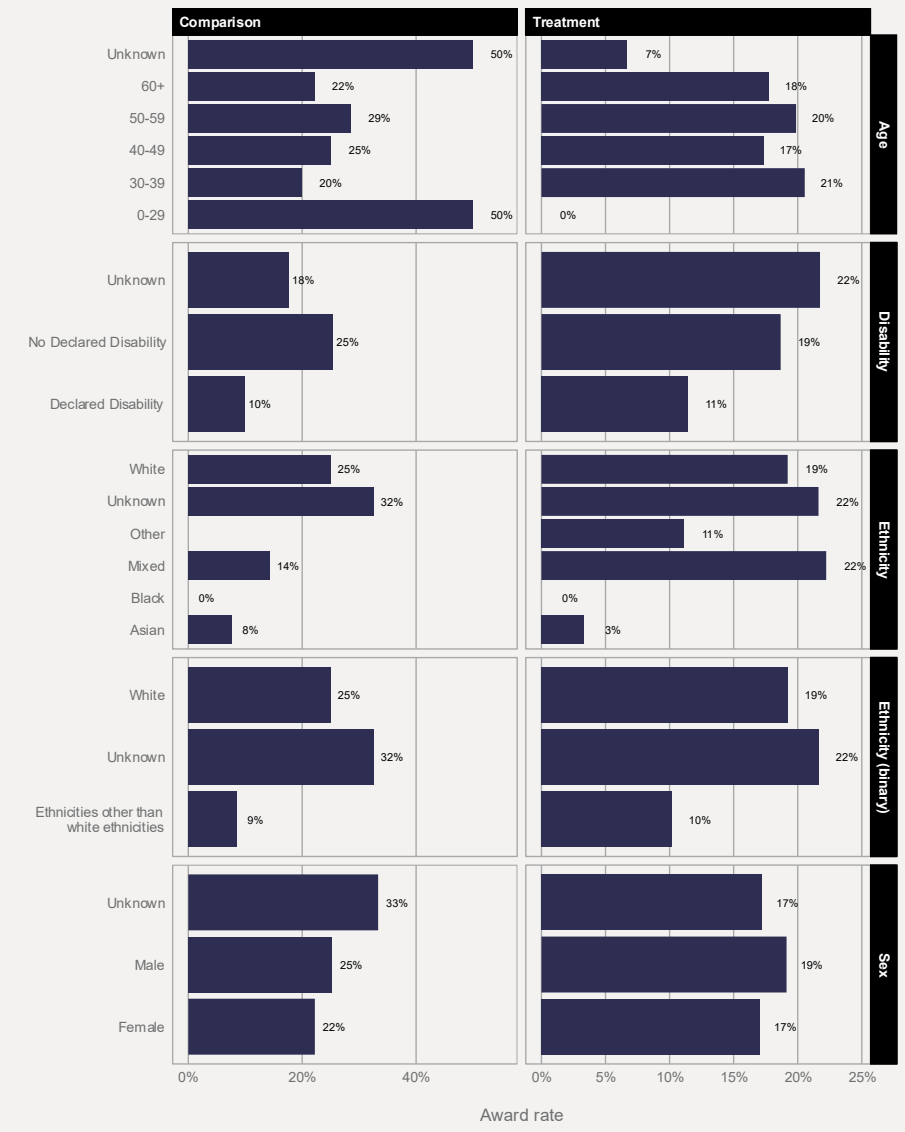


**Figure 5** Award rates for the five categorisations (rows) across four characteristics, for each of the Treatment and Comparison groups (columns). Note the different x-axis scales.

Taking each of the five categorisations in turn:

- There is no sign of an age-related award rate gradient in either the Comparison or the Treatment group
- Applicants not declaring a disability had higher award rates in both Comparison and Treatment groups

- There are indications that White applicants did particularly well in both Comparison and Treatment groups
- Male applicants had higher award rates in both Comparison and Treatment groups, but the difference was not large.

Figure 6 looks more directly at differences in award rates across binary groups, where this is possible. Age data is dealt with separately to deal with the five groups present in the category.

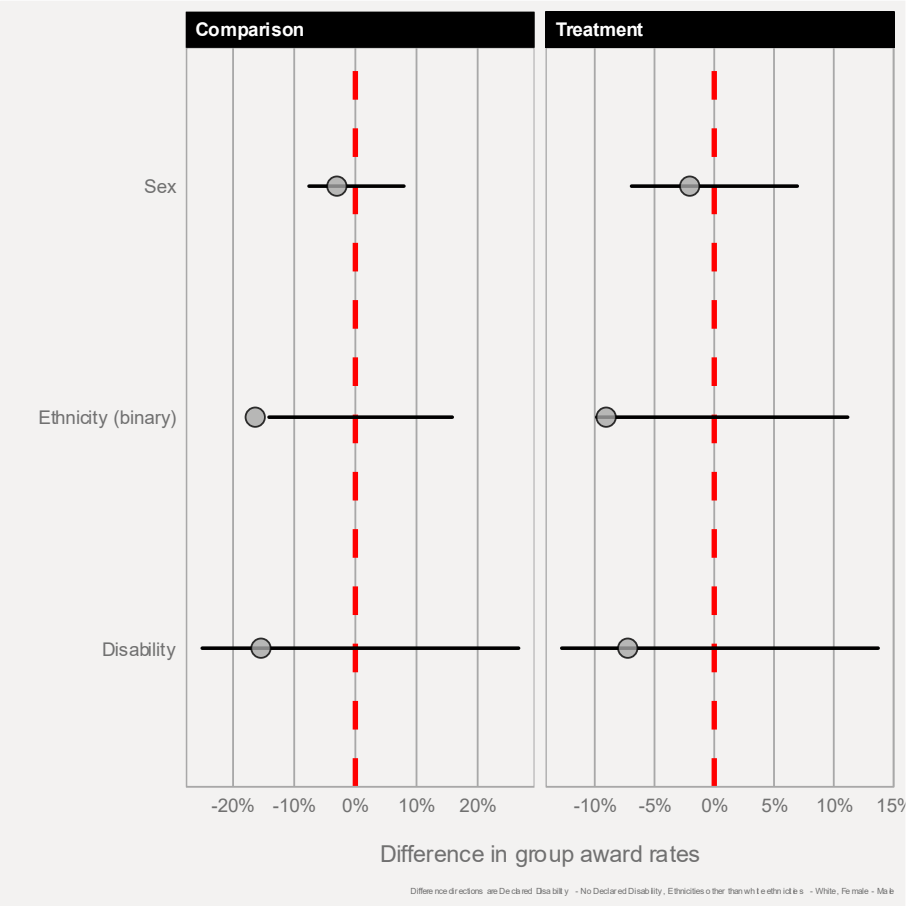


**Figure 6** Differences in award rates across binary characteristics for disability, ethnicity and sex. Direction of difference shown in the caption. Note the different x-axis scales in the two panels.

While outcomes tended to favour applicants who were male, those declaring a White ethnicity and those not declaring a disability, for most differences the observed value was within the expected range. The exception is for the difference in award rates using a binary ethnicity categorisation in the Comparison group, where the observed difference (~16 percentage points) was outside the expected range.

Even within the statistically non-significant results it should be noted that the award rate differences are in directions seen for many UKRI activities and for UKRI as a whole. It would be unwise to discount them purely because they fall within the intervals calculated, or to take that as evidence of a lack of bias.

We can extend this logic to ask whether the differences in award rates themselves differ across the Comparison and Treatment groups. Figure 7 shows these ‘differences-in-differences’ along with their expected ranges.

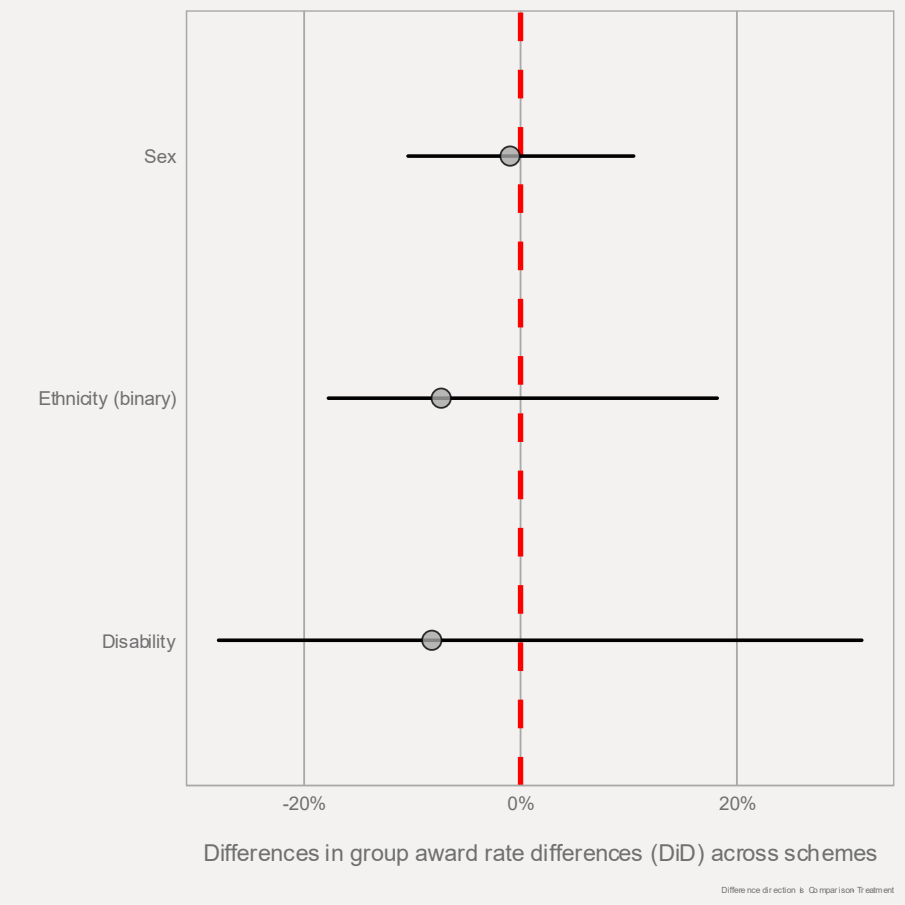


**Figure 7** Differences between inter-group differences for disability status, ethnicity (binary) and sex, across Comparison and Treatment groups

The differences-in-difference are all negative, showing that any bias is stronger in the Comparison group. None of them is outside the expected range, indicating that, statistically speaking, the bias was no greater or less following the move to PtF. However, it is notable that the Comparison data consistently shows a greater bias[1].

Figure 8 shows award rates by age group for Comparator and Treatment groups along with the range of rates we would expect to see if there was no association between outcome and age.

[1] There is one chance in eight that we would see all three biases stronger in the Comparison group if there was in fact no difference between the groups.

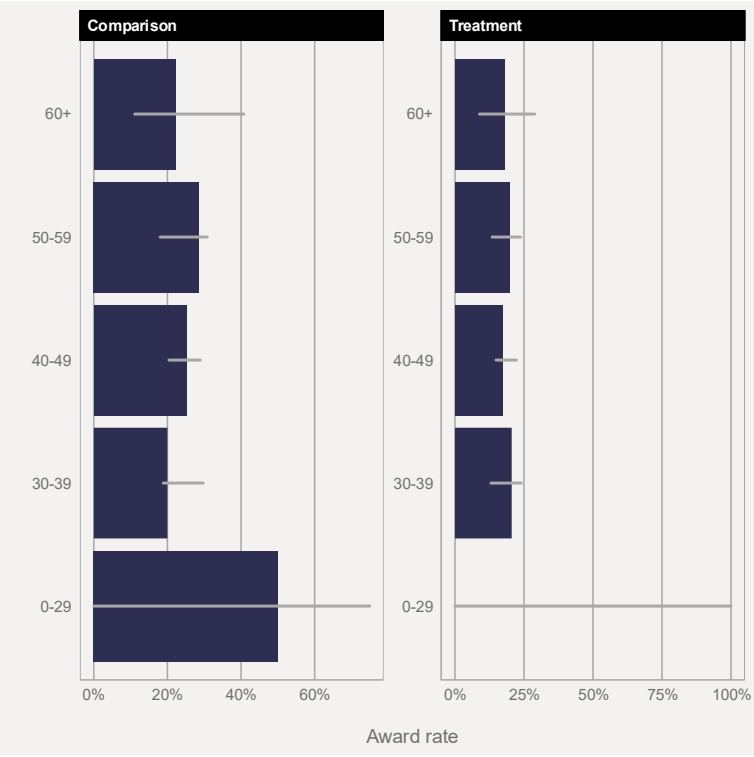


**Figure 8** Award rates by age group, with expected ranges (grey horizontal lines). Note the different x-axis scales.

All the observed rates are within their expected ranges, usually very comfortably so. Those in the Treatment group who fell into the 0-29 age range had an award rate of zero. However, there were fewer than 5 people in this category (table 1), and so the expected range is very wide (0% to 100%), meaning that it is not possible to reach any conclusions about potential bias against the very youngest applicants.

It is possible to compare directly the award rates of each age group (Figure 9.) If there is no difference between each group's probability of being successful, the observed difference in rates ought to be within a range centred on 0%.

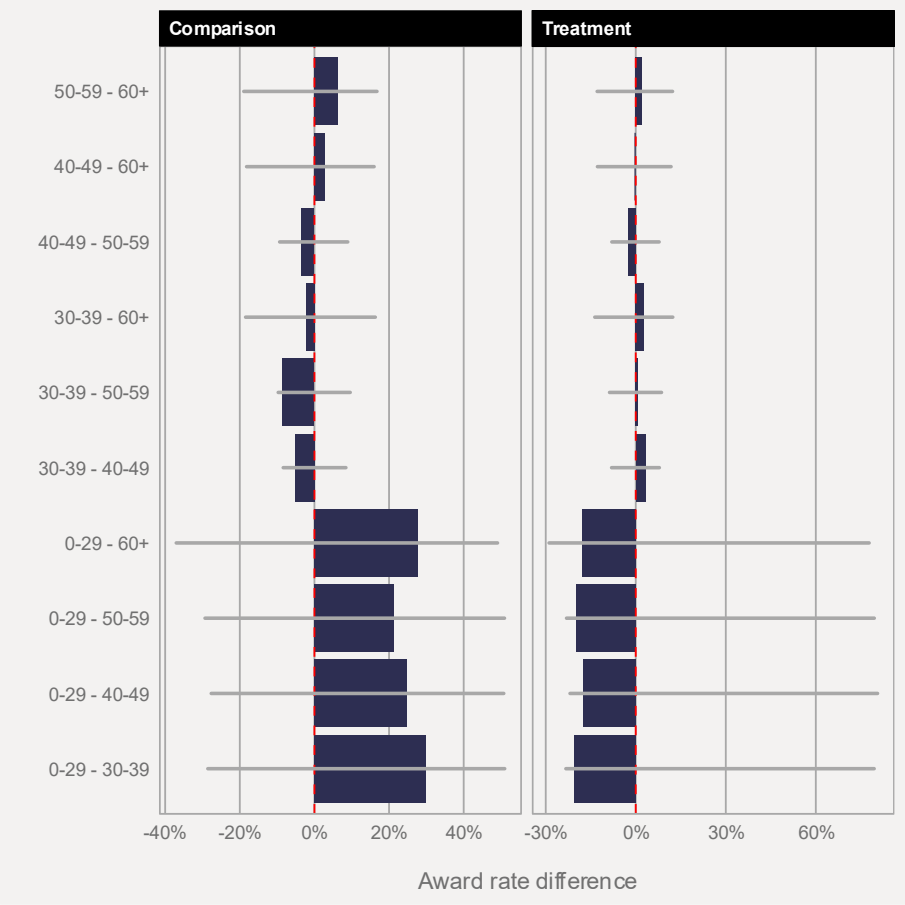


**Figure 9** Differences in award rates across age categories for Comparison and Treatment groups. Note the different x-axis scales.

For most age group pairings, the observed rate is indeed near zero and well within expected ranges. The pattern associated with decisions affecting those in the '0-29' age category is interesting, although also within expected ranges, which are wide, given the very small numbers of applicants in this group.

## What was the ability of the randomisation scheme used, or other possible randomisation schemes, to influence those same outcomes?

Turning now to the question of what might have happened if the randomisation had led to a different outcome, Figure 10 shows award rates by group overlaid with the range of rates that were possible under three different randomisation schemes. These schemes are:

- 'Partial' – this is the scheme actually implemented, in which only those highly rated but not receiving the very highest rating were placed in the randomisation pool
- 'Full' – these are the outcomes likely if all the highest-rated applications were placed in the randomisation pool, not just those in the second tier of scoring. It is the combination of those actually funded and those in the randomisation pool.
- 'All fundable' – the outcomes likely if all proposals deemed to be fundable had been placed in the randomisation pool. This is all proposals which had a score of at least 7.

As only the Treatment group underwent randomisation, only its results can be shown.

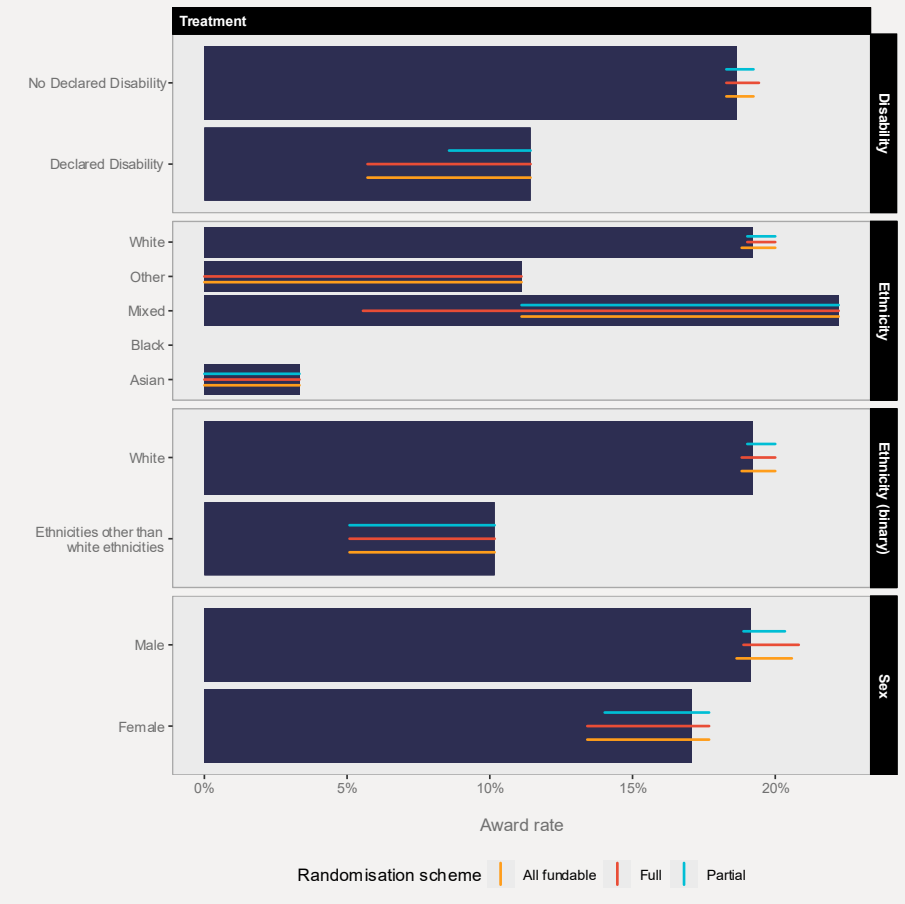


**Figure 10** Likely (95% ranges) award rates by group based on three different randomisation schemes: partial (the randomisation scheme actually implemented, pale blue), 'Full' (orange) and 'All fundable' (yellow).

Even the most comprehensive randomisation scheme ('All fundable') had little potential to reduce substantially the differences in award rates across groups.

To the extent to which it led to outcomes that differed from the mean value, the use of randomisation tended to favour applicants with declared disabilities, declaring an ethnicity other than White, or who were female. For the first two characteristics, the actual outcome was at the extreme end of the likely range. For sex the outcome was towards the top end of the likely range, but not as extreme.

The degree to which the results were at the ends of their likely ranges is more clear in Figure 11, which shows the inter-group rate differences and their likely ranges under the partial randomisation scheme that was actually used.

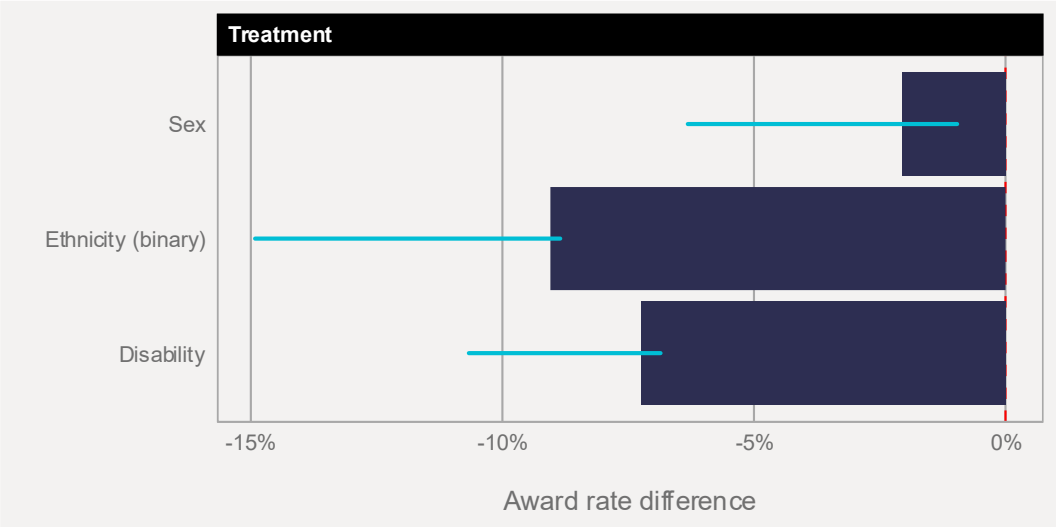


**Figure 11** Award rate differences across binary disability, ethnicity and sex categories (blue bars) and their possible ranges under partial randomisation (horizontal lines)

Had the randomisation outcome been somewhat more towards the middle of the likely range of outcomes, we might have seen award rate differences of ~4, 12 and 8 percentage points for sex, ethnicity and disability status. Instead we saw differences of ~2, ~9 and ~7 percentage points respectively: closer to parity.

Finally, Figure 12 shows the range of award rates in each age group that were likely under the partial randomisation process.

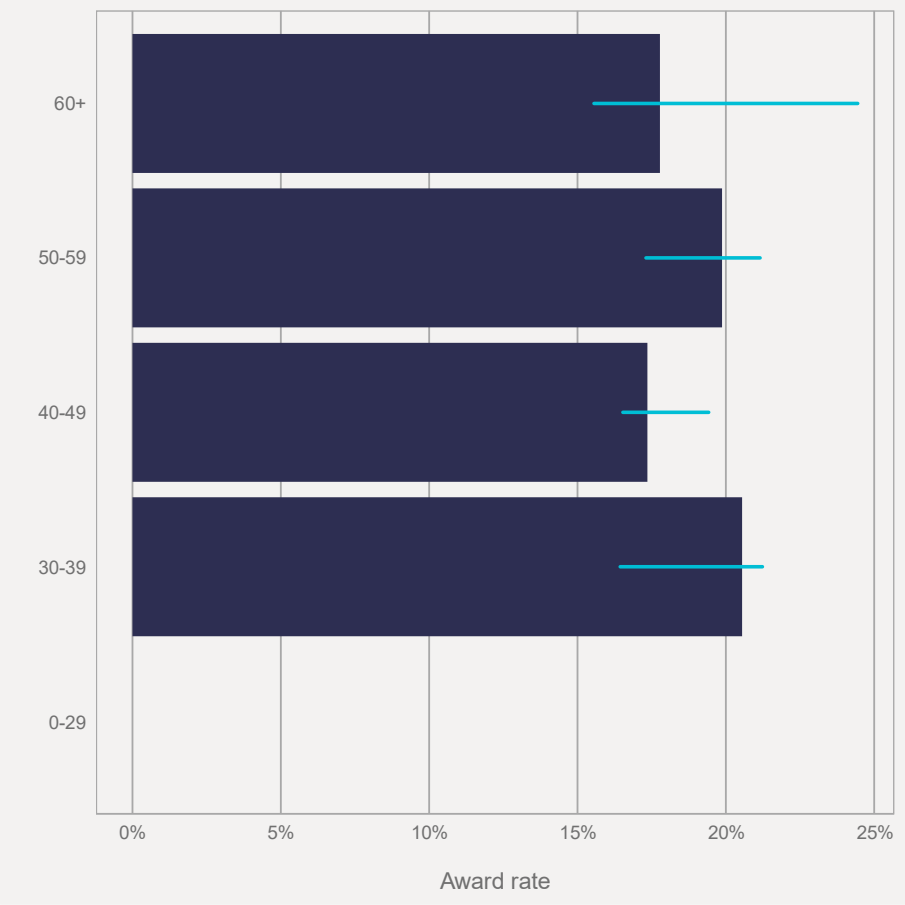


**Figure 12** Range of award rates likely under randomisation (horizontal lines) and observed values (bars), by age group

For the four age groups that saw at least one award made, the actual rate was well within the range of likely outcomes under randomisation. It was not possible for randomisation to have made any difference to those in the '0-29' group because none of them was ranked high enough to make it into the randomisation pool.

## Annex – explanation of statistical tests

### Permutation testing

The data we have is not a sample of all possible data. It includes the whole ‘population’ of funding decisions, and there is nothing more that can be known. The most familiar statistical tests are used, broadly speaking, to infer something about a population from a sample, given that the process of sampling leads to uncertainty about the population. Here there is no uncertainty about the population.

Instead we are faced with uncertainty about how these outcomes might have played out if things had been different. If we know the range of possible outcomes, we can place the actual outcome within that range and see whether the observed outcome was unusual. This is what the tests applied in this analysis do. They have been used in two ways.

For the analysis of success rates within and across groups sharing a characteristic, and their differences, we are interested in the question of whether outcomes are independent of that characteristic. For example, does an applicant’s ethnicity matter, in the sense of being associated with the outcome they experienced? We can test for this by randomly assigning ethnicities to applicants in the data (in the proportion found in the data) and then calculating the award rate for each group, or the award rate difference.

That is one instance of the sort of award rate (or difference in award rates) we might expect to see for that group if ethnicity had no influence on outcomes. If we repeat this process many times we create a distribution of the outcomes possible, conditional on the assumption that the characteristic is not associated with the outcome.

We can use this distribution to identify a range of outcomes that is compatible with a belief that outcome and characteristic are not associated. By comparing the observed result with this range of plausible outcomes, we can decide whether we believe that the observed outcome is too unusual for us to discount the possibility that outcome and characteristic are associated. To ensure that the range is a reliable reflection of what might have been, the sampling process has been repeated 10,000 times.

This is known as randomisation or permutation testing. The horizontal lines on the charts are a visual representation of the range of outcomes plausible under the null hypothesis of no association between outcome and characteristic. Sometimes these outcomes are award rates, sometimes they are differences in award rates and sometimes they are differences in differences. The logic of the test is the same in each case.

An interval covering 95% of the possible outcomes is used in the charts as it has some familiarity in relation to the traditional $p < .05$ statistical testing threshold. But it is an arbitrary choice, and others could be used which might be more or less conservative.

For tests associated with a partial randomisation process, the intervals show the range of outcomes that would arise if the randomisation had been repeated many times rather than just once. These intervals then show the extent to which (partial) randomisation of funding decisions might have had the ability to influence outcome, and where the actual randomisation used to make funding decisions sits in that spectrum.

Permutation tests make few assumptions, the main one being that of ‘exchangeability’. This is simply the requirement that the labels associated with a person can be swapped freely with the labels associated with another person. The data we have does not strictly meet this requirement because if the same person appears in the data more than once, presumably with the same age etc, each instance of their appearance ought to have the same label. The same person cannot have two different ages, and their characteristics may be associated with each other. In reality though, instances of duplicated applicants will be rare, and the calculated ranges will be affected only imperceptibly by this violation (which will tend to make the ranges larger than they ought to be, making the error a conservative one.)