BMJ Open

Gender and ethnic differences in publication of BMJ letters to the editor: an observational study using machine learning

Mohamad Zeina ,[1] Alfred Balston,[1] Amitava Banerjee ,[2] Katherine Woolf [3]

To cite: Zeina M, Balston A, Banerjee A, *et al*. Gender and ethnic differences in publication of BMJ letters to the editor: an observational study using machine learning. *BMJ Open* 2020;**10**:e037269. doi:10.1136/bmjopen-2020-037269

► Prepublication history and supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-037269).

Received 26 January 2020
Revised 29 July 2020
Accepted 23 September 2020

[1]Barts Health NHS Trust, London, UK
[2]Farr Institute of Health Informatics Research, University College London, London, UK
[3]Research Department of Medical Education, University College London, London, UK

Correspondence to
Dr Katherine Woolf;
k.woolf@ucl.ac.uk

ABSTRACT

Objectives To analyse the relationship between first author's gender and ethnicity (estimated from first name and surname), and chance of publication of rapid responses in the *British Medical Journal* (BMJ). To analyse whether other features of the rapid response account for any gender or ethnic differences, including the presence of multiple authors, declaration of conflicts of interests, the presence of Twitter handle, word count, reading ease, spelling and grammatical mistakes, and the presence of references.

Design A retrospective observational study.

Setting Website of the BMJ (BMJ.com).

Participants Publicly available rapid responses submitted to BMJ.com between 1998 and 2018.

Main outcome measures Publication of a rapid response as a letter to the editor in the BMJ.

Results We analysed 113 265 rapid responses, of which 8415 were published as letters to the editor (7.4%). Statistically significant univariate correlations were found between odds of publication and first author estimated gender and ethnicity, multiple authors, declaration of conflicts of interest, the presence of Twitter handle, word count, reading ease, spelling and grammatical mistakes, and the presence of references. Multivariate analysis showed that first author estimated gender and ethnicity predicted publication after taking into account the other factors. Compared to white authors, black authors were 26% less likely to be published (OR: 0.74, CI: 0.57–0.96), Asian and Pacific Islander authors were 46% less likely to be published (OR: 0.54, CI: 0.49–0.59) and Hispanic authors were 49% less likely to be published (OR: 0.51, CI: 0.41–0.64). Female authors were 10% less likely to be published (OR: 0.90, CI: 0.85–0.96) than male authors.

Conclusion Ethnic and gender differences in rapid response publication remained after accounting for a broad range of features, themselves all predictive of publication. This suggests that the reasons for the differences of these groups lies elsewhere.

Strengths and limitations of this study

- This study uses corpus of publicly available data to analyse correlations between first author's characteristics and chance of publication of letters to the editor in the *British Medical Journal*.
- Multivariate analysis allowed us to account for a range of other features of submitted letters.
- To the best of our knowledge, this is the largest ever analysis of a scientific corpus that looks at associations with publication rate.
- The nature of this data means that only associations can be inferred, and not causation.
- We highlight automated techniques that scientific journals can use to look for associations between ethnicity and gender in their own publication rates.

INTRODUCTION

Much has been written about the 'attainment gap' or 'differential attainment', the observation that many fields exhibit discrepancies in achievement based on personal attributes such as gender and ethnicity. In medicine, for example, students from black and minority ethnic (BME) groups achieve poorer marks and are more likely to fail, on average, than their white counterparts.[1] As they progress in their careers, BME doctors also more often fail their specialty training exams,[1 2] earn a lower average salary than others at the same level of seniority[3] and are less likely to be awarded funding grants.[4] There also remain discrepancies in the representation of women in medical leadership and faculty despite a long history of roughly equal proportions of male and female medical students.[5 6]

Another specific area where the effects of gender have been studied thoroughly is academic publishing. A survey of 1065 authors from different backgrounds found that women were under-represented in the scientific literature, along with certain ethnic minorities.[7] In a group of high-impact medical journals, including the *British Medical Journal* (BMJ), the number of articles with female first authors has increased over time; however, the gender balance of last authors, who are typically senior researchers or heads of departments, has not followed this trend.[8]

A larger scale analysis of 1.8 million scholarly articles indexed in JSTOR found female authors are poorly represented in the prestigious first and last author positions in the majority of academic disciplines.[9] The same result was observed in an analysis of 21 million articles indexed in Medline, which also found BME authors are similarly less likely to be in the last author position even when accounting for seniority.[10]

The cause of these gender and ethnic differences remains the subject of debate. Experimental studies have shown that identical submissions randomly assigned to have a male or female name are ranked differently depending on the gender of the applicant's name, favouring men.[11 12] For example, applications for a laboratory manager position that were assigned a male name were rated as being significantly more competent and hireable by faculty members compared with identical applications assigned a female name.[11] Similarly, a study presented graduate students with a sample of abstracts from an international conference, where abstracts were randomly shown to have male or female authors. The study found that abstracts presented as having female authors were deemed to have lower 'Scientific Quality'.[12] Both papers found that the gender of the reviewer did not affect how applicants were rated, and concluded that pervasive gender stereotypes create a subtle but significant bias against women.[11 12] Indeed, the bias was reduced when reviewers' attitudes towards gender roles were taken into account, with higher support for gender equality being associated with higher ratings for female authors.

Some argue that gender disparities arise from men and women choosing different career paths.[13] For example, women may opt to prioritise flexibility or take time out of their career to have children. However, studies that incorporate these factors into multivariate statistical models fail to fully account for discrepancies in pay.[14 15] One such study found a $14 581 yearly salary difference between male and female hospital physicians in the USA, which remained after accounting for differences in job satisfaction priorities between genders.[16]

The underlying causes of these differences are likely to be complex and multifactorial, but identifying and characterising disparities in new specific situations might hint at potential solutions. These may be broadly applicable, especially because the causative issues are likely to compound each other. For example, lower average pay for women and BME doctors may be partly due to lower chances of scientific publication, especially in a work environment where publication in the scientific literature is important for attaining certain senior academic and leadership positions.

Though many studies mentioned here find group differences based on personal characteristics, they rarely have access to raw data from the journals that would be necessary to quantify publication rate. For example, a finding that women are under-represented in authorship of medical journal papers compared with in the medical workforce is not enough to draw conclusions about discrimination or bias, as it may be due to differences in priorities and the number of submissions sent. For a study to draw meaningful conclusions regarding discrepancies in acceptance rates, it must be able to quantify the percentage of submitted scientific works that are accepted, and this submission data is seldom released by scientific journals.

Letters in the BMJ are derived from rapid responses, which are available online freely and in their entirety, and, therefore, they may provide a valuable perspective for looking at this issue. Moreover, publication of rapid responses is of importance since letters to the editor carry PubMed identifiers and thus discrepancies in their publication may have knock-on effects for jobs in academia where PubMed indexed publications play an important role in candidate selection.

We aimed to compare the corpus of available rapid responses with published letters to the editor to look for correlations between ethnicity, gender and odds of publication.

METHODS

Data acquisition and processing

An automated script was used to download every BMJ.com online rapid response between 25 April 1998 and 23 March 2018, as well as every letter to the editor that was published in the same timeframe.

To minimise the impact on BMJ servers, webpage requests were only sent every 15 s, and each request explicitly stated a full name and contact email address of the researcher carrying out the automated data collection, so that they could easily be contacted if the BMJ wished this collection to stop. Furthermore, we only collected publicly available data that can be accessed without a login to a BMJ account.

Once collected, every available field from the rapid response was extracted. This included: title, title of article being responded to, body of text, first author name, first author title, other authors, date of submission and the presence of Twitter handle. Further processing with software packages mentioned below allowed us to look at a richer set of features, including word count, the presence of references, number of references, Flesch reading ease (a measure of complexity of language, with a higher value meaning easier to read), number of spelling and grammatical mistakes, gender of first author, ethnicity of first author and the presence of multiple authors.

The position of the author was extracted by looking for the presence of each of the words 'Consultant', 'Professor', 'Senior' and 'Student' in the self-declared occupation field of submitted rapid responses, for example, someone who had the word 'Consultant' anywhere in their occupation field was classed 'Consultant'.

We did not expect a linear relationship between publication and word count or Flesch reading ease, because the most successful letters are likely to be long enough to offer a meaningful insight into the topic, but not

too long as to be unsuitable for the short letter to the editor format. Thus, we created two additional features from these, 'Near ideal word count' and 'Near ideal Flesch reading ease' to reflect whether a rapid response was within the 50% of rapid responses which are closest in word count and Flesch reading ease to the numbers which have historically been associated with higher rates of publication.

Some rapid responses (528, or 0.46%) could not be collected automatically due to errors in their formatting which prohibited their automated collection. These rapid responses were omitted from analysis. Regarding collected rapid responses, the absence of data was itself useful information (eg, the absence of second authors was processed as there being no second authors) and so no analysed data point was considered missing.

As there is a lag between submission of a rapid response and publication of the response as a letter, we excluded all rapid responses that were within 66 days of our data collection window (ie, submitted after 16 January 2018). This value was based on preliminary analysis that found a vast majority (80%) of letters were published within 66 days of the rapid response submission date.

Matching protocol

Although both rapid responses and published letters are available freely on BMJ.com, they are available on different parts of the website, and the vast majority of published letters do not link to the specific rapid response that was initially submitted. The task of finding out which rapid responses have been accepted is further complicated by the fact that many editorial changes are made between the submission of a rapid response, and it being printed in the BMJ. Therefore, finding the corresponding rapid response for a letter is not as trivial as looking for a rapid response with identical text content.

To carry out this correspondence, a hierarchical matching protocol was used, which we summarise here. For each published letter, we search the corpus of rapid responses for those by the same first author. To make this possible, author names were standardised by removing middle names or initials. When a first author was only associated with a single rapid response, and a single letter to the editor, these were designated as the same submission. When the author of a letter to the editor has submitted numerous rapid responses, one was chosen where the first 50 characters had the highest similarity with the letter to the editor.

If no rapid response could be found with the same first author as the published letter to the editor, author name was ignored and rapid responses submitted recently before publication of the letter to the editor were searched for one with the highest similarity in the first 50 characters. A subset of 600 matched rapid responses and letters was checked manually by AB (Alfred Balston) and MZ and found to be 85.3% accurate.

Classifying ethnicity and gender

Authors of rapid responses are not asked to disclose their ethnicity or gender, and the number of rapid responses involved was too great to individually contact the authors and ask this sensitive information. An automated method was used that could determine ethnicity of a name, for many tens of thousands of names, quickly and with little manual input. This took the shape of a previously published machine learning algorithm, nameprism.com, that has been trained to classify ethnicity on 74 million names, as well as being externally validated on datasets other than those which it was trained on.[17] To the best of our knowledge, it has demonstrated the highest classification accuracy of any publicly available tool for this task, with an F1 score of 0.795. We follow previous medical research that has used name to classify ethnicity,[18 19] as well as a validation study that suggested name analysis is accurate enough to be used to aid health research.[20]

This ethnicity classification tool was trained on a large, diverse set of names which the authors claim cover 90% of the world's names.[17] It was developed in the USA and so the six ethnicity categories used are American: white, black, Asian and Pacific Islander, Hispanic, American Indian and Alaska Native and more than two races. Ethnicity was estimated using the first and last name of authors, which aims 'to reduce errors when names are mixtures because of immigration or cross-nationality marriages'. This tool is designed and tuned to infer on a world population, and so is well suited to a journal such as the BMJ with a worldwide authorship.

Gender of first author name was determined using a tool called Gender Guesser, which uses a database of approximately 40 000 common names and their corresponding gender.[21] The first names of rapid response authors are checked against this database, and placed into one of the following categories: male, female, mostly male, mostly female, androgynous (equal probabilities of being male or female) and unknown (not in the database). In an independent validation, on a manually labelled dataset of 7076 names, it was compared with four other such gender inference tools, and was found to achieve 'the lowest misclassification rate without parameter tuning for the entire dataset, introducing also the smallest gender bias'.[22] We also validated the Gender Guesser tool on a public dataset of 29 872 names extracted from Wikipedia. The tool was able to infer gender for 82.76%. The names inferred as 'male' were 99.2% accurate, and those inferred as 'female' were 95.6% accurate. Overall, this tool was 98.4% accurate when detecting 'male' or 'female' names in our validation dataset.

The ethnicity and gender classification tools provide an estimated ethnicity and gender that, for the purpose of this study, is assumed to be analogous to the ethnicity and gender that a reader or reviewer would assign to an author.

Statistical analysis

Univariate associations between author and rapid response features, and publication was carried out by calculating χ^2 test and t-test scores. Hierarchical binary logistic regression was used to look at the correlation between ethnicity and publication, taking into account other author and rapid response features.

Software used

Gender of first author names was classified using Gender Guesser.[21] Ethnicity was classified using nameprism.com.[17] Flesch reading ease score was calculated using an open source library called textstat.[23] Spelling and grammatical mistakes were quantified using a tool called language-check.[24]

All code was written in 'Python 3' in the 'Jupyter notebook' text editor.[25] Data collection used an automated script, using the open source Python libraries 'Requests' and 'BeautifulSoup'.[26 27] Further processing and data manipulation used the Python libraries 'NumPy', 'Pandas' and 'SciKit learn'.[28–30] Statistical analysis was carried out in the IBM SPSS V.25 package and in Python.

RESULTS

Baseline data

Analysis was performed on 113 265 rapid responses, of which 8415 (7.4%) were published as letters to the editor. Of all submitted rapid responses, 83% had first authors with names classed as 'white'; 62% of first authors were classed as 'male'. See table 1 for baseline author and rapid response features. We also performed an analysis of the characteristics of submissions, broken down by inferred gender and ethnicity. These can be found in online supplemental tables 1 and 2.

Univariate analysis

Univariate associations were analysed and included in table 1.

Multivariate analysis

All variables above were used in a hierarchical, binary logistic regression with two blocks. The first block included all variables except first author's gender and ethnicity. The second block additionally contained first author's gender and ethnicity.

First author's gender and ethnicity remained statistically significant after accounting for measured confounders. In the second block, incorporating this information significantly improved the model (omnibus test of model coefficients, χ^2=4648.412, degrees of freedom=43, p<0.0005), the pseudo R^2 value was 0.098, up from 0.088 in the first block.

Table 2 shows the results of the second, complete logistic regression and ORs for each variable.

DISCUSSION

Statement of principal findings

The estimated gender and ethnicity of first author names of BMJ rapid responses were predictive of publication, even when other features of the rapid response and the author were taken into account.

Strengths and limitations of the study

To the best of our knowledge, this is the largest ever analysis of a scientific corpus that looks for associations with publication rate. This was possible because of the open nature of BMJ rapid responses, and through automation at various stages, including data gathering and processing, which includes the use of validated machine learning algorithms for automatic ethnicity classification. This allowed us to analyse over 100 000 submissions, a feat which would not have been possible manually.

One of the largest limitations of this study is that it is only sensitive to associations. It is not possible to infer causality from this data, when the exact mechanisms for the observed discrepancy are not known. There could be other unmeasured factors accounting for the discrepancy in publication rates such as subtle differences in communication style, which have been posited to explain at least partly the ethnicity attainment gap in medical school clinical examinations.[31] It is worth noting, however, that in clinical examinations it is male students that are consistently found to underperform relative to their female counterparts,[32 33] while we found that female first authors were under-represented.

Though the tool for classifying ethnicity from name has been validated on a global population, it was developed in the USA and uses ethnicity categories that closely resemble those officially used within the USA. This is not ideal for names outside the USA, where different categories are defined officially. The categorisation of gender resulted in a fairly large proportion (17.3%) of authors with unknown gender, and they were less likely to be published. Another similar tool for inferring gender from name was shown to have an overall 93.8% accuracy in classifying author names in an analysis in the journal *Science*.[34] This high accuracy is due to these techniques' ability to quantify their uncertainty; for example, if they believe there is a roughly equal chance of the name being male or female, it is classed as 'androgynous'. Only names which are very likely to be a specific gender are inferred as such.

Although there was an option to add location to the gender tool to determine the likely gender of a name in a specific country, this was not done for two reasons. First, the location data extracted from rapid responses was highly heterogeneous with some authors providing countries, cities or institution names, or multiple addresses, without consistent spelling or abbreviations. Second, of the 45 376 names in the Gender Guesser dictionary, only 286 (0.6%) names are influenced sufficiently by location such that the estimate is changed from 'male' to 'female' or vice versa.

In this study, gender and ethnicity were estimated from the author's name, which provides a proxy for the gender and ethnicity that a reviewer would assign to an author given the same information. While assigning ethnicity

Table 1 Characteristics of published and unpublished rapid responses submitted to BMJ.com between 25 April 1998 and 23 March 2018

Characteristic	All submissions (n=113 265)	Published (n=8415)	Unpublished (n=104 850)	Statistical significance
Author gender (number (%))				χ^2=181,p<0.0005
Male	70 256 (62.0)	5636 (67.0)	64 620 (61.6)	
Female	18 592 (16.4)	1409 (16.7)	17 183 (16.4)	
Mostly male	2434 (2.1)	171 (2.0)	2263 (2.2)	
Mostly female	1321 (1.2)	98 (1.2)	1223 (1.2)	
Androgynous	1021 (0.9)	82 (1.0)	939 (0.9)	
Unknown	19 641 (17.3)	1019 (12.1)	18 622 (17.8)	
Author ethnicity (number (%))				χ^2=267, p<0.0005
White	94 077 (83.1)	7492 (89.0)	86 585 (82.6)	
API	15 759 (13.9)	726 (8.6)	15 033 (14.3)	
Hispanic	1903 (1.7)	90 (1.1)	1813 (1.7)	
Black	1204 (1.1)	64 (0.8)	1140 (1.1)	
AIAN	2 (0.0)	0 (0.0)	2 (0.0)	
Unknown	320 (0.3)	43 (0.5)	277 (0.3)	
Word count	314±318	410±278	307±319	t=−32.4, p<0.0005
Flesch reading ease	50.3±16.3	47.6±12.1	50.5±16.5	t=20.8 p<0.0005
Has references (number (%))	40 173 (35.5)	4445 (52.8)	35 728 (34.1)	χ^2=1196, p<0.0005
Number of references	1.3±3.0	2.2±3.3	1.3±2.9	t =−26.6, p<0.0005
Author position (number (%))				128, p<0.0005
Consultant	16 291 (14.4)	1592 (18.9)	14 699 (14.0)	
Professor	9959 (8.8)	1110 (13.1)	8849 (8.4)	
Senior	4491 (4.0)	523 (6.2)	3968 (3.8)	
Student	3080 (2.7)	143 (1.7)	2937 (2.8)	
Other	79 444 (70.1)	5047 (60.0)	74 397 (71.0)	
Twitter handle present (number (%))	1868 (0.2)	254 (0.3)	1614 (0.2)	χ^2=105, p<0.0005
US spelling and grammar errors	27.0±44.6	35.2±39.2	26.3±44.9	t=−19.9, p <0.0005
UK spelling and grammar errors	8.9±18.0	9.4±15.3	8.9±18.2	t=−3.0, p=0.003
Multiple authors (number (%))	19 256 (17.0)	2914 (34.6)	16 342 (15.6)	χ^2=2002, p<0.0005
Competing interests declared (number (%))	6184 (5.5)	924 (11.0)	5260 (5.0)	χ^2=537, p<0.0005

Plus–minus values are means±SD. Percentages may not sum to 100% due to rounding.
AIAN, American Indian and Alaska Native; API, Asian and Pacific Islander.

and gender based on name may not always match the self-identified ethnicity and gender of the author, it is practical and necessary in this setting. A difficulty in classifying both ethnicity and gender into discreet categories is that the nuances of these complex social identities is lost, for example, in the gender tool there is no categorisation for non-binary genders or transgender individuals, and limited provision for mixed race individuals in the ethnicity tool.

A limitation in this work is that for all analysed letters, the corresponding rapid response had to be imputed using the protocol mentioned previously. A small minority of letters seem not to have been submitted as rapid responses, which may represent either direct publications from the editor or direct correspondences between a paper author and the editor. It is worth noting, however, that recently published letters link directly to the original rapid response that was submitted. This would allow future analysis to have ground truth data on which rapid responses were published and which were not.

Strengths and limitations in relation to other studies

A recent report by *Science* found no statistically significant evidence that their editorial process under-represented

Table 2 ORs with 95% CIs and p values in multivariate analysis

Variable	Block 1: ORs (95% CI)	Block 1: p value	Block 2: ORs (95% CI)	Block 2: p value
Gender				
Male	–	–	–	<0.0005
Female	–	–	0.90 (0.85–0.96)	0.002
Mostly male	–	–	0.97 (0.83–1.14)	0.727
Mostly female	–	–	0.98 (0.80–1.22)	0.882
Androgynous	–	–	1.02 (0.80–1.29)	0.901
Unknown	–	–	0.75 (0.69–0.81)	<0.0005
Ethnicity				
White	–	–	–	<0.0005
API	–	–	0.54 (0.49–0.59)	<0.0005
Hispanic	–	–	0.51 (0.41–0.64)	<0.0005
Black	–	–	0.74 (0.57–0.96)	0.023
Unknown	–	–	1.39 (0.99–1.96)	0.057
Word count	1.00 (1.00–1.00)	<0.0005	1.00 (1.00–1.00)	<0.0005
Flesch reading ease	1.00 (1.00–1.00)	0.016	1.00 (1.00–1.00)	0.042
Has references	0.71 (0.67–0.75)	<0.0005	0.70 (0.67–0.74)	<0.0005
Number of references	1.01 (1.00–1.02)	0.290	1.00 (0.99–1.01)	0.475
Author position				
Consultant	0.69 (0.65–0.73)	<0.0005	0.70 (0.65–0.74)	<0.0005
Professor	0.75 (0.70–0.81)	<0.0005	0.74 (0.69–0.79)	<0.0005
Senior	0.73 (0.66–0.81)	<0.0005	0.74 (0.67–0.81)	<0.0005
Student	1.64 (1.38–1.95)	<0.0005	1.54 (1.30–1.83)	<0.0005
Twitter handle present	0.69 (0.60–0.79)	<0.0005	0.69 (0.60–0.79)	<0.0005
US spelling and grammar errors	1.00 (1.00–1.00)	<0.0005	1.00 (1.00–1.00)	0.001
UK spelling and grammar errors	1.00 (0.99–1.00)	<0.0005	0.99 (0.99–1.00)	<0.0005
Multiple authors	0.47 (0.44–0.49)	<0.0005	0.44 (0.42–0.46)	<0.0005
Competing interests declared	0.56 (0.52–0.60)	<0.0005	0.57 (0.53–0.61)	<0.0005
Recent submission	1.00 (1.00–1.00)	<0.0005	1.00 (1.00–1.00)	<0.0005
Near ideal word count	0.54 (0.51–0.57)	<0.0005	0.54 (0.51–0.57)	<0.0005
Near ideal Flesch reading ease	0.80 (0.76–0.85)	<0.0005	0.80 (0.76–0.85)	<0.0005

'American Indian and Alaska Native' was removed from the ethnicity figures due to an absence of published letters from that ethnic group.
API, Asian and Pacific Islander.

female authors.[35] In this report, gender was identified manually, which limited their sample size to a small random selection of submissions. The gender disparity we observed is relatively subtle compared with that which we observed for ethnicity, and this might mean that larger sample sizes are needed to elucidate any discrepancies.

The under-representation of ethnicity has so far been less thoroughly studied in comparison to gender. Nonetheless, our findings are in line with published data that BME authors are under-represented in published articles[10]; however, our study is further able to identify that submissions from BME authors are less likely to be accepted for publication relative to similar submissions by their white counterparts.

Implications

In our hierarchical logistic regression, gender and ethnicity explain a small amount of additional variance (0.01 increase in the pseudo R^2) compared with the other features alone; however, this is considerable compared with a low pseudo R^2 baseline of 0.088.

Given the likely complexity of the selection process, it is unsurprising that the pseudo R^2 is low, as there are many unmeasured factors. Indeed, publication is determined by the expert opinion of the editor, on things that are impractical or impossible to quantify in a study like this, including clarity, style and interest to the potential reader.

Nonetheless, one factor which may play a role is that of unconscious gender or racial bias. Implicit bias has

been documented in clinical decision-making,[36] medical school admissions[37] and selection of junior doctors.[38] It is important to state, however, that the current study design does not provide causal evidence of bias, which would require a prospective experimental study design to fully account for other unmeasured factors.

Our results suggest that scientific journals should look for such discrepancies in all forms of submissions, including opinion pieces and research papers, which are not posted publicly. Such analysis should include looking at differences in submission and publication rates by author ethnicity as well as by author gender.

This study demonstrates the ability of data science and machine learning techniques to rapidly extract and analyse a large and complex dataset with relative ease. Without these techniques, this analysis would not have been possible. Being able to automate the process of feature extraction, including gender and ethnicity, opens avenues for further observational studies of open access data. It has also opened countless possibilities in medicine across the entire patient journey, from triage and improving attendance, to automated disease diagnosis, prognostication, management and even the discovery or repurposing of new medications.[39–45]

Unanswered questions and future research

This work highlights important associations in past data; however, more research is necessary to draw concrete conclusions regarding the reasons for these associations. For example, other confounders might be considered, including communication style, field of study of the rapid response author and of the article being responded to, and the locations of institutes of submitted pieces. It may also be interesting to see how these discrepancies change over time, if at all. Additionally, studies have demonstrated unconscious gender biases in science,[12] and unconscious racial biases in other areas,[36–38] but far less research has studied unconscious racial biases in academia.

It is important to establish whether the discrepancies we found in BMJ letters to the editor are present in other journals, and for other scientific manuscript types such as original research. Though trends have been studied for published papers, quantifying the rate of acceptance is an invaluable way to eliminate the confounder that is the number of submissions, and we hope that future research in this field can either be done by journals themselves, or by researchers in close collaboration with journals to ensure that this submission data is included in any analysis.

CONCLUSION

A number of variables were identified that correlated with the acceptance rate of rapid responses. Discrepancies in the publication rates between genders and ethnicities remained significant after accounting for other factors. The cause of these discrepancies is unclear and may in part be explicable by implicit bias. Regardless of the cause, it is evident that female and BME voices are underrepresented, and efforts should be made to identify these causes and rectify them.

Twitter Amitava Banerjee @amibanerjee1 and Katherine Woolf @kathwoolf

Contributors MZ: conceptualisation, design and coordination of project, lead in programming and statistical analysis, and writing first draft. ABal: contribution to programming, statistical analysis and writing first draft. ABan: aiding in conceptualisation, offering statistical advice and proofreading. KW: conceptualisation and design, offering statistical advice, and proofreading.The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data may be obtained from a third party and are not publicly available. The first author (mohamad.zeina@nhs.net) will share any publicly available data if requested by email. Some data was obtained through third parties, and may be limited or omitted at the third parties' discretion.

ORCID iDs
Mohamad Zeina http://orcid.org/0000-0002-8119-1931
Amitava Banerjee http://orcid.org/0000-0001-8741-3411
Katherine Woolf http://orcid.org/0000-0003-4915-0715

REFERENCES

1 Woolf K, Potts HWW, McManus IC. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. *BMJ* 2011;342:d901.
2 General Medical Council. The state of medical education and practice in the UK, 2015. Available: https://www.gmc-uk.org/-/media/documents/somep-2015_pdf-63501874.pdf [Accessed cited 2019 Jul 20].
3 Appleby J. Ethnic pay gap among NHS doctors. *BMJ* 2018;362:k3586. doi:10.1136/bmj.k3586
4 Ginther DK, Schaffer WT, Schnell J, *et al*. Race, ethnicity, and NIH research awards. *Science* 2011;333:1015–9.
5 Aamc. Analysis in Brief - July 2005: The changing representation of men and women in academic medicine [Internet], 2005. Available: www.aamc.org/data/aib [Accessed 14 Mar, 2019].
6 Rochon PA, Davidoff F, Levinson W. Women in academic medicine leadership. *Academic Medicine* 2016;91:1053–6 http://insights.ovid.com/crossref?an=00001888-201608000-00014
7 Hopkins AL, Jawitz JW, McCarty C, *et al*. Disparities in publication patterns by gender, race and ethnicity based on a survey of a random sample of authors. *Scientometrics* 2013;96:515–34 http://link.springer.com/
8 Sidhu R, Rajashekhar P, Lavin VL, *et al*. The gender imbalance in academic medicine: a study of female authorship in the United Kingdom. *J R Soc Med* 2009;102:337–42 http://journals.sagepub.com/doi/

9 West JD, Jacquet J, King MM, *et al*. The role of gender in scholarly authorship. Hadany L, editor. *PLoS One*2013;8:e66212 https://dx.plos.org/10.1371/journal.pone.0066212
10 Marschke G, Nunez A, Weinberg BA, *et al*. Last place? The intersection of ethnicity, gender, and race in biomedical. *AEA Pap Proc* 2018;108:222–7 https://www.aeaweb.org/doi/
11 Moss-Racusin CA, Dovidio JF, Brescoll VL, *et al*. Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci U S A* 2012;109:16479http://www.ncbi.nlm.nih.gov/pubmed/22988126
12 Knobloch-Westerwick S, Glynn CJ, Huge M. The Matilda effect in science communication. *Sci Commun* 2013;35:603–25 http://journals.sagepub.com/doi/
13 Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. *Proc Natl Acad Sci USA* 2011;108:3162http://www.ncbi.nlm.nih.gov/pubmed/21300892
14 Boll C, Leppin J, Rossen A, *et al*. Magnitude and impact factors of the gender pay gap in Eu countries [cited 20 Jul, 2019], 2016. Available: www.fondazionebrodolini.it
15 Blau FD, Kahn LM. The gender wage gap: extent, trends, and explanations. *J Econ Lit* 2017;55:789–865 http://pubs.aeaweb.org/doi/
16 Weaver AC, Wetterneck TB, Whelan CT, *et al*. A matter of priorities? Exploring the persistent gender pay gap in hospital medicine. *J Hosp Med* 2015;10:486–90 http://www.journalofhospitalmedicine.com/jhospmed/article/127833/priorities-and-gender-pay-gap
17 Ye J, Han S, Hu Y, *et al*. *Nationality Classification Using Name Embeddings. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management - CIKM '17 [Internet]*. New York, New York, USA: ACM Press, 2017: 1897–906. http://dl.acm.org/citation.cfm?doid=3132847.3133008
18 Mateos P. A review of name-based ethnicity classification methods and their potential in population studies. *Popul Space Place* 2007;13:243–63 http://doi.wiley.com/
19 Banda Y, Kvale MN, Hoffmann TJ, *et al*. Characterizing Race/Ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* 2015;200:1285–95 http://www.ncbi.nlm.nih.gov/pubmed/26092716 doi:10.1534/genetics.115.178616
20 Lakha F, Gorman DR, Mateos P. Name analysis to classify populations by ethnicity in public health: validation of Onomap in Scotland. *Public Health* 2011;125:688–96 https://www.sciencedirect.com/science/article/pii/S0033350611001508
21 Dukebody. Gender Guesser [Internet]. GitHub, 2018. Available: https://github.com/lead-ratings/gender-guesser
22 Santamaría L, Mihaljević H. Comparison and benchmark of name-to-gender inference services. *PeerJ Comput Sci* 2018;4:e156.
23 Bansal S, Aggarwal C. *textstat*, 2018.
24 Myint. *language-check*, 2017.
25 Project Jupyter. *Jupyter Notebook [Internet]*, cited 2019 Apr 16. https://jupyter.org/index.html
26 kennethreitz. Requests: HTTP for HumansTM [Internet], cited 2019 Apr 16. Available: http://docs.python-requests.org/en/master/
27 Richardson L. Beautiful Soup [Internet]. [cited 2019 Apr 16]. Available: https://www.crummy.com/software/BeautifulSoup/
28 Oliphant T. NumPy [Internet]. [cited 2019 Apr 16]. Available: http://www.numpy.org/
29 McKinney W. Python Data Analysis Library [Internet]. [cited 2019 Apr 16]. Available: https://pandas.pydata.org/
30 Cournapeau D. scikit-learn: machine learning in Python [Internet]. [cited 2019 Apr 16]. Available: https://scikit-learn.org/stable/
31 Wass V, Roberts C, Hoogenboom R, *et al*. Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study. *BMJ* 2003;326:800–3 http://www.ncbi.nlm.nih.gov/pubmed/12689978
32 Woolf K, Haq I, McManus IC, *et al*. Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations. *Adv Health Sci Educ Theory Pract* 2008;13:607–16 http://link.springer.com/
33 Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting early performance on the medical course. *Med Educ* 2004;38:1002–5 http://doi.wiley.com/
34 New tools for gender analysis | Sciencehound [Internet]. [cited 2020 Feb 14]. Available: https://blogs.sciencemag.org/sciencehound/2019/01/03/new-tools-for-gender-analysis/
35 Berg J. Looking inward at gender issues. *Science* 2017;355:320–329 http://www.sciencemag.org/lookup/doi/
36 Green AR, Carney DR, Pallin DJ, *et al*. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *J Gen Intern Med* 2007;22:1231–8 http://www.ncbi.nlm.nih.gov/pubmed/17594129
37 Capers Q, Clinchot D, McDougle L, *et al*. Implicit racial bias in medical school admissions. *Acad Med* 2017;92:365–9 http://insights.ovid.com/crossref?an=00001888-201703000-00032
38 Esmail A, Everington S. Racial discrimination against doctors from ethnic minorities. *BMJ* 1993;306:691–2.
39 De Fauw J, Ledsam JR, Romera-Paredes B, *et al*. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50 http://www.nature.com/articles/s41591-018-0107-6
40 Nelson A, Herron D, Rees G, *et al*. Predicting scheduled Hospital attendance with artificial intelligence. *NPJ Digit Med* 2019;2:1–7.
41 Rajpurkar P, Irvin J, Zhu K, *et al*. CheXNet: Radiologist-Level pneumonia detection on chest x-rays with deep learning. [cited 2018 Dec 15], 2017. Available: http://arxiv.org/abs/1711.05225
42 Alaa AM, van der Schaar M. Prognostication and risk factors for cystic fibrosis via automated machine learning. *Sci Rep* 2018;8:11242 www.nature.com/scientificreports
43 Longmire M. US8548828B1 - Method, process and system for disease management using machine learning process and electronic media - Google Patents [Internet]. [cited 2020 Jul 26]. Available: https://patents.google.com/patent/US8548828B1/en
44 Stokes JM, Yang K, Swanson K, *et al*. A deep learning approach to antibiotic discovery. *Cell* 2020;180:e13:688–702.
45 Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. vol. 20, the Lancet oncology. *Lancet Publishing Group* 2019:e262–73.