ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

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

Executive Summary

UKRI Engineering and Physical Sciences Research Council

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

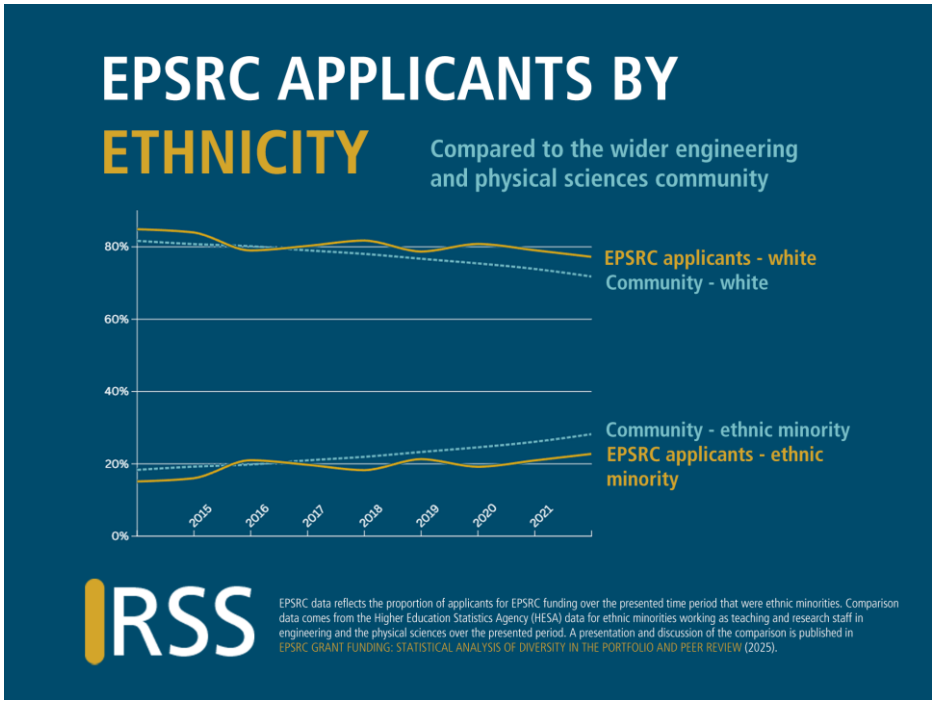
Potential sources of bias and disparity in the application and funding process can come from a variety of sources. The focus of our research was split between three broad conceptual "stages" of the process that each have the potential to be one such source, starting with the representativeness of different 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.

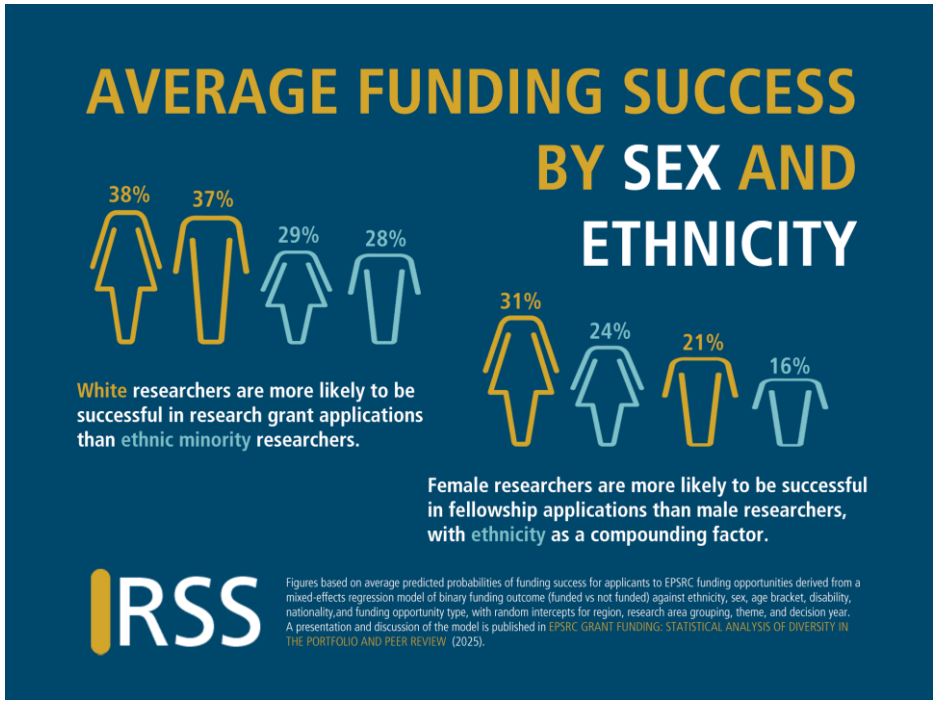
Additionally, as part of this study, the Alan Turing Institute conducted a research community survey into perceptions of bias in the EPSRC review and award processes. Higher perceptions of bias in the EPSRC peer review process were reported by male respondents, by ethnic minority researchers and by individuals 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).

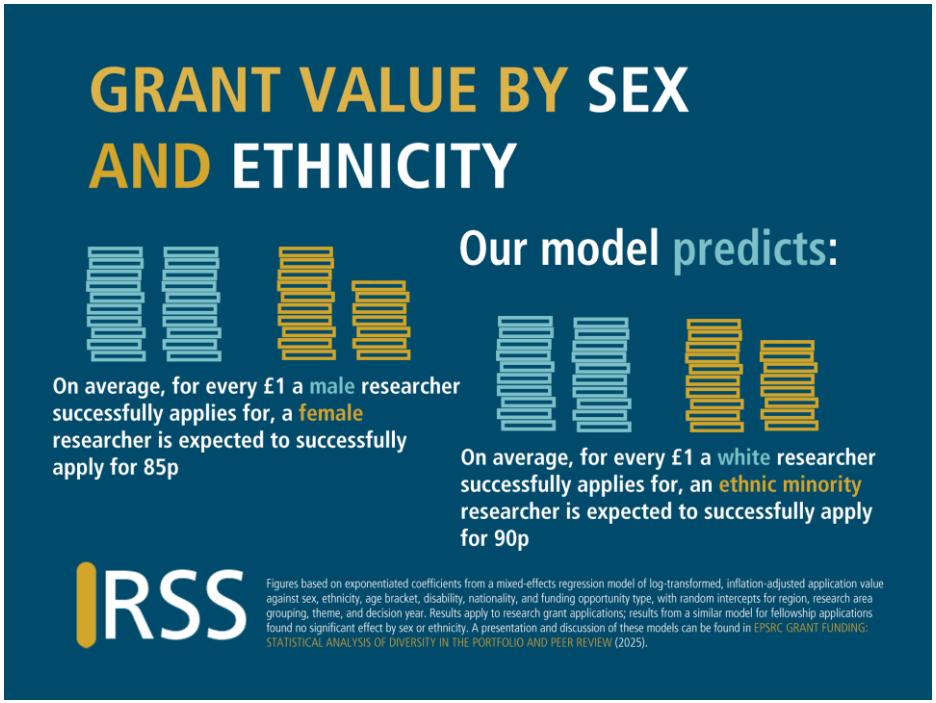
In a parallel descriptive analysis, we focussed on the characteristics which differentiate those applications which were funded, those which EPSRC considered to be fundable, but which were ultimately unsuccessful and those which were considered to be not fundable. This analysis, which separates fellowship 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%]

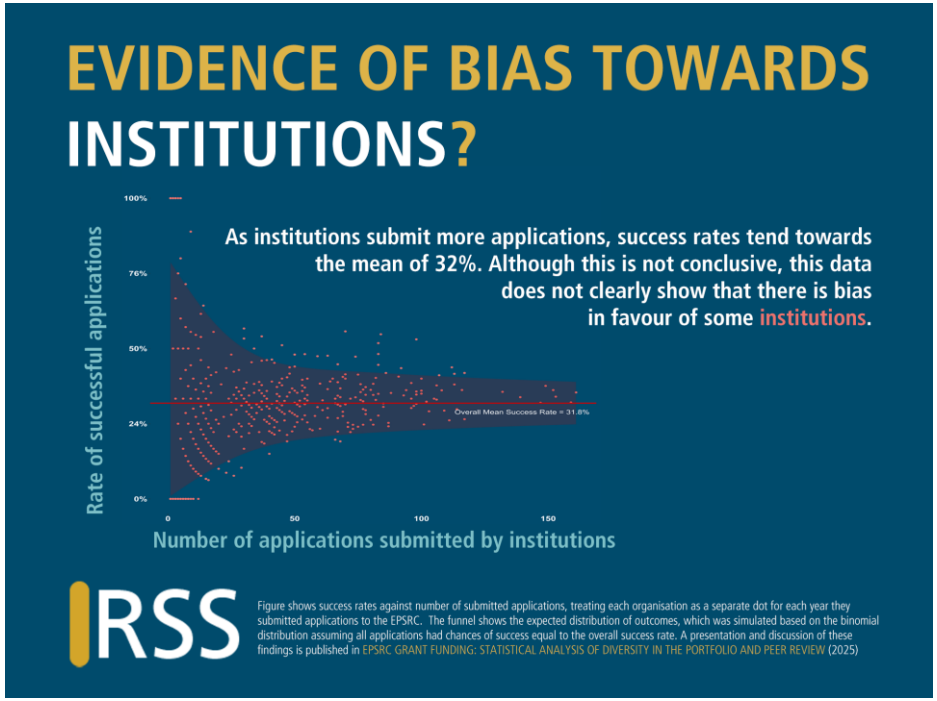
of the value of those in the 36-55 age range while those for PIs in the >55 age range estimated to be 129.5% [119.9%,139.8%] of the value of those in the 36-55 age range.

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

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.