



Supporting Online Material for

Race, Ethnicity, and NIH Research Awards

Donna K. Ginther,* Walter T. Schaffer, Joshua Schnell,
Beth Masimore, Faye Liu, Laurel L. Haak, Raynard Kington

*To whom correspondence should be addressed. E-mail: dginther@ku.edu

Published 19 August 2011, *Science* **333**, 1015 (2011)
DOI: 10.1126/science.1196783

This PDF file includes:

Materials and Methods
Figs. S1 to S4
Tables S1 to S15
References

Supplemental Information for Ginther et al., 2011

Materials and Methods

Details of Data Linkage:

In this supplement we provide a description of how the data sets were developed, an explanation of methods used in the analysis, and additional estimation results.

We created the analysis samples by combining data elements from several different sources. The initial sample dataset was derived from IMPAC II and stored in a Microsoft SQL Server 2005 database. These data were then matched to US Department of Education IPEDS organizational information, US National Science Foundation Doctoral Record File (DRF) PhD degree information, DRF and AAMC Faculty Roster race/ethnicity information, MEDLINE publication data, and Web of Science citation information to create the analysis sample (1).

We matched IPEDS data to IMPAC II data using an algorithm that compared standardized name fields, institutional identifiers such as DUNS, EIN numbers, and Federal Interagency Committee on Education (FICE) codes, and institution locations. We manually reviewed a random sample of these records to determine the quality of the match and did not find any mismatches. For the 100 institutions not linked to IPEDS by our algorithm, we matched these records using a manual process. In total, we matched 487 organizations within IMPAC II to records in the IPEDS database. These 487 organizations were the sponsoring organization on 67.2% of the applications in our full dataset.

Since 1993, NIH has matched DRF person records with IMPAC II profiles going back to 1958. To link these records, NIH employed a five step matching process that uses an individual's name, address and research expertise to determine a similarity score. When this similarity score went above a threshold, the records in the two databases were considered matched. Data from the DRF were used to confirm degree-related information in IMPAC II and to add information about the educational history of the applicant pool, such as BA-granting, and PhD-degree granting institutions. In the full data set, the above process provided a strong match for 55.1% of applicants in the dataset, comprising 56.8% of the applications.

We developed a set of algorithms to determine the publication history of applicants at the time of application. These algorithms combined a series of name-based matches with email and ZIP code matches. We tested these algorithms using an initial sample of 800 applicants from all racial/ethnic groups. Approximately 13,000 candidate papers were identified for this sample and partitioned into High Confidence matches (exact name matches combined with at least one additional attribute) and Low Confidence matches based on name matching standards and name frequency values derived using the IMPAC II database as a reference set. The Low Confidence group matches were sub-sampled at intervals based on author name frequency, and a name frequency threshold was identified, below which the publication matching precision was at least 81%. A review of approximately 1,500 publications from the High Confidence group whose author names were common determined the precision for this group to be 99% (false positive error approximately 1%). These same algorithms were then applied to the full dataset of 79,218 applicants, with 317,326 publications identified in the Low Confidence group for author names below the frequency threshold representing at least 81% precision, and 273,286 identified in the High Confidence group. Names that are shorter and more common (e.g. Asian names) are more likely to overstate publications since methodologies to match authors rely on shorter name strings. However, our conservative matching approach required more attributes than just name for inclusion (email and ZIP code) and we eliminated the most frequent names for which our matching precision was below 81%. We discuss the impact of measurement error on our estimates below.

Variable Derivation:

The following provides additional information about how selected variables were derived in the analysis sample.

1. Race/Ethnicity: An applicant's race/ethnicity was derived from IMPAC II with additional information provided from the DRF and/or the AAMC Faculty Roster if IMPAC II data were not complete.
2. Nativity and Citizenship: An applicant's citizenship was derived from the DRF, and the citizenship variables refer to citizenship status at the time of receipt of the doctorate. Nativity was derived from the DRF citizenship variable along with information from IMPAC II. If an NIH applicant was not matched to the DRF, but received all of their degrees from non-US institutions, they were classified as a non-citizen.
3. Degree Information: An applicant's degree type was derived from IMPAC II with additional information provided from the DRF if IMPAC II data were not complete.
4. Organization Type: Organization type was derived from IMPAC II and used to broadly classify the type of organization sponsoring the applicant's submission.
5. Carnegie Classification: The classification of the applicant's sponsoring institution was provided through IPEDS and linked through the sponsoring organization information from IMPAC II.
6. Prior Training Support: We used IMPAC II to determine if an applicant was the principal investigator on any fellowship (F) or career (K) awards, or if the applicant was appointed as a trainee on a training grant (T), prior to the first submission for a grant.
7. NIH Institutional Funding: To determine the NIH funding rank of an organization, we averaged the annual grant support received for each sponsoring institution from FY2000 to FY2006 and ranked them in descending order of the total grant dollars received. These institutions were then categorized into 4 categories, Top 30, 31 to 100, 100 to 200, and 200+.
8. Total Grant Applications and Awards: We counted the number of R01 applications and awards by individual investigators going back to fiscal year 1980 to get an accurate count of grants submitted over an investigator's career. Of course some individuals were still active researchers and may have submitted grants past FY2006.

Missing data and/or data conflicts:

We have used IMPAC II to create a person-record for the individual applicant that includes time-invariant demographic characteristics. However, individual applicants can and do change their gender, race/ethnicity, and birth date in IMPAC II. Where there were multiple observations in these fields, we used the most-frequently reported race and ethnicity, sex, and birth date. We dropped 5,853 individuals with missing age because they had missing information for multiple variables such as age, race, and sex.

When we had missing information for variables other than age, we included dummy variables that took on a value equal to one in the specification to control for that fact. Our models included dummy variables to account for missing information in the sample on race (9% missing), citizenship status (15%), PhD field (27%), rank of PhD institution (36%), human subjects (.03%), and publications (16%).

Sample selection:

Our sample included R01 grant applications associated with new projects (Type 1) submitted between FY2000 and FY2006. We did not evaluate continuing R01 grants (Type 2). Grant applications in our sample could have been resubmitted as revised applications multiple times. Information about the application and its review was derived from the last funded or unfunded application submitted. We restricted this sample to individuals with a PhD residing in the United States. This sample included

83,188 observations with non-missing data for the explanatory variables. This sample was further limited to US citizens for specific aspects of the analysis.

Table S1 provides information on the number, percentage, and award probability of all RPG applications for FY2000 – FY2006, as well as all R01 applications, and PhDs residing in the US, broken down by race and ethnicity. We examined whether award probabilities differed between Whites and other race/ethnicities and found that Asians, Blacks, Hispanics, and those with unknown race were significantly less likely to receive an R01 award than Whites.

Award Probabilities by Priority Score:

We examined the relationship between race, priority score, and award probability in our funding sample. **Table S2** shows the distribution of priority scores by race and ethnicity in the analysis sample. We divided the priority score into 25 point bins across the full 100 to 500 scoring range and graphed the award probability in each interval by race in **Figure S1**. These results indicated that NIH funding was largely determined by priority scores. The lower the priority score, the higher the likelihood of being funded for all race and ethnic groups. As priority scores exceed 200, award probabilities drop close to zero and equal zero past 400.

Success Rates by Race 2006 - 2010:

In 2009 the NIH review process substantially changed. Consequently, our sample was adjusted to include only Type 1 applications and their revisions that were submitted between FY 2000 and FY 2006 and observed through 2008. To evaluate whether funding rates were stable across race/ethnicity groups and time, we examined success rates by race. The NIH tracks success rates for each fiscal year. NIH success rates are defined as the percentage of reviewed grant applications that receive funding. They are computed on a fiscal year basis and include applications that are peer reviewed and either scored or unscored by an Initial Review Group. The NIH success rate differs from the funding rate used in this analysis because a grant could be unfunded in one fiscal year and then be resubmitted and funded in a subsequent fiscal year. The success rate would include the grant application in the denominator in the first year and in both the numerator and denominator in the second fiscal year. Applications submitted multiple times in a single year include the grant in the success rate denominator only once. In our funding rate analysis, this grant application would be counted in the numerator and denominator only in the last year it was submitted.

Table S3 shows the distribution of grants submitted by race/ethnic group each year. We examined race/ethnicity differences in relative success rates from FY 2000 to FY 2008 to determine whether success rates fluctuated across race and ethnicity categories over the sample time frame. The relative success rate is the success rate for a particular race divided by the success rate for all applicants in a given year. **Figure S2** shows that relative R01 success rates by race and ethnicity were fairly stable over time.

Citizenship by Race/Ethnicity:

We examined the percentage of citizens by race/ethnicity at the time of PhD receipt. **Figure S3** shows that 87% of Asian applications were from non-US citizen investigators. 45% of Black and 56% of Hispanic applications were from non-citizen investigators, as were 25% of White applications. For that portion of each group (15 – 22%) unable to be classified, we assumed they were not citizens if they did not appear in the DRF (a census of PhDs in the US).

Multivariate Regressions:

We used probit models to test the effect of various investigator and organization characteristics on award probability. We used heteroskedasticity-robust standard errors that were clustered on the individual applicant in order to adjust for the fact that applicants could submit more than one proposal in our sample (2). P-values for the race variables were corrected for multiple comparisons using the Bonferroni method.

The Bonferroni method adjusts the p-values to reduce the likelihood of rejecting the null hypothesis when it is true. Coefficients were transformed to be marginal effects and can be interpreted as the change in probability resulting from an infinitesimal change in the independent variable. When the independent variable is a dummy variable (as was the case with many of our covariates), the coefficient reports the percentage point change in the probability given that the dummy variable changes from zero to one where all other covariates were evaluated at the mean of the predictors.

The probit models were specified as follows. Model 1 included gender, race/ethnicity, age, naturalized citizen, non-citizen, and foreign PhD degree variables. This baseline model provided information on the association between race and ethnicity and funding without any other explanatory variables. Model 2 included the covariates from Model 1 plus degree type, previous NIH training support, PhD field, and the NIH funding rank of the PhD institution. Model 2 variables controlled for the effect of NIH training on funding probability. To the extent that NIH training provides good preparation for a research career, we expected these variables to have a positive association with receiving an award. Model 3 included variables from Model 2 and employer characteristics such as organization type (university, hospital, research institute), the Carnegie ranking of higher education institutions, the NIH funding rank, and the geographical region of the employer. We expected that the more research-intensive the institution, the higher the likelihood of receiving funding. Model 4 included Model 3 variables and indicator variables for NIH funding experience and environment including the applicants' prior receipt of NIH grants and NIH review committee experience, the NIH institute receiving the application, the fiscal year of grant award, and whether the grant involved human subjects. We included these controls to adjust for differences in funding rates across NIH institutes, annual differences in success rates, and differences in experience with the NIH funding process. Model 5 included all variables from Model 4 and controlled for investigator productivity at the time of application, such as publication and citation counts, median and maximum journal impact factor, and percentage of publications where the grant applicant was a single author, first author and last author. **Table S4** shows the covariates included in each probit model. **Table S5** shows our main results.

Goodness of Fit:

Probit models do not have a single measure of goodness of fit like the R^2 in the linear regression model. We evaluated the goodness of fit of the probit models by comparing predicted outcomes with actual outcomes in the data. Assuming a symmetric loss function, the fitted outcome, $\hat{y} = 1$, if the predicted probability was greater than .5, and $\hat{y} = 0$ if the predicted probability was less than or equal to .5. We compared the fitted outcomes of award probability from the model, \hat{y} , to the observed outcomes in the data. Our models fit the data well, correctly predicting R01 award outcomes for 71-72% of the observations in the sample.

Measurement Error in Probit Models:

Our data matching process may have underreported publication and related information for a given applicant. This measurement error may have biased the coefficients in the model. While there were fewer results for the implications of measurement error in probit and related models, this research suggested that, at least qualitatively, the results were similar in sign and magnitude to those found in linear models (3, 4). The sign and size of the bias on the mismeasured variables depended on the relative magnitude of the average underreporting and the variance of the underreporting. Hence it was not clear if the coefficients on publications and related information would be under- or overestimated. The sign and magnitude of the bias on other correctly measured variables depended on the covariance between publications, citations and other variables included in the specification. In practice, previous researchers have found that measurement error bias is often not particularly large. Moreover, the bias is typically smaller than the omitted variable bias had these variables been left out (5,6). Given the relevance of prior

research for R01 awards it is important to include these variables in the analysis despite the presence of measurement error.

Robustness Tests:

We conducted a series of robustness tests to examine whether our basic results were sensitive to variables included in the models.

- (1) We examined whether there were race/ethnicity differences in award probabilities for grant applications submitted the first time and upon resubmission (**Tables S6 and S7**). **Table S6** shows the number of grants submitted and awarded and the percentage of grants awarded by the number of times submitted. Award probabilities increase for all race/ethnic groups with the number of times a grant is submitted. **Table S7** shows the effect of controlling for resubmissions on race/ethnicity coefficients in the probit model. The first column presents the results of Model 5 from **Table S5**. In the second column we added dummy variables for the number of times a grant was submitted (once, twice, or three times or more). Controlling for resubmissions in Model 5 did not change the estimated results for Blacks and Asians. In the third column, we estimated Model 5 for those grants submitted only once. We still found significant differences for Asians and Blacks in R01 award probability compared to Whites, but the estimated effects were about one percentage point smaller. The fourth column shows the race/ethnicity differences for grants submitted twice. The estimated effects for Asians were very similar to those found in Model 5 (first column). However, Blacks who submitted grants twice were 13.5 percentage points less likely to receive an R01 award compared to Whites who submitted twice. A much smaller proportion of proposals were submitted three or more times (**Table S6**). In this case the difference between Black and White award probabilities dropped by 10 percentage points and was no longer statistically significant. However, Asians were still almost 4 percentage points less likely to receive R01 awards upon the third submission than Whites. This suggested that race/ethnicity differences for Blacks and Asians relative to Whites in award probabilities were partially explained by the probability of submitting a revised grant and lower award probabilities upon resubmission for these two groups.
- (2) Only US Citizens and Permanent Residents qualify for NIH training, so award probabilities may have been affected by citizenship status. To explore this possibility we limited the sample to US citizens only. **Figure S3** shows the percent of applications by citizenship status at time of PhD. **Table S8** shows the distribution of training by race/ethnicity. **Table S9** shows the effect of race and selected coefficients on the award probability for a sample limited to US citizens. In contrast to the results in **Table S5**, **Table S9** shows that the Asian coefficient was no longer statistically significant. This indicated that the difference in award probability for Asians may be driven by non-citizen Asians who did not have access to formal NIH training programs.

Figure 2 and **Table S10** examined the effect of training on R01 award probabilities. In **Figure 2A** we limited the sample to US Citizens and Permanent residents and regressed R01 award on a dummy variable that equaled one if the application received F or T training; for applicants, we regressed whether the person has ever received an R01 award on training. We performed separate regressions for the full sample and each race. Each bar in **Figure 2A** shows the effect of receiving training relative to not receiving training by race/ethnicity. F or T training significantly increased the probability of an NIH applicant ever receiving an R01 award. In **Figure 2B** we limited the sample to those who received F or T training only and regressed R01 awards (for applications) and ever received an R01 award (for applicants) on dummy variables for race/ethnicity with White as the omitted race category. Each bar in **Figure 2B** shows the effect of race on R01 award probability relative to Whites with training. We also estimated the effect of

race on R01 awards using Model 5 for those that received training only and found that applications from Blacks were -0.115, ($p < .001$) less likely to receive funding and Black applicants were -0.163, ($p < .001$) less likely to ever receive an R01 award compared to Whites after controlling for the full set of covariates. Despite the positive association between training and R01 awards within race groups (**Figure 2A**), **Figure 2B** shows that training did not narrow differences in R01 awards across race/ethnic groups.

- (3) We examined how grant applications by race/ethnicity differed with respect to the observable characteristics of NIH training, prior research experience, and affiliation variables for the full sample and the unscored sample. **Table S11** and **Table S12** showed that there are significant differences between Blacks, Asians, Hispanics, and Whites in NIH training, prior NIH grants, and researcher productivity in both samples. **Table S13** shows the number of applications by race/ethnicity that had prior NIH training, prior NIH grants, NIH review committee experience, employment at research organizations and organizations ranked 1-30 and 31-100 in NIH funding, and had citations above the median.
- (4) We tested whether the estimated effects in **Figure 3** differed significantly by race by interacting all of the variables in Model 5 with race and testing the joint significance of selected interaction terms. For example, to test whether estimated effects of research organization, 1-30 NIH funded organization, 31-100 NIH funded organization, prior NIH award, NIH review committee service, last authored publications, and >24 citations for Blacks differed from the full sample, we interacted the Black variable with all of the covariates in Model 5, and then tested whether these interaction terms for the subset of variables in **Figure 3** were jointly significantly different from zero. The heteroskedasticity-robust standard errors in this regression were clustered on the individual applicant. We failed to reject the null hypothesis that the interaction terms were equal to zero, indicating no significant differences in these coefficients by race.

By fully interacting model variables with race to perform the hypothesis test, we assumed that the residual variation for all race and ethnic groups was the same. This assumption was necessary because probit coefficients are scaled by unobserved residual variation (7, 8). Coefficients from linear probability estimates are not scaled by residual variation. As a robustness check, we performed the hypothesis test using a linear probability model and found that coefficients for research organization, 1-30 NIH funded organization, 31-100 NIH funded organization, prior NIH award, NIH review committee service, last authored publications, and >24 citations for Whites ($p < .01$) and Asians ($p < .05$) differed significantly from the full sample.

- (5) We estimated Model 5 for the full sample and by race for the probability of receiving a priority score. We tested whether the coefficients for Blacks, Asians, Hispanics, and Whites were jointly significantly different from those for the full sample by fully interacting them with the race variables and testing the joint hypothesis that the race interaction terms were significantly different from zero (as described above). We found that the coefficients for Blacks were significantly different from the full sample for the priority score model (**Table S14**) using both the probit and linear probability models. **Figure S4** shows how the marginal effects of selected variables on the probability of receiving a priority score differed between the full sample and Blacks.
- (6) Prior to the enhanced review criteria announced in 2009 (9), a large proportion of grant applications were streamlined during the review process and did not receive priority scores. We found significant race and ethnicity differences in the likelihood of having an unscored application. 40% of applications from Whites were unscored, while grant applications from Hispanics (42%, $p < .01$), Asians (46%, $p < .001$), and Blacks (59%, $p < .001$) were more likely to be

unscored. If the analysis is restricted to US citizens, likelihood of an unscored application shows little change for Blacks (60% $p < .001$) and Hispanics (44% $p < .01$), whereas the proportion of unscored applications for Asians is no longer different from Whites. Thus, a significant number of the unscored Asian applications were submitted by investigators who did not receive a PhD from a US institution.

- (7) To test whether unscored applications explained the observed race/ethnicity differences in R01 award probability, we estimated the probit models after eliminating all unscored applications from the sample, leaving us with 48,226 scored proposals (**Table S15**). While the marginal effect for Black applications in Model 5 was reduced significantly from -10.4 to -6.5 percentage points ($p < .05$), unscored applications did not fully explain differences in award probability. We still found significant differences for Asians and Blacks in R01 award probability, however the sizes of these effects were smaller than what we observed in the full sample (**Table S5**). After controlling for all covariates, the coefficients for Blacks and Asian fell by about half.

REFERENCES

1. Publicly available data include: Information on funded awards from IMPAC II; IPEDS micro data from the National Center for Education Statistics (<http://nces.ed.gov/ipeds/datacenter/>); MEDLINE (<http://www.ncbi.nlm.nih.gov/pubmed/>). Proprietary data include: AAMC Faculty Roster (<https://www.aamc.org/data/facultyroster/>); DRF: Micro data from the Survey of Earned Doctorates is available by site license from the National Science Foundation. Access to sensitive and personally identifiable information, such as names, is not granted to individual researchers. Web of Science: Citation and impact factor data are available for a fee from Thomson Reuters.
2. StataCorp. 2009. *Stata Release 11 User's Guide*. Statistical Software. College Station, TX: StataCorp LP, pp. 297-302.
3. Hausman, J. A., Abrevaya, J., Scott-Morton, F. M. (1998). Misclassification of the Dependent Variable in a Discrete-Response Setting. *Journal of Econometrics*, 87(2): 239-6.
4. Bollinger, C.R. and David, M. (1997). [Modeling Food Stamp Program Participation in the Presence of Reporting Errors](#). *Journal of the American Statistical Association*, 92(439): 827-835.
5. Bollinger, C.R. (2003). Measurement Error in Human Capital and the Black-White Wage Differential. *Review of Economics and Statistics*, 85(3): 578-585.
6. Lubotsky, D. and Wittenberg, M. (2006). Interpretation of Regressions with Multiple Proxies. *Review of Economics and Statistics*, 88(3): 549-62.
7. Allison, P. (1999). Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods and Research*, 28(2): 186-208.
8. Williams R.W. (2009). Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients. *Sociological Methods and Research*, 37(4): 531-559.
9. The National Institutes of Health (2008). Enhancing Peer Review: The NIH Announces New Scoring Procedures for Evaluation of Research Applications Received for Potential FY2010 Funding. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-024.html>.

Figure S1. Award Probability by Priority Score

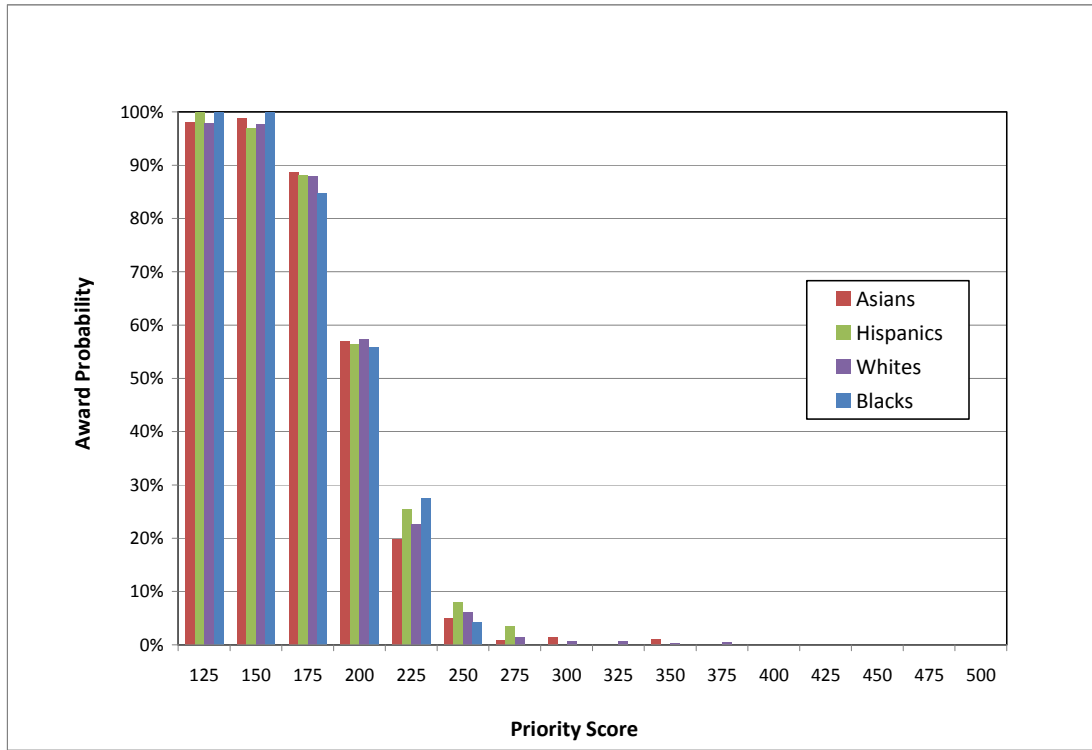


Figure S2. Relative R01 Success Rates by Race, FY 2000 – FY 2008.

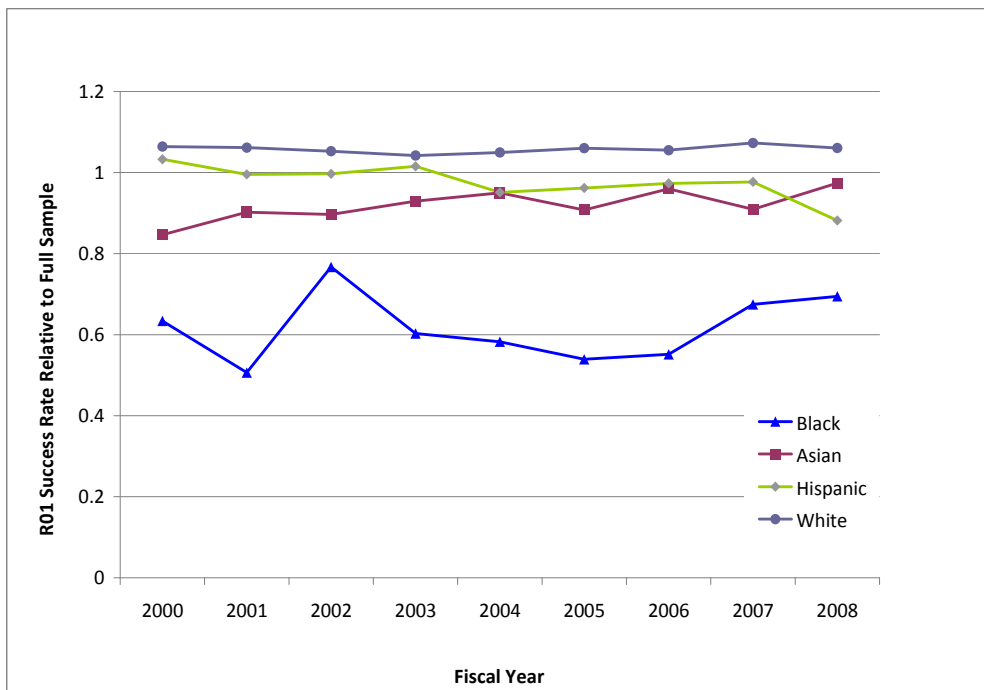


Figure S3. Percentage of applications from US native and foreign-born investigators, by race/ethnicity. Numbers within the boxes are numbers of applications.

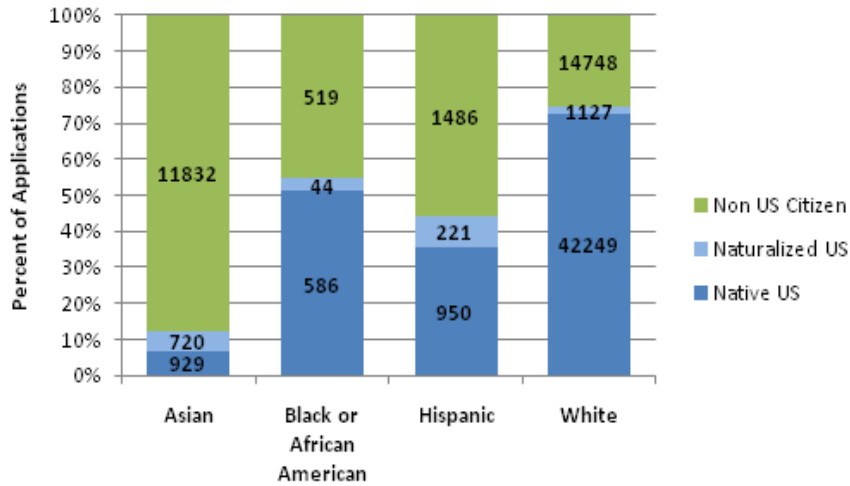


Figure S4. Marginal Effects of Factors Associated with Scored R01 Applications for the Full Sample and for Blacks. $p < .001$ ‡, $p < .01$ **, and $p < .05$ *.

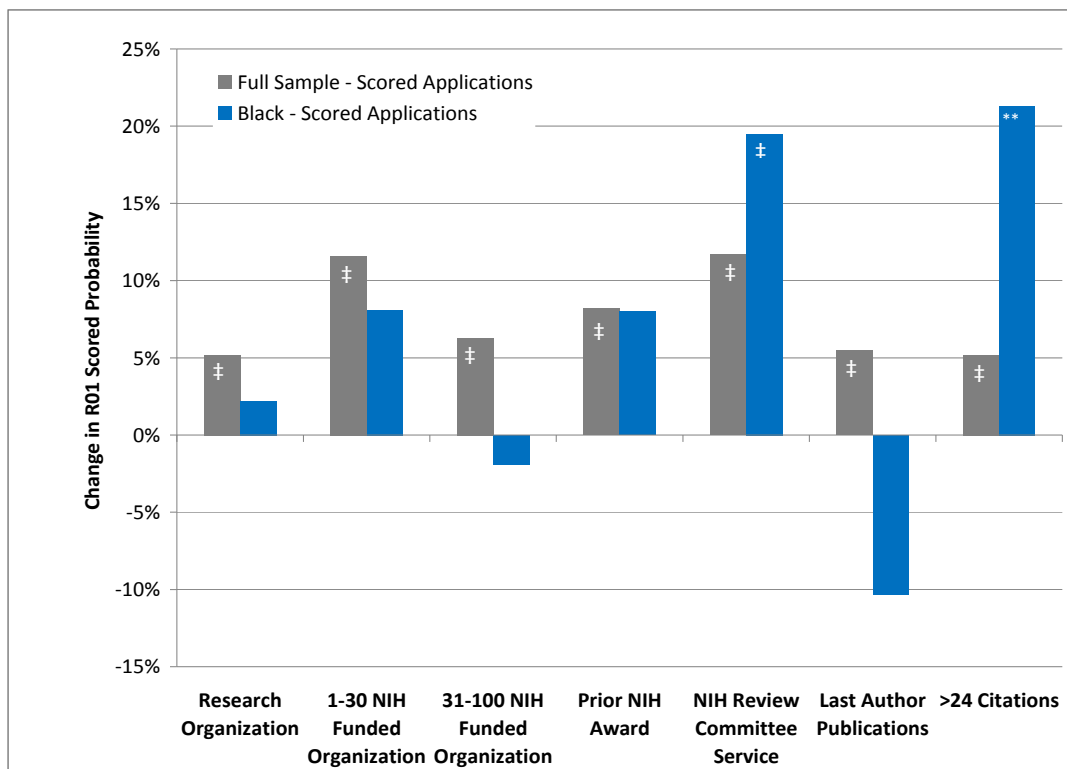


Table S1: Distribution of Priority Scores on R01 Applications by Race/Ethnicity 2000-2006

Race	Priority Scores					Number Scored 500 or Unscored	Total
	Number Scored 100-150	Number Scored 151-200	Number Scored 201-250	Number Scored 251-300	Number Scored 301-459		
Native American	s	s	s	s	s	s	41
Asian	993	2,940	1,949	1,047	340	6,212	13,481
Black	65	140	128	82	43	691	1,149
Hispanic	231	601	415	214	70	1,126	2,657
White	5,462	13,690	9,019	4,827	1,689	23,437	58,124
Other	s	s	s	s	s	47	99
Unknown Race	654	1,585	1,049	532	176	3,641	7,637
Total	7,418	18,987	12,582	6,708	2,327	35,166	83,188

Notes: s = suppressed for confidentiality (n<50).

Table S2. NIH Research Grant Applications and Awards FY2000-FY2006, by Race and Ethnicity. SOURCE: NIH IMPAC II, DRF.

All Competing RPG Applications (2000-06)				
Race	Number of Applications	Percent of Applications	Number of Awards	Award Probability
Native American	152	0.1%	48	31.6%
Asian	28,274	13.8%	8,040	28.4%***
Black	2,942	1.4%	585	19.9%***
Hispanic	6,954	3.4%	2,223	32.0%***
White	135,594	66.1%	49,303	36.4%
Other	227	0.1%	68	30.0%
Unknown	30,963	15.1%	8,404	27.1%***
Total	205,106	100%	68,671	33.5%
Competing New (Type 1) R01s (2000-06)				
Race	Number of Applications	Percent of Applications	Number of Awards	Award Probability
Native American	67	0.1%	18	26.9%
Asian	15,739	14.8%	4,006	25.5%***
Black	1,466	1.4%	250	17.1%***
Hispanic	3,705	3.5%	1,006	27.2%
White	70,773	66.5%	20,710	29.3%
Other	99	0.1%	27	27.3%
Unknown	14,519	13.7%	3,347	23.1%***
Total	106,368	100%	29,364	27.6%
Competing New (Type 1) R01s (2000-06) PhD Analysis Sample				
Race	Number of Applications	Percent of Applications	Number of Awards	Award Probability
Native American	41	0.0%	12	29.3%
Asian	13,481	16.2%	3,430	25.4%***
Black	1,149	1.4%	185	16.1%***
Hispanic	2,657	3.2%	746	28.1%
White	58,124	69.9%	17,017	29.3%
Other	99	0.1%	27	27.3%
Unknown	7,637	9.2%	1,964	25.7%***
Total	83,188	100%	23,381	28.1%

Notes: Standard errors are clustered on the individual applicant. P-values are adjusted for multiple testing using the Bonferroni method. We test whether award probabilities by race are significantly different from Whites, $p < .001$ ***, $p < .01$ ** , and $p < .05$ *.

Table S3: Number of Competing Type 1 R01s Submitted by Race/Ethnicity and Year

Race	Year							Total
	Number 2000	Number 2001	Number 2002	Number 2003	Number 2004	Number 2005	Number 2006	Number 2000 -2006
Native American	s	s	s	s	s	s	s	41
Asian	1,404	1,525	1,555	1,833	2,064	2,050	3,050	13,481
Black	138	141	138	158	188	159	227	1,149
Hispanic	310	297	314	340	390	426	580	2,657
White	7,712	7,464	7,031	7,736	8,105	8,333	11,743	58,124
Other	s	s	s	s	s	s	s	99
Race Unknown	487	540	620	952	1,244	1,436	2,358	7,637
Total	10,073	9,988	9,674	11,039	12,008	12,419	17,987	83,188

Notes: s = suppressed for confidentiality (n<50).

Table S4: Covariates Included in Probit Models

Variables	Model 1	Model 2	Model 3	Model 4	Model5
Demographics					
Race/Ethnicity	X	X	X	X	X
Gender	X	X	X	X	X
Age, Age-Squared ^a	X	X	X	X	X
Naturalized Citizen	X	X	X	X	X
Non-Citizen	X	X	X	X	X
Race Unknown	X	X	X	X	X
Foreign PhD	X	X	X	X	X
Nativity/Citizenship Missing	X	X	X	X	X
Education and Training					
Degree Type (PhD, MD/PhD)		X	X	X	X
NIH Training: (F, T, K)		X	X	X	X
PhD Major Field: (Biomedicine, Chemistry, Physics, Engineering, Psychology, Field Missing)		X	X	X	X
NIH Funding Rank of PhD Institution: (Top 30, 31-100, 100-200) ^b		X	X	X	X
Employer Characteristics					
Employer NIH Funding Rank: (Top 30, 31-100, 100-200) ^b			X	X	X
Employer Organization Type: (Research Institute, Hospital, Higher Education, Other)			X	X	X
Higher Education Carnegie Class: (Research Very High, Research High Research, Medicine, BA or MA Inst., Other, Carnegie Rank missing)			X	X	X
Region: (Midwest, South, West)			X	X	X
NIH Experience					
NIH Institute Code: (21 Indicators for IC receiving proposal)				X	X
Prior NIH Grant				X	X
NIH Review Committee Member				X	X
Grant uses Human Subjects				X	X
Human Subject Code Missing				X	X
Fiscal Year (2001 - 2006)				X	X
Research Productivity					
Publication Quartiles (4-7, 8-18, >18) ^b					X
Citation Quartiles (6-24, 25-84, >84) ^b					X
Maximum Impact Factor of Publications ^a					X
Median Impact Factor of Publications ^a					X
Ratio of First author/ Total Publications ^a					X
Ratio of Last author/ Total Publications ^a					X
Ratio of Single author/ Total Publications ^a					X
Publication information missing					X

Notes: Variables are indicator variables (0,1) unless otherwise indicated. ^aContinuous variables. ^bCategorical variable definition.

**Table S5: Probit Estimates of the Effect of Race/Ethnicity on R01 Funding Award
Marginal Effects, Standard Errors in Brackets, FY 2000 - FY 2006**

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Native American	0.036 [0.076]	0.034 [0.077]	0.027 [0.077]	0.054 [0.081]	0.063 [0.084]
Asian	-0.054*** [0.006]	-0.054*** [0.006]	-0.051*** [0.006]	-0.040*** [0.006]	-0.042*** [0.006]
Black	-0.131*** [0.011]	-0.131*** [0.011]	-0.119*** [0.012]	-0.110*** [0.012]	-0.104*** [0.012]
Hispanic	-0.027* [0.010]	-0.027* [0.010]	-0.023 [0.010]	-0.014 [0.010]	-0.012 [0.010]
Other	-0.020 [0.049]	-0.020 [0.047]	-0.010 [0.047]	0.013 [0.048]	0.021 [0.049]
Unknown Race	-0.049*** [0.006]	-0.044*** [0.006]	-0.040*** [0.006]	0.012 [0.007]	0.016 [0.007]
Non-citizen	0.004 [0.006]	0.021** [0.007]	0.016* [0.007]	0.024*** [0.006]	0.018** [0.006]
Naturalized Citizen	-0.020 [0.011]	-0.022* [0.011]	-0.028* [0.011]	-0.015 [0.011]	-0.018 [0.011]
Foreign-PhD	0.071*** [0.008]	0.071*** [0.013]	0.081*** [0.014]	0.093*** [0.014]	0.093*** [0.014]
Foreign-Born Missing	-0.055*** [0.005]	-0.049*** [0.012]	-0.046*** [0.012]	-0.019 [0.013]	-0.019 [0.013]
F Recipient		0.029*** [0.005]	0.029*** [0.005]	0.011* [0.005]	0.008 [0.005]
T Recipient		0.018*** [0.004]	0.011** [0.004]	0.012** [0.004]	0.009* [0.004]
K Grant Recipient		0.049*** [0.006]	0.038*** [0.006]	0.018** [0.006]	0.017** [0.006]
Employer Ranked 1-30 NIH Funding			0.133*** [0.008]	0.112*** [0.008]	0.097*** [0.008]
Employer Ranked 31-100 NIH Funding			0.088*** [0.007]	0.072*** [0.007]	0.061*** [0.007]
Employer Ranked 101-200 NIH Funding			0.069*** [0.007]	0.055*** [0.007]	0.049*** [0.007]
Employer Hospital			0.042** [0.014]	0.036* [0.014]	0.032* [0.014]
Employer Research Institute			0.074*** [0.014]	0.048*** [0.014]	0.042** [0.013]
Prior NIH Grants				0.088*** [0.004]	0.081*** [0.004]
Served on NIH Review Committee				0.091*** [0.004]	0.082*** [0.004]
Citations 3rd Quartile (24 - 84 citations)					0.032*** [0.007]
Citations 4th Quartile (>84 citations)					0.068*** [0.009]
Ratio of First Author to Total Publications					-0.009 [0.008]
Ratio of Last Author to Total Publications					0.021* [0.008]
Ratio of Single Author to Total Publications					0.027* [0.011]

Notes: Numbers in table are marginal effects which report change in probability of receiving an R01 award given an infinitesimal change in continuous independent variables. Marginal effects on dummy variables report change in probability of receiving an R01 award given a change in the dummy from 0 to 1. Multiply marginal effects by 100 to obtain percentage points. Robust standard errors clustered on individual applicant in brackets. P-values on race adjusted for multiple comparisons using Bonferroni method. p<.001***, p<.01**, p<.05*. N = 83, 188 Observations.

Table S6: Number and Percentage of R01 Application Submissions, Resubmissions and Awards by Race/Ethnicity 2000-2006

Race	Total Submissions 2000-2006			Grants Submitted Once			Grants Submitted Twice			Grants Submitted 3 or More Times		
	Number Submitted	Number Awarded	Percentage Awarded	Number Submitted	Number Awarded	Percentage Awarded	Number Submitted	Number Awarded	Percentage Awarded	Number Submitted	Number Awarded	Percentage Awarded
Native American	41	12	29.3%	s	s	16.0%	s	s	54.5%	s	s	40.0%
Asian	13,481	3,430	25.4%	8,452	1,563	18.5%	3,523	1,244	35.3%	1,506	623	41.4%
Black	1,149	185	16.1%	782	74	9.5%	257	65	25.3%	110	46	41.8%
Hispanic	2,657	746	28.1%	1,667	334	20.0%	704	281	39.9%	286	131	45.8%
White	58,124	17,017	29.3%	36,216	7,991	22.1%	15,539	6,010	38.7%	6,369	3,016	47.4%
Other	99	27	27.3%	s	s	5.1%	s	s	34.1%	s	s	53.3%
Race Unknown	7,637	1,964	25.7%	5,008	880	17.6%	1,894	721	38.1%	735	363	49.4%
Total	83,188	23,381	28.1%	52,208	10,851	20.8%	21,954	8,341	38.0%	9,026	4,189	46.4%
	Ratios of Resubmissions											
Per Applicant Comparisons	Resubmitted/ Awarded	Resubmitted/ Unfunded										
Native American	2.24	73.2%										
Asian	1.85***	62.4%										
Black	2.01	45.3%***										
Hispanic	1.37*	55.7%***										
White	1.58	64.3%										
Other	1.58	63.4%										
Race Unknown	1.17***	38.8%***										

Notes: s = suppressed for confidentiality (n<50). P-values on race adjusted for multiple comparisons using Bonferroni method. p<.001***, p<.01**, p<.05*.

Table S7: Probit Estimates of the Effect of Race/Ethnicity on R01 Funding Award, Controlling for Resubmitted Applications, Marginal Effects, Standard Errors in Brackets FY 2000 - FY 2006

Race	Model 5	Model 5 with Controls for Number of Resubmissions	Model 5 Grants Submitted Once	Model 5 Grants Submitted Twice	Model 5 Grants Submitted Three or More Times
Native American	0.063 [0.084]	0.048 [0.087]	0.018 [0.084]	0.228 [0.158]	-0.129 [0.272]
Asian	-0.042*** [0.006]	-0.041*** [0.006]	-0.038*** [0.006]	-0.041** [0.012]	-0.037* [0.018]
Black	-0.104*** [0.012]	-0.101*** [0.012]	-0.089*** [0.011]	-0.135*** [0.027]	-0.036 [0.048]
Hispanic	-0.012 [0.010]	-0.011 [0.010]	-0.017 [0.010]	0.011 [0.020]	-0.010 [0.032]
Other	0.021 [0.049]	0.001 [0.045]	-0.091* [0.038]	0.193 [0.101]	0.086 [0.116]
Unknown Race	0.016 [0.007]	0.023** [0.007]	0.012 [0.008]	0.043* [0.015]	0.059* [0.022]
Observations	83188	83188	52199^a	21947^a	9026

Notes: Numbers in table are marginal effects which report change in probability of receiving an R01 award given an infinitesimal change in continuous independent variables. Marginal effects on dummy variables report change in probability of receiving an R01 award given a change in the dummy from 0 to 1. Multiply marginal effects by 100 to obtain percentage points. Robust standard errors clustered on individual applicant in brackets. P-values on race adjusted for multiple comparisons using Bonferroni method. p<.001***, p<.01**, p<.05*. ^a 16 observations were dropped from these regressions because missing human subjects data predicted the probability of award perfectly.

Table S8: Probit Estimates of the effect of Race/Ethnicity on R01 Funding Award, US Citizen Sample, Marginal Effects, Standard Errors in Brackets FY 2000 - FY 2006

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Native American	-0.022 [0.114]	-0.026 [0.115]	-0.031 [0.115]	-0.015 [0.128]	-0.008 [0.133]
Asian	-0.014 [0.013]	-0.023 [0.013]	-0.027* [0.012]	-0.021 [0.012]	-0.023 [0.012]
Black	-0.130*** [0.016]	-0.132*** [0.015]	-0.123*** [0.016]	-0.115*** [0.017]	-0.107*** [0.017]
Hispanic	-0.040 [0.015]	-0.040 [0.015]	-0.037 [0.015]	-0.030 [0.015]	-0.025 [0.015]
Other	-0.045 [0.051]	-0.047 [0.048]	-0.039 [0.048]	-0.013 [0.050]	-0.007 [0.051]
Unknown Race	-0.014 [0.016]	-0.010 [0.016]	-0.006 [0.016]	0.007 [0.016]	0.010 [0.016]
Naturalized Citizen	-0.032** [0.012]	-0.030* [0.012]	-0.034** [0.012]	-0.020 [0.012]	-0.026* [0.012]
F Recipient		0.025*** [0.005]	0.025*** [0.005]	0.008 [0.005]	0.004 [0.005]
T Recipient		0.022*** [0.005]	0.014** [0.005]	0.017** [0.005]	0.012* [0.005]
K Grant Recipient		0.048*** [0.008]	0.037*** [0.008]	0.020** [0.008]	0.017* [0.007]
Employer Ranked 1-30 NIH Funding			0.141*** [0.010]	0.118*** [0.010]	0.101*** [0.010]
Employer Ranked 31-100 NIH Funding			0.092*** [0.009]	0.078*** [0.009]	0.066*** [0.009]
Employer Ranked 101-200 NIH Funding			0.072*** [0.010]	0.058*** [0.010]	0.050*** [0.010]
Employer Hospital			0.028 [0.019]	0.026 [0.019]	0.018 [0.019]
Employer Research Institute			0.072*** [0.018]	0.051** [0.018]	0.041* [0.018]
Prior NIH Grants				0.083*** [0.006]	0.073*** [0.006]
Served on NIH Review Committee				0.100*** [0.005]	0.086*** [0.005]
Citations 3rd Quartile (24 - 84 citations)					0.049*** [0.009]
Citations 4th Quartile (>84 citations)					0.092*** [0.013]
Ratio of First Author to Total Publications					-0.005 [0.010]
Ratio of Last Author to Total Publications					0.023* [0.011]
Ratio of Single Author to Total Publications					0.018 [0.013]
Observations	47890	47890	47890	47890	47890

Notes: Numbers in table are marginal effects which report change in probability of receiving an R01 award given an infinitesimal change in continuous independent variables. Marginal effects on dummy variables report change in probability of receiving an R01 award given a change in the dummy from 0 to 1. Multiply marginal effects by 100 to obtain percentage points. Robust standard errors clustered on individual applicant in brackets. P-values on race adjusted for multiple comparisons using Bonferroni method. p<.001***, p<.01**, p<.05*.

**Table S9: Number and Percentage of Applications with NIH Training by Race/Ethnicity,
US Citizen Sample, 2000-2006**

Race	<u>T, K or F</u>		<u>T</u>		<u>F</u>		<u>K</u>		<u>Total</u>
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number
Native American	s		s		s		s		s
Asian	1,138	69.0%	958	58.1%	370	22.4%	280	17.0%	1,649
Black	340	54.0%	277	44.0%	98	15.6%	63	10.0%	630
Hispanic	726	62.0%	524	44.7%	259	22.1%	183	15.6%	1,171
White	26,700	61.6%	18,718	43.2%	11,583	26.7%	4,967	11.5%	43,376
Other	45	50.6%	26	29.2%	23	25.8%	13	14.6%	89
Race Unknown	366	38.3%	185	19.4%	153	16.0%	113	11.8%	956
Total	29,318	61.2%	20,688	43.2%	12,489	26.1%	5,621	11.7%	47,890

Notes: s = suppressed for confidentiality (n<50).

**Table S10: Number of Applications and Applicants by Race/Ethnicity,
F or T Training and R01 Awards 2000-2006**

Race	Applications						Applicants					
	No Training			Trained			No Training			Trained		
	Number Submitted	Number Awarded	Percentage Awarded	Number Submitted	Number Awarded	Percentage Awarded	Number Submitted	Number Awarded	Percentage Awarded	Number Submitted	Number Awarded	Percentage Awarded
Native American	s	s	18.8%	s	s	33.3%	s	s	71.4%	s	s	100.0%
Asian	552	131	23.7%	1,097	331	30.2%	283	133	47.0%	513	306	59.6%
Black	311	46	14.8%	319	57	17.9%	185	51	27.6%	171	67	39.2%
Hispanic	501	124	24.8%	670	178	26.6%	270	124	45.9%	304	173	56.9%
White	18,520	4,849	26.2%	24,856	7,821	31.5%	9,753	5,349	54.8%	12,086	8,041	66.5%
Other	s	s	11.5%	s	s	43.2%	s	s	36.7%	s	s	62.5%
Race Unknown	649	145	22.3%	307	88	28.7%	406	171	42.1%	154	89	57.8%
Total	20,601	5,304	25.7%	27,289	8,492	31.1%	10,934	5,844	53.4%	13,254	8,693	65.6%

Notes: s = suppressed for confidentiality (n<50).

Table S11: Distribution and Averages by Race/Ethnicity for Selected Covariates Full and Unscored Samples

Full Sample										
Race	Prior Grant (Percent)	Prior K (Percent)	Prior F (Percent)	Prior T (Percent)	Prior F or T (Percent)	Research Org. (Percent)	Higher Education (Percent)	Review Committee (Percent)	1-30 Fund Rank (Percent)	31-100 Fund Rank (Percent)
Native American	70.7%	9.8%	12.2%	29.3%	41.5%	0.0%	90.2%	39.0%	48.8%	29.3%
Asian	72.8%***	8.5%***	9.1%***	19.8%***	25.5%***	8.6%	83.0%	40.0%***	32.4%	38.7%**
Black	68.9%***	10.6%	12.9%***	34.0%	40.0%**	4.9%**	86.2%	47.8%	27.1%*	34.0%
Hispanic	69.9%***	12.4%	13.7%***	26.6%***	35.0%***	7.7%	81.7%	46.3%***	32.6%	35.2%
White	77.8%	11.2%	22.4%	36.5%	48.7%	8.6%	82.7%	53.2%	33.2%	35.4%
Other	70.7%	13.1%	23.2%	26.3%	37.4%*	3.0%*	84.8%	47.5%	31.3%	30.3%
Unknown Race	57.6%***	5.4%***	4.8%***	9.8%***	13.3%***	8.5%	78.4%***	31.5%***	32.0%*	34.3%
Total	74.8%	10.3%	18.2%	31.0%	41.1%	8.5%	82.4%	48.8%	32.8%	35.8%
Race	Publications (Mean) ^a	Citations (Mean) ^a	Citations > Median (>24) (Percent) ^a	Maximum Impact Factor (Mean) ^a	Median Impact Factor (Mean) ^a	Single Author (Percent) ^a	First Author (Percent) ^a	Last Author (Percent) ^a		
Native American	11.0	67.3	43.5%	6.8	3.6	4.3%	38.5%	32.9%		
Asian	28.8***	143.4***	64.5%***	13.6***	4.7***	5.4%***	32.8%***	34.2%***		
Black	13.7	40.1***	39.7%	9.2*	3.2***	11.2%	40.8%	22.4%***		
Hispanic	17.8**	90.3	49.3%	11.7***	4.2	7.6%***	37.1%	30.3%		
White	14.3	77.7	45.9%	10.0	4.2	10.3%	38.2%	30.4%		
Other	9.1**	44.51*	40.0%	8.2	3.6	8.4%	49.1%	27.3%		
Unknown Race	17.6***	90.4	53.3%***	11.1***	4.5***	7.7%***	37.3%	27.7%***		
Total	17.2	90.1	49.8%	10.8	4.3	9.1%	37.2%	30.7%		

Notes: ^aEstimates limited to those with valid match to publications. Standard errors are clustered on the individual applicant. P-values are adjusted for multiple testing using the Bonferroni method. We test whether these variables are significantly different from Whites, p<.001***, p<.01**, and p<.05*.

Table S12: Distribution and Averages by Race/Ethnicity for Selected Covariates Full and Unscored Samples

Unscored Sample										
Race	Prior Grant (Percent)	Prior K (Percent)	Prior F (Percent)	Prior T (Percent)	Prior F or T (Percent)	Research Org (Percent)	Higher Education (Percent)	Review Committee (Mean)	1-30 Fund Rank (Mean)	31-100 Fund Rank (Percent)
Native American	63.6%	0.0%	0.0%	45.5%	45.5%	0.0%	81.8%	36.4%	36.4%	27.3%
Asian	67.2%***	7.0%***	8.3%***	18.7%***	24.0%***	9.2%	82.4%	33.6%***	27.5%	38.7%
Black	63.1%**	9.2%	13.9%**	31.5%	37.8%*	5.0%*	86.4%	39.5%	20.8%*	35.1%
Hispanic	63.7%***	11.7%	13.3%***	26.3%**	34.0%***	6.5%	81.4%	39.6%	27.2%	37.4%
White	71.9%	9.5%	21.6%	34.0%	46.2%	8.1%	83.3%	44.9%	27.6%	36.2%
Other	68.9%	15.6%	15.6%	13.3%**	22.2%**	4.4%	77.8%	44.4%	22.2%	28.9%
Unknown Race	49.4%***	4.2%***	3.9%***	9.1%***	11.7%***	8.0%	76.9%***	25.5%***	26.4%	33.7%
Total	68.3%	8.6%	17.0%	28.4%	38.1%	8.1%	82.5%	40.6%	27.3%	36.4%
Race	Publications (Mean) ^a	Citations (Mean) ^a	Citations > Median (>24) (Percent) ^a	Maximum Impact Factor (Mean) ^a	Median Impact Factor (Mean) ^a	Single Author (Percent) ^a	First Author (Percent) ^a	Last Author (Percent) ^a		
Native American	7.1	28.6	14.3%*	7.1	2.2	14.3%	38.0%	34.0%		
Asian	27.3%***	119.0***	61.0%***	12.5***	4.3***	5.0%***	34.9%***	32.0%***		
Black	14.3	35.3***	35.4%	9.3	3.1***	10.3%	41.9%	23.2%*		
Hispanic	16.8	68.6	43.5%	10.4**	3.8	7.2%*	38.4%	28.0%		
White	13.5	56.6	40.7%	8.8	3.7	9.7%	39.8%	28.7%		
Other	9.3	49.0	38.9%	7.8	3.6	13.2%	52.8%	21.5%		
Unknown Race	16.6*	75.3**	49.6%***	10.1***	4.1***	7.2%***	39.3%	25.0%***		
Total	16.6	70.5	45.5%	9.7	3.9	8.5%	38.8%	28.9%		

Notes: ^aEstimates limited to those with valid match to publications. Standard errors are clustered on the individual applicant. P-values are adjusted for multiple testing using the Bonferroni method. We test whether these variables are significantly different from Whites, p<.001***, p<.01**, and p<.05*.

Table S13: Number of Applications by Race/Ethnicity for Selected Covariates Full and Unscored Samples

Full Sample										
Race	Prior Grant (Number)	Prior K (Number)	Prior F (Number)	Prior T (Number)	Prior F or T (Number)	Research Org (Number)	Review Committee (Number)	1-30 Fund Rank (Number)	31-100 Fund Rank (Number)	Citations > Median (>24) (Number)
Native American	s	s	s	s	s	s	s	s	s	s
Asian	9,816	1,149	1,229	2,663	3,435	1,158	5,388	4,364	5,215	7,824
Black	792	122	148	391	460	56	549	311	391	347
Hispanic	1,856	330	365	708	931	204	1,231	865	936	1,081
White	45,232	6,507	13,018	21,203	28,298	4,998	30,927	19,290	20,600	22,290
Other	70	s	s	s	37	s	47	31	30	32
Unknown Race	4,398	416	369	748	1,013	648	2,408	2,445	2,616	3,418
Total	62,164	8,524	15,129	25,713	34,174	7,064	40,550	27,306	29,788	34,992
Unscored Sample										
Race	Prior Grant (Number)	Prior K (Number)	Prior F (Number)	Prior T (Number)	Prior F or T (Number)	Research Org (Number)	Review Committee (Number)	1-30 Fund Rank (Number)	31-100 Fund Rank (Number)	Citations > Median (>24) (Number)
Native American	s	s	s	s	s	s	s	s	s	s
Asian	4,157	436	513	1,160	1,487	567	2,079	1,705	2,395	3,353
Black	431	63	95	215	258	34	270	142	240	179
Hispanic	710	130	148	293	379	72	441	303	418	383
White	16,742	2,215	5,039	7,931	10,773	1,876	10,468	6,441	8,434	7,670
Other	s	s	s	s	s	s	s	s	s	s
Unknown Race	1,790	153	141	328	423	291	923	955	1,219	1,435
Total	23,830	2,997	5,936	9,927	13,320	2,840	14,181	9,546	12,706	13,020

Notes: s = suppressed for confidentiality, (n<50).

Table S14: Probit Estimates of the Effect of Selected Covariates on Receiving a Priority Score, Marginal Effects, Standard Errors in Brackets--FY 2000 - FY 2006

Variable	Model 5				
	Full Sample	Blacks	Asians	Hispanics	Whites
Employer Ranked 1-30 NIH Funding	0.116*** [0.008]	0.081 [0.076]	0.131*** [0.020]	0.104** [0.038]	0.107*** [0.009]
Employer Ranked 31-100 NIH Funding	0.063*** [0.007]	-0.019 [0.067]	0.082*** [0.019]	0.017 [0.037]	0.056*** [0.009]
Employer Research Institute	0.052*** [0.013]	0.022 [0.093]	-0.018 [0.039]	0.067 [0.060]	0.046** [0.016]
Prior NIH Grants	0.082*** [0.005]	0.080 [0.045]	0.089*** [0.012]	0.071* [0.029]	0.081*** [0.006]
Served on NIH Review Committee	0.117*** [0.005]	0.195*** [0.040]	0.114*** [0.012]	0.097*** [0.026]	0.118*** [0.005]
Citations above the median (>24 citations)	0.052*** [0.007]	0.213*** [0.065]	0.047* [0.019]	0.084* [0.043]	0.057*** [0.009]
Ratio of Last Author to Total Publications	0.055*** [0.010]	-0.103 [0.089]	0.055* [0.024]	0.044 [0.055]	0.053*** [0.012]
Observations	83188	1143	13481	2651	58124
Chi-squared test that Coefficients are same as the full sample		14.03*	5.63	3.85	5.23

Notes: Numbers in table are marginal effects which report change in probability of receiving an R01 award given an infinitesimal change in continuous independent variables. Marginal effects on dummy variables report change in probability of receiving an R01 award given a change in the dummy from 0 to 1. Multiply marginal effects by 100 to obtain percentage points. Robust standard errors clustered on individual applicant in brackets. P-values on race adjusted for multiple comparisons using Bonferroni method. p<.001***, p<.01**, p<.05*.

**Table S15: Probit Estimates of the Effect of Race/Ethnicity on R01 Funding Award, Sample Omits Unscored Applications
Marginal Effects, Standard Errors in Brackets, FY 2000 - FY 2006**

Race	Model 1	Model 2	Model 3	Model 4	Model 5
Native American	-0.049 [0.099]	-0.052 [0.101]	-0.056 [0.102]	-0.066 [0.104]	-0.059 [0.106]
Asian	-0.036*** [0.008]	-0.035*** [0.008]	-0.034*** [0.008]	-0.027** [0.008]	-0.028** [0.008]
Black	-0.093*** [0.023]	-0.094*** [0.023]	-0.084** [0.023]	-0.069* [0.024]	-0.065* [0.024]
Hispanic	-0.021 [0.013]	-0.021 [0.013]	-0.018 [0.013]	-0.011 [0.013]	-0.010 [0.013]
Other	0.016 [0.068]	0.015 [0.066]	0.022 [0.065]	0.052 [0.068]	0.059 [0.067]
Unknown Race	-0.023 [0.009]	-0.018 [0.009]	-0.017 [0.009]	0.038*** [0.010]	0.040*** [0.010]
Observations	48226	48226	48226	48226	48226

Notes: Numbers in table are marginal effects which report change in probability of receiving an R01 award given an infinitesimal change in continuous independent variables. Marginal effects on dummy variables report change in probability of receiving an R01 award given a change in the dummy from 0 to 1. Multiply marginal effects by 100 to obtain percentage points. Robust standard errors clustered on individual applicant in brackets. P-values on race adjusted for multiple comparisons using Bonferroni method. p<.001***, p<.01**, p<.05*.