ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

EPSRC Grant Funding: Statistical Analysis of Diversity in the Portfolio and Peer Review

Technical Report

UKRI Engineering and Physical Sciences Research Council

OPENING REMARKS

Sarah Cumbers, chief executive of the RSS

The RSS was very pleased to be commissioned by the EPSRC to independently investigate their grant portfolio and review process to see if – by using advanced statistical and machine learning techniques – it was possible to identify any evidence of bias. EPSRC, as far as we are aware, are the first research council to share their funding data with an external and independent body with a view to determining whether there are ways in which their processes entrench inequality and we warmly welcome that move.

The RSS is also grateful to the Alan Turing Institute, who we partnered with on this work. They developed a mixed methods approach drawing on qualitative research, linguistic analysis and machine learning, which were applied to understand perceptions and how demographic characteristics relate to reviewer scores and comments. This was an exciting and novel approach and shed light on the review process and how it is perceived by applicants.

This work is important. In 2024, the RSS published our new five-year strategy. Our first goal is to enable a strong and diverse profession. For the profession – especially in the context of academic statisticians working in universities – to be diverse, it is important that research councils' funding practices do not inadvertently entrench bias and make it harder for female and minority statisticians to pursue research careers.

In 2022 the RSS – along with the London Mathematical Society, Edinburgh Mathematical Society and Institute for Mathematics and its Applications – wrote to UKRI as it was developing its EDI (equality, diversity and inclusion) strategy. We expressed our concern that the way that grant submissions are assessed by reviewers focuses on the perceived quality of the researcher rather than the quality of research, which disadvantages women and minorities. We suggested that UKRI share the data that they have collected to allow the research community to examine the current funding system and interrogate hypotheses that have been proposed to explain the observed inequalities. This project arose from that correspondence.

Our research has identified some key areas where there is evidence of differential outcomes for people with different characteristics. The picture it paints is nuanced, but there is evidence that white researchers are more likely to be successful in applications than ethnic minority researchers; female researchers are more likely to be successful in fellowship applications than male researchers – but male researchers request and receive larger rewards when applying for research grants. This is a rich area of work that would benefit from further investigation and system change to ensure that in future all can be confident that decisions are based on merit.

Finally, I would like to acknowledge the work of the RSS volunteers who have made this work possible. Their time and effort has been crucial to the development of this report – making sure that the analysis is robust, is communicated clearly and can be helpful to EPSRC in their efforts to build a more diverse and inclusive research system.

CONTENTS

EXECUTIVE SUMMARY

EXPLORING POTENTIAL BIAS IN PEER REVIEW PROCESSES

In this report, we present the findings of an exploratory study commissioned by the Engineering and Physical Sciences Research Council (EPSRC) to the Royal Statistical Society (RSS). The project set out to investigate potential bias in the EPSRC peer review processes using structured data analysis and textual analysis of reviewers' comments and scores, this latter done in partnership with the Alan Turing Institute.

BACKGROUND

The RSS has been commissioned by the EPSRC to carry out an independent exploration of its portfolio of research grants to investigate potential bias in the EPSRC peer review processes in two ways: (1) focussing on the characteristics of the individuals' awarded grants and (2) through textual analysis of reviewers' comments and scores. The study uses data collected over the financial years 2014-2023. It aims to establish whether there are demographic disparities in its peer review processes and awards, including potential effects due to intersectionality.

There is a large body of work regarding bias and peer review, across both grant funding and academic publications[1]. Sex[2] has been a particular area of focus, generally due to the availability of data with other protected characteristics receiving less attention. Findings can vary by country, discipline, or institution. For instance, there is evidence from the USA of variation in grant funding by ethnicity,[3] which has been flagged in previous internal work by the EPSRC. Although the practices of different funding agencies around the world vary, they usually involve similar stages: 1) funding call; 2) application; 3) peer review to assess each application, often using both numerical scoring and free text comments; 4) panel, where panellists meet to discuss all applications under the same call/theme, producing rankings or recommendations; 5) funding outcome. Scholars with particular protected characteristics could become disadvantaged during any stage.

The literature review suggests that there is evidence of bias on the basis of sex, ethnicity , career stage, disability and institution. Previous internal work by EPSRC also indicated some evidence of bias on the basis of differential sex and ethnicity. In particular, previous internal work by the EPSRC[4], based on data from the funding years 2014-2019, returned a mixed picture regarding sex, evidencing that award rates by number for research grants led by male and female Principal Investigators (PIs) were very similar during this period, while award rates by number for fellowship applications by females have been higher than funding rates for fellowship applications by males by about 20 percentage points since 2015/16. For research grants, mean and median values are higher for male PIs than female PIs, and women are underrepresented both as applicants and as awardees for very large grants (> £2.5 million).

Regarding ethnicity, previous work by the EPSRC[5] flagged up consistent disparities with ethnic minority researchers underrepresented in the grant portfolio, as well as in award rates for PIs, co-investigators and fellowship applicants. White applicants were consistently proportionately more highly represented than those applicants from ethnic minority groups, with median award value for these researchers consistently lower than median award value for white[6] researchers. Significant lack of trust in the peer review process has been highlighted by ethnic minority researchers, primarily due to feeling they experience bias at the peer review stage.

Building on this previous work, this project aimed to:

- Survey the research population to understand their impressions about disparities in the funding process.

[1] See eg, EDICa https://edicaucus.ac.uk/peer-review-bias/
[2] Throughout the executive summary and technical report, we follow the dataset in referring to sex rather than gender and use the provided male and female labels accordingly.
[3] See, eg, Race, Ethnicity, and NIH Research Awards (Ginther et al, 2011), https://doi.org/10.1126/science.1196783

[4] Understanding our portfolio: a gender perspective – UKRI
[5] See, for instance, https://www.ukri.org/publications/epsrc-detailed-ethnicity-analysis/
[6] Throughout this report, we follow the BBC style with regards to the capitalisation of ethnicities. See https://www.bbc.co.uk/newsstyleguide/all/#r.

- Use statistical inference methods to understand which disparities or changes in disparities observed were likely to have arisen as the result of expected random variations in the process and which could indicate issues with the process.
- Investigate dimensions of difference that have not previously been studied in the data, particularly institution, region, research area group and theme, and funding mode.
- Carry out multivariate analyses of the data in order to understand how a range of factors in combination influence funding outcomes.
- Undertake separate analyses of different parts of the funding process, including applications, reviewer comments, reviewer scores, panel rankings and final funding outcomes to investigate evidence of disparities at each stage of the pipeline.

HEADLINES

Our research has found some evidence that ethnic minority researchers in the engineering, mathematical and physical sciences don't get the same level of funding as white researchers. The odds of ethnic minority researchers being successful in a funding application overall are around 32% lower than for white researchers (odds ratio of 0.684, 95% confidence interval [0.629, 0.744]). Notably, we found that there were meaningful interactions between ethnicity and age that help contextualise this difference—average predicted rates of funding success for white applicants in the under 36, 36-55, and over 55 age brackets were 37.4%, 32.7%, and 32.1%, respectively, while for ethnic minorities they were 34.2%, 24.4%, and 21.3%. Our research also found that there is a higher predicted success rate for having UK nationality as a white applicant (35.1% for UK nationalities compared to 29.7% for non-UK nationalities[7]), but this is considerably weaker for ethnic minority applicants (25.9% compared to 25.0%).

Our modelling also suggests that when they are successful in research grant applications, ethnic minority researchers applied for less funding – for every £1 a white applicant successfully applies for and receives, a successful ethnic minority applicant is expected to apply for and receive 90p.

It should be noted that the EPSRC generally awards the values for which researchers apply and does not cut value from grants. Differences in expected funding amounts for successful applications may reflect different amounts applied for on average between the demographic groups or bias implicit in a panel's funding decisions, or both, depending on both applied-for funding value and demographic characteristics.

The picture is more nuanced in the case of sex. We found that female applicants are more likely to be successful in funding applications, with their overall odds of success being around 13% higher than for male applicants (odds ratio 1.13 [1.042, 1.232]). The effect is driven primarily by fellowship applications, for which the odds of success for female applicants are 80% higher than for males (odds ratio 1.80 [1.426, 2.279]). For research grant applications, we did not find evidence of differential odds of funding success between the sexes (female odds ratio 1.026 [0.916, 1.148]). Proportionately far more females apply for research grants than fellowships, which helps explain why the overall odds of funding success are only 13% higher for females.

However, male applicants, when successful, are likely to apply for and receive a higher value award than female applicants: our models suggest that on average, where a successful male applicant would apply for and receive £1 in funding, a female applicant would be expected to apply for and receive 85p. Again, these differences may at least partially driven by discrepancies in the award amounts that are requested by the applicants themselves, rather than a simple reflection of reviewer or panel biases.[8]

Disparities of any sort may feed into perceptions of bias, particularly as researchers have very personal experiences of the review process. So, with the Alan Turing Institute, we looked further into this. The RSS commissioned a survey of EPSRC applicants to get their perspectives on the process. We found that ethnic minority researchers were the

[7] When reporting on nationality, we follow the terminology given in the shared data, which uses "UK" and "non-UK" as the categories.

[8] The EPSRC examined differences in requested funding amounts in previous work available here: https://www.ukri.org/publications/gender-diversity-in-our-portfolio-survey-findings/

most likely group to perceive bias in the peer review process. However, the most common reason given for the perceived bias was institution rather than ethnicity – indicating a more nuanced picture than simply processes that inadvertently introduce bias against ethnic minority applicants.

Another key component of this analysis was to examine reviewer and panellist's scores and comments more closely to determine more precisely where there may be sources of bias in the review process itself, rather than in its outcomes or the perceptions of bias that permeate the research community.

While this portion of the analysis found that the sentiment reviewers expressed in their comments generally tracked their scores well, it also found modest but significant negative effects of reviewer and applicant ethnicity on scores (0.23 points [0.20,0.26] higher both for white reviewers and white applicants). This overshadows effects based on reviewer and applicant sex (0.07 points [0.04,0.10] higher both for male applicants and male reviewers). This is reflective of the picture that we have found throughout: there is evidence that ethnic minority researchers are not achieving the same level of application success as white researchers and a complicated picture around the effect of researchers' sex on outcomes.

Reviewers with white and Chinese ethnicity appeared to give higher scores to applicants that share their ethnicity; white reviewers gave white applicants scores that were 0.239 ([0.106, 0.373]) points higher than other applicants, and Chinese reviewers gave scores that were 0.567 ([0.351, 0.783]) and 0.560 ([0.019, 1.10]) points higher to applicants with Chinese and "other" ethnicities, respectively.[9] Other combinations of reviewer and applicant ethnicities did not show significant associations with score, although small sample sizes for many of these combinations limits the precision of our estimates, so some caution in interpretation is warranted.

Finally, the analysis of the review process found evidence that the composition of panels matters. Specifically, female applicants are ranked 8.5% higher (ie more favourably) than male applicants in interviews when at least one panellist was female but 7.9% lower (ie less favourably) than males when this was not the case. This gives evidence for the likely effectiveness of policies like a Mixed Gender Panel Policy, which has been run by the EPSRC since 2016—the policy has led to the panellist population being more than 30% female in each year since 2017, roughly 10 to 15 percentage points higher than the female share of the reviewer, applicant, or awardee populations each year. One caveat to this is that, because nearly all of the panels that did not have females would have taken place before this policy was enacted, the results may be sensitive to changes over time as well as panel composition.

OUR FINDINGS

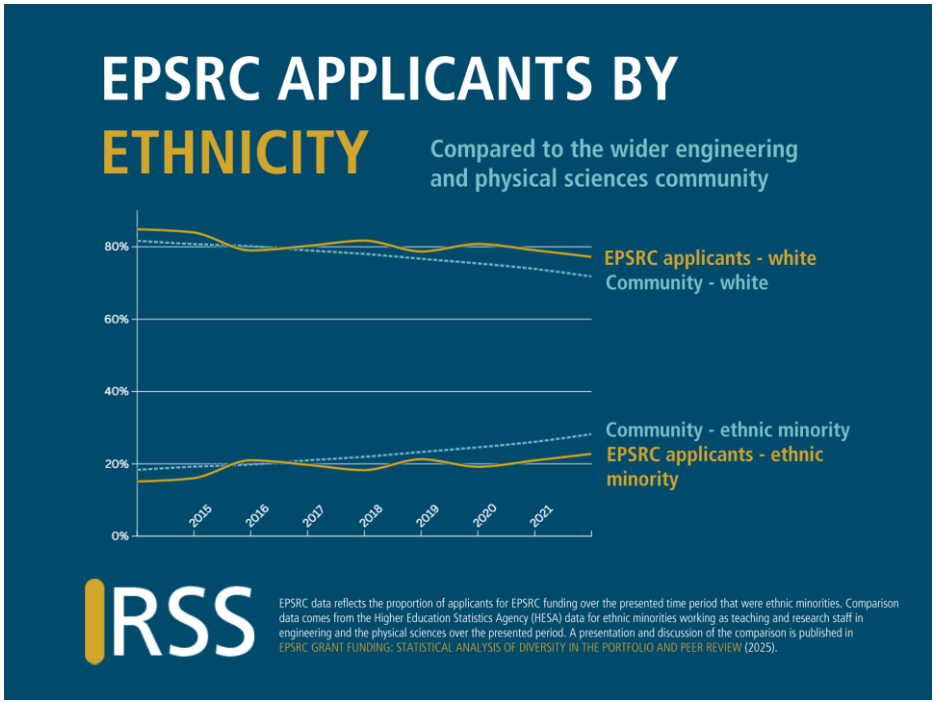
Potential sources of bias and disparity in the application and funding process can come from a variety of sources. The focus of our research was split between three broad conceptual "stages" of the process that each have the potential to be one such source, starting with the representativeness of different groups applying for EPSRC grants, then moving into the peer review process itself, and finally examining outcomes.

As mentioned earlier, previous research had found that females and ethnic minorities are underrepresented in the EPSRC applicant data. Part of our research sought to update this analysis and consider whether there has been change over time. To do this, the population of researchers applying for EPSRC grants was compared with the HESA "Teaching and Research" population (across Engineering and Physical Sciences subject areas). Over the 10 years the female proportion of the EPSRC applicant pool has roughly tracked the HESA female population, both of which have seen small increases. For ethnicity the picture is different with the small increases over time in the ethnic minority EPSRC application population not matching the more substantial increases in the HESA population, widening the gap to around 5.5% by 2022.

Perhaps unsurprisingly the age profile of EPSRC applicants is less reflective of the HESA population with older researchers (>55) being under-represented in the EPSRC applicant population. Individuals with declared disabilities are consistently

[9] See Table 4 in the Alan Turing Report available on the GitHub repository: https://github.com/alan-turing-institute/equity-in-grant-funding

underrepresented in the EPSRC applicant population (by about 1% across the 10 years of the data) relative to the HESA population.



We also investigated the role that protected characteristics might play in the grant review process itself. This analysis focused on reviewers and panellists' evaluations of applications, as well as the perceptions of applicants regarding bias in the process. We analysed grant reviews using sentiment analysis and modelled the dependence of reviewer scores on review sentiment and of panel outcomes on both reviewer scores and review sentiment.

Overall, this analysis indicated that assessment scores are strongly associated with the sentiment in reviewers' comments. Longer reviews and reviews including more standout adjectives tend to accompany higher scores, although their effects are much weaker than sentiment. There was no evidence to indicate that words in reviews associated with male gender stereotype were preferred over stereotypical female traits during scoring: the use of both masculine and feminine words on average are associated with slightly higher reviewer scores.

We explored the joint effect of applicant and reviewer characteristics on review scores. This analysis indicated that on average reviewers nominated by the applicant rated that application 0.723 [0.697,0.748] points (out of 6) higher than other reviewers of the same application. We also found that overseas reviewers on average produced scores 0.325 [0.287,0.363] points higher than UK-based reviewers. In this model, the ethnicity of the reviewer and the applicant have a larger influence (0.23 points [0.20,0.26] higher both for white reviewer and white applicant) on reviewer score than their sex (0.07 points [0.04,0.10] higher both for male applicant and male reviewer), although both effects are statistically significant at conventional significance levels. Investigating interaction, there is no apparent interaction between sex of applicant and reviewer. However, there is some evidence that white and Chinese reviewers give higher scores to applicants who share their ethnicity, while we cannot say the same for other combinations of reviewer and applicant ethnicities.

We also investigated the influence of review narratives and scores on panel rankings. Here, interview panels exhibit different dynamics, possibly because those panels have extra information in the form of the interview itself. Considering reviewer score only, the ranking from interview panels is much more weakly predicted by reviewer scores compared with panels not conducting interviews. There also appears to be more variations associated with demographic characteristics in interview panels: male applicants were ranked 7.3% [3.6%,11.1%] lower and UK nationals 4.9% [1.6%, 8.2%] higher on average in interview panels, but a similar effect was not observed in panels not involving interviews. White applicants were ranked significantly higher than ethnic minorities in both types of meetings, but the effect size is smaller in non-interview panels (1.4% [0.4%, 2.4%]) than in interview panels (5.7% [1.6%,9.8%]).

Including the language features of the comments (averaged over all reviews received) the effect of applicant ethnicity (higher ranking for white applicants) becomes statistically significant for both non-interview (1.1% [0.2%,2.1%]) and interview (5.0% [0.9%,9.2%]) panels. Sex and nationality are still associated with different ranking outcomes, but only in interviews (male 7.1% [3.3%,10.8%] lower, and UK nationals 5% [1.7%,8.3%] higher).

We also investigated interactions between protected characteristics of the panel and the applicants. A binary indicator of whether there were any female panellists present appears more influential compared to the sex of the chair of the panel or the proportion of female panellists. In particular, female applicants were ranked 8.5% higher than male applicants in interviews when at least one panellist was female but 7.9% lower than males when this was not the case.

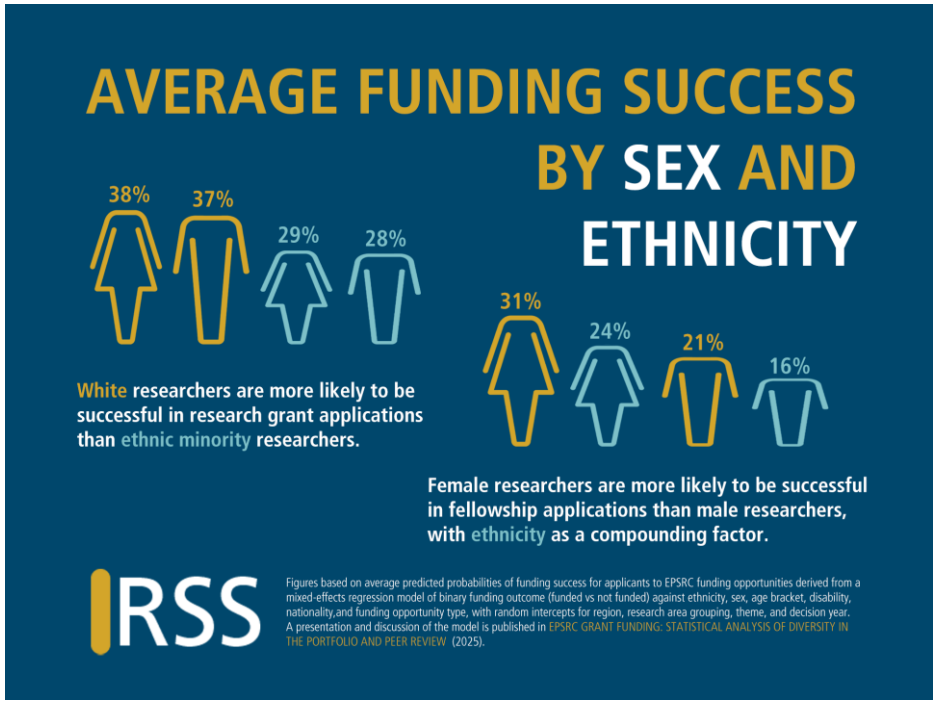
Additionally, as part of this study, the Alan Turing Institute conducted a research community survey into perceptions of bias in the EPSRC review and award processes. Higher perceptions of bias in the EPSRC peer review process were reported by male respondents, by ethnic minority researchers and by individuals in the age range 56 or above. Applicants indicated that perceived bias adversely affected career progression, such as delays in promotions and missed funding opportunities, and impacted mental health, including increased stress and demoralisation. Institutional prestige was the most commonly cited source of perceived bias, followed by factors such as sex, nationality/language, ethnicity, and age.

The final "stage" of the EPSRC process that we examined was its outcomes, focussing especially on rates of funding success and the value of successful applications by demographic group. The analysis of binary funding success used logistic regression models to investigate how grant success depended on protected characteristics – age, disability, ethnicity and sex – controlling for funding type and mode, region, time, research area and theme.

The effect of any given explanatory variable in a logistic regression model can be presented as an odds ratio, the relative odds of success (odds being the success rate divided by the failure rate) for different demographic groups, with an odds ratio of 1 corresponding to identical rates. Where the odds ratio in favour of one group over another is greater (less) than one then the success rates is higher (lower) in that group. We also present 95% confidence intervals for odds ratios which can be interpreted as the range of possible population values which are supported by the observed data.

This analysis found a strong significant effect of ethnicity on grant success. The odds of ethnic minority researchers being successful in a funding application are around 32% lower than for white researchers (odds ratio of 0.684, 95% confidence interval [0.629, 0.744]). To help interpret these figures, we calculated the corresponding estimates for success rates for individual proposals by sex and ethnicity across a population whose other characteristics matched the EPSRC applicant population. We found that for a white male the mean chance of success in a research grant proposal was 36.5%, whereas for an ethnic minority male it was 27.5%. For females, the mean chance of success for a research proposal was 38.2% for white applicants and 28.7% for ethnic minority applicants.

There is a more nuanced picture in the case of sex. We found that female applicants are more likely to be successful in funding applications, with their odds of success being around 13% higher than for male applicants (odds ratio 1.13 [1.042, 1.232]). The effect is more marked for fellowship applications where the odds of success for female applicants are 80% higher than for males (odds ratio 1.80 [1.426, 2.279]). We have estimated that this corresponds to a 38% mean predicted chance of success for white female applicants for research grants compared to 37% for white males (with a similar gap between female and male ethnic minority applicants).



For fellowships, though, the corresponding figures are a 31% chance of success for white female and a 24% chance of success for ethnic minority females compared to a 21% chance of success for white males and a 16% chance of success for ethnic minority males. However, as detailed below, while males are less likely to be successful than females in applications, they also apply for and can expect to receive more funding when they are successful.

Focusing for now on application success, we found that UK applicants are more likely to be successful than non-UK applicants (with an odds ratio 0.818 [0.761, 0.879]). We also found evidence of enhanced success rate for younger applicants (35 and under) estimated as an odds ratio of 1.332 [1.227,1.447] compared to applicants in the age range 36 to 55 and a similar value relative to applicants in the age range 56 and older. No significant effect of disability was identified.

We examined intersectionality by investigating the significance of interactions between protected characteristics. This analysis identified potential interactions between ethnicity and nationality and ethnicity and age. The predicted rates of funding success for ethnic minority lead applicants with UK and non-UK nationalities being 25.9% and 25.0%, respectively, while the predicted rates for white lead applicants with UK and non-UK nationalities being 35.1% and 29.7%, respectively. We also found that white lead applicants had a 37.4%, 32.7%, and 32.1% chance of success in the under 36, 36 to 55, and over 55 age brackets, respectively. This compares to 34.2%, 24.4%, and 21.3% for the same age brackets for ethnic minority lead applicants. The age-related interaction may indicate some level of convergence between white and ethnic minority applicants with each generation, although there may genuinely be bias based on the intersection of ethnicity with age that we will continue to observe with time.

We conducted a similar analysis for panel ranking. As part of EPSRC's peer review process, applications are sent to a panel to be reviewed and ranked. Our analysis looked at the characteristics of applicants and calculated the probability of applicants being ranked in the top quarter of the rank-ordered list. This analysis uncovered similar, but not identical results. We found a weaker effect of ethnicity (odds ratio 0.794 [0.719, 0.876] of appearing in the upper quarter of the list for ethnic minority relative to white applicants), but no significant dependence on nationality (odds ratio 0.965 [0.887, 1.049] non-UK to UK) or sex (odds ratio 1.088 [0.985, 1.201] female to male). For age, the odds of appearing in the upper half of the list decrease significantly with increasing age (0.633 [0.577, 0.695] and 0.529 [0.462, 0.604] odds ratio for applicants 36-55 and 56 and older respectively relative to applicants 35 and under).

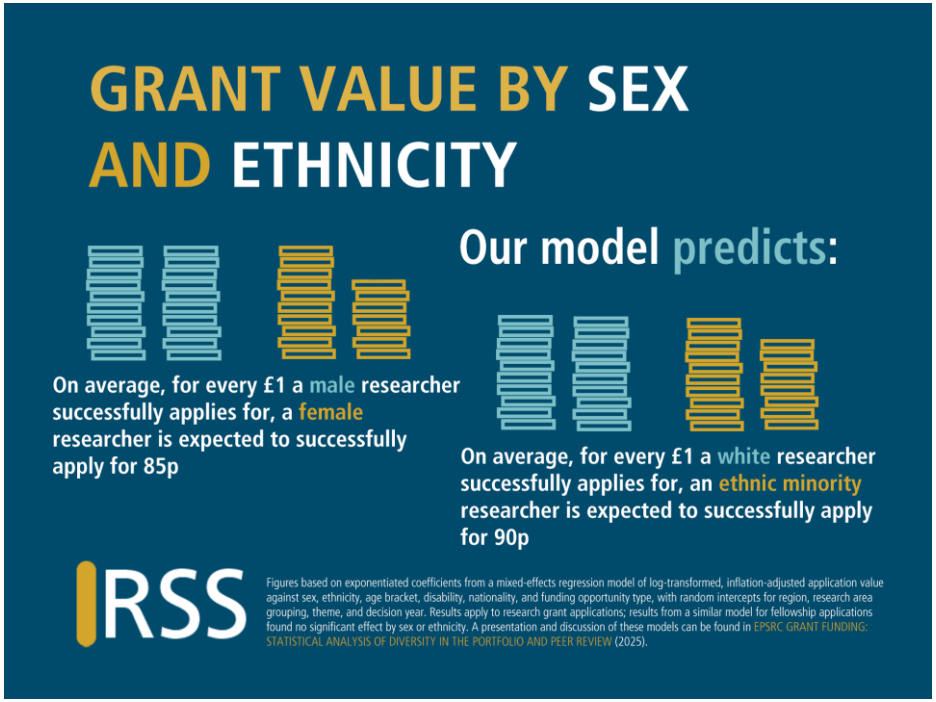
In a parallel descriptive analysis, we focussed on the characteristics which differentiate those applications which were funded, those which EPSRC considered to be fundable, but which were ultimately unsuccessful and those which were considered to be not fundable. This analysis, which separates fellowship applications from other grant funding sheds some further light on some of the differences identified in the main analysis above. In particular, any differences in overall funding success between male and female applicants largely seem to arise from fellowships, where female applicants are funded at a higher rate than male applicants. A similar effect is observed for nationality with higher success rates for fellowships for UK applicants. Applications by ethnic minority applicants, both for fellowships and for other funding opportunities are disproportionately assessed as unfundable. Similarly, analysing the difference between success rates for responsive mode and strategic funding opportunities, we find that females and younger (<35) applicants fare better in responsive mode with no significant differences by sex or age for strategic funding opportunities. On the other hand, the observed effect of non-UK and ethnic minority applicants experiencing lower success rates is roughly consistent across responsive mode and strategic funding opportunities.

We also investigated the association of applied for and awarded grant value for successful grants with the protected characteristics of the Principal Investigator, controlling for funding type and mode, region, time, research area and theme. Again, we analysed fellowships and other funding opportunities separately. For non-fellowship grants, our analysis used linear models for grant value on a logarithmic scale, so the identified effects are multiplicative. Significant effects of sex, ethnicity, nationality and age (but not disability) were identified. Successful grants by female PIs are estimated to be 84.7% [78.3%,91.6%] of the value of successful grants by male PIs; successful grants by ethnic minority PIs are estimated to be 90.1% [82.9%,97.8%] of the value of successful grants by white PIs.

Successful research grant applications (non-fellowships) by non-UK PIs are estimated to apply for and receive 74.6% [69.5%,80.0%] of the value of successful grants applied for and received by UK PIs. Grant value tends to increase with age of PI, with successful grants for PIs in the <36 age range estimated to be 49.5% [45.6%,53.7%]

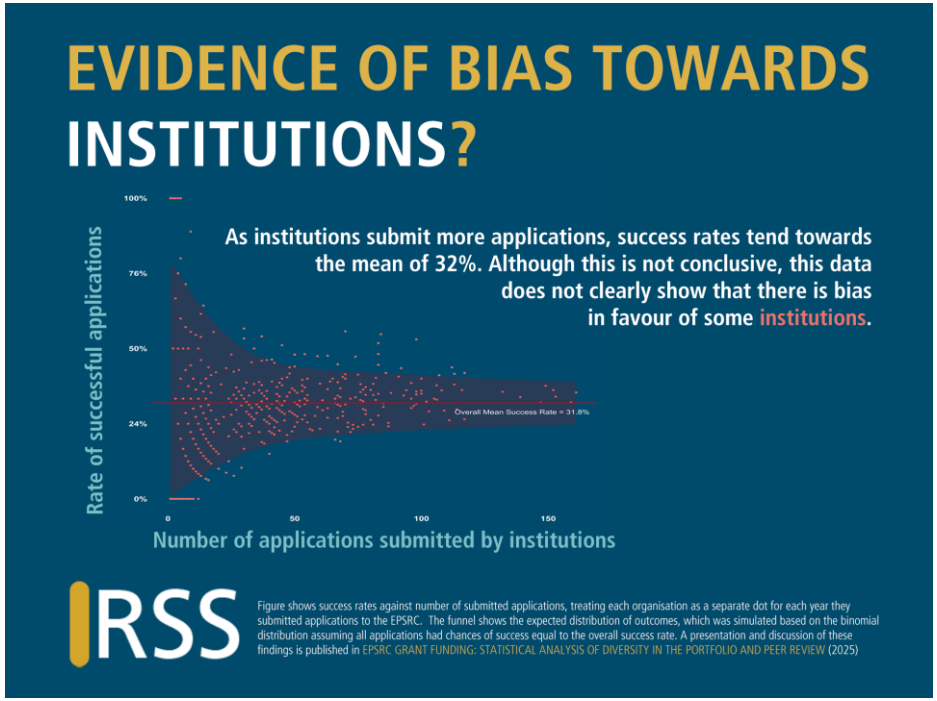
of the value of those in the 36-55 age range while those for PIs in the >55 age range estimated to be 129.5% [119.9%,139.8%] of the value of those in the 36-55 age range.

For fellowships, our analysis used linear models for applied for and awarded grant value on its original scale. The only protected characteristic where we see a significant difference in successful fellowship value is age. Fellowship value tends to increase with age of PI, with successful grants for PIs in the <36 age range estimated to be £466K [£391K,£541K] lower than those in the 36-55 age range while those for PIs in the >55 age range estimated to be £363K [£153K,£572K] higher than those in the 36-55 age range.



The dependence of grant success and the size of successful grant funding on protected characteristics was also investigated at the level of EPSRC research areas. This analysis demonstrated that while the majority of research areas shared similar characteristics in terms of dependence of funding success and average successful award value on protected characteristics, there were two (anonymised) research areas which exhibited more substantive differences. Further investigation of these differences by EPSRC has the potential to shed light on underlying drivers of difference.

In addition to protected characteristics, we also investigated the dependence of success rate on previous success rates, research institution and region. Our analysis did not identify strong regional effects. Furthermore, while institutions differed greatly in terms of the number of applications submitted, institutions submitting greater numbers of applications do not achieve a substantively higher success rate. And, as institutions submit more applications, success rates tend to concentrate around the mean – this is not conclusive, but the data does not clearly show that there is bias in favour of some institutions.



We did find that applicants who had experienced success as an EPSRC PI in the previous six years had a higher success rate than those who had success only as an

EPSRC Co-Investigator (CI) who respectively had a higher success rate than those with no EPSRC funding over that period.

RECOMMENDATIONS

Our analysis has provided some new insights around the diversity of funding in the mathematical, engineering and physical sciences. While we have been able to answer some key questions, we have also identified some areas that would benefit from further investigation. We make the following recommendations:

- We identified preliminary evidence of a link between prior success in an application and improved outcomes in subsequent applications. Further investigation would be useful to unpick the extent to which this is due to the strength of applications and the extent to which it is due to established researchers tending to disproportionately attract funding.
- Review panel composition seems to be associated with differential outcomes by sex – female applicants receiving more favourable outcomes when the review panel contains at least one woman. The EPSRC has operated a Mixed Gender Panel Policy since 2016, which requires all prioritisation and interview panels to be mixed gender.[10] We recommend that more detailed work is conducted to understand the effects of panel composition in terms of ethnicity and evaluate whether a similar policy for ethnicity is appropriate.
- Further work on perceptions of bias by different ethnic minority groups would be helpful. There is likely to be variance between the experiences of different ethnic minority groups and this would benefit from further examination. More qualitative approaches may help further illuminate the nuances in perceptions between ethnic groups.
- We observed differences between some research areas in terms of application success rates and the value of awarded grants. Further investigation of these differences has the potential to shed light on underlying drivers of difference. This is something that we were not able to do as research areas were anonymised.
- Due to small numbers of individuals in the data with known disabilities, we are unable to draw firm conclusions about this group. Future investigation could take a qualitative approach to understanding potential bias along the lines of disability.
- Unfortunately, it was not possible to directly control for application or researcher quality with the available data. This is a key limitation to the present report, and future research may wish to employ experimental methods that can help eliminate variation in quality.
- Further research could investigate interactions between protected characteristics and requested funding amount in success rates and panel rankings, as well as investigating how much of the difference in requested funding amount is associated with different average salaries for applicants from different groups.

ACKNOWLEDGEMENTS

Prepared by the Royal Statistical Society and The Alan Turing Institute

Project Researchers: Ged Hiscoke, Ruoyun Hui, Yesim Kakalic, Dakota Langhals, and Lucy Teece

Project Steering Group: Anjali Mazumder, Eugenie Hunsicker, Jat Singh, Jonathan Everett, Jonathan Forster, Sara Lombardo, and Tricia Dodd

Project Advisory Group: Katie Severn, Robin Mitra, and Veronica Bowman

EPSRC Facilitators: Christina Hurrell, Caitlin McAuley, Paula Bailey, and Alison Wall

[10] See: Evolving and upholding fairness in peer review – UKRI

INTRODUCTION

BACKGROUND

The goal of this project, and of previous diversity and inclusion projects by EPSRC, is to investigate what evidence there is of demographic disparities in its grant funding. Previous investigative work by the EPSRC team, based on data from the financial years 2014/15 to 2018/19, gave a mixed picture regarding sex[11].

- Award rates by number (ie the number of successfully funded applications over the total number of applications) for research grants led by male and female principal investigators (PI)[12] were very similar during this period.
- Award rates by number for fellowship applications by females have been higher than funding rates for fellowship applications by males by about 20 percentage points since 2015/16.
- For research grants, mean and median values are higher for male PIs than female PIs.
- Women are underrepresented both as applicants and as awardees for very large grants (that is, awards with values greater than £2.5 million), relative to their representation for grants below that threshold.

Regarding ethnicity, previous work by the EPSRC[13] flagged up consistent disparities:

- Ethnic minority researchers were underrepresented in the grant portfolio.
- Award rates for PIs, co-investigators and fellowship applicants from white ethnic groups were consistently higher than those applicants from ethnic minority groups.
- Median award value for ethnic minority researchers has been consistently lower than median award value for white researchers.
- Almost every institution interviewed raised that their ethnic minority researchers had a significant lack of trust in the peer review process. This was primarily due to the researchers' feeling they experienced bias, particularly at the fellowship reviewer stage.
- While there has been an increase in the proportion of ethnic minority researchers participating in peer review, this is still not representative of the engineering and physical sciences academic population.

The EPSRC contracted the Royal Statistical Society in late 2023 to use our expertise in analysis and knowledge of the academic community and peer review process to independently investigate in detail key aspects of the review process and disparities between demographic groups using a more recent dataset. EPSRC provided us with data regarding all applications from the financial year 2014/15 to 2022/23. The data was anonymised so that individual researchers could not be identified—anonymous IDs were provided for individuals, projects, institutions, regions, themes, and research areas. The anonymisation was important, but the process did introduce limitations in the analyses that could be conducted.

Building on the previous work, this project aims to:

- Survey the research population to understand their impressions about disparities in the funding process,

[11] EPSRC have performed specific investigations (quantitative and qualitative) exploring disparities in their portfolio and inequities. The associated reports are as follows:

- Gender disparity – Understanding our portfolio: a gender perspective: https://www.ukri.org/publications/epsrc-understanding-our-portfolio-a-gender-perspective/
- Gender disparity - community engagement and findings: https://www.ukri.org/publications/gender-diversity-in-our-portfolio-survey-findings/
- EPSRC detailed ethnicity analysis: https://www.ukri.org/publications/epsrc-detailed-ethnicity-analysis/
- Ethnicity inequity - community engagement and findings: https://www.ukri.org/publications/ethnicity-and-race-inequity-in-our-portfolio/
- Peer review participation: https://www.ukri.org/publications/epsrc-peer-review-diversity-data-2014-15-to-2019-20/

[12] Lead applicants to research grants are labelled "Principal Investigators" in the data, whereas lead applicants to fellowships are labelled "Fellows." In this report, we use "lead applicant," "principal investigator," or "PI" interchangeably to refer to lead applicants of either application type unless we specify otherwise.

[13] See, for instance, https://www.ukri.org/publications/epsrc-detailed-ethnicity-analysis/

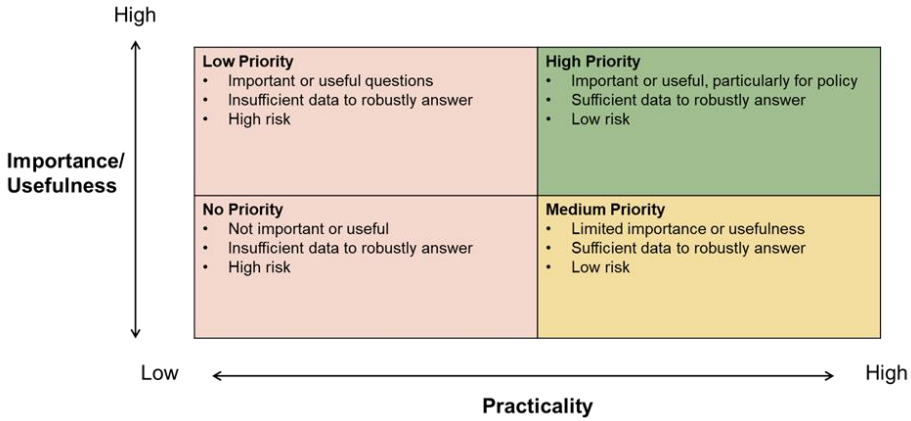
- Use statistical inference methods to understand which disparities or changes in disparities observed were likely to have arisen as the result of expected random variations in the process and which could indicate issues with the process,
- Investigate dimensions of difference that have not previously been studied in the data, particularly institution, region, research area group and theme, and funding mode,
- Carry out multivariate analyses of the data in order to understand how a range of factors in combination influence funding outcomes, and
- Undertake separate analyses of different parts of the funding process, including applications, reviewer comments, reviewer scores, panel rankings and final funding outcomes to investigate evidence of disparities at various stage of the pipeline.

There are also new questions that this work aims to investigate related to evidence in the broader research literature on funding and peer review, evidence from the survey, particular aspects of the EPSRC funding portfolio and initiatives that EPSRC has undertaken to address previously observed inequities:

- Is there evidence that reviewer demographics influence the content of reviews or scores proposed? Is there evidence that panel demographics influence panel rankings?
- Is there evidence of disparity in funding by institution or geographic region?
- Is there evidence of disparity in outcomes in certain themes or research area groups?
- Are there differences in demographics and outcomes by funding opportunity type (responsive or strategic)?
- Is there evidence of the Matthew Effect, whereby previous grant success increases subsequent grant success?
- Is there evidence of demographic differences in reapplication rates after an unsuccessful application?
- Are the answers to these questions different for grants over £2.5 million than for grants under this value?

This project is, as far as we know, the first time that a funding body has made its data available for this type of analysis by an external organisation. The data is very rich and there was a wide variety of types of investigation that we could have conducted. After compiling a list of questions, one might wish to answer, based on previous work, evidence in the literature, and experience of the particulars of the EPSRC funding process, questions were prioritised according to the scheme presented in Figure 1. Some of these questions were investigated quite fully, while in other areas the investigation had a more exploratory nature that helps point the way to future work. The full list of questions that were considered for analysis can be found in the GitHub repository.[14]

FIGURE 1: QUESTION PRIORITISATION DIAGRAM



Where previous work had already given a first idea about disparities, this project aimed to investigate more thoroughly their combined effect on outcomes using statistical models. Where the research questions were new, the project prioritised giving a first exploration of disparities associated with these differences.

The majority of the work undertaken in this project is exploratory in nature. However, one part of the study was pre-registered[15], which models the influence of PI or fellow protected characteristics on two outcomes of proposals: funding success and panel ranking.

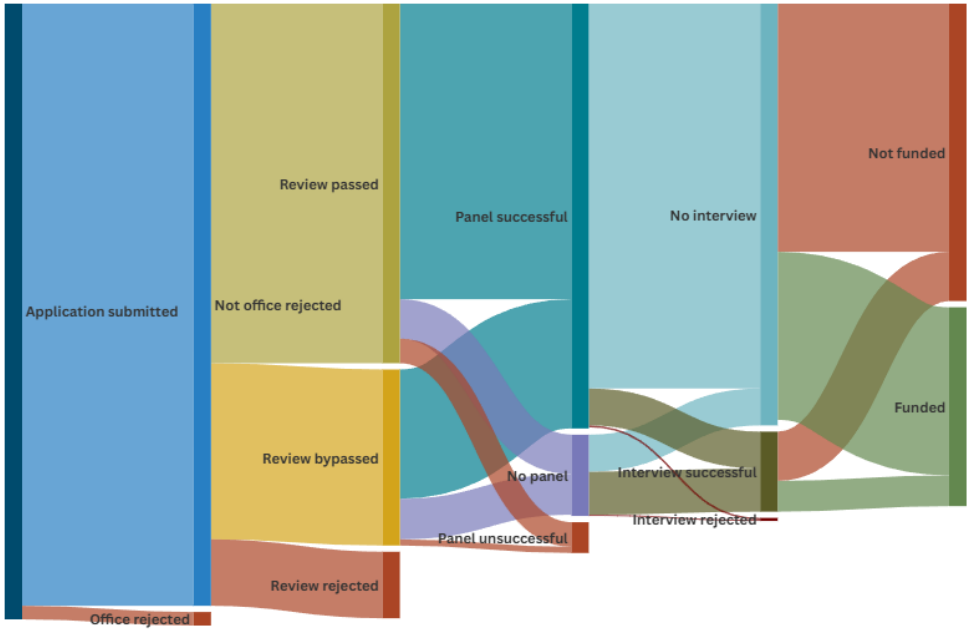
[14] This can be accessed at: https://github.com/alan-turing-institute/equity-in-grant-funding

[15] Hiscoke, Michael G. "Estimating the Effect Size of Applicant Characteristics on EPSRC Grant Application Outcomes." OSF, 30 May 2024. Web: https://doi.org/10.17605/OSF.IO/UMFZW.

In the interests of transparency, the analysis code, a complete set of outcome tables and diagrams, and a full description of the datasets used are available in a GitHub repository.

A summary of the steps in the review process and with illustrative success rates at each stage is summarised in Figure 2.

FIGURE 2: STAGES OF THE EPSRC REVIEW PROCESS



OVERVIEW OF EPSRC DATA

The dataset contains information from 20,420 unique applications (17,510 Research Grant applications and 2,920 Fellowship applications) that were submitted between financial years 2013/2014 and 2022/2023 and reviewed between 2014/2015 and 2022/2023. These were led by 11,250 unique individuals whose demographic characteristics are shown in Figures 3, 4, 5, 6 and 7.

Each application is characterised by the demographics of its PI. Although applications may have multiple contributing individuals (Co Investigators, Co-Is), we take this approach for simplicity when running statistical models. Unless otherwise stated, for charts of a more descriptive nature, we use the full population of applicants to give a more complete picture of the applicant pool. Applications are also characterised by the financial year in which they are submitted and each of the following:

- The region associated with the organisation of the PI.
- Research areas and a research theme related to the topic of the application. Research areas fall into one of ten Research Area Groups (RAGs) in the data provided to us. There are also 22 themes that act as another grouping for applications.
- The type of funding opportunity to which the application was made, of which there are two types: strategic and responsive. Strategic funding opportunities are "top-down" in the sense that they pertain to topics set by EPSRC. Responsive opportunities are "bottom-up," as they pertain to researcher-generated topics and often run on a rolling basis.

In this report, some of the outcomes of interest relate to the review process directly, such as whether applications from PIs with certain demographic characteristics are more likely to be "fundable" but not ultimately funded compared to others, or whether reviewer scores are commensurate with the sentiments they express in their reviews. Other outcomes of interest have to do with the final decision (ie, whether an application is funded or not), while others have to do with the outcomes of particular stages of the process, such as whether there is bias in the language used by reviewers.

Of all applications, 82.7% (14,190 research grant and 2,700 fellowship applications) could be linked to a total of 61,540 review reports from 17,860 reviewers. 79.8% of applications (13,910 research grant and 2,400 fellowship applications) were discussed during at least one interview or proposal panel meeting.

FIGURE 3

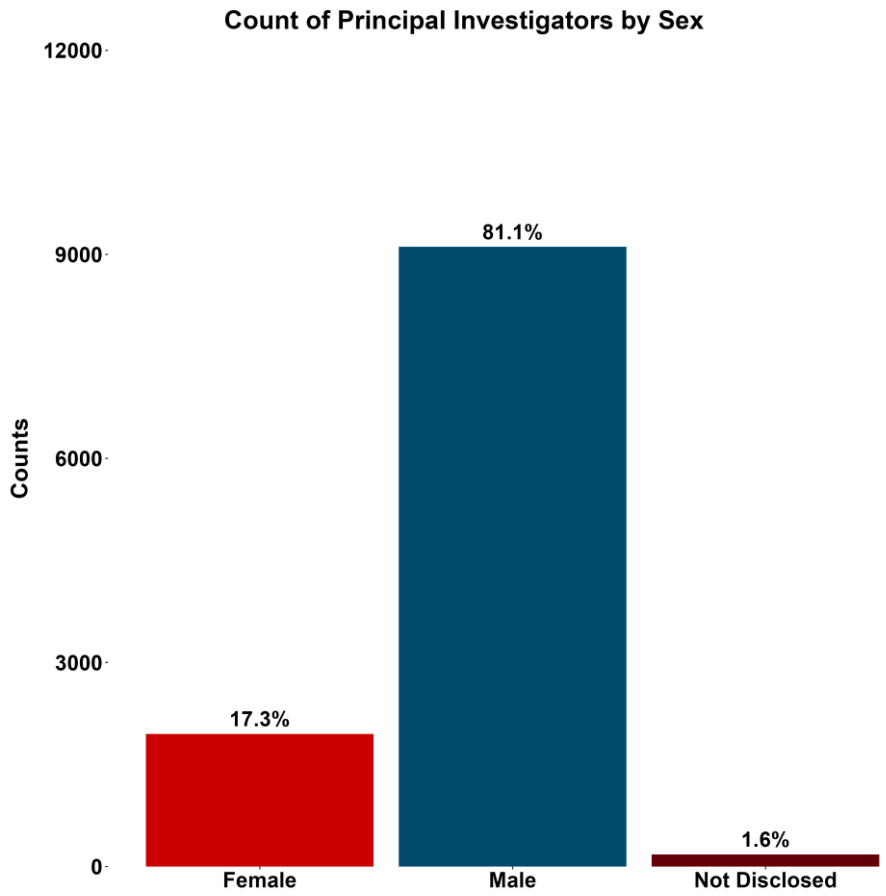


FIGURE 4

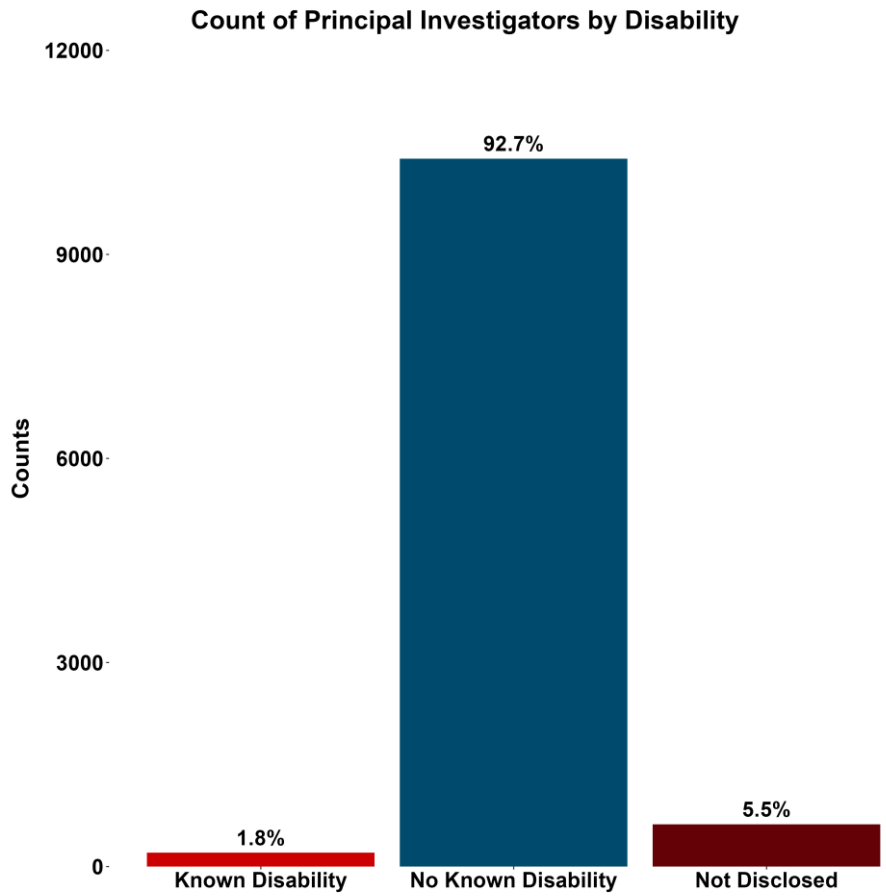


FIGURE 5

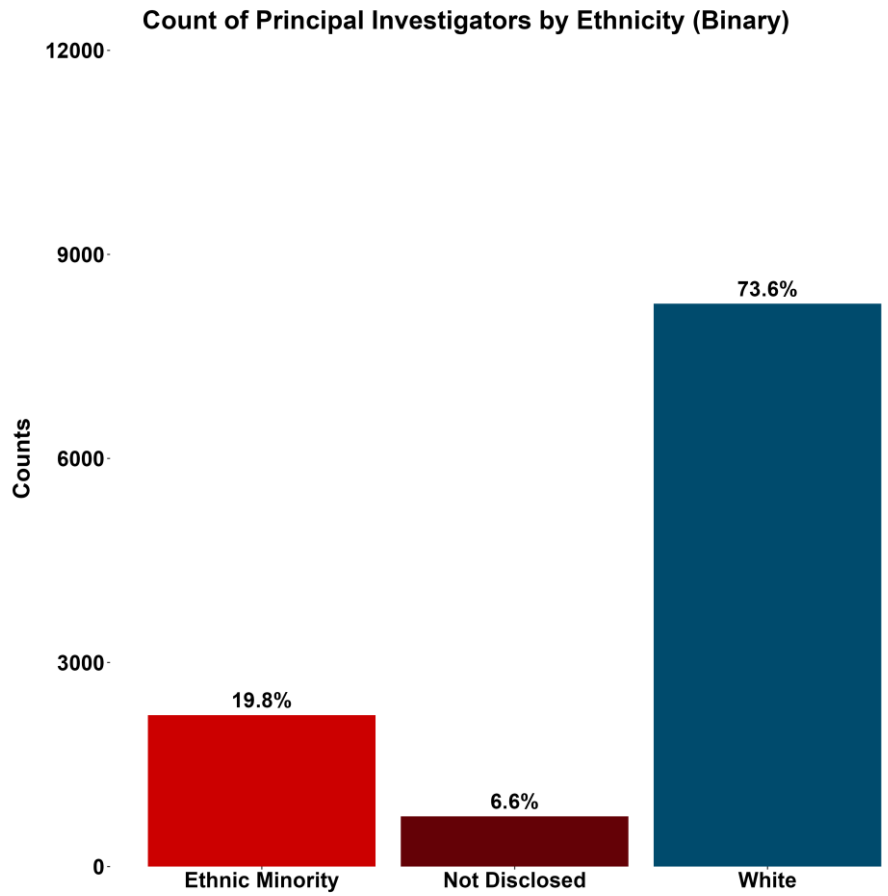


FIGURE 6

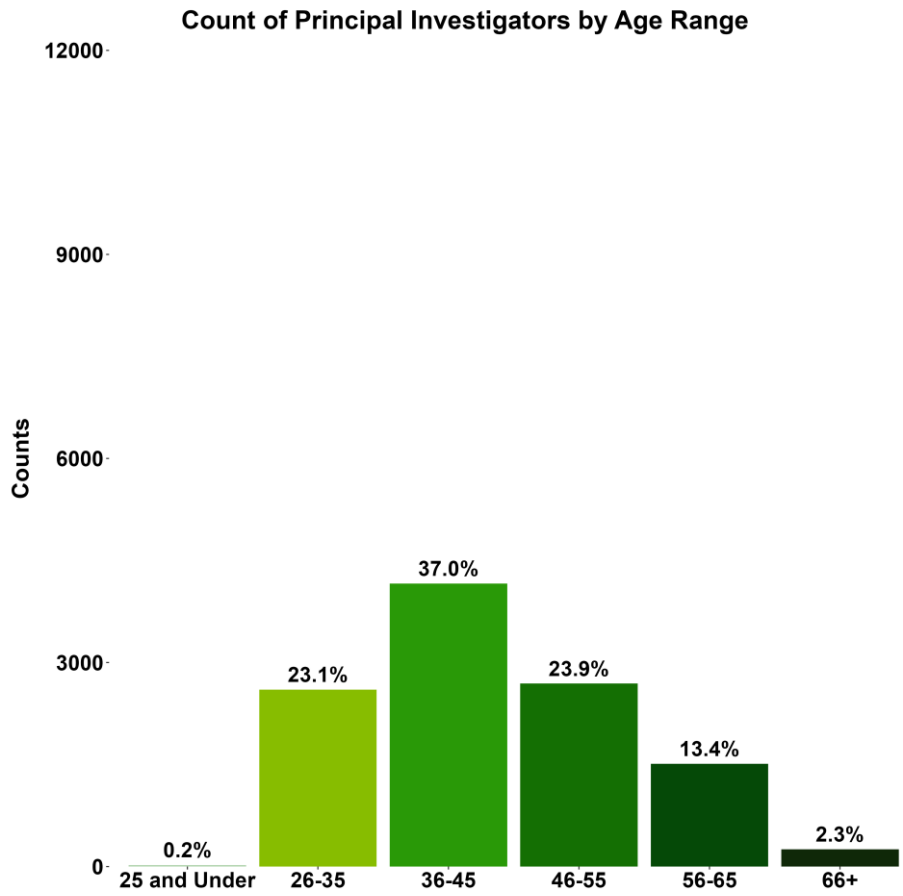
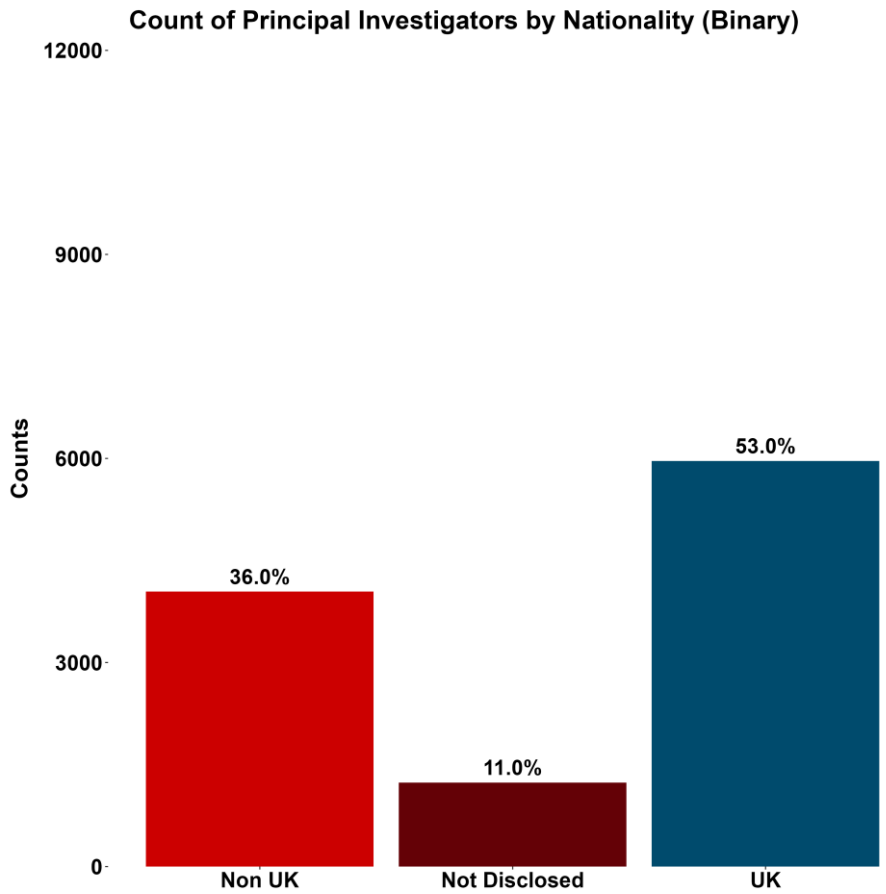


FIGURE 7



Throughout this report, we pay particular attention to the available protected characteristics in the data– namely sex, ethnicity, age, and disability – along with a fifth characteristic – nationality – that collectively will paint a picture of whether there are disparities for certain types of applicants.

One objective of this project is to better understand associations between intersectional characteristics and disparities in representation or outcomes. The advantage of considering intersectionality is that it may provide new insights that a one-dimensional approach cannot. The key limitation of an intersectional approach, particularly from a statistical perspective, is that it can dramatically increase the complexity of constructing and interpreting models, sometimes to the point that models fail to produce any results at all.

In the case of this project, the dataset is complex and intersecting multiple applicant characteristics at the same time sometimes resulted in overly small or even empty demographic groupings. Thus, it was not always possible to create models that included all characteristics and their intersections, particularly when focussing on smaller subsets of the available grants (which will also typically mean looking at a smaller subset of applicants). Indeed, for this reason, we prioritised using higher-level groupings for some variables like ethnicity (ie white vs ethnic minority), although we recognise that experiences will differ between members of more granular categories.

To handle both the complexity of an intersectional lens and the complexity of a large set of research questions each with many possible approaches, we generally prioritised examining three sets of interactions: sex with ethnicity, sex with age, and ethnicity with age in our exploratory analyses.

Additionally, we use the less detailed versions for these variables unless otherwise stated. That is, ethnicity is collapsed to white and ethnic minority, while age is collapsed to three age brackets: 35 and under, 36 to 55, and 56 and over – these brackets were chosen to capture relatively young and old groups while also accounting for the low number of applicants at the extremes of the age distribution.[16] More detail can be found in the relevant methodological sections corresponding to the different analyses.

[16] As individuals who make multiple applications may fall into different age brackets for each application, we use the highest age individuals reached in the data.

MISSING AND NON-DISCLOSED DATA

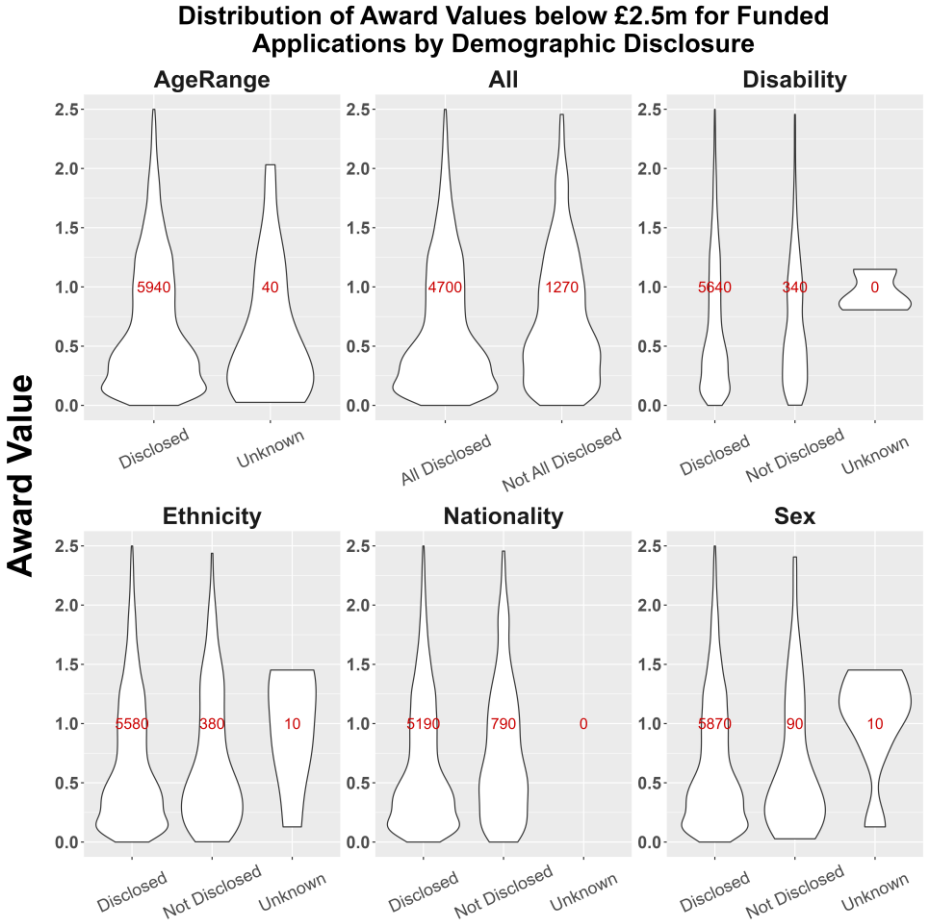
The central focus throughout our research is on demographic characteristics and how they may be associated with outcomes and decisions in the funding process. However, demographic details are not complete for all records in the data – this can be due to a number of reasons, including mistakes or gaps in the data collection process or applicants simply choosing not to disclose certain details when making an application.

There are at least two reasons to be interested in missing and non-disclosed data. Firstly, perceptions or expectations of bias in the grant funding process may lead some groups to withhold their demographic information more often than others. Identifying these behavioural differences may be informative in its own right, although it is not the central focus of this report.

Secondly, it is possible that outcomes may differ systematically between the groups that disclose their information and those that do not. Ordinarily, filtering out records that contain missing or non-disclosed information can make results more easily interpretable, particularly when trying to understand how the intersections of demographic variables are associated with outcomes in more complex statistical models. However, if missingness and non-disclosure are systematically correlated with the outcomes, then filtering for complete records will bias the results.

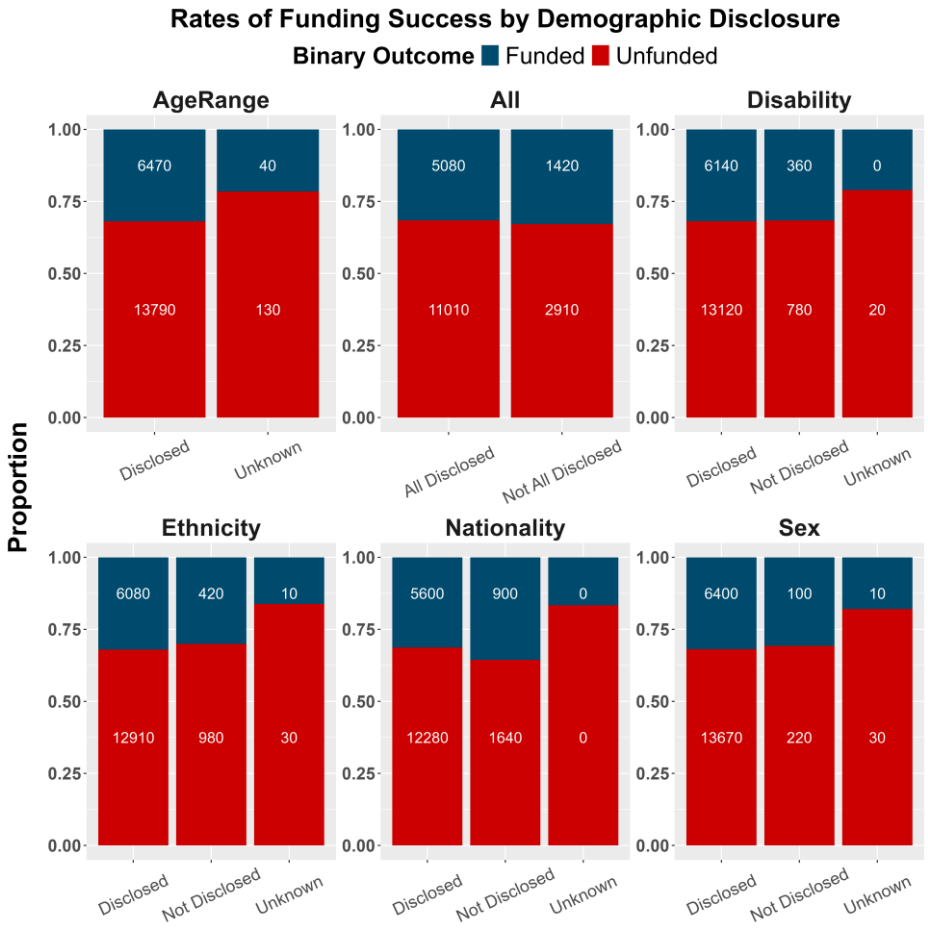
We conducted descriptive analyses comparing the rates of non-disclosure and missingness between the demographic groups, as well as comparing the distributions of outcomes by whether certain demographic details were disclosed. Figures 8 and 9 in this report provide a sampling of the findings using grant applications as the unit of analysis. Figure 8 shows the distributions of award values for funded applications below £2.5 million[17] grouped by whether demographic details for the PI were provided or not. Figure 9 compares the rates at which applications with provided details for the PI were funded or not funded. Both outcomes look largely similar for disclosing and not disclosing groups. While applications with "Unknown" demographic details show different outcomes, they also account for a very small portion of the total sample.

FIGURE 8



[17] Award values are right skewed. We follow previous work from the EPSRC in applying this cutoff here, as it makes visualisation easier. However, we also compared award values above this threshold and found that the conclusions for missingness and non-disclosure were unaffected. Those charts can be found in the GitHub repository, which can be accessed at: https://github.com/alan-turing-institute/equity-in-grant-funding

FIGURE 9



The demographic category that shows the greatest variation between the disclosing and non-disclosing groups is nationality, where the non-disclosing group appears somewhat more likely to be funded and somewhat more highly funded than the group that discloses this information. Except where stated otherwise, we choose to include records from individuals and applications when nationality is not disclosed, exclude records when other demographics are not disclosed, and exclude records where demographic details are marked as "Unknown" or missing. For details of the full analysis of missing and non-disclosed data, please refer to the GitHub repository.[18]

HOW REPRESENTATIVE IS THE EPSRC APPLICANT POPULATION OF THE BROADER RESEARCH COMMUNITY AND WIDER SOCIETY?

Here, we are interested in understanding the diversity of people applying for EPSRC funding, both fellowships and research grants combined, and then examining how this compares to the broader research community and the broader working population in the UK.

We proxy the research community at large with data from the Higher Education Statistics Agency (HESA). The HESA data contains demographic information about academic research staff in Engineering and Physical Sciences (EPS) fields between academic years 2014/15 to 2022/23. In principle, if the EPSRC applicant pool is perfectly representative of the broader research community, its demographic composition will closely track that of the HESA data.

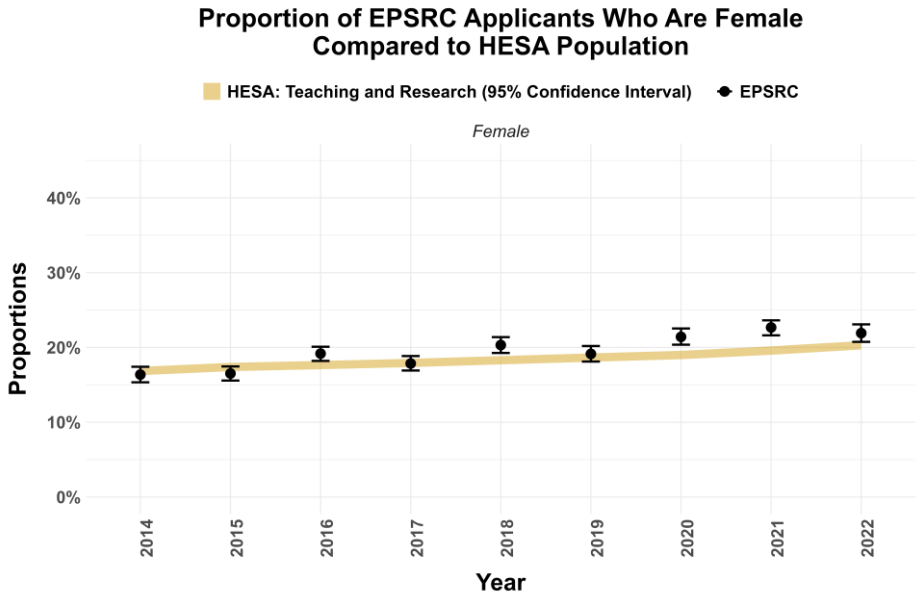
Academic staff in the HESA data are split into categories that correspond to the kind of work that they do: Teaching Only, Teaching and Research and Research Only. For the purposes of comparison, we considered the category of Teaching and Research staff to be the most likely to apply for EPSRC funding so used that as our comparator. Our view was that Research Only staff were less likely to be required to apply for funding than academics in teaching and research roles. This may not be the case, and we have no means to test this assumption. From this point on, where we refer to the HESA population, we are referring to the Teaching and Research portion of the HESA population working in EPS, unless otherwise stated.

The characteristics of the working population in the UK are taken from the publicly available UK Annual Population Survey between calendar years 2014 and 2022.

[18] Further details include analysis of the rates of non-disclosure and missingness by demographic group, outcomes for award values above £2.5 million and panel ranking. Consideration of these other analyses does not affect our decision regarding missingness.

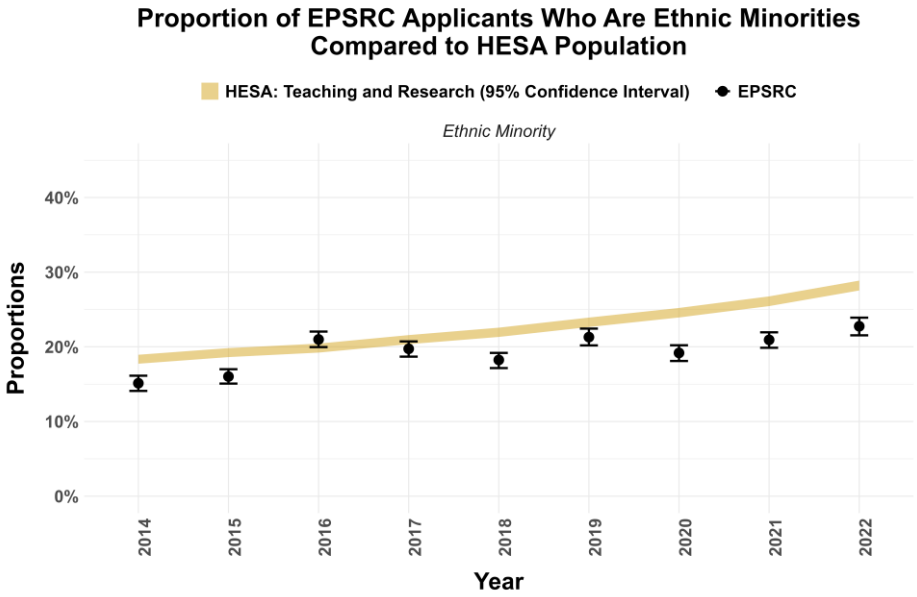
Figures 10, 11, and 12 show the trends over time for sex, ethnicity, and age compared to the HESA population.[19] [20] [21] In terms of sex, the EPSRC applicant population in the more recent years has had roughly the same proportion of females as the overall HESA population (shown by the yellow line).

FIGURE 10 [22]



For ethnicity, the EPSRC applicant population has a lower proportion of ethnic minority investigators than the wider research community, particularly in recent years, although the gap is small. However, both females and ethnic minorities have seen gradual increases in their relative shares of the EPSRC applicant and HESA populations since 2014/2015.

FIGURE 11 [23]



For age, the EPSRC applicant population diverges from the HESA population more consistently: all age brackets are significantly different from their comparators in the HESA population. Applicants aged 35 and younger are slightly over-represented in EPSRC data, while applicants from the aged 36-55 and aged 56 and older are slightly over and under-represented, respectively.

[19] These charts all calculate proportions based only on the HESA and EPSRC applicants for whom demographic details are known.

[20] For all comparisons to HESA data, it is important to note that we do not have a way to determine which researchers have or have not applied for EPSRC funding.

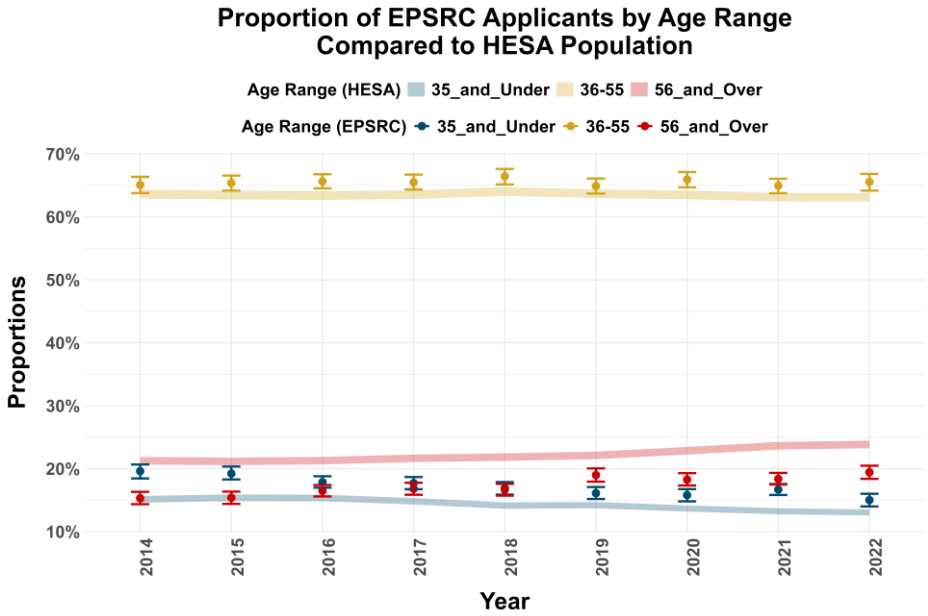
[21] HESA data are calculated for academic years, while EPSRC data are calculated for financial years. In both cases, we report the academic or financial year beginning in the designated calendar year on the comparison charts.

[22] Proportions are calculated for males and females only and exclude "Unknown" and "Not Disclosed" categories. Thus, the male proportions are simply 1 minus the female proportion and have been left off the graph for the sake of legibility.

[23] Proportions are calculated for white and ethnic minority applicants only and exclude "Unknown" and "Not Disclosed" categories. Thus, the white proportions are simply 1 minus the ethnic minority proportion and have been left off the graph for the sake of legibility.

Similarly, the EPSRC applicant pool is consistently but slightly under-representative of the HESA population in terms of its share of investigators with non-UK nationality.

FIGURE 12



For disability, the proportion of EPSRC applicants with known disabilities is consistently about 1% lower than the HESA EPS population. While the overall rates of known disability are very low in both datasets (from about 1.5% to 4%, depending on the year), the consistent gap relative to the overall size of the population with known disabilities may be indicative of strategic behaviour by grant applicants. This might arise if, for instance, applicants perceive that they may be disadvantaged by declaring a disability.

However, it is important to note that protected characteristics are not directly shared with reviewers and staff during the EPSRC peer review process. Still, some disabilities (and other protected characteristics) may be visible during panel interviews, and in any case, perceptions of bias can influence behaviour regardless of the process.

When comparing with the UK labour force data, the overall labour force is close to evenly split between male and female, while the EPSRC applicant population is not. The EPSRC applicant population appears to be more ethnically diverse than the overall workforce, likely because of the higher number of international researchers in academia. Full details with graphs are available in the GitHub repository.[24]

HOW DO THE EPSRC APPLICANT, REVIEWER, AWARDEE, AND PANELLIST POPULATIONS COMPARE OVER TIME?

This portion of the report seeks to answer: are the proportions of various demographic groups in the population of unique PIs for research applications and for fellowships applications comparable to the proportions of these groups in the awardee and reviewer populations each year? Additionally, we are interested in whether there have been changes over time towards greater or lesser parity between the populations.

We investigate this question by constructing 95% confidence intervals for each population and their demographics in each year.

For sex, panellists are consistently more female than any of the other groups. There appears to be a noticeable shift during the observable period around 2016, when the proportion of female panellists jumped significantly. The EPSRC operates a Mixed Gender Panel Policy[25] that aims for no single-gender panels and 30% representation of the underrepresented gender across all panels in a financial year. The policy was initiated in 2016 and appears to be achieving its objective.

There is also a gradual tendency for each group to become more female over time, but nothing as pronounced as the change for panellists. See Figure 13.

In terms of ethnicity, there is more variation between the groups; a larger relative share of applicants are ethnic minorities than awardees and panellists in most years. Reviewers in most of the recent years have been close to proportional to the applicant

[24] This can be accessed at: https://github.com/alan-turing-institute/equity-in-grant-funding

[25] For more information, see: Evolving and upholding fairness in peer review – UKRI

population. There is a gradual growth in the ethnic minority population for all four groups over time. See Figure 14.

FIGURE 13

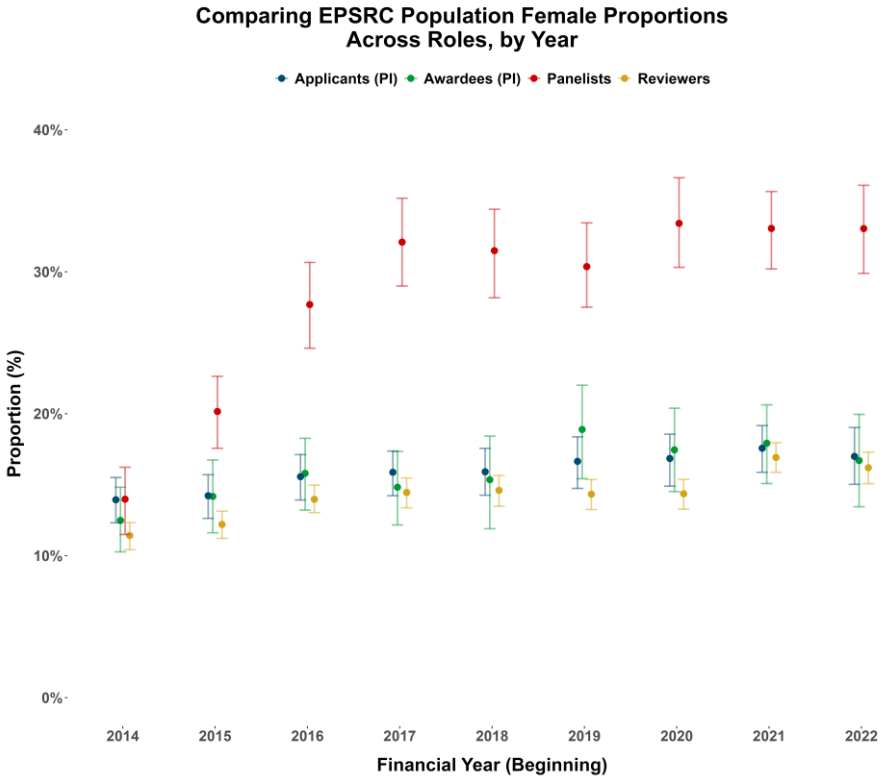
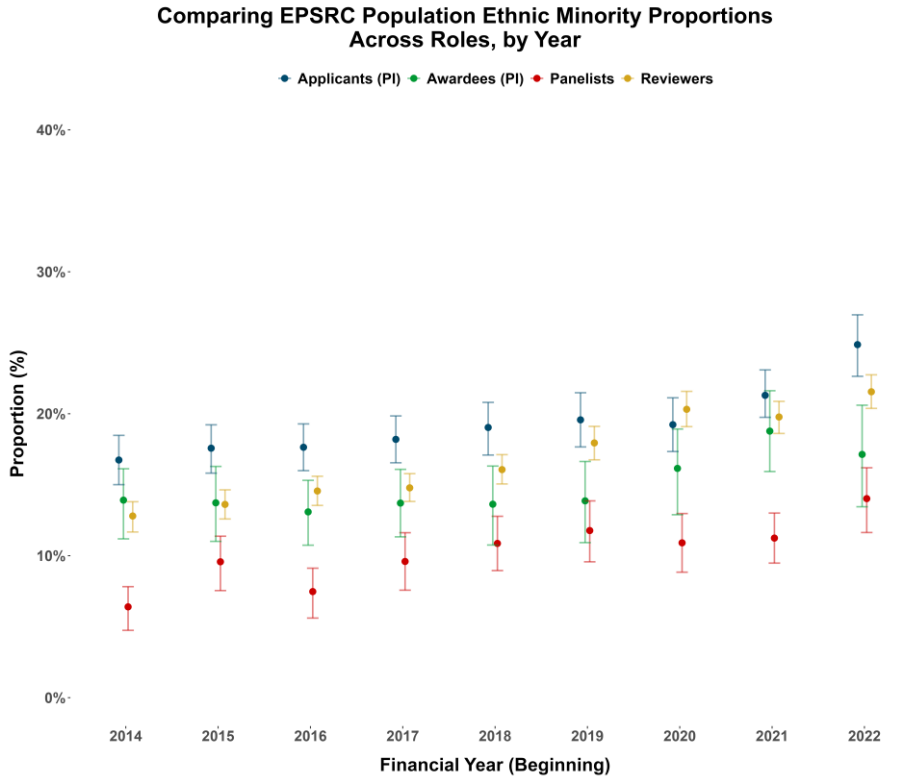


FIGURE 14



There is very little variation between the populations in terms of their relative share of individuals with known disabilities. However, there is differentiation in terms of nationality, where panellists consistently have non-UK nationality less often than the other groups, and in terms of age, where panellists and reviewers tend to be older than applicants and awardees. See Figures 15, 16, and 17.

FIGURE 15

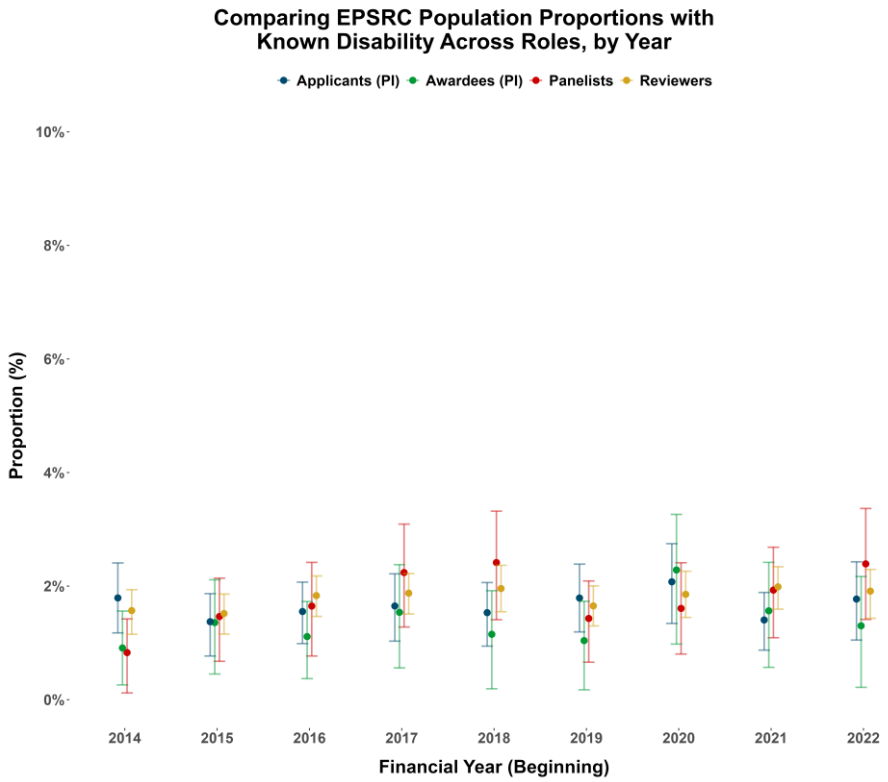


FIGURE 16

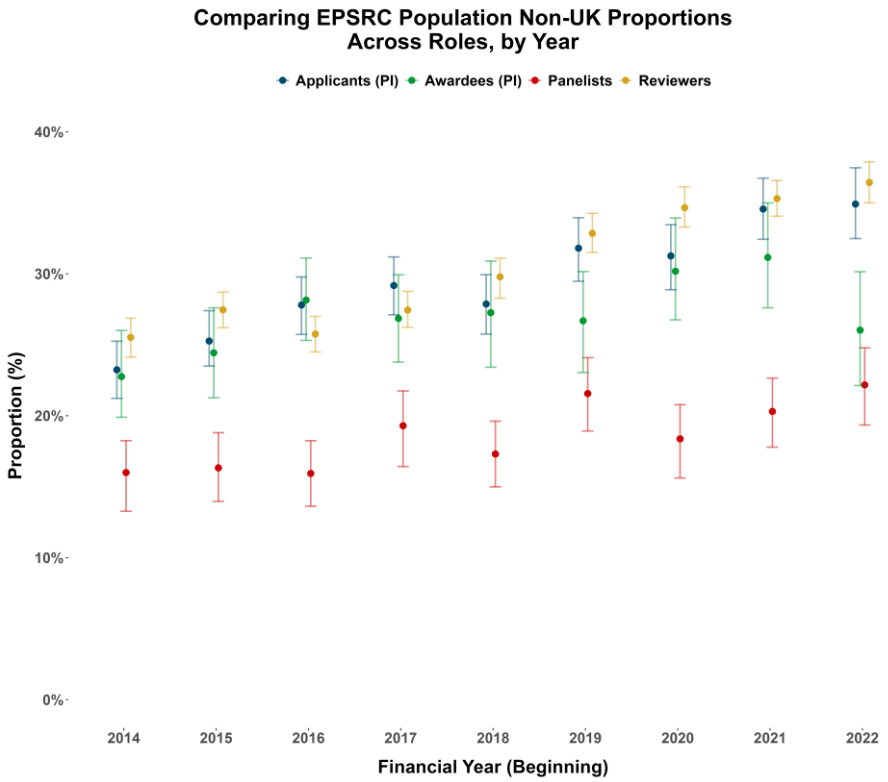
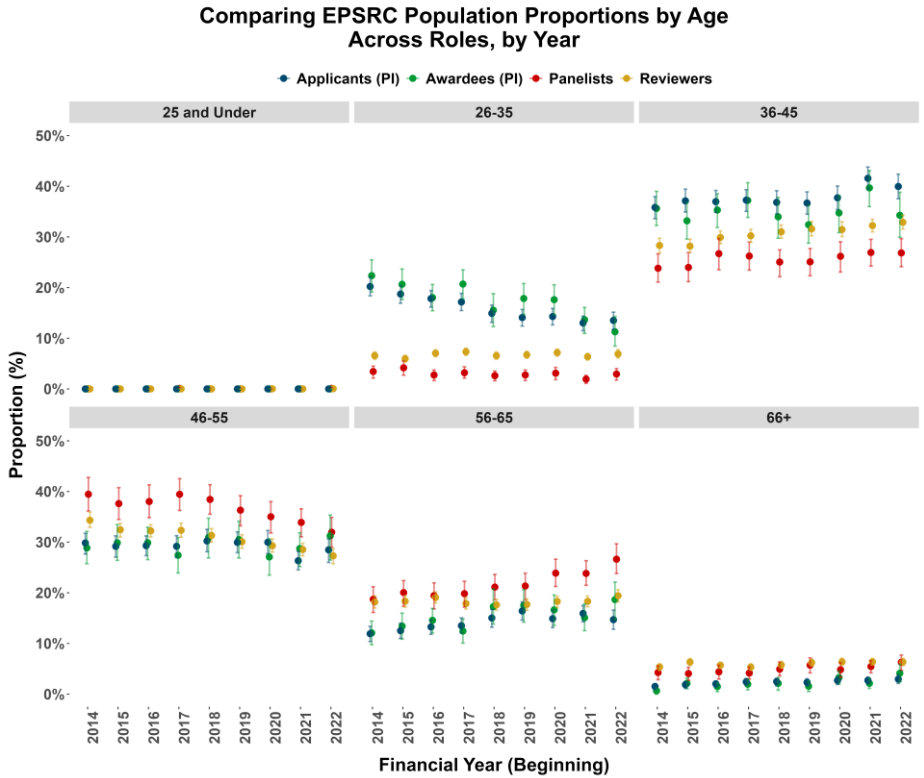


FIGURE 17


Comparing EPSRC Population Proportions by Age Across Roles, by Year
Applicants (PI)
Awardees (PI)
Panelists
Reviewers
25 and Under
26-35
36-45
46-55
56-65
66+
Proportion (%)
50%
40%
30%
20%
10%
0%
2014
2015
2016
2017
2018
2019
2020
2021
2022
Financial Year (Beginning)

IS THERE EVIDENCE OF AN ASSOCIATION BETWEEN PROTECTED CHARACTERISTICS AND DECISION OUTCOMES?

Fundamental to this project was understanding if there is evidence of disparity in outcomes of funding applications associated with demographic characteristics of applicants. As noted, previous work by EPSRC already investigated award rates by sex and by ethnicity, and found two differences of relevance:

- Award rates for fellowship applications by females are higher than for fellowship applications by males.
- Award rates for fellowships and standard grants led by a white investigator are higher than for fellowships and standard grants led by an ethnic minority investigator.

In this part of the current study, the goal was to understand the combined effects of sex, ethnicity, age and disability while controlling for other factors. The research plan for this investigation was pre-registered.[26]

As the plan was registered before the data was examined, it was necessary to put in place an updated approach to ensure that the most informative model would be completed that had sufficient data for the models to successfully run. Issues arise when relevant subgroups to the model are too small to obtain reliable estimates for model parameters.

Thus, the models for reporting were determined according to the following:

- The model was fit at the level of application rather than investigator, where the demographic variables are those associated with the investigator leading the proposal. It was anticipated that adding demographic information for Co-Investigators would add too much complexity to permit convergence of the model fitting process, and that Principal Investigator characteristics were likely to be the most influential on outcome.
- The outcome variables selected for the model were:
    - Funding success: Funded, Not Funded.
    - Panel ranking: Top Quartile, Not Top Quartile (top 25% of panel ranking in the final panel at which the application was judged). It is important to note that not all applications receive a panel ranking. This can happen if, for instance, the application gets rejected before reaching a panel. Rather than exclude such applications entirely, we simply considered such applications to be outside the top quartile.
- The demographic variables were combined into fewer categories to prevent small sub-populations. In each variable, the Unknown category included both "Not Disclosed" and missing. The base levels are listed first for each variable.
    - Sex: Male, Female, Unknown
    - Ethnicity: White, Ethnic Minority (combining "Asian", "Black", "Chinese", "Mixed" and "Other"), Unknown
    - Age at time of submission: 36-55, Under 35, Over 56, Unknown
    - Disability: Known disability, No known disability, Unknown
    - Nationality: UK, Non-UK[27]
- Control variables used in the models were:
    - Funding mode (Responsive, Strategic)
    - Application type (Fellowship, Research Grant)
    - Funding Year (2014/15 to 2022/23)

[26] Hiscoke, Michael G. "Estimating the Effect Size of Applicant Characteristics on EPSRC Grant Application Outcomes." OSF, 30 May 2024. Web: https://doi.org/10.17605/OSF.IO/UMFZW.

[27] When reporting on nationality, we follow the terminology given in the shared data, which uses "UK" and "non-UK" as the categories.

- Research Theme: these are set by the EPSRC as cross-cutting, potentially multi-disciplinary topics of research interest. Themes change over time. There are currently ten research themes, but there have been 22 over the past ten years. The (anonymised) research themes relevant within this data set are: digital economy, energy and decarbonisation, engineering, healthcare technologies, information and communication technologies (ICT), manufacturing the future, mathematical sciences, physical sciences, quantum technologies, and research infrastructure.
- Research Area Group: this is the academic discipline of the work. More than 100 research areas are grouped together into 11 research area groups, which are anonymised in the data provided by the EPSRC.
- Geographical Region: 11 anonymised regions

- Mixed-effect logistic regression models were used, with Region, Funding Year, Research Theme and Research Area Group treated as random effects and the remaining predictors treated as fixed effects.
- The interaction terms to consider were fixed effects for pairwise combinations of Sex, Ethnicity, Disability, Nationality, and Age. A forward method of model development was employed to build up the final models, with terms added if they produced significant improvements in model performance, following likelihood ratio tests.
- Random slopes were to be considered for Sex and Ethnicity with Funding Year to investigate evidence of changes in effects across the study period. However, due to convergence issues, the inclusion of these random slopes was abandoned.
- Standard confidence intervals for the main effect parameters of the demographic variables were calculated. These and the estimated coefficients were exponentiated so that they can be interpreted as odds ratios rather than effects along the log-odds scale.

Table 1 shows the observed counts and percentages of different demographic groups achieving funding success and a top-quartile panel ranking.

TABLE 1

Rates of Funding Success and Top Quartile Ranking by Demographic Group

Applications

Demographic Group	Funded (N)	Funded (%)	Top Quartile (N)	Top Quartile (%)
Male	5,320	31.6%	2,925	17.4%
Female	1,075	33.1%	610	18.7%
36 to 55	4,085	31.5%	2,200	17.0%
Over 55	1,030	34.0%	455	15.0%
Under 36	1,350	31.9%	915	21.6%
White	5,095	33.9%	2,760	18.4%
Ethnic Minority	980	24.7%	600	15.1%
UK	3,780	33.4%	2,010	17.8%
Non UK	1,815	27.7%	1,145	17.5%
No Known Disability	6,400	31.9%	3,530	17.6%
Known Disability	105	28.9%	60	16.5%

Descriptively, Table 1 shows that the rate of funding success for applications that are led by ethnic minority and non-UK PIs are considerably lower than the rates for applications led by PIs from other demographic groups. For receiving a top-quartile panel ranking, the rates are also comparatively quite low for applications led by ethnic minority PIs, but nationality appears less important. Rather, applicants aged 56 and

over have the lowest percentage of their applications receiving a top-quartile panel ranking.

FUNDING SUCCESS AS A FUNCTION OF SEX, AGE, ETHNICITY, AND DISABILITY, CONTROLLING FOR FUNDING MODE, APPLICATION TYPE, REGION, RESEARCH AREA GROUP, THEME, AND FUNDING YEAR

Odds ratios and 95% confidence intervals for the logistic regression model predicting binary funding success, which included only main effects and no interactions, are presented here in Table 2.

Note that when there is no difference between outcomes for a given level over the base level for the variable, the odds ratio will equal 1. When the odds of funding success are better for the given level than for the base level, the odds ratio will be greater than 1, and when they are worse, the odds ratio will less than 1. There is evidence of a statistically significant difference between outcomes for the levels when the 95% CI does not contain 1.

We see from Table 2 that the difference in odds of funding success between female and male PIs is marginally significant, with proposals having a female PI having slightly better odds of funding success than those with a male PI. The difference in odds between white and ethnic minority PIs is the largest of all observed between known demographic categories, with proposals with a white PI having better odds of funding success than proposals with ethnic minority PIs.

Additionally, applications for strategic funding opportunities and for research grants had higher odds of funding success than responsive opportunities and fellowships, respectively. There are also significant differences in odds of funding success on the basis of age, with grants led by older PIs having lower odds than grants with younger PIs. There is no significant difference on the basis of PI disability status.

TABLE 2 [28]

Mixed-Effects Model of Binary Funding Success Against Predictors

Odds Ratios and 95% Confidence Intervals for Fixed-Effect Predictors[1]

Predictor	Odds Ratio	CI Lower Limit	CI Upper Limit
(Intercept)	0.316*	0.150	0.665
Female	1.133*	1.042	1.232
Ethnic Minority	0.684*	0.629	0.744
Age > 55	0.974	0.891	1.065
Age < 36	1.332*	1.226	1.447
Known Disability	0.795	0.626	1.011
Non-UK Nationality	0.818*	0.761	0.879
Strategic Funding Opportunity	1.165*	1.059	1.282
Research Grant	2.024*	1.821	2.250

[1] * = *significant at 5% significance level*

To assist with interpretability, we used the models to predict the probability of funding success for a specific set of characteristics, namely each combination of sex and ethnicity, for a population whose distribution across all other variables matched their

[28] The models for this section were fit including applications led by individuals with "Unknown" or "Not Disclosed" information. Results shown here exclude these categories to aid legibility.

distribution in observed data. We then averaged those predictions and present the results in Table 2A.

TABLE 2A

Average Predicted Probabilities of Funding Success, Selected Demographics

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Sex and Ethnicity	Mean Prediction
Male, White	33.1%
Female, White	35.7%
Male, Ethnic Minority	25.7%
Female, Ethnic Minority	28.0%

We also ran the model with all pairwise interactions between sex, ethnicity, nationality, age, and disability in sequence. Only the interactions of ethnicity with nationality gave statistically significant improvements to the model. Including both of these interactions in the model simultaneously resulted in highly imprecise coefficient estimates due to high collinearity. However, we present predictions derived from these models for selected demographics in Tables 2B, 2C, and 2D. The predictions for combinations of sex and ethnicity are largely the same as in the model without interaction terms.

Applications led by white PIs with non-UK nationalities had lower predicted probabilities of funding success than white PIs with UK nationality. By contrast, there was virtually no difference for ethnic minority PIs based on nationality.

The interaction between ethnicity and age group indicates that both white and ethnic minority PIs in the under 36 age bracket have higher predicted rates of funding success than their older colleagues, but the rate for white PIs in this bracket is still higher than for ethnic minorities, and there appears to be a stronger effect of ethnicity as age increases. There is a modest decline in predicted success rate with age for white PIs but a much more substantial decline for ethnic minority PIs. This could indicate improvement in equity over time, as younger generations see convergence in their funding success rates, but we cannot rule out that there is genuine bias against ethnic minorities who are also older that could persist with time.

TABLE 2B

Average Predicted Probabilities of Funding Success with Significant Interactions

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Sex and Ethnicity	Mean Prediction
Male, White	33.1%
Female, White	35.8%
Male, Ethnic Minority	25.1%
Female, Ethnic Minority	27.4%

TABLE 2C

Average Predicted Probabilities of Funding Success with Significant Interactions

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Nationality and Ethnicity	Mean Prediction
UK, White	35.1%
Non UK, White	29.7%
UK, Ethnic Minority	25.9%
Non UK, Ethnic Minority	25.0%

TABLE 2D

Average Predicted Probabilities of Funding Success with Significant Interactions

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Age and Ethnicity	Mean Prediction
Under 36, White	37.4%
36 to 55, White	32.7%
Over 55, White	32.1%
Under 36, Ethnic Minority	34.2%
36 to 55, Ethnic Minority	24.4%
Over 55, Ethnic Minority	21.3%

When considering the model results for funding success, it is important to note a key limitation of the models used in this section, namely, that while they include a control for grant category, they do not include an interaction term between grant category and other predictors. This is an important caveat, because, as we show later, there are considerable differences in rates of funding success for different demographic groups by grant category. This is particularly true for sex, where applications for fellowships are much more likely to receive funding when they are led by females than males. Results in Tables 5, 6, 7 and 8 later in this report show the results when stratifying the data by grant category.

PANEL RANKING AS A FUNCTION OF SEX, AGE, ETHNICITY, AND DISABILITY, CONTROLLING FOR FUNDING MODE, APPLICATION TYPE, REGION, RESEARCH AREA GROUP, AND FUNDING YEAR

The other outcome we examined in this portion of the analysis was the binary outcome of receiving a top-quartile panel ranking from the final panel where an application was considered. Panel rankings do not directly translate to funding success or failure, but they are a core component of the review process.

We regressed top-quartile panel rankings against nearly the same set of predictors used for binary funding success, with the caveat that we had to exclude the control for application theme, since results could not successfully be computed when including it. Odds ratios and their confidence intervals for the model, with only main effects and no interactions, are in Table 3.

TABLE 3

Mixed-Effects Model of Top Quartile Ranking Against Predictors

Odds Ratios and 95% Confidence Intervals for Fixed-Effect Predictors[1]

Predictor	Odds Ratio	CI Lower Limit	CI Upper Limit
(Intercept)	0.099*	0.072	0.137
Female	1.088	0.985	1.201
Ethnic Minority	0.794*	0.719	0.876
Age > 55	0.834*	0.745	0.934
Age < 36	1.579*	1.438	1.733
Known Disability	0.933	0.701	1.242
Non-UK Nationality	0.965	0.887	1.049
Strategic Funding Opportunity	0.862*	0.784	0.947
Research Grant	2.092*	1.843	2.375

[1] * = *significant at 5% significance level*

As for the model for funding success, an odds ratio greater than 1 indicates better odds than the base level in each demographic variable, odds ratio less than 1 indicates worse odds, and confidence intervals that do not include 1 indicate a significant difference in odds under the conventional 95% confidence threshold.

There is not a significant difference in odds of appearing in the top 25% of the panel rankings based on sex or disability. As in the funding success model, applications led by ethnic minority PIs have worse odds of appearing in the top 25% of the rankings, and the odds of an application appearing in the top 25% of panel rankings decreases with each older age group.

For interpretability, a selection of average predicted probabilities of receiving a top-quartile panel ranking is given in Table 3a. We also ran the model after separately adding all pairwise interactions between the demographic variables. None of the interaction terms produced significant improvements in the model performance.

TABLE 3A

Predicted Probability of Top Quartile Panel Ranking, Selected Demographics

Average Predictions Based on Mixed-Effects Model for Top Quartile Ranking

Age and Ethnicity	Mean Prediction
Under 36, White	24.7%
36 to 55, White	17.3%
Over 55, White	14.9%
Under 36, Ethnic Minority	20.7%
36 to 55, Ethnic Minority	14.3%
Over 55, Ethnic Minority	12.2%

The preregistered analysis yielding the results in this and the previous subsection does not give a full picture of the process for two reasons. First, the preregistered models do not include all desired interaction terms due to sample size issues. This means that the analysis cannot detect differences in success rates for different demographic groups according to different application characteristics. Second, the preregistered models consider only the overall results for applications in terms of funding success and achieving a top quartile panel rank, so it does not investigate differential outcomes in individual parts of the process. In order to examine the questions that can't be answered by the registered study, additional unregistered analyses were undertaken.

In order to understand how different groups may fare in different types of funding application, we undertook exploratory analyses where different types are investigated separately. Beginning on page 32, we present the results from analyses of funding success and award value undertaken separately for fellowships and research grants. From page 45, we present the results from analyses based on whether a funding opportunity was classed as strategic or responsive.

In order to understand what differential outcomes may exist at different stages of the process, the next section presents findings from an exploratory analysis of panel rankings and reviewer scores and comments.

ANALYSIS OF REVIEW PROCESS OUTCOMES[29]

FURTHER ANALYSIS OF PANEL RANKING

Further analysis of the review process considered all panel rankings rather than the final ones, and also controlled for a different set of covariates: we explored whether panel ranking was influenced by features of the reviews prior to panel meetings (such as reviewer scores), applicant characteristics, the type of panel meeting (ie proposal or interview), and panel characteristics using linear regression models.[30] [31]

We found that, on average, male applicants and applicants from the UK are ranked about 7% lower (less favourably) and 5% higher (more favourably) in interview panels compared to female and non-UK applicants, respectively, but not in proposal panels. When accounting for linguistic features of reviewer comments rather than just controlling for reviewer scores, white applicants score about 5% higher than ethnic minority applicants in panel interviews and about 1% higher than ethnic minority applicants in proposal prioritisation panel meetings.[32]

There is also evidence that the panel composition matters. In particular, female applicants are ranked about 8% higher than male applicants in interviews when at least one panellist was female, compared to 7.9% lower than males when no female was present. Panel composition would be a rich area for future detailed research. It would also be useful to expand the analysis of applicant characteristics to consider the full composition of the applicant team, rather than just the characteristics of the lead applicant.

RELATIONSHIP BETWEEN REVIEWER SCORES AND COMMENTS

A substantial portion of this further analysis was centred around text-based analysis of reviewer comments. Machine learning and natural language processing approaches were used to classify comments according to their sentiment (positive, negative, or neutral). Other language features that have previously been used in the literature on scientific grant review were derived from how often certain types of words appear in the comments. This allowed comments to be categorised by how often they used words that corresponded to masculine or feminine gender stereotypes, among other features.

A central insight from this approach is that reviewer sentiment scores correlate better with the overall assessment score than other language features do. Using regression techniques to predict reviewer scores with sentiment scores and other linguistic features, the sentiment of reviewer comments is the most influential predictor—a change in reviewer sentiment from negative to neutral or a change from neutral to positive is associated with an increase of about one point out of seven: 0.947. Other linguistic features of reviewer comments had much smaller effects on score.

The upshot of this analysis is that, overall, the assessment scores appear to align with the sentiment in reviewers' comments. Longer reviews and reviews including more standout adjectives tend to accompany higher scores, although their effects are much weaker than the sentiment. This analysis did not find evidence that traits associated with the male gender stereotype are preferred over stereotypically female traits during scoring: the use of both masculine and feminine words on average are associated with slightly higher reviewer scores.

IMPLICIT BIASES IN REVIEWER SCORES

We considered whether there is association between demographic characteristics of the reviewers, applicants, or their interactions and the reviewer scores. To answer this question, we fit a series of regression models of reviewer scores against demographic

[29] To find more detail for these analyses than is provided in this report, we encourage interested readers to explore the write-up from the Alan Turing Institute, which is available in the GitHub repository at: https://github.com/alan-turing-institute/equity-in-grant-funding.
[30] Panel characteristics are examined using the demographics of the panel chair alongside the proportion of the whole panel that was female, ethnic minority, or not from the UK.
[31] Full results from the regression models can be found in Tables 7, 8, 9, and 10 in the ATI report on GitHub.
[32] Other results were found to be statistically significant, but their effect sizes are small. Note that in the absence of adjustments for multiple hypothesis testing, spurious results in terms of statistical significance are more likely.

characteristics of reviewers and applicants, alongside language features of the comments and other predictors.

Table 4 shows one of the regression models that were used, which does not include linguistic features or an intersectional component. Characteristics of reviewers giving on average higher assessment scores include: white, male, based outside of the UK, nominated by the applicant, and not working in an interdisciplinary field, relative to females, ethnic minorities, non-UK applicants, not nominated by the applicant, and working in an interdisciplinary field, respectively. Characteristics of applicants receiving higher scores include: white, male, and under 35 years old, relative to ethnic minorities, females, and older age groups.

TABLE 4: SAMPLE REGRESSION RESULTS FROM ANALYSIS OF REVIEWER SCORE AGAINST APPLICANT AND REVIEWER CHARACTERISTICS

Characteristic	Regression Coefficient[1]	95% CI[2]
(Intercept)	5.04*	4.98, 5.10
ApplicantSex		
Female	—	—
Male	0.071*	0.043, 0.100
ReviewerSex		
Female	—	—
Male	0.070*	0.040, 0.101
ApplicantEthnicityBinary		
Ethnic minority (excluding white minority)	—	—
White	0.231*	0.205, 0.255
ReviewerEthnicityBinary		
Ethnic minority (excluding white minority)	—	—
White	0.229*	0.202, 0.255
ReviewerSource		
Applicant	—	—
Other	-0.723*	-0.748, -0.697
Country		
Non UK	—	—
UK	-0.325*	-0.363, -0.287
MultiDisc		
FALSE	—	—
TRUE	-0.155*	-0.239, -0.072
ApplicantAge		
35 and under	—	—
36-55	-0.027*	-0.053, -0.001
56+	-0.017	-0.053, 0.019
R^2	0.082	
Log-likelihood	-83,713	
No. Obs.	51,790	

[1] * = Significant at 5% significance level
[2] CI = Confidence Interval

The various regression results suggest that part of the variation in reviewer score was related to the ethnicity and sex of reviewers and applicants. Ethnicity seems to have more substantial influence on reviewers' scores than sex, with some evidence that reviewers scored applicants higher from the same ethnic background as their own. White reviewers gave white applicants scores that were 0.239 points higher than other applicants, and Chinese reviewers gave scores that were 0.567 and 0.560 points

higher to applicants with Chinese and "other" ethnicities, respectively.[33] Other combinations of reviewer and applicant ethnicities did not show significant results, although interpretation should be made with caution as many of these combinations will have very low sample sizes. However, whether the reviewer was nominated by the applicant and whether they were based in the UK have even stronger impact than the combination of reviewer and applicant ethnicities.

IMPLICIT BIASES IN REVIEWER COMMENTS

In addition to potential disparities in the scores given by reviewers based on demographic characteristics, we also examined whether the linguistic features of reviewer comments are associated with demographics. To investigate this, we considered eight different language features[34] and fit regression models with each one as the outcome against the demographic characteristics of applicants and their reviewers while controlling for reviewer scores.

The findings from this analysis suggests that there are modest associations between demographic characteristics and the linguistic features of comments. For instance, words related to female stereotypes are more common both when the reviewer is female and when the applicant is female. Words related to masculine stereotypes are more common when the reviewer is male, but not when the applicant is male.

Reviewers from ethnic minority backgrounds tended to use more words related to ability, achievement, agentic traits, and research than white reviewers. Additionally, white reviewers tended to give comments that had slightly more negative sentiments, especially to ethnic minority applicants.[35]

[33] See Table 4 in the Alan Turing Report available on the GitHub repository: https://github.com/alan-turing-institute/equity-in-grant-funding

[34] The features were: sentiment, ability, achievement, agentic, research, standout adjective, feminine, and masculine.

[35] Table 6 in the ATI report on GitHub (not shown here) shows the predictors found to be statistically significant when looking at regression models with various language features as the outcome of interest.

EXPLORATORY ANALYSES

METHODOLOGICAL BASICS, APPROACH, & LIMITATIONS

Throughout the exploratory portion of the analysis for this report, we have attempted to follow a consistent approach when answering our research questions. When fitting regression models, we first specify a model without interaction terms between the demographic characteristics of interest, then compare it to models that include interaction terms between sex and age, sex and ethnicity, and ethnicity and age using likelihood ratio tests and comparison of the AIC and BIC[36]. We prefer the simpler models unless improvements are substantial with interaction terms.

Regression models have been followed by diagnostic residual plots that help determine whether there are severe violations of model assumptions. In the presence of moderate or severe heteroskedasticity for linear models, we use sandwich approaches to compute robust standard errors and construct 95% confidence intervals around the estimates.

Our approach to variable selection has been driven primarily by an interest in testing the full set of demographic characteristics (sex, age, ethnicity, nationality, and disability) in each of our research questions, and as such have included each in the models on the basis that they carry their own independent theoretical interest. However, in the interest of parsimony and interpretability, we restrict our intersectional analysis wherever it is relevant to all two-way interactions between sex, age, and ethnicity.

When taking a more descriptive view of the EPSRC populations, we use bootstrapping methods to compute confidence intervals around our statistics of interest and have incorporated these into the relevant visualisations. Bootstrapping is a method by which the available data are resampled with replacement a large number of times, and the statistic of interest is recalculated with each iteration. This process produces a distribution of the statistic from which we can treat the middle 95% as the confidence interval for the statistic. This can be computationally expensive, and we have used 1,000 iterations as a standard to compute the bootstrapped confidence intervals.

There are some limitations that span across most of the exploratory analyses in this report that are worth noting. The first is that the nature of the data and the observational nature of our analyses generally limit us to discussion of associations, correlations, and relationships, rather than causal explanations. This means that, while we can identify disparities between groups, we are not able to determine the precise causes of those disparities. Given the larger literature and discourse around inequity throughout society, not least in STEM fields, it would be natural to interpret disparities as evidence of discriminatory bias in one way or another. Such explanations are often plausible, but we cannot directly substantiate those interpretations above others.

Additionally, we do not have a way of determining individuals' career stage precisely in the data. The best proxy we have are relatively crude age bands, but these are imperfect and will not account for the many ways in which career paths may vary.

Another limitation is that, in certain portions of the analysis, such as some instances of intersectionality or very high value grants, our sample sizes rapidly become too small to sufficiently power our models. As is often the case, larger sample sizes in future research may help alleviate this issue to some extent.

We also have not undertaken adjustments of our confidence intervals to account for multiple hypothesis testing throughout this study. Without such adjustments, confidence intervals can be underestimated, since the chances of finding a statistically significant finding purely by chance increase with the number of tests that are run. We have tried to take a relatively practical approach to inference, one that places less emphasis on whether a result is marginally statistically significant or insignificant. Rather, we try to pay more attention to whether the estimates are clearly different from 0 and also of practical significance in terms of their size.

Importantly, readers should note a key caveat to interpretation of our findings with respect to award values for successful applications throughout this report. The EPSRC

[36] Akaike Information Criterion and Bayes Information Criterion. These are commonly used metrics for comparing statistical models according to how well they fit the data.

does not generally cut award values, and except in very rare circumstances will award successful applications with the amounts that were requested in their original applications. This means that discrepancies picked up by our models in terms of award value for successful applications can be driven in part by differences in the amounts that different demographic groups request on average.

To the extent possible, we have controlled for factors that may be associated with different requested amounts, such as research areas, which can have structurally different cost outlays for research projects based on how much equipment they require, for example. Still, there are many things that we cannot control for with the data. Practically, when modelling application and award values, we must recognise that observed disparities between groups can have at least two explanations that are not mutually exclusive:

1. Applicants from some groups could unknowingly or even strategically request lower award values relative to other groups.
2. Reviewers and decision panels could unconsciously view applications from certain groups less favourably, and this may be reflected in the award values of applications that they decide to fund.

Previous work by the EPSRC found that females, for instance, generally apply for lower award values than males. See their report titled "Understanding our portfolio: a gender perspective."

RESULTS

ASSOCIATIONS BETWEEN FUNDING SUCCESS AND DEMOGRAPHICS, BY GRANT CATEGORY

The results from our pre-registered analysis indicated modest overall associations between sex and the probability of funding success. However, those models did not account for possible interactions between sex and grant category, and descriptive exploration of that interaction indicate that it is likely to be an important one.[37] Therefore, as part of a follow-up exploratory analysis, we examined whether there were associations between funding success and demographics when splitting the sample into two groups: fellowship and research grant applications.

Besides stratifying by grant category, the models used for this supplementary exploration are the same as what was used for the pre-registered portion, except that we apply the approach outlined in the previous section to the data, particularly with regards to unknown and non-disclosed data, and inclusion of interactions terms.

Reiterating the model specification here, we use a mixed-effects model predicting the binary outcome of funding success (funded vs not funded) against sex, ethnicity, age range, disability, nationality, and funding opportunity type (responsive or strategic), controlling for region, research area group, theme, and year as random intercepts.

The results from these models split by research grants and fellowships are presented in Tables 5 and 6. In the case of research grant applications, likelihood ratio tests indicated significant improvement in model performance when including pairwise interactions between sex, age, and ethnicity, while this was not the case for fellowship applications. These differences are reflected in the outputs accordingly.

The outputs from our models have been converted to odds ratios. Odds ratios larger than 1 indicate increased odds of an application being successfully funded when led by a PI with the associated characteristic. Confidence intervals that include 1 indicate a result that is not significant at the conventional 5% significance level. Tables 7 and 8 present the average predicted probabilities of funding success for selected demographics for each grant category in the interests of easier interpretability.[38]

The results indicate some interesting differences between the outcomes by grant category. For research grants, there is no significant association between funding success and having a female lead applicant. The results for fellowships, by contrast,

[37] Further details of that exploration can be found in Figure 18 later in this report.
[38] As elsewhere when presenting these average predictions, we first fit the model to the data, then use it to make predictions for all applications in the dataset, then average those predictions for the specified demographics. In this case, the dataset used for prediction was filtered to include only the relevant grant category.

indicate that applications led by females are significantly more likely to receive funding than applications led by males.

TABLE 5

Mixed-Effects Model of Binary Funding Success Against Predictors, Research Grants

Odds Ratios and 95% Confidence Intervals for Fixed-Effect Predictors[1]

Predictor	Odds Ratio	CI Lower Limit	CI Upper Limit
(Intercept)	0.650	0.324	1.302
Female	1.026	0.916	1.148
Ethnic Minority	0.637*	0.571	0.711
Age < 36	1.390*	1.243	1.554
Age > 55	1.027	0.932	1.131
Non-UK Nationality	0.865*	0.802	0.934
Undisclosed Nationality	1.087	0.984	1.200
Known Disability	0.789	0.621	1.003
Strategic Funding Opportunity	1.345*	1.220	1.483
Ethnic Minority Female	0.986	0.783	1.241
Female 35 or Younger	1.055	0.842	1.322
Female 56 or Older	1.280	0.983	1.666
Ethnic Minority 35 or Younger	1.257*	1.018	1.553
Ethnic Minority 56 or Older	0.826	0.637	1.072

[1] * = significant at 5% significance level

TABLE 6

Mixed-Effects Model of Binary Funding Success Against Predictors, Fellowships

Odds Ratios and 95% Confidence Intervals for Fixed-Effect Predictors[1]

Predictor	Odds Ratio	CI Lower Limit	CI Upper Limit
(Intercept)	0.360*	0.163	0.793
Female	1.803*	1.426	2.279
Ethnic Minority	0.671*	0.519	0.868
Age < 36	0.794*	0.647	0.975
Age > 55	0.957	0.540	1.697
Non-UK Nationality	0.530*	0.431	0.652
Undisclosed Nationality	1.174	0.753	1.831
Known Disability	1.076	0.599	1.934
Strategic Funding Opportunity	0.501*	0.332	0.756

[1] * = significant at 5% significance level

There is also an interesting age dynamic in the results—applications led by investigators in the youngest age bracket are more likely to receive funding for research grants, but somewhat less likely for fellowships.

Ethnic minorities have lower odds of receiving funding for both grant categories, but being both an ethnic minority and in the youngest age bracket moderates this effect for research grants. This could indicate that, over time, there may be a reduction in disparities between ethnic groups but confirming that would require further exploration.

TABLE 7

Average Predicted Probabilities of Funding Success for Research Grants, Selected Demographics

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Sex and Ethnicity	Mean Prediction
Male, White	36.5%
Female, White	38.2%
Male, Ethnic Minority	27.5%
Female, Ethnic Minority	28.7%

TABLE 8

Average Predicted Probabilities of Funding Success for Fellowships, Selected Demographics

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Sex and Ethnicity	Mean Prediction
Male, White	20.7%
Female, White	30.5%
Male, Ethnic Minority	15.5%
Female, Ethnic Minority	23.6%

ASSOCIATIONS BETWEEN AWARD VALUE AND DEMOGRAPHICS, BY GRANT CATEGORY

We split the sample into applications for fellowships and applications for research grants, and then fit a series of linear mixed-effects regression models for each grant category that predict award value against sex, ethnicity, nationality, disability, and age as fixed effects. We approach the question first at the level of applications, and then at the level of unique individuals.[39]

When looking at the application level, we add a fixed effect for funding opportunity type (responsive or strategic) and include random intercepts for Research Area Group (RAG), theme, region, and year. As these variables are associated with the grant and application and are not consistent for individuals across applications, they are not included when looking at the individual level. Additionally, the individual level averages the award values across all of an individual's successful applications and uses that as the outcome of interest.

Table 9 shows estimates derived from our model for associations between demographics and successfully applied-for-and-received award values for research grant applications, along with 95% confidence intervals. Here, confidence intervals that include 1 indicate an insignificant effect, and effects greater than 1 indicate higher award values relative to the baseline categories for each predictor. The effects can be interpreted as percentage changes, such that an estimated effect of 0.5 indicates that expected award values for an application led by a member of the indicated group are 50% of the value for a member of the baseline group.

When they are successful, applications led by females and ethnic minority PIs are estimated to apply for and receive awards worth 84.7% and 90.1% of the award values of successful applications led by males and white PIs, respectively. Additionally, we see that having a lead applicant under the age of 36 is associated with a large reduction in estimated award value—applications led by PIs in this age bracket are

[39] In all cases, models including interaction terms were compared to those without, and in each case the interaction terms did not produce significant improvements in model performance.

estimated to apply for and receive less than half the amount compared to applications led by PIs between 36 and 55. Applicants in the oldest age bracket apply for and receive the largest award values.

Applications from PIs with non-UK nationalities apply for and receive less funding than PIs with UK nationality. The estimated effect of having a disability is not significant. Again, it should be noted that these discrepancies are likely to reflect both differences in applied-for values between groups, and any bias in the decisions of panels.

TABLE 9 [40]

Mixed-Effects Model of Award Value Against Predictors

Exponentiated Coefficients and 95% CI (Research Grants, Application Level)[1]

Predictor	Effect (multiplicative)	CI Lower Limit	CI Upper Limit
(Intercept)	1.131	0.528	2.421
Female	0.847*	0.783	0.916
Ethnic Minority	0.901*	0.829	0.978
Age < 36	0.495*	0.456	0.537
Age > 55	1.295*	1.199	1.398
Non-UK Nationality	0.746*	0.695	0.800
Undisclosed Nationality	1.104*	1.012	1.205
Known Disability	1.160	0.927	1.451
Strategic Funding Opportunity	1.024	0.940	1.116

[1] * = *significant at 5% significance level*

Results for fellowships differ considerably from those for research grants. Results from the models for that set of applications are presented in Table 10.

TABLE 10 [41]

Mixed-Effects Model of Award Value Against Predictors

Estimated Coefficients and 95% CI (Fellowships, Application Level)[1]

Predictor	Effect (additive)	CI Lower Limit	CI Upper Limit
(Intercept)	1.222*	1.041	1.403
Female	0.001	−0.080	0.082
Ethnic Minority	0.023	−0.072	0.118
Age < 36	-0.466*	−0.541	−0.391
Age > 55	0.363*	0.153	0.572
Non-UK Nationality	-0.040	−0.115	0.036
Undisclosed Nationality	0.068	−0.087	0.224
Known Disability	-0.223*	−0.425	−0.020
Strategic Funding Opportunity	-0.022	−0.159	0.116

[1] * = *significant at 5% significance level*

For our model for fellowships, confidence intervals that include 0 indicate a finding that is not significant at the 5% significance level. Effects above 0 indicate an increase in award value. Award values are measured in millions of pounds, so an estimated effect of 1 suggests an increase in award value of £1 million.

[40] For research grants, the models were fit after log-transforming the award value. Coefficients were then exponentiated to aid interpretability. This results in estimated effects that are interpreted multiplicatively, rather than additively.

[41] The models for fellowships were fit against award value without a log-transformation, so these results are interpreted additively. Application value is measured in the millions, so an effect size of 0.5 indicates an increase in award value of £500,000.

We see from Table 10 that the effects associated with having a female or ethnic minority PI are no longer significant. Similarly, non-UK nationality is no longer significant. However, there is still an estimated premium for being in the oldest age bracket, and having a disability now shows a significant negative effect on applied-for-and-received award values, with a moderate-to-large effect size.

Results from the analysis at the unique individual level rather than the application level show few differences of practical significance, although it is important to reiterate that the models control for fewer potential factors at this level. These results are shown in Tables 11 and 12.

Here, the main difference is that the effect of having an ethnic minority PI is now marginally insignificant, although the effect size is largely similar to the application-level result. Additionally, there is a larger estimated premium for being in the oldest age bracket. Still, the directions of all of the estimated associations are the same.

Qualitatively, the results for fellowships when considering unique individuals is practically the same as when looking at applications. While in some cases the confidence intervals shift from marginally significant to marginally insignificant or vice-versa, the estimated effect sizes are very close to one another, so interpretation of such changes should be made with caution.

Previous research from the EPSRC[42] indicated that there are substantial differences between male and female applicants above and below £2.5 million. We added to that finding using a logistic regression predicting whether an award was above or below that threshold against applicant characteristics, which found that having a PI who was white, older, and without a known disability were strong predictors of a successful application being above £2.5 million. As this was a tertiary piece of analysis not in the prioritised research questions, it is left for reference in the GitHub repository.[43]

Following that analysis, the preceding models that predict award value do not split the data by that threshold, although a supplementary analysis available in the GitHub repository does. Results from models above the cutoff suffer from small samples that limit what can be said about associations above £2.5m when considered in isolation.

TABLE 11

Mixed-Effects Model of Award Value Against Predictors

Exponentiated Coefficients and 95% CI (Research Grants, Individual Level)[1]

Predictor	Effect (multiplicative)	CI Lower Limit	CI Upper Limit
(Intercept)	0.565*	0.528	0.605
Female	0.875*	0.785	0.976
Ethnic Minority	0.924	0.828	1.032
Age < 36	0.499*	0.447	0.557
Age > 55	1.518*	1.357	1.698
Non-UK Nationality	0.733*	0.669	0.804
Undisclosed Nationality	1.154*	1.011	1.318
Known Disability	1.186	0.859	1.636

[1] * = *significant at 5% significance level*

[42] For instance, see: https://www.ukri.org/wp-content/uploads/2022/03/EPSRC-070322-UnderstandingOurPortfolio-AGenderPerspective.pdf

[43] This can be accessed at: https://github.com/alan-turing-institute/equity-in-grant-funding

TABLE 12

Mixed-Effects Model of Award Value Against Predictors

Estimated Coefficients and 95% CI (Fellowships, Individual Level)[1]

Predictor	Effect (additive)	CI Lower Limit	CI Upper Limit
(Intercept)	1.276*	1.191	1.361
Female	-0.048	−0.148	0.052
Ethnic Minority	0.013	−0.096	0.121
Age < 36	-0.494*	−0.582	−0.406
Age > 55	0.322	−0.029	0.673
Non-UK Nationality	-0.037	−0.127	0.054
Undisclosed Nationality	0.127	−0.075	0.330
Known Disability	-0.219	−0.447	0.009

[1] * = *significant at 5% significance level*

ASSOCIATIONS BETWEEN DEMOGRAPHICS AND BEING FUNDED VS FUNDABLE BUT NOT FUNDED

Reviewers and panels make decisions about whether applications can progress to further stages of the review process. Early in the process, applications can receive office rejections or review rejections, which entail that they are not considered fundable. Applications can also be deemed not fundable at the panel stage. The dataset includes 13,760 fundable applications, of which 6,500 were ultimately funded.

Exploring differences between the demographics in terms of how often they are funded, fundable, or not fundable is an interesting addition to analyses that simply look at funding rates, because it suggests more specific areas where disparities arise.

Figure 18 shows the rates at which applications led by each sex are determined to be funded, fundable but not funded, and not fundable by the end of the review process. For research grants, there is virtually no meaningful difference between the sexes for any of these categories. However, females are funded for fellowships at a rate almost 50% higher than their male counterparts.

By contrast, males are more likely to be fundable but not ultimately funded for fellowships. These figures align with our earlier findings from models predicting funding success when stratifying the data by grant category. The results from those analyses can be found in Tables 5, 6, 7, and 8 of this report.

FIGURE 18

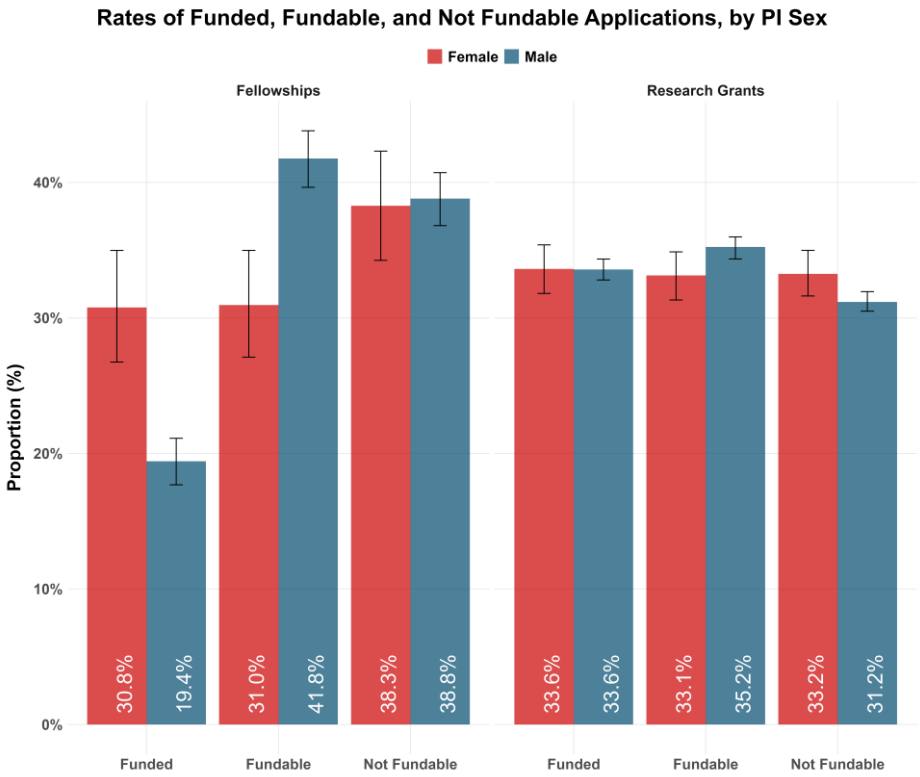
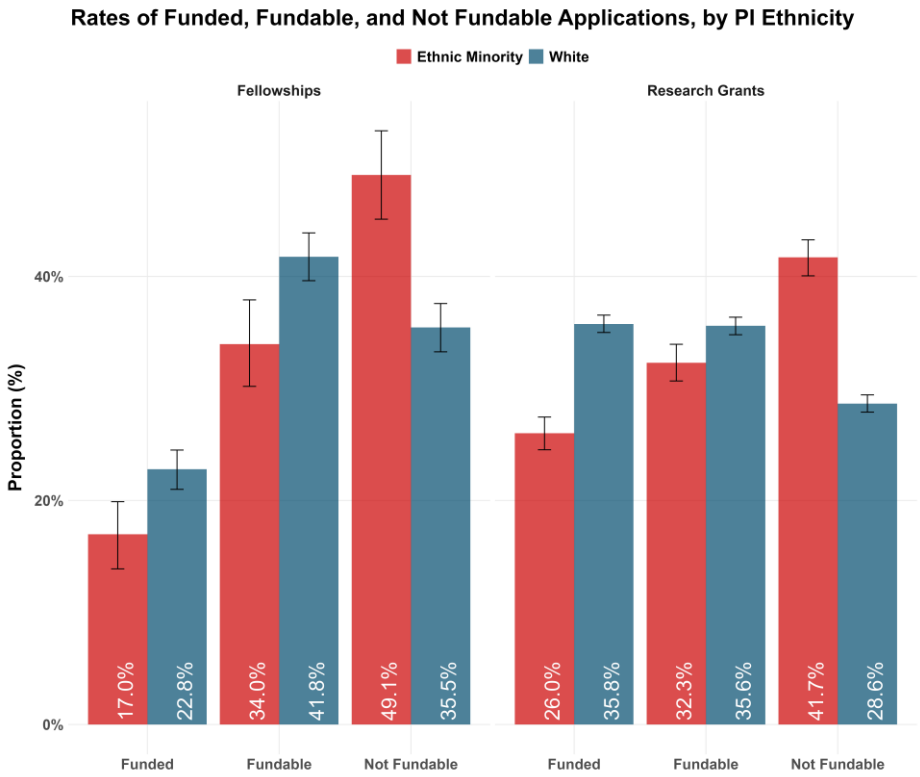


Figure 19 shows same rates for ethnicity. There is a clear pattern for applications led by ethnic minorities to be more likely to be not fundable than applications led by white PIs. Conversely, applications with ethnic minority PIs are less often funded or fundable.

FIGURE 19



Similarly, for nationality and disability, we see disparities in funding and fundable rates for lead applicants with non-UK nationalities (Figure 20) and known disabilities (Figure 21), although in the latter case the group with known disabilities is so small that our confidence intervals are too wide to draw meaningful conclusions.

FIGURE 20

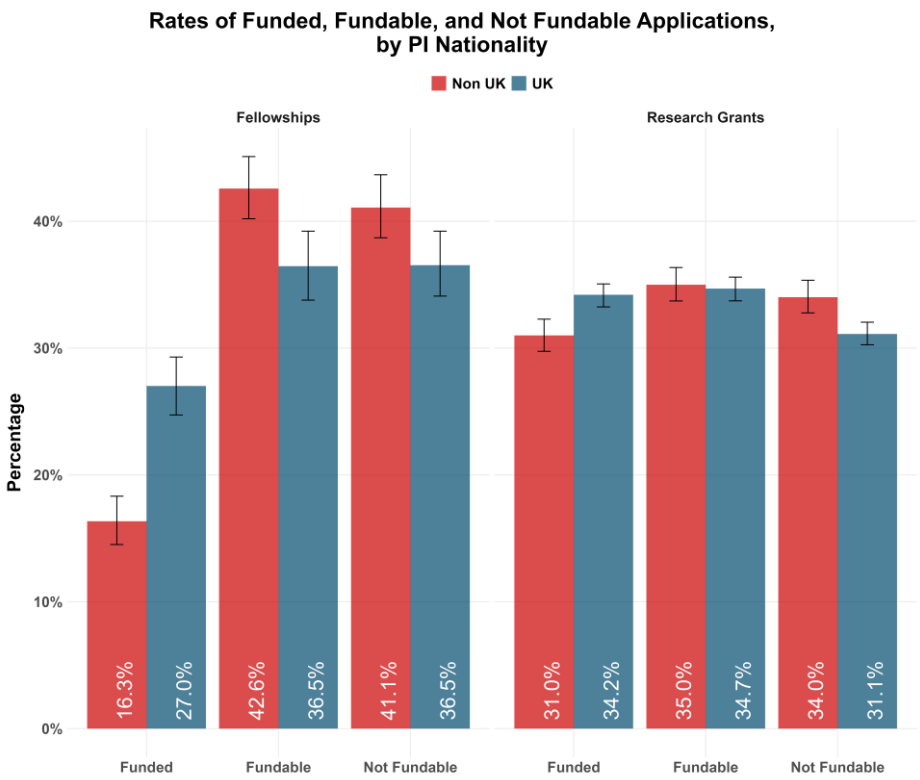
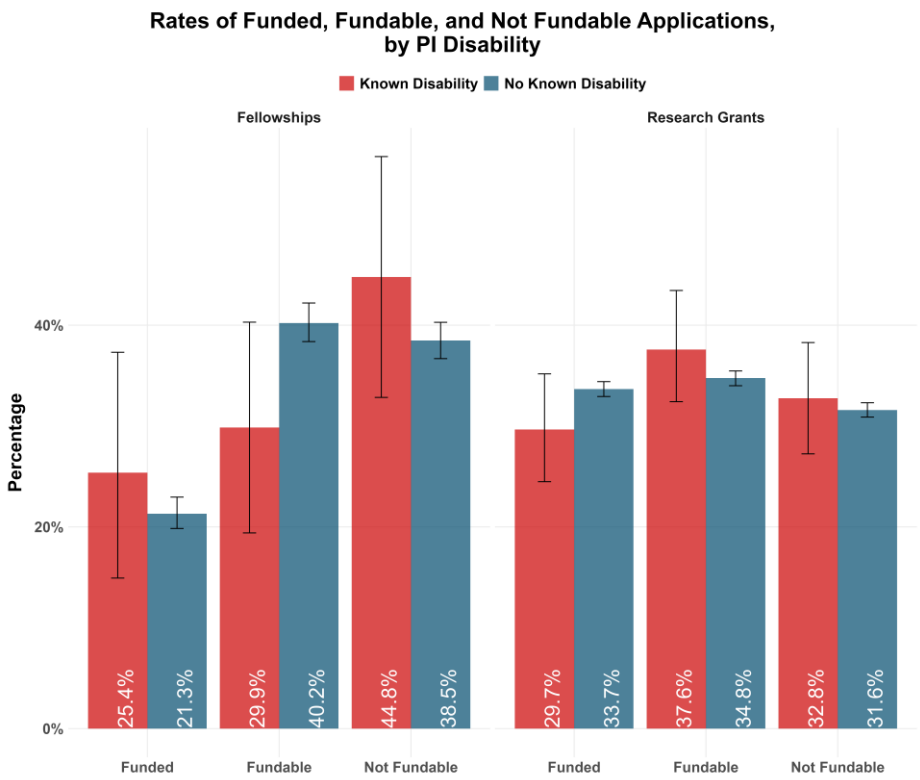
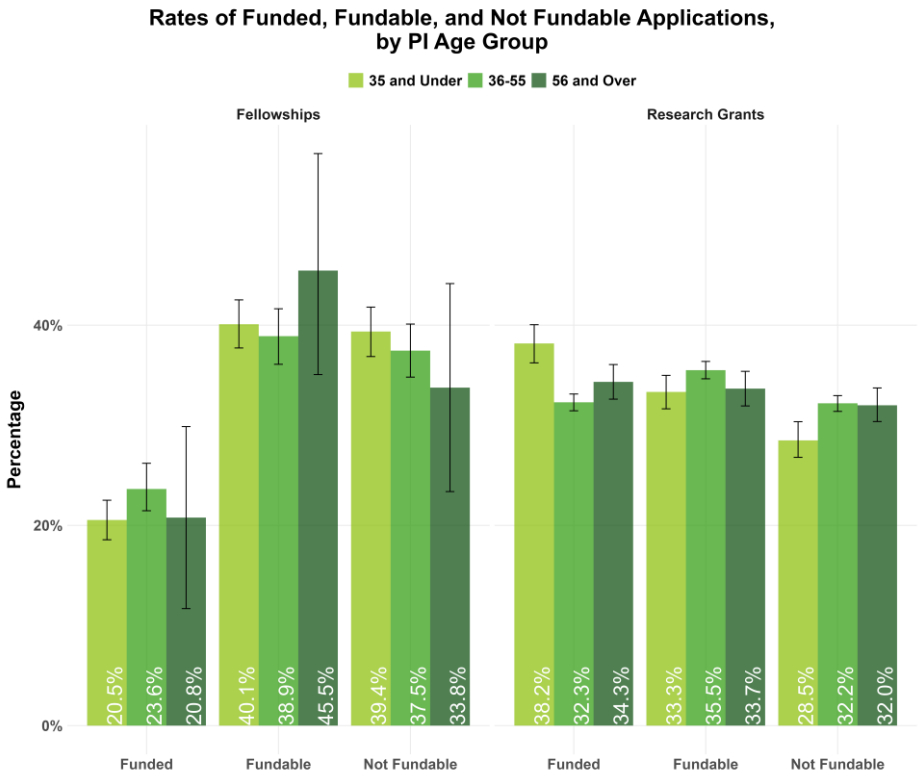


FIGURE 21



When looking at age group in Figure 22, we see that there is relatively little variation between the groups and their fundability rates for either grant category, although we still observe that applications are funded at substantially lower rates overall in fellowships than in research grants.

FIGURE 22



COMPARING APPLICANTS THAT SUBMIT APPLICATIONS AFTER REJECTIONS TO THOSE THAT DO NOT

Another way to break down the data is by whether individuals submit additional applications after receiving a rejection.[44] In total, there are 13,770 individuals in the dataset who received at least one rejection as a co-investigator or PI. Among that group, 4,340 chose to submit another application at a later date as (again as either a co-investigator or PI).

Figure 23 shows the proportions of these groups by demographic characteristics. The baseline proportion of the population that receives at least one rejection is also given by the circular points to help indicate the direction of over or under-representation. The key comparisons are between the shares of each population that are accounted for by each demographic group. For instance, the chart shows that males account for a slightly larger share of the population that submits again at a later date following a rejection relative to what might be expected if they submitted after rejection proportionally to their baseline representation in the rejected population.

Compared to the population that does not submit following rejection, the submitting population is more often from the UK, more often between the ages of 36 and 55, less often under 35, more often male, and more often white.

When considering all three-way intersections of sex, ethnicity, and age, an interesting picture emerges—all groups account for a slightly smaller share of the population that submits again after rejection than the population that does not submit again, except for two groups: white males aged 56 and over, and white males aged 36 to 55. The latter group in particular accounts for about 30% of the non-submitting population, but about 45% of the submitting population, which is the largest disparity by overall percentage. By relative proportions, female ethnic minorities aged 36-55, white females under 35, and ethnic minority males under 35 see large disparities as well, although the differences in terms of overall share are small. The comparisons can be seen in Figure 24.

There are more limitations to an analysis of submission after rejection than other breakdowns. Because the data do not cover the full EPSRC grant history, we can't account for rejections that may have happened prior to 2013. Additionally, we can't account for submissions that have happened or will happen since 2023. The lack of

[44] We refer to individuals who submit any application after a rejection as the "resubmitting" population. The "resubmissions" are not necessarily revised versions of their rejected applications.

complete information can lead to misclassification—inevitably, some individuals we have marked as non-submitting will eventually submit again.

FIGURE 23

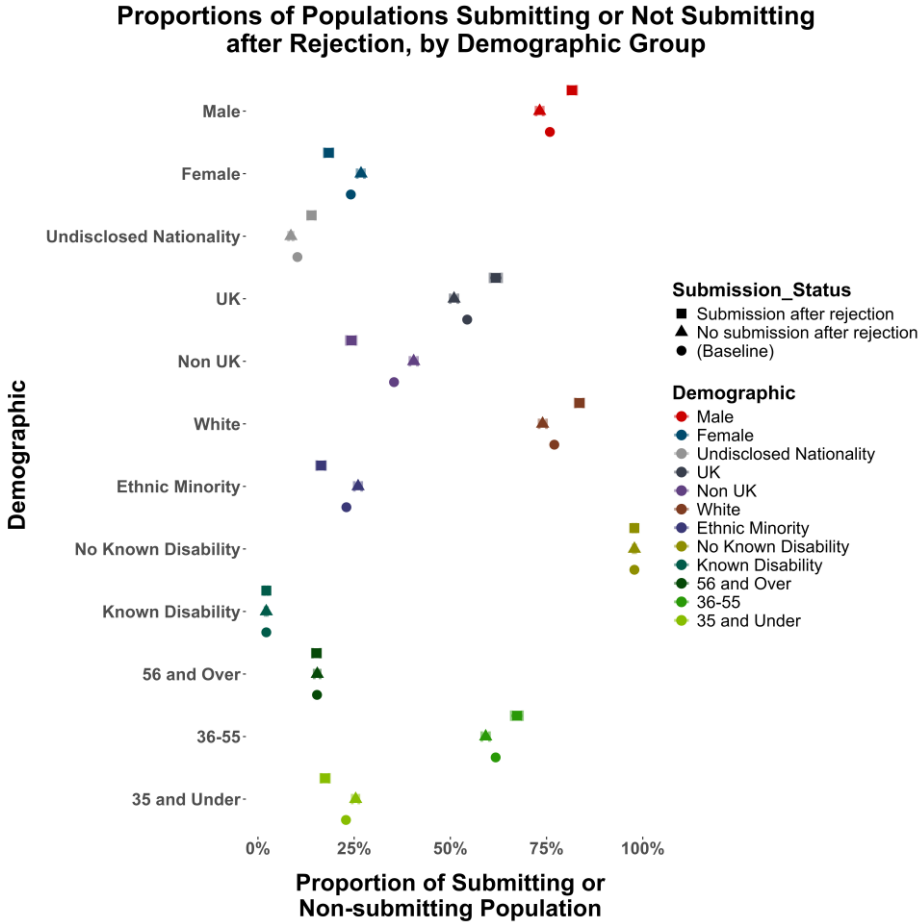
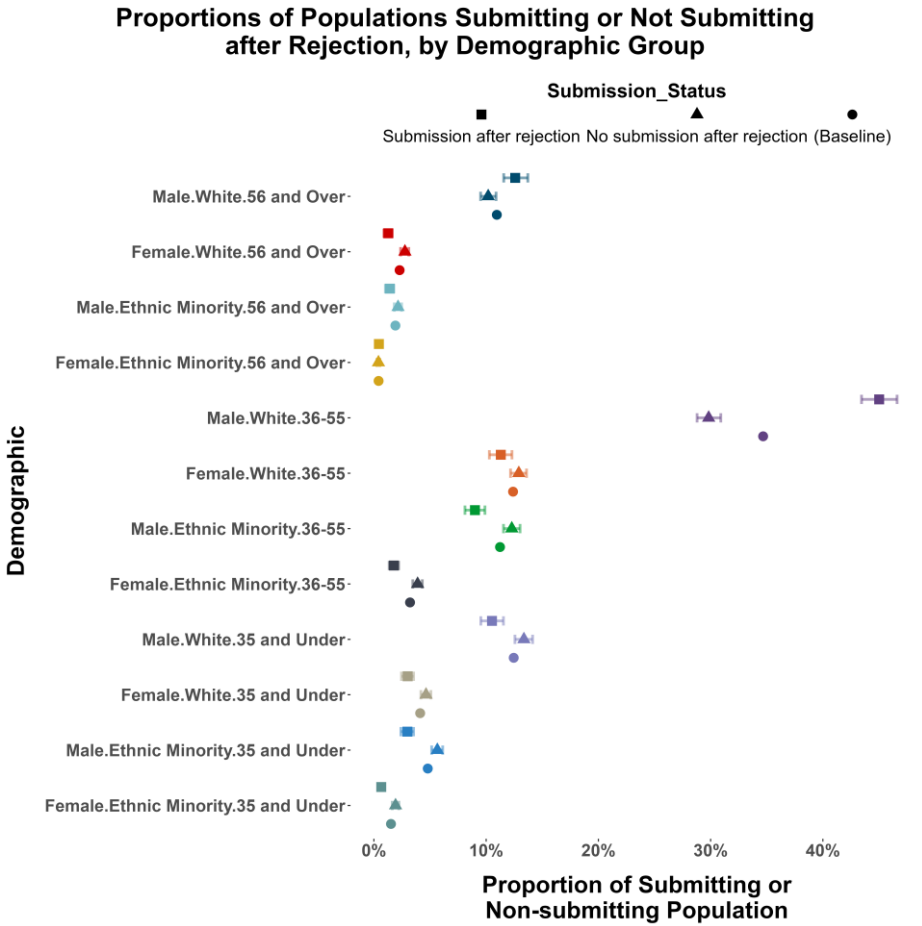


FIGURE 24



Additionally, the motivations behind choosing to submit following rejection are likely to be complex and may be influenced by factors unrelated to the EPSRC peer review process, including prior rejections or applications to other funders. We also cannot account for factors such as systemic differences between the demographic groups, such as differences in access to mentoring or coaching, or individual life circumstances that may incline someone to spend less time on submitting grant applications.

Further investigation would do well to consider these more systemic factors and approach the question using a mixture of methodologies to better understand gaps that are incompletely answered by the available data.

DOES HAVING A SUCCESSFUL GRANT APPLICATION CORRELATE WITH DIFFERENT OUTCOMES IN SUBSEQUENT APPLICATIONS?

We are interested in exploring whether earlier success is predictive of success in future applications. Unfortunately, the available data do not lend themselves to causal explorations of that question, not least because we cannot control for researcher ability that would increase the chances of success throughout time. Another limitation in our ability to explore this question comes from having only nine years of data about funding decisions. Thus, it is not possible to determine from the available data which investigators had successful EPSRC grant funding awarded before the 2014/2015 funding year. Yet another limitation is that there are several sources of funding other than EPSRC from which applicants may have previously held funding, and there is no information on which investigators had received other funding at any point.

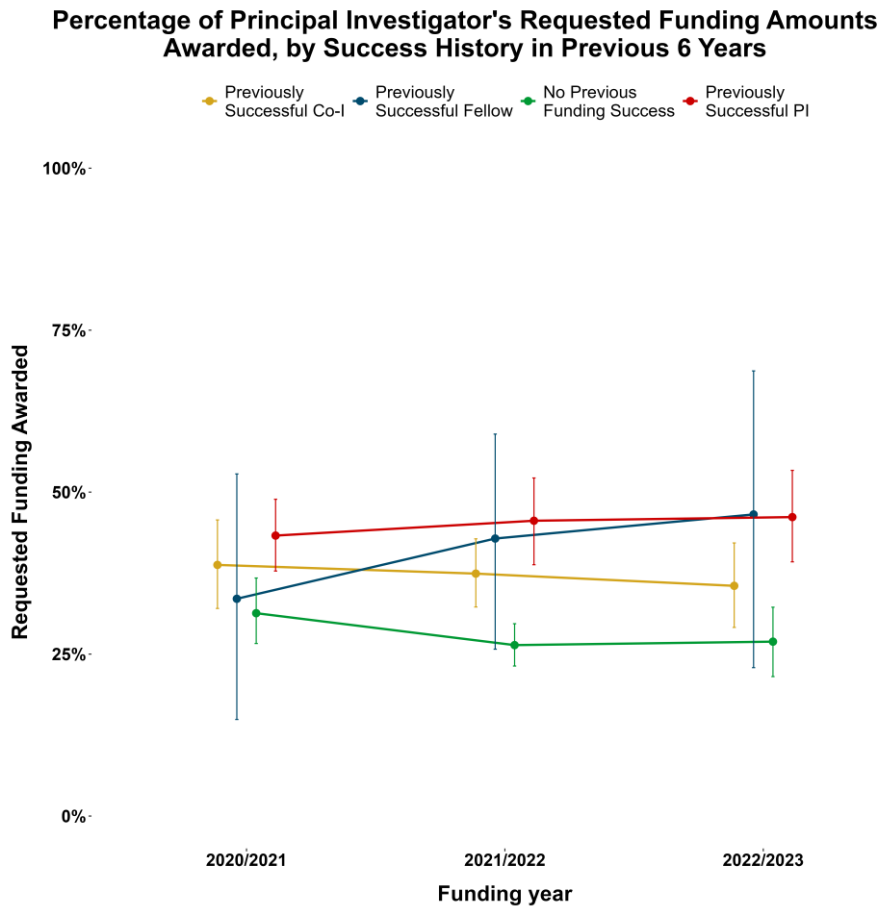
Nevertheless, due to general interest in the question, a first exploration was undertaken, exploring the impact on success rates of recent EPSRC funding success.

For the final three years in the dataset, we calculate the success rates by number and by value of applications with Principal Investigators (PI or fellow) who belong in each of the following sets:

- Previous Fellow: principal applicant received fellowship funding in the previous six years,
- Previous PI: principal applicant received funding as a PI but not as a fellow in the previous six years,
- Previous Co-I: principal applicant received funding as a Co-I, but not as a PI or fellow, in the previous six years,
- No previous funding: principal applicant did not receive any EPSRC funding in the previous six years.

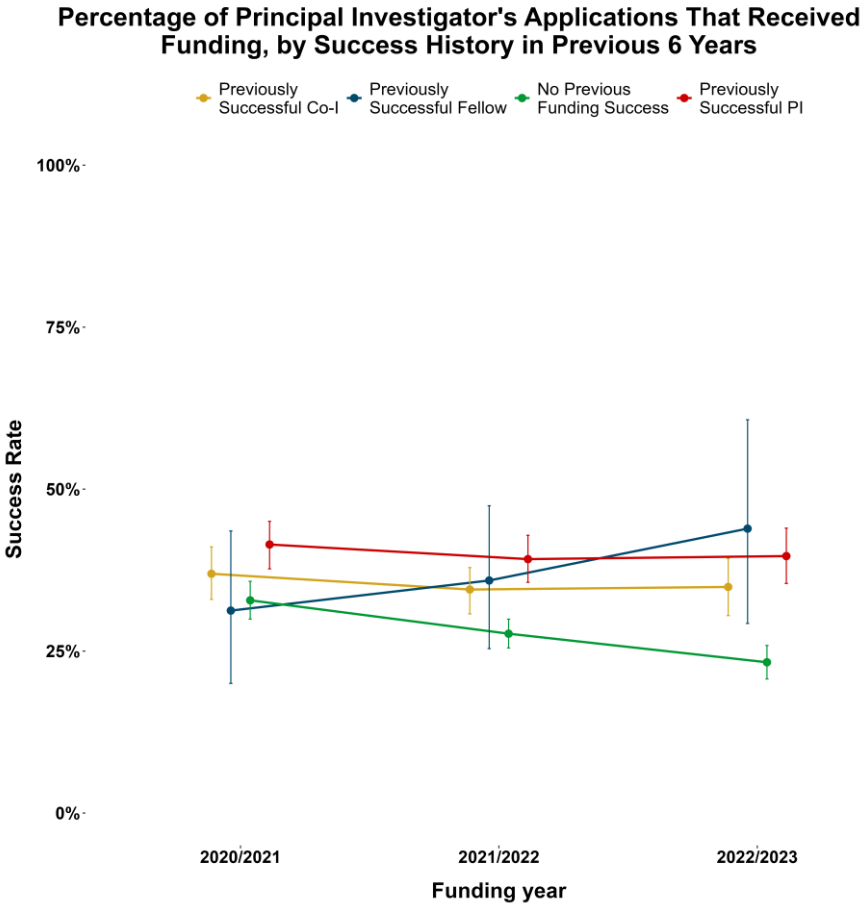
The results from this approach are presented in Figures 25 and 26.

FIGURE 25



Excluding former success as fellows, these indicate that applications whose main investigator had success as a PI in the previous six years have the highest success rate both by value and number, followed by applications whose main investigator had recent success as a Co-I, with those whose main investigator had no recent previous success having the lowest success rate.

FIGURE 26



The magnitude of this effect does not seem to be decreasing over the three-year period investigated. The picture for fellows is somewhat less clear, though this should be interpreted with caution as there are very few fellows in each year, and consequently very few subsequent applications with principal investigators who are recent fellows.

IS THERE EVIDENCE THAT INSTITUTIONS HAVE DIFFERENT SUCCESS RATES IN FUNDING APPLICATIONS?

In the EPSRC community survey analysed by the Alan Turing Institute as part of this project (the details of which can be found in their full report available in the GitHub repository[45]), 32.4% of respondents who indicated that they perceived bias in the review process identified "the prestige of academic institutions" as the primary reason for the bias. This was the most commonly perceived source of bias in the survey.

This study therefore attempts a first investigation of the impact of institution on grant success rates. In the dataset provided by EPSRC, organisations are anonymised, so it is not possible to attach an external measure of prestige to use in analysis. Additionally, it is not entirely clear what external measure of prestige would be appropriate, as in most cases there are causal influences between these measures and funding success in both directions. Information was provided about which organisations were EPSRC partners in a given year, where EPSRC partners are those institutions with the 36 largest portfolios with the EPSRC.[46] Again, as partner status is determined to a large extent by funding outcomes, it is not useful for understanding influence on outcomes.

For this reason, the investigation instead focuses on variation in institutional success rates and on funding concentration over time at the institutional level. We consider the applying institution to be the institution associated with the application's principal investigator. We seek to address the following subquestion:

- How are institutional success rates and the overall number of applications that are decided in a given year related to each other?

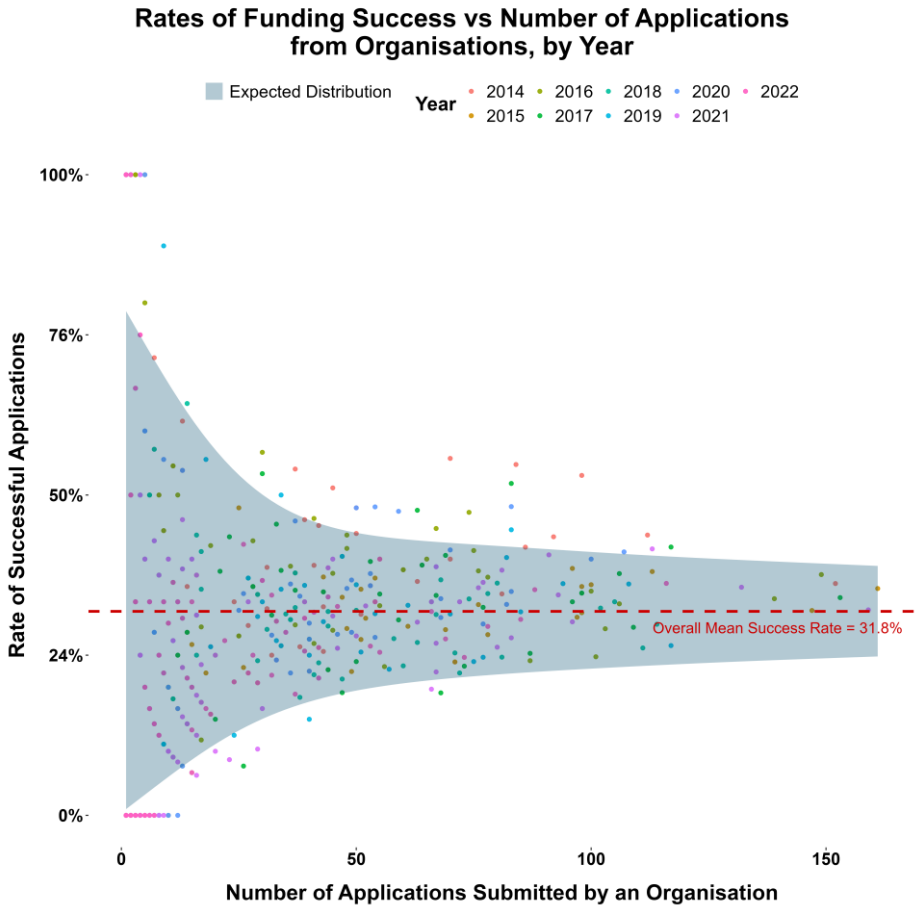
A combined plot covering all the years where we have data of the number of proposals submitted by institutions vs their success rate by number is presented in Figure 27. The graph indicates that, as statistically expected, the institutions with very high and very low success rates by number in any given year correspond to those

[45] This is available at https://github.com/alan-turing-institute/equity-in-grant-funding

[46] See https://www.ukri.org/who-we-are/epsrc/relationships/working-with-universities/

where the number of submissions was small, and where consequently success rates are more likely to be extreme. Although there are some exceptions scattered over the years, success rates tend to concentrate around the mean, particularly as the number of applications increase.

FIGURE 27



Additionally, we seek to answer the following:

- Are there institutions that consistently have high success rates?

Here, we investigate how often institutions appear in the top quartile of institutional success rates. In any given year, there are some institutions with no applications decided. They are in a fifth "NA" category for that year. Some institutions were never in the top quartile, and some were in the top quartile in six of the nine decision years. There were no institutions in the top quartile more than 6 times across the nine funding decision years. The number of institutions appearing n times, n = 0 to 6, is given in Figure 28.

Finally, we can evaluate funding concentration by asking in each year what percent of highest funded institutions together received 50%, 75% and 95% of overall funding awarded in that year. The results from this angle of analysis are in Figure 29.

We observe that in each year, 90% of funding is concentrated around 25-30% of institutions, indicating a fair degree of funding concentration at the institution level.

However, some caution is required in interpretation, as institutions also have different sizes, and it is likely that the top funded quarter of institutions represent substantially more than a quarter of the eligible research population. Thus, this concentration at the institution level does not necessarily correspond to an analogous concentration at the level of eligible researchers. It also does not pick up collaborations between investigators from different institutions, as we only consider the institution of the principal investigator on each application.

FIGURE 28

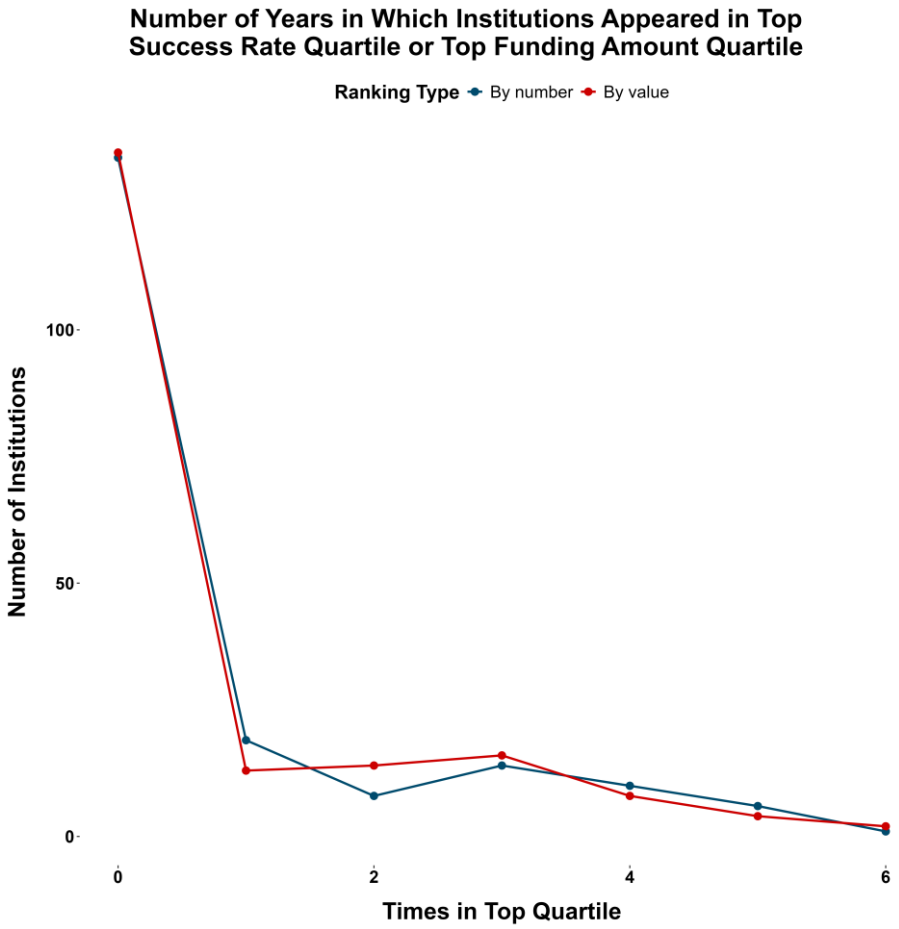
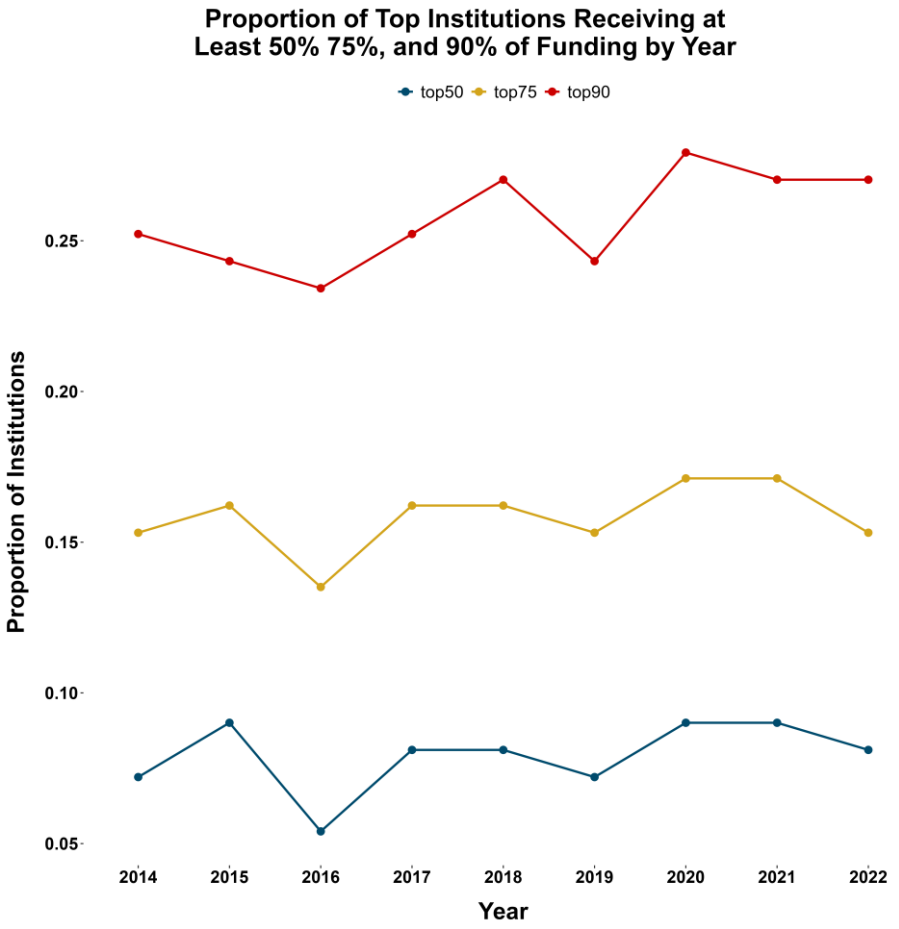


FIGURE 29



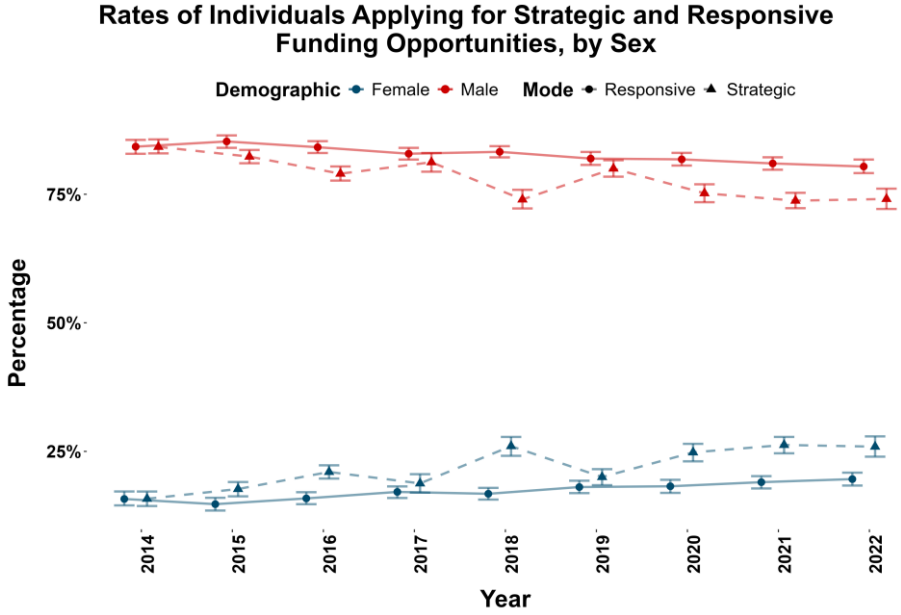
HOW DO APPLICANTS TO STRATEGIC AND RESPONSIVE FUNDING OPPORTUNITIES DIFFER, AND HOW HAS THIS CHANGED OVER TIME?

In the EPSRC data, applications are made for grants that fall into one of two bins: strategic funding opportunities and responsive opportunities. Strategic funding opportunities are "top-down" in the sense that they pertain to topics set by the EPSRC. Responsive opportunities are "bottom-up," as they pertain to researcher-generated topics and often run on a rolling basis Additionally, strategic funding

opportunities can vary greatly from year to year, whereas responsive ones tend to be relatively similar over time, with comparatively minor modifications. Thus, observed irregularities across years for strategic funding opportunities could relate to idiosyncrasies of the particular grants.

Figure 30 shows the proportions of all individuals applying to strategic and responsive funding opportunities by sex, along with 95% confidence intervals around the proportions in each year. There has been some convergence between the sexes over time for both funding opportunity types, although the trend appears to be stronger for strategic opportunities.

FIGURE 30



The gap between the funding types within the sexes is very small in most years, although we do observe that relatively more women apply to strategic opportunities than responsive ones in recent years compared to the earlier years, when there were no substantial differences. However, the confidence intervals frequently overlap, making it difficult to conclude that there are structural differences between the funding opportunity types.

Figure 31 breaks down the applicant pools to each funding opportunity type by ethnicity. We see some convergence between white and ethnic minority applicants in terms of their share of applications to both types. Within the ethnic groups, we see that there are not large differences in the relative share of individuals applying for either funding opportunity type in most years, with 2016, 2017, and 2019 being notable exceptions when ethnic minorities applied more often for strategic opportunities than responsive opportunities.

FIGURE 31

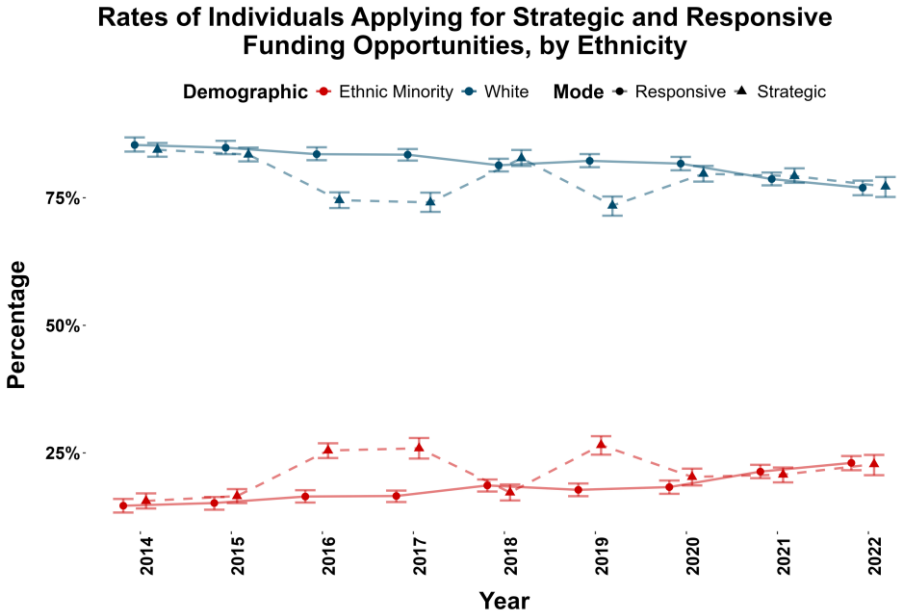


Figure 32 shows the applicants to each funding opportunity type by age. There is some small differentiation within the under 36 and 36-55 age groups between the relative share of applicants to each funding opportunity type in early years in the data. However, in more recent years these gaps have mostly vanished in practical terms.

FIGURE 32

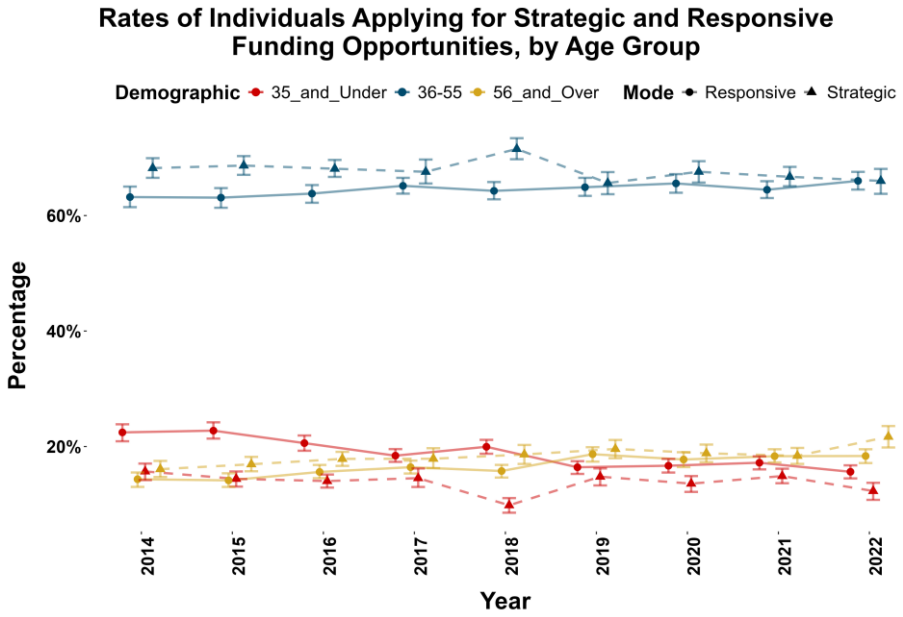
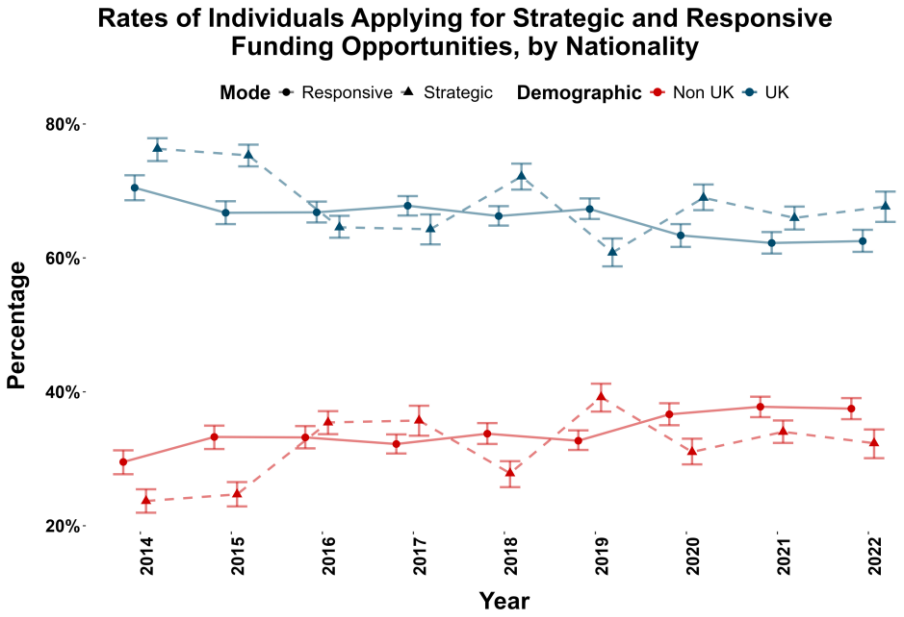


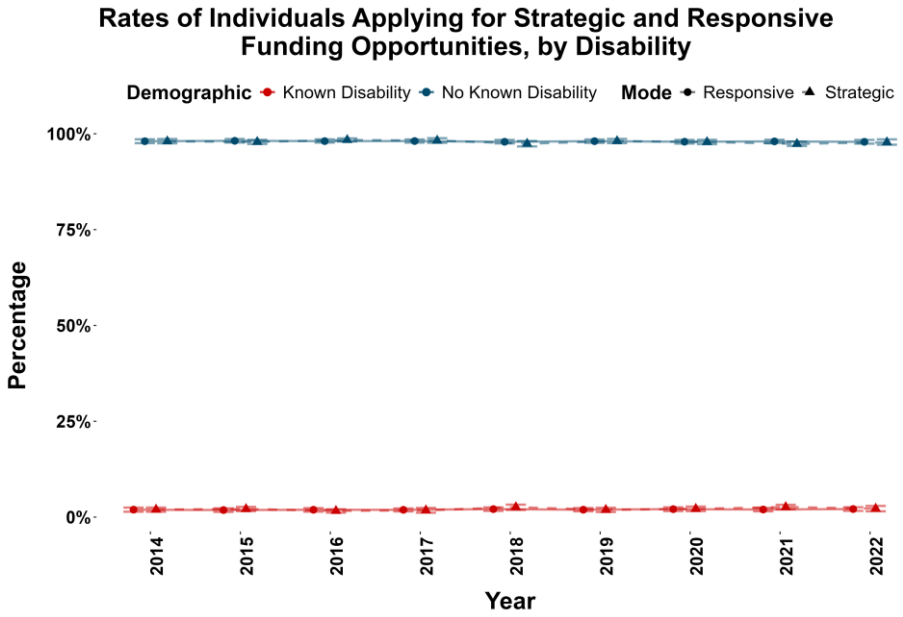
FIGURE 33



The picture for nationality shows more instability in the composition of the applicant pool of strategic opportunities than the other demographic comparisons presented here. The breakdown by nationality is presented in Figure 33. Among applicants with non-UK nationalities, there is a trend towards a greater share of both funding opportunity types. However, there does not appear to be a consistent difference between the opportunity types over time.

There are no notable differences between the share of applicants with and without known disabilities within either funding opportunity type. Figure 34 shows this comparison. The vast majority of applicants to both types do not have a known disability, and the share of such applicants is effectively stable over the analysable time period.

FIGURE 34



DO OUTCOMES FOR STRATEGIC AND RESPONSIVE FUNDING OPPORTUNITIES DIFFER ACROSS DEMOGRAPHICS?

We consider two outcomes: inflation-adjusted award amount and a binary funding decision (ie, an application was ultimately funded or not funded). We consider these outcomes at the application level. We fit a series of mixed-effects regression models similar to others that have been run elsewhere in this report; we regress the outcomes of interest against sex, ethnicity, disability, age group, nationality, and grant category as fixed effects, with random intercepts for region, theme, research area group, and the year that a decision was made for the applications. However, for this research question, we stratify the sample by funding opportunity type (ie responsive or strategic opportunities) and run the models separately for each type.

Tables 13 and 14 present the results of our models for binary funding success for responsive and strategic opportunities, respectively. Tables 15 and 16 present the corresponding average predicted probabilities of funding success for selected demographics.

For both models, we compared the baseline models without interaction terms to models that included all pairwise interactions between sex, age, and ethnicity using likelihood ratio tests. In both cases, the interaction terms were found to significantly improve performance. We report the results from those models here. Odds ratios greater than 1 indicate higher odds of a successful application, while confidence intervals that include 1 indicate a finding that is not significant at the 5% significance level.

Applications led by ethnic minority applicants are less likely to be successful than applications led by white applicants, regardless of funding opportunity type. This finding is offset somewhat for ethnic minorities in the youngest age group strategic opportunities.

TABLE 13

Mixed-Effects Model of Binary Funding Success Against Predictors, Responsive Opportunities

Odds Ratios and 95% Confidence Intervals for Fixed-Effect Predictors[1]

Predictor	Odds Ratio	CI Lower Limit	CI Upper Limit
(Intercept)	0.305*	0.125	0.745
Female	1.090	0.956	1.243
Ethnic Minority	0.698*	0.613	0.794
Age < 36	1.247*	1.118	1.392
Age > 55	1.006	0.892	1.134
Non-UK Nationality	0.838*	0.772	0.911
Undisclosed Nationality	0.961	0.851	1.084
Known Disability	0.694*	0.516	0.934
Research Grant	2.579*	2.282	2.915
Ethnic Minority Female	0.754*	0.581	0.980
Female 35 or Younger	1.292*	1.035	1.613
Female 56 or Older	1.217	0.847	1.748
Ethnic Minority 35 or Younger	1.176	0.950	1.456
Ethnic Minority 56 or Older	0.767	0.545	1.081

[1] * = significant at 5% significance level

TABLE 14

Mixed-Effects Model of Binary Funding Success Against Predictors, Strategic Opportunities

Odds Ratios and 95% Confidence Intervals for Fixed-Effect Predictors[1]

Predictor	Odds Ratio	CI Lower Limit	CI Upper Limit
(Intercept)	0.396*	0.167	0.939
Female	1.114	0.917	1.353
Ethnic Minority	0.552*	0.458	0.665
Age < 36	0.987	0.758	1.286
Age > 55	0.946	0.802	1.116
Non-UK Nationality	0.760*	0.657	0.880
Undisclosed Nationality	1.345*	1.137	1.591
Known Disability	1.075	0.753	1.533
Research Grant	2.548*	1.841	3.528
Ethnic Minority Female	1.390	0.937	2.061
Female 35 or Younger	0.542*	0.320	0.918
Female 56 or Older	1.244	0.839	1.845
Ethnic Minority 35 or Younger	1.687*	1.055	2.696
Ethnic Minority 56 or Older	0.998	0.669	1.489

[1] * = significant at 5% significance level

Looking at the average predicted probabilities of success in Tables 15 and 16, we can see an interesting intersectional picture emerging: for responsive opportunities, the sexes are relatively close in their predictions across white and ethnic minority applicants. However, for strategic opportunities, while white males and females are close, there is a roughly 9% gap between the average predictions for ethnic minority females compared to ethnic minority males, in favour of females.

TABLE 15

Average Predicted Probabilities of Funding Success for Responsive Funding Opportunities, Selected Demographics

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Sex and Ethnicity	Mean Prediction
Male, White	31.0%
Female, White	34.6%
Male, Ethnic Minority	24.3%
Female, Ethnic Minority	22.6%

TABLE 16

Average Predicted Probabilities of Funding Success for Strategic Funding Opportunities, Selected Demographics

Average Predictions Based on Mixed-Effects Model for Binary Funding Success

Sex and Ethnicity	Mean Prediction
Male, White	42.8%
Female, White	45.0%
Male, Ethnic Minority	31.7%
Female, Ethnic Minority	40.3%

There are other observable differences by funding opportunity type that emerge. For instance, applications led by PIs with non-UK nationality have lower odds of success than their UK-led counterparts for strategic opportunities, but not for responsive ones. Similarly, applications that have a PI with a known disability are less likely to be successful for responsive opportunities, but not for strategic ones.

We also looked at the award values that were applied for (and received—EPSRC does not generally adjust the amounts requested) using the same set of predictors. The results from those models are presented in Tables 17 and 18. Note that the outcome variable was log-transformed, and the coefficients were exponentiated to improve interpretability, so the estimated effects should be interpreted multiplicatively rather than additively. In other words, an estimated effect of 1.5 suggests a 50% increase in applied-for-and-received award value relative to the relevant reference group. Confidence intervals including 1 indicate the result was not significant at the 5% level.

For these models, likelihood ratios tests did not indicate that the inclusion of interaction terms significantly improved model performance. We therefore present the results from the models with no interactions.

In these models, we see that applications to responsive opportunities led by females and ethnic minorities are expected to request and receive lower award values than applications led by males and white investigators, respectively. The estimated effects for strategic opportunities were not significant, although the directions were the same.

The model for responsive opportunities generally shows more variables with significant effects than the strategic opportunity models—applied-for-and-received award values appear to increase with age group, and non-UK nationalities are associated with lower award values as well. However, these effects are not significant in the strategic models, with the exception of the effect for applications with PIs in the youngest age group, for which applied-for-and-received award values are lower throughout.

TABLE 17

Mixed-Effects Model of Award Value Against Predictors, Responsive Opportunities

Exponentiated Coefficients and 95% CI (Application Level)[†]

Predictor	Effect (multiplicative)	CI Lower Limit	CI Upper Limit
(Intercept)	1.852*	1.298	2.644
Female	0.868*	0.795	0.948
Ethnic Minority	0.898*	0.819	0.984
Age < 36	0.514*	0.474	0.556
Age > 55	1.521*	1.380	1.676
Non-UK Nationality	0.733*	0.680	0.790
Undisclosed Nationality	1.191*	1.070	1.327
Known Disability	0.954	0.724	1.258
Research Grant	0.378*	0.338	0.423

[†] * = significant at 5% significance level

TABLE 18

Mixed-Effects Model of Award Value Against Predictors, Strategic Opportunities

Exponentiated Coefficients and 95% CI (Application Level)[1]

Predictor	Effect (multiplicative)	CI Lower Limit	CI Upper Limit
(Intercept)	1.746	0.725	4.210
Female	0.890	0.790	1.001
Ethnic Minority	0.955	0.840	1.086
Age < 36	0.665*	0.558	0.793
Age > 55	1.075	0.964	1.199
Non-UK Nationality	0.911	0.809	1.026
Undisclosed Nationality	0.989	0.875	1.118
Known Disability	1.208	0.915	1.596
Research Grant	0.678*	0.513	0.895

[1] ** = significant at 5% significance level*

IS THERE EVIDENCE OF DIFFERENT OUTCOMES BY UK GEOGRAPHICAL REGION?

Alongside an exploration of demographic characteristics and their associations with outcomes for various subsets of the EPSRC grant data, we also did some preliminary analysis on whether there is variation in outcomes by UK geographical region, again focusing particularly on rates of applications receiving funding and the award value of successful applications. These outcomes are displayed in Figure 35. Unfortunately, to protect the identity of individual researchers, we used anonymised region codes, which makes practical interpretation challenging. Nevertheless, we present our observations and leave more substantive interpretation for future work from the EPSRC.

By region, we present the density of award values requested and received by applicants for awards less than £5 million.[47] We also plot the computed means for the overall distribution with the corresponding 95% confidence intervals. For award rates, we plot the proportion of successfully funded applications per region with 95% confidence intervals.

The mean award values in each region are quite similar, with most means falling within the confidence intervals of all other regions. The notable outlier is Region 06, which has a higher observed mean than all other regions, although its confidence intervals still contain the means of most other regions except Regions 03, 04, and 09.

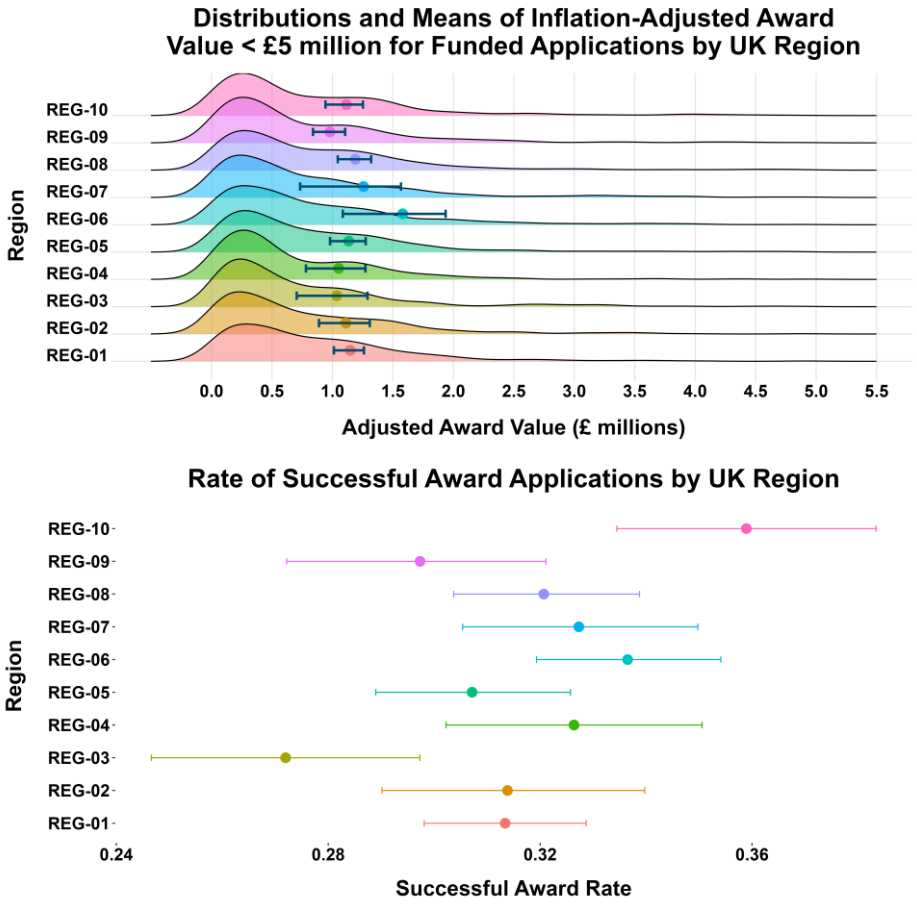
There is some more variation in the successful award rates between regions. Region 10 is a clear leader, with Region 03 lagging behind the others. While there is also minor variation among the remaining regions, the observed success rates are not practically different from one another.

Future research may find it prudent to explore differences between regional outcomes in greater detail. Although the results here do not suggest major differences, a more

[47] Award values range into the hundreds of millions of pounds, so this restriction was imposed to help preserve legibility. Only 2.6% of grants had award values above £5 million.

detailed analysis that goes beyond the scope of this report may be helpful in confirming that finding.

FIGURE 35



HOW DO APPLICANT, AWARDEE, REVIEWER AND PANELLIST POPULATIONS COMPARE ACROSS BOTH THEME GROUP AND RESEARCH AREA GROUP?

This question analyses unique individuals to understand whether demographic groups disproportionately are associated with any of the 22 themes or 9 research area groups (RAGs).[48] Demographics are considered for the applicant, awardee, reviewer, and panellist populations. For conciseness, this report discusses the demographic comparisons for the RAGs, and the comparisons for themes are available in the GitHub repository.

Unfortunately, as with regions, we do not have access to non-anonymised RAG or theme identifiers. This limits our ability to meaningfully interpret patterns of over or under-subscription by demographic group. That task is left to the EPSRC in their future investigations.

Figure 36 shows comparisons between each RAG in terms of the share of their applicants with certain demographic characteristics. There is relatively little practical variation between the RAGs in terms of the share of their applicants who are in the over 56 age bracket, but some RAGs do see a considerably higher proportion of the applicants aged under 36 and a lower proportion of applicants from the 36-55 age bracket. This is particularly true of RAG 06, for which more than 25% of the applicants were aged 35 or younger.

RAGs 03, 06, and 09 stand out for having the highest share of applicants who are white, while RAG 03 also has the highest share of applicants who are male. Other RAGs show mostly similar proportions of ethnic minority or female applicants, though for females RAG 03 appears to have a noticeably lower share than the others. There is

[48] Applications can have multiple RAGs. In such cases, the RAGs are each given a weighting that reflects the proportion of the project that corresponds to it. For our analysis, we take the RAG for each application to be the RAG with the highest weighting. For ties, we keep all of the most highly weighted RAGs. Thus, our analysis is based on the primary RAG(s) for each application.

a little more variation between the RAGs by nationality. RAG 06 is unique in that a nearly 50% of its applicants have non-UK nationalities.

The awardee population shows similar variation within each demographic group and RAG as the applicant population. This is shown in Figure 37.

FIGURE 36

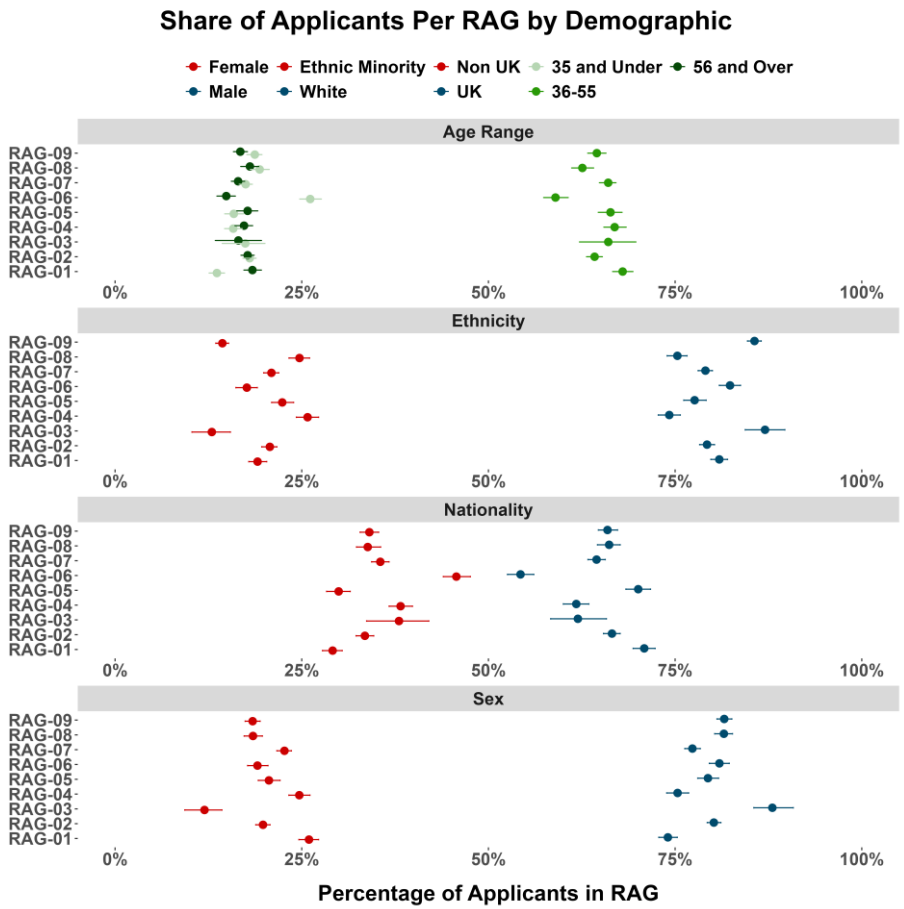
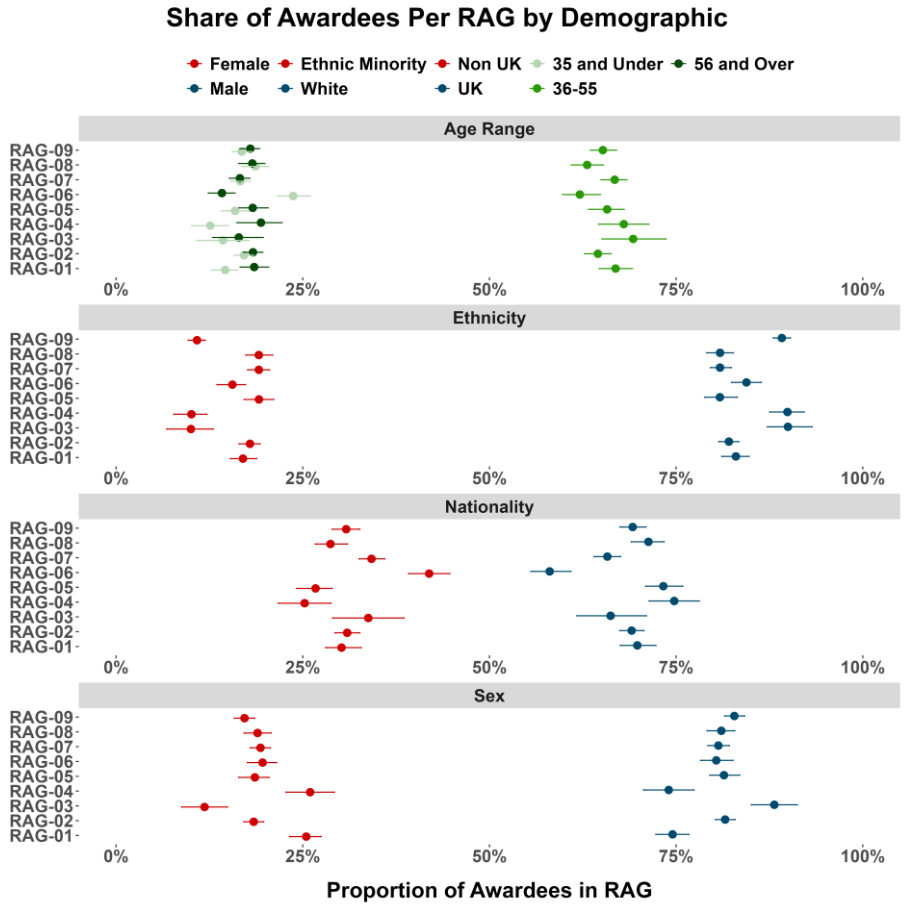
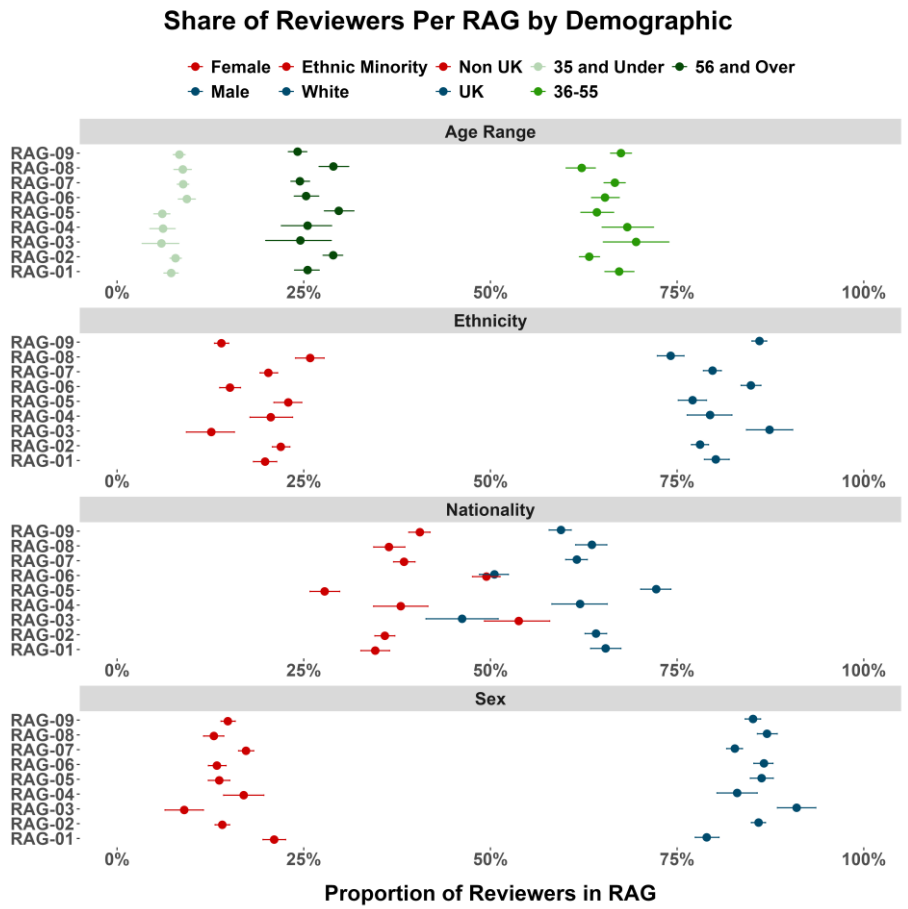


FIGURE 37



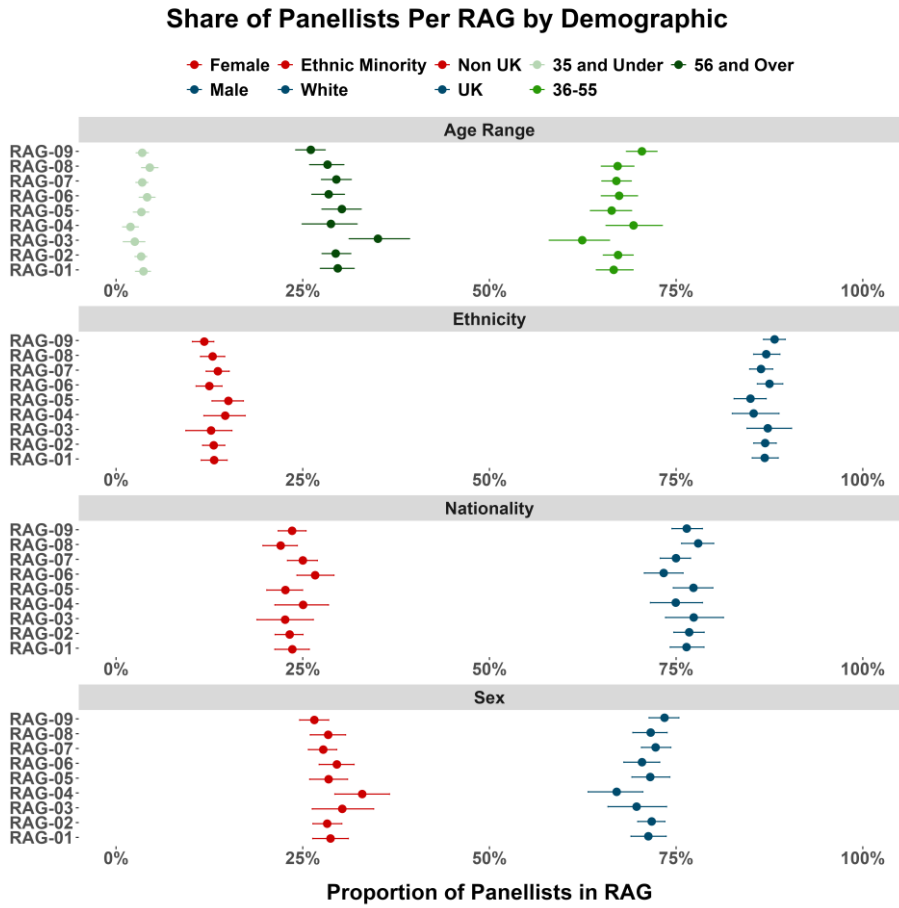
By comparison, reviewers are slightly more likely to be 56 or older than applicants and awardees, and also less likely to be 35 or younger, but they have similar patterns of variation in terms of ethnicity, sex, and nationality. Comparisons for the reviewer population are available in Figure 38.

FIGURE 38



The panellist population shows the lowest amount of variation in demographic shares across the RAGs. This is shown in Figure 39. Additionally, comparisons of the panellist population with the applicant population in each RAG reveals common patterns of misalignment between the demographic characteristics of both groups. In particular, comparing panellists to applicants, panellists are more likely to be 56 or older, while applicants are much more likely to be 35 or younger, panellists are somewhat more likely to be white than applicants in most RAGs, panellists are much more likely to be from the UK, and also somewhat more likely to be female. These differences may reflect a combination of differences in career stage (since senior researchers may be more often eligible to be panellists) and changing trends in the demographics of the research community more generally.

FIGURE 39



ACROSS BOTH THEME GROUP AND RESEARCH AREA GROUP HOW DO FUNDING OUTCOME, PANEL RANKING, AND AWARD VALUE DIFFER ACROSS APPLICATIONS?

We approach this question at the level of applications, and we explore the degree to which demographic characteristics may be differentially associated with each outcome of interest in different themes or RAGs. Our statistical model allows us to estimate a "baseline" association between the demographics and outcomes across all RAGs and themes, and then estimate additional effects for each demographic that are RAG or theme specific.

To do this, we fit a series of mixed effects regression models of the outcomes against sex, ethnicity, age, disability, nationality, funding opportunity type, and grant category with random intercepts by RAG or theme and random slopes for sex, ethnicity, and age that are allowed to vary with RAG or theme. Unlike our other models which included random intercepts for year and region, we did not include these here since adding the random slopes for demographics as well can cause results to fail to compute.

Here, we present the additional effects for each RAG and theme visually, since these are the parts of the outputs that are most directly associated with the question at hand.

Figure 40 presents the RAG-specific effects to the associations between demographics and award values. There is variation in the baseline award values for each RAG, represented by the Intercept values, and the demographic characteristics, although this appears to be concentrated in RAG-04 and RAG-03.

These two RAGs appear to be somewhat worse for the youngest applicants, somewhat better for the oldest applicants, worse for females, and somewhat worse for ethnic minorities in terms of award values requested and received relative to the other RAGs.

Figure 41 shows the additional effects made within each RAG for the associations between demographics and the likelihood of funding success (that is, the odds of being funded as opposed to not funded). Here, we see that variations for the baseline effect from RAGs as well as the added, RAG-specific effect for sex, ethnicity, and age are most pronounced in RAG-04 and RAG-03.

FIGURE 40: RANDOM EFFECTS BY RAG, AWARD VALUE

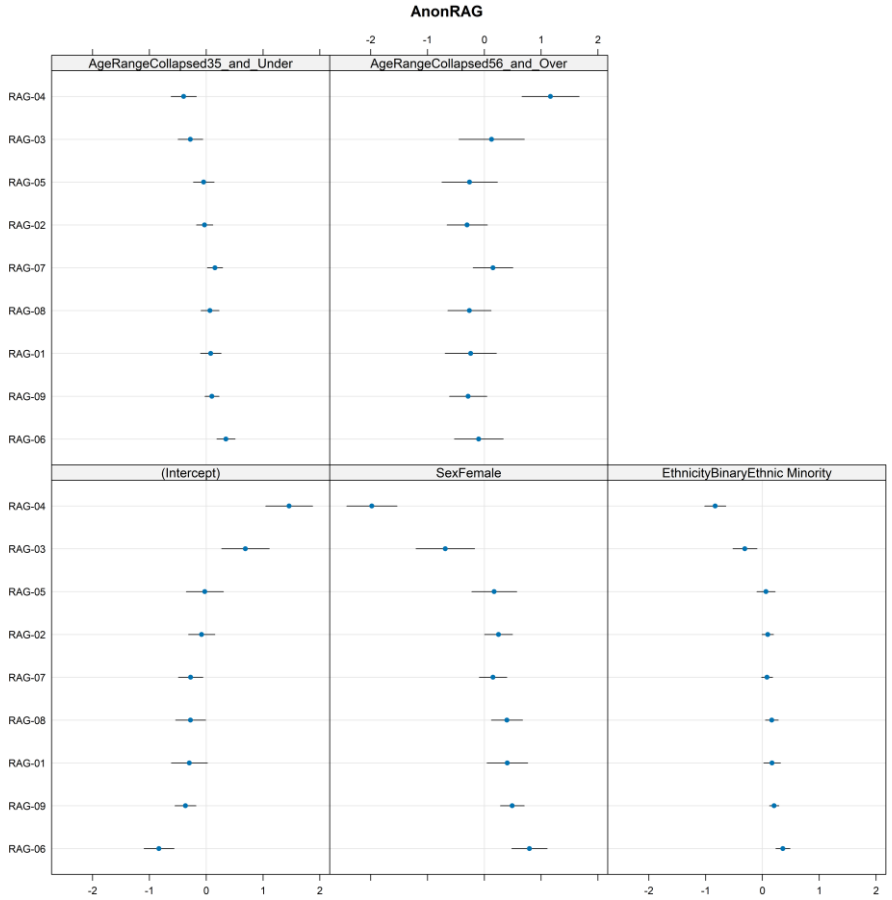
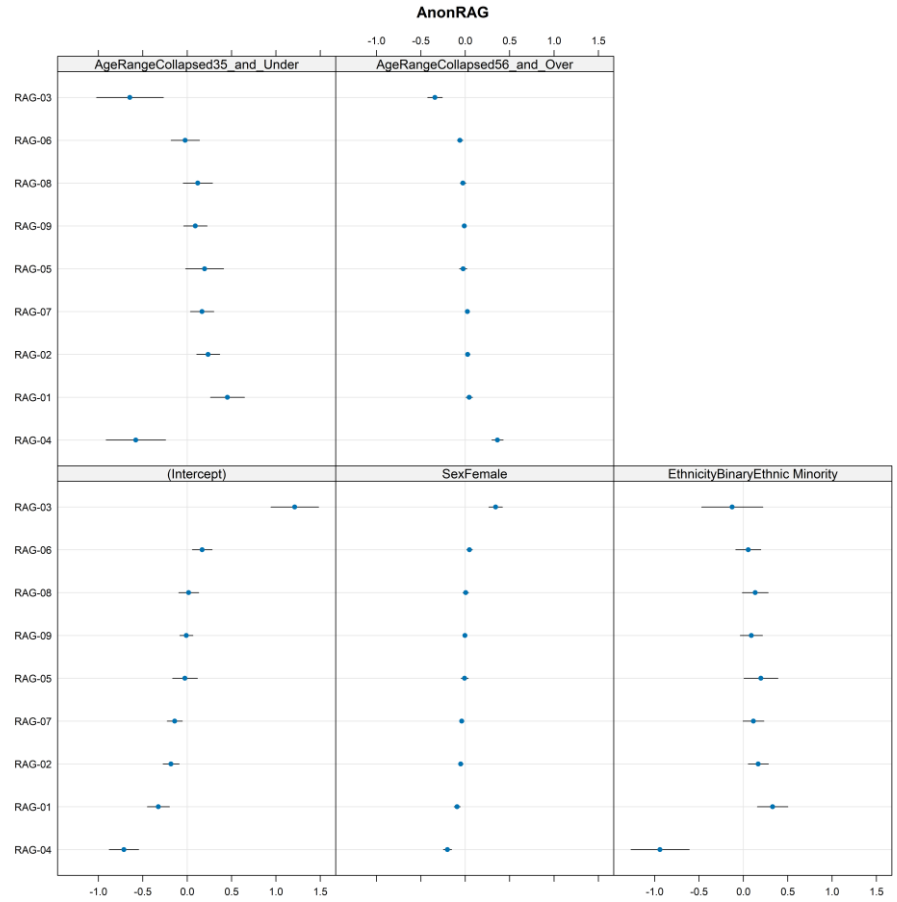


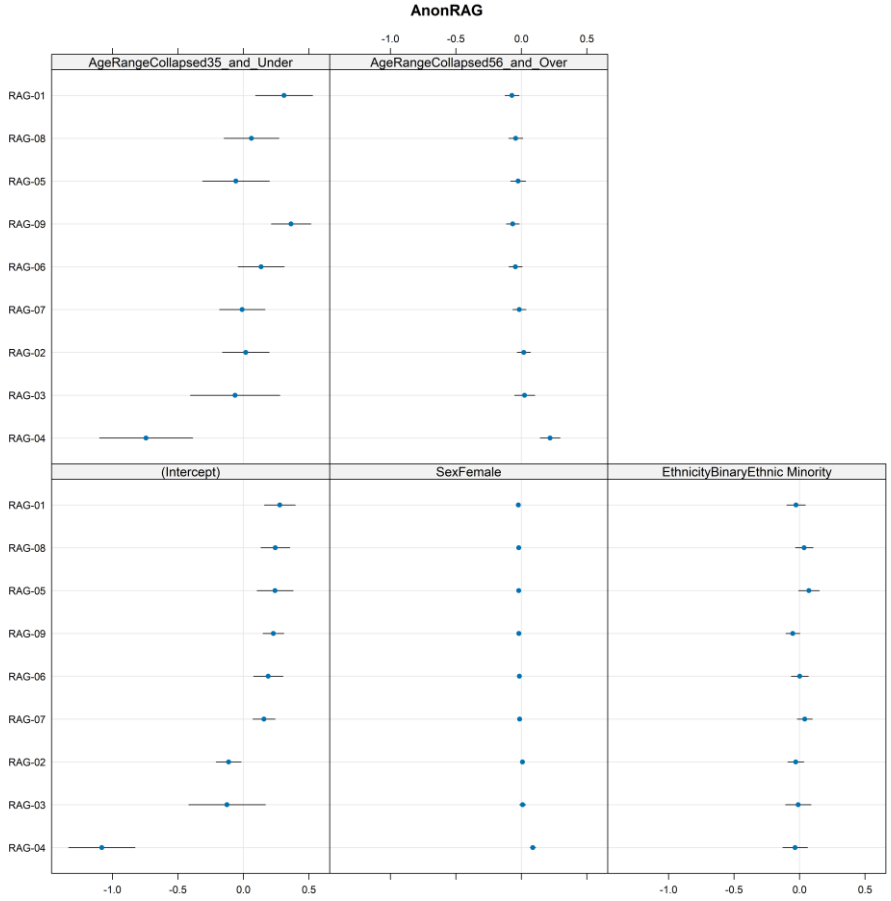
FIGURE 41: RANDOM EFFECTS BY RAG, FUNDING SUCCESS



For panel ranking, we consider the likelihood that an application is ranked in the top 25% of all applications during the final panel meeting of an application's journey through the peer review process. This is typically the point at which a final decision is made regarding the funding success of a project, but an application does not need to be in the top quartile to be funded.

Figure 42 shows the RAG-specific additional effects on the estimated associations between demographic characteristics and the likelihood of a top quartile ranking. We see that RAG-04 has a higher baseline effect than the other RAGs, followed by RAG-03, and both of these have negative effects on the odds of a top panel ranking. This relative pattern is also apparent for age, but the direction of the adjustment varies. For the youngest age group, RAG-04 is associated with the lowest likelihood of a top quartile panel ranking, whereas it is only marginally associated with the highest likelihood for the oldest age group. Variation by sex and ethnicity is practically small or non-existent.

FIGURE 42: RANDOM EFFECTS BY RAG, TOP-QUARTILE PANEL RANKING



We omit the results for theme from this report, as the same general patterns are visible there as well: there is some variation in baseline outcomes between the themes, but between most of them there is not much variation in terms of their relationships to demographic characteristics, particularly sex and ethnicity, for which variations are weak or practically very small. However, the full results are available in the GitHub repository.

One possible interpretation of these results is that different RAGs and themes have structurally different funding needs and norms, as well as different reviewing and assessment cultures that are prevalent in different research communities. Future research may wish to dig into the reasons behind variation between the different research areas and themes, potentially with a qualitative lens that can more directly reflect the experiences of those within them.

CONCLUSION AND DISCUSSION

This study considered a wide range of questions relating to potential bias in the EPSRC's peer review process. The pre-registered analysis found significant associations between ethnicity and grant success, with applications led by white principal investigators more likely to be successful than those led by ethnic minority PIs.

There is some evidence that the intersections of ethnicity with nationality and ethnicity with age matter, as white applicants with UK nationality have higher predicted success rates than other white applicants—but for ethnic minority applicants there is no substantial difference—and ethnic minorities under the age of 36 see funding success levels that are not far below their white colleagues in the same age group (older groups see larger differences). However, intersectional analysis is limited by small populations for some categories.

Similar to funding success, both our pre-registered analysis further exploratory analysis found that applications led by ethnic minorities were less likely to receive a top-quartile panel ranking than applications led by white PIs.

Applications for research grants that were led by ethnic minority investigators tended to request and receive lower award values than applications led by white investigators when they were successful. This was not true for fellowship applications, where we did not find a significant association between ethnicity and the award values requested and received.

Negative associations between outcomes and being an ethnic minority were also visible throughout the exploratory analysis of the review process, where reviewer scores tended to be lower for ethnic minority applicants. Additionally, there is some evidence of homophily in reviewer scores, where reviewers from the same ethnic background score applicants who share their ethnicity more highly.

The findings for ethnicity in terms of outcomes are in some respects mirrored in more descriptive views of how demographic groups are represented throughout the EPSRC data. For instance, comparisons of the EPSRC data with HESA data, which we take to be the closest available comparator for determining proportional representation, indicate small but persistent under-representation of ethnic minority applicants for grant funding.

Moreover, applications led by ethnic minorities account for a higher share of applications deemed "unfundable" at some point in the review process relative to their share of the "funded" and "fundable-but-not-funded" applications to both research grants and fellowships.

The picture for sex is more mixed. Applications led by female PIs were slightly more likely to receive funding than applications led by male PIs overall, although there is an important nuance to this that is revealed when considering fellowship and research grant applications separately: the odds of funding success for fellowships is far higher when applications are led by females than by males. For research grant applications, there is no association between funding success and the sex of the lead applicant.

Our pre-registered analysis did not find significant associations between sex and the odds of a top panel ranking, although analysis from the Alan Turing Institute did find a small association in favour of females during interview panels. This was particularly true when at least one of the members of the panel was female. Nevertheless, these findings should be considered in the context of a small but significant finding from the ATI that reviewer scores tended to be higher for males than females.

When applications for research grants led by females were successful, they tended to request and receive award values lower than when applications led by males were successful. The same was not true for fellowship applications, where there was no significant sex-based association. This aligns with our finding that suggests a disparity in favour of females in funding success for fellowships, but not for research grants.

Representationally, female applicants are generally representative of the HESA population, although it is worth noting that there has been very little convergence between the sexes in both the EPSRC data and HESA data; males still account for the vast majority of both. Females also make up a larger share of the funded population than they do of the fundable or not fundable populations, which is in line with the models estimating somewhat higher funding odds for women.

Applications led by investigators with non-UK nationality tended to receive worse outcomes than those led by UK applicants. This manifested in terms of lower odds of

funding success, lower odds of a top panel ranking, and lower award values for successful applications to research grants, although there was not a difference in award values for fellowships.

Results for disability were frequently insignificant in our models, possibly due to the very small sample size for people with known or disclosed disabilities. Nevertheless, we did find a significant negative association between having a disability and requested-and-received award values for fellowships, although the precision of that estimate is low, ranging from £20,000 to £425,000.

Age featured frequently in the list of demographic characteristics that were significantly associated with outcomes and differential representation. The under 36 and over 55 age brackets were less likely than the 36-55 group to have successful funding applications when they led them, and the same was true for receiving a top-quartile panel ranking. However, once again, there is evidence that the grant category matters, at least for funding success.

For award value, applications led by investigators from the youngest age bracket requested and received substantially less than when they were led by investigators aged 36-55. In contrast to the findings for success rates and panel rankings, though, the over 55 age bracket requested and received higher award values than the 36-55 group. This was true both for fellowship and research grant applications.

There is clear variation in the baseline outcomes (odds of funding success, odds of a top panel ranking, and award value) across RAGs and themes, which likely reflects their different structural needs and norms. In terms of variations in association between demographics and the outcomes, there is clear evidence that age is differentially associated, but the variation is weaker when looking at sex and ethnicity for most RAGs and themes. Still there are exceptional cases where demographics do present differential outcomes. Substantive interpretation of those cases is left to EPSRC, as non-anonymised groupings were not available for this analysis.

We considered whether there was evidence of prior success in applications being associated with better outcomes in subsequent applications. While there are considerable limitations to our ability to answer this question with the data, we did find preliminary evidence that such an association exists.

Finally, we considered broadly whether there was evidence of institutional concentration in terms of funding value and success rates. In the case of award value, there was evidence of moderate-to-large concentration of award value, although we can't control for things like institutional size to adjust our measures.

On the other hand, we found that, although there are a few exceptions, success rates for institutions tend to converge towards the overall success rate across all institutions as their number of applications increases. This finding is consistent with what we might expect if there was not bias in favour or against certain institutions in terms of the likelihood of success, but once again this finding is inconclusive.

Triangulating these findings holistically within the context of survey analysis conducted by the Alan Turing Institute for this project[49], we can glean some insights that help to contextualise these findings. The survey was intended to explore perceptions and experiences of bias amongst the community of EPSRC applicants. Qualitative analysis of this type allows for the exploration of different dimensions of this research topic that quantitative analysis alone cannot touch.

Firstly, the survey found that male respondents were more likely to perceive bias in the EPSRC review process than females. It is impossible to say precisely what is driving this difference, but results from our analysis of outcomes indicate that funding success for fellowships is significantly higher for females than for males. It could be the case that, although award value tends to be higher for males who are successful at research grants, perceptions of bias are more sensitive to the binary "pass or fail" outcomes. Another plausible explanation is that the socio-cultural expectations among sexes differ, leading to the gaps observed in the survey.

[49] The full analysis of this survey is detailed in section 2 of the ATI report, available in the GitHub repository at: https://github.com/alan-turing-institute/equity-in-grant-funding

Secondly, besides age, ethnicity was the most consistently significant demographic characteristic in our analyses, with the overall picture generally showing poorer outcomes for ethnic minorities. This is consistent with the survey responses, where ethnic minority respondents were the most likely to report perceived bias in the review process. However, it is important to recognise that there are differences between the groups that are classed as ethnic minorities—although our analyses focus on the collapsed categories for modelling reasons, it is clear that perceptions of bias are not uniform.

Thirdly, the leading stated reason for perceptions of bias was institution. This is an interesting finding because our analysis of institution did not clearly indicate systematic disparities between institutions in terms of their success rates after accounting for the number of applications they made, but we did find that there were fairly high levels of concentration for award value. Still, our findings are only preliminary and indicative rather than final, and there are likely to be complex nuances to the story about perceived bias and institutions.

Importantly, with the exception of our findings for ethnicity, the juxtaposition of the survey findings with the results of our analyses of outcomes suggests that perceptions of bias are more complex than we can adequately account for based solely on the observed differences in outcomes. Indeed, quantitative analysis of outcomes are unlikely to be able to explain the complex perceptions people hold about the process, and further exploration using other methods would be a welcome expansion of the present research.

This is to be expected; individual experiences are always much more varied and nuanced than a relatively high-level, abstracted dataset can tell. The lesson for further research is that, while there is clear value in quantitative, observational research, a deeper understanding of potential bias in the review process and how it may affect perceptions, and behaviour would do well to incorporate other methods. These can be experimental and related to studying the effects of specific policy changes on the review process, or they can be qualitative, such as focus groups of applicants and panellists to get a deeper understanding of why certain perceptions exist.

Particularly in light of the suggestive finding that panel composition can be associated with differential outcomes by sex (ie, female applicants receive better outcomes when the panel consists of at least one female), the space for fruitful experimental work on panel composition is expansive. Future research may help determine whether policies like the Mixed Gender Panel Policy[50] should be expanded to target other protected characteristics.

Furthermore, while the present project helps progress our understanding of the relationships between various outcomes and protected characteristics and does so along new dimensions not covered by prior research, it also has some limitations that should be kept in mind. Among the most important of these is that our methodologies cannot establish causal relationships between our variables of interest. While it would be natural to interpret discrepancies in outcomes or representation as the result of bias, particularly given the widespread perceptions of bias in the sector, our methods cannot support that explanation over others.

An additional limitation is that we could not control for application quality, which is potentially the most important determinant of outcomes. Reviewer scores are perhaps the closest approximation to a measure of quality, but these are themselves subject to investigation of bias in the review process, which partially undermines their usefulness for this purpose. Future research may consider experimental designs that can more directly compare applications of identical or similar quality.

[50] See: Evolving and upholding fairness in peer review – UKRI

From past to present...

The image of the wheatsheaf first appeared in our original seal. Being the end product of the harvesting and bundling of wheat, it was a pictorial way of expressing the gathering and analysis of data: the foundations of statistical work.

It also implied that statistical practice comprises more than the collection of data: it consists of active interpretation and application as well (threshed for others, if the rural analogy is sustained). Rigorous data gathering is still at the heart of modern statistics, but as statisticians we also interpret, explain and present the data we collect.