Government Equalities Office

THE BEHAVIOURAL INSIGHTS TEAM

Increasing applications from women through targeted referrals

Research report

June 2021

Leonie Nicks, Vivek Roy-Chowdhury, Tim Hardy, Filip Gesiarz, and Hannah Burd - The Behavioural Insights Team

Contents

Acknowledgements

We wish to acknowledge the roles that the following individuals had in this report:

- Paula Fredersdorff at the Ministry of Defence (MOD) as MOD project lead.
- Jake MacNaughton and Jo Debnam in the recruitment team at MOD for implementing the trial.
- Abigail Fox and Rayman Sandhur at MOD for management of the data.
- Natalia Shakhina at the Behavioural Insights Team (BIT) for writing up the trial into this report.
- Alasdair Smith at BIT for building the Excel macro to semi-automate the implementation of the trial.
- Tiina Likki at BIT for setting up the partnership.
- Jessica Barnes, James Lawrence, Rony Hacohen, Tom O'Keefe and Ariella Kristal at BIT for their quality assurance input.

Executive summary

Executive Summary

Occupational segregation is one of the key factors explaining the gender pay gap in the UK.[1] Women make up the majority in lower paying occupations and the minority in higher paying occupations.[2] One of the barriers for more women to enter higher paying occupations is that they have smaller networks in these fields, which reduces the likelihood of being referred for a job.[3] Personal recommendations are one of the key hiring channels,[4] but it can limit diversity if employees only refer people who are similar to them.[5]

We partnered with the Ministry of Defence (MOD) to run a two-armed randomised controlled trial (RCT) testing whether using targeted referrals would increase the referrals, applications and hires of women. As hiring managers advertised new vacancies in areas of MOD where women had been historically underrepresented, they were randomly allocated to either the control group or the intervention group.

Managers in the intervention group received an email inviting them to challenge their team to share the role with five women. Managers in the control group did not receive this communication. We ran the trial for five months, from August 2020 to December 2020, and the sample consisted of all 784 hiring managers for 1,052 vacancies created in the trial period, which received 18,841 applications.

Targeted referrals improved the gender balance among referrals (54% women), while referrals in the control group reflected the gender make-up of the organisation (40% women). Twice as many women were referred to a vacancy in the intervention group.[6] The intervention increased the number of applications from women and the number of offers made to women. However, it also resulted in more men applying and receiving offers. As a result, the intervention did not lead to significant changes in the share of applications from women and the share of offers made to women. There was no backfire effect on the share of applications from other minority applicants. Finally, the intervention had a positive impact on the quality of applicants as vacancies in the intervention group were significantly more likely to find a suitable candidate compared to the control group.

Our results suggest that targeted referrals can rebalance existing inequality in informal referrals. We recommend that organisations consider testing encouraging their employees to share roles with people they know from a wider range of underrepresented groups.

---

[1] Olsen, W., Gash, V., Kim, S., & Zhang, M. (2018). The gender pay gap in the UK: evidence from the UKHLS
[2] Francis-Devine, B., Ferguson, D. (2020). The Gender Pay Gap. Briefing paper Number 7068.
[3] Das, S., & Kotikula, A. (2019). Gender-based employment segregation: Understanding causes and policy interventions. World Bank.
[4] LinkedIn. (2017). Global Recruiting Trends 2017.
[5] Brown, M., Setren, E., & Topa, G. (2016). Do informal referrals lead to better matches? Evidence from a firm's employee referral system. *Journal of Labor Economics*, 34(1), 161-209.
[6] Control variables include vacancy's region, business area, grade and manager type (could be the line manager or the person who created the vacancy).

Introduction

Introduction

One of the causes of the gender pay gap (GPG) is occupational segregation: the share of women is greater in lower-paying occupations and smaller in higher-paying ones compared to men.[7] Occupational segregation accounted for 19% of the GPG in 2014/2015.[8] In a report prepared for the House of Commons, pay data was analysed for 350 different occupations categorised into four pay groups. This revealed that 30% of female employees worked in the lowest-paying occupations compared to 17% of men. At the same time, 18% of female employees worked in the highest paying occupations compared to 23% of men.[9] One of the contributing factors is that women may have smaller networks and fewer role models in occupations traditionally dominated by men.[10] This in turn reduces the likelihood of women entering these occupations. To mitigate this, it is important to find ways to increase applications from women into male-dominated occupations and roles.

A significant route for recruiting new employees is through personal recommendations.[11] Many organisations have formal referral schemes in place to encourage and incentivise this way of finding new applicants. However, people in our personal networks are likely to be similar to us[12] and employees tend to refer those who are similar to them in terms of gender, age and other characteristics.[13] With or without a formal referral scheme, employees are likely to share vacancies with people who are similar to them, affecting the diversity of the applicant pool and ultimately of the company.

Inviting employees to refer more diverse candidates could be an effective way to rebalance personal recommendations. In one study, among participants asked to refer someone for a job vacancy, men referred more men (77%) while women referred women at the same rate at which they applied themselves (43%). However, when asked to refer women, men referred women in similar numbers and of similar quality to the male candidates they referred when gender was not specified.[14] To address the underrepresentation of certain groups in engineering roles, the tech company Pinterest asked their employees to refer women and people from underrepresented ethnic backgrounds. They claimed that there was a 24% increase in the percentage of women referred and a 55-times increase in the percentage of applicants from underrepresented ethnic backgrounds.[15]

[7] Francis-Devine, B., Ferguson, D. (2020). The Gender Pay Gap. Briefing paper Number 7068.
[8] Olsen, W., Gash, V., Kim, S., & Zhang, M. (2018). The gender pay gap in the UK: evidence from the UKHLS
[9] Francis-Devine, B., Ferguson, D. (2020). The Gender Pay Gap. Briefing paper Number 7068.
[10] Das, S., & Kotikula, A. (2019). Gender-based employment segregation: Understanding causes and policy interventions. World Bank.
[11] LinkedIn. (2017). Global Recruiting Trends 2017.
[12] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415-444.
[13] Brown, M., Setren, E., & Topa, G. (2016). Do informal referrals lead to better matches? Evidence from a firm's employee referral system. *Journal of Labor Economics*, 34(1), 161-209.
[14] Beaman, L., Keleher, N., & Magruder, J. (2018). Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi. *Journal of Labor Economics*, 36(1), 121–157.
15 Pinterest (2016) Diversifying Engineering Referrals at Pinterest. It is unclear whether this initiative was rigorously evaluated.

Part of the reason that targeted recruitment approaches can be effective is that they invite managers to be part of the solution. Such engagement approaches have been shown to be more effective than compliance-based approaches.[16] Key features of these approaches are that they are voluntary and engage managers in active problem-solving. Organisations that implement targeted recruitment see increases in women in management after five years.[17]

We partnered with the Ministry of Defence (MOD) to test whether targeted referrals would increase the gender diversity of their workforce. The trial included the MOD Top Level Budget (TLB)[18] civilian workforce, numbering more than 39,000 employees.[19] Men made up 57% of the TLB workforce in 2019.[20] The civilian MOD TLB had a median pay gap of 10.5%[21] in 2019 (and a mean of 9.6%) compared to 11.1% across the whole of the Civil Service.[22]

This research is part of a three-year collaboration between the Behavioural Insights Team (BIT) and the Government Equalities Office (GEO): the Gender and Behavioural Insights (GABI) programme. The programme aims to generate evidence for what works to improve gender equality in the workplace.

---

16 Dobbin, F., & Kalev, A. (2017). Are Diversity Programs Merely Ceremonial? Evidence-Free Institutionalization. In *The SAGE Handbook of Organizational Institutionalism* (pp. 808–828).
17 Dobbin, F., Schrage, D., & Kalev, A. (2015). Rage against the Iron Cage: The Varied Effects of Bureaucratic Personnel Reforms on Diversity. *American Sociological Review*, 80(5), 1014–1044.
18 MOD is organised into seven TLBs: Navy Command, Army Command, Air Command, Joint Forces Command, The Defence Infrastructure Organisation, The Defence Nuclear Organisation, Head Office and Corporate Services
19 MOD (2021) MOD Workforce return to the Cabinet Office
20 MOD (2019) MOD Gender Pay Gap report 2019
21 MOD (2019) MOD Gender Pay Gap report 2019
22 Civil Service blog (2020) Update on the gender pay gap 2019

Research aims and trial methodology

Research aims and trial methodology

Exploratory data research and findings
Before deciding where to focus, we analysed MOD's HR data on recruitment, pay, bonuses, progression, promotions, retention, performance, sick leave, parental leave and more from 2013-2018 to examine the drivers of its GPG.[23] In addition to examining 'raw' data, we compared similar employees by accounting for a number of factors (e.g. grade, tenure, job type, region, working pattern) that might otherwise vary between men and women. Our key findings are summarised below:

- Almost all grades were male-dominated at MOD apart from one of the lowest grades (E1). A 50:50 gender split and equalised pay at each grade for almost all grades[24] would decrease their GPG by between 1.02 and 6.96 percentage points (pp) depending on the grade.

- In most grades, women were less than 40% of applicants and the share of women applying was lower than the share of women working in those grades at MOD. Even though women were less likely than men to apply, they were more 2pp successful in getting hired when they did.

- Once in MOD, women in similar roles to men were promoted at 1.6pp higher rates after a similar length of time and 7pp more likely to receive the highest performance score.

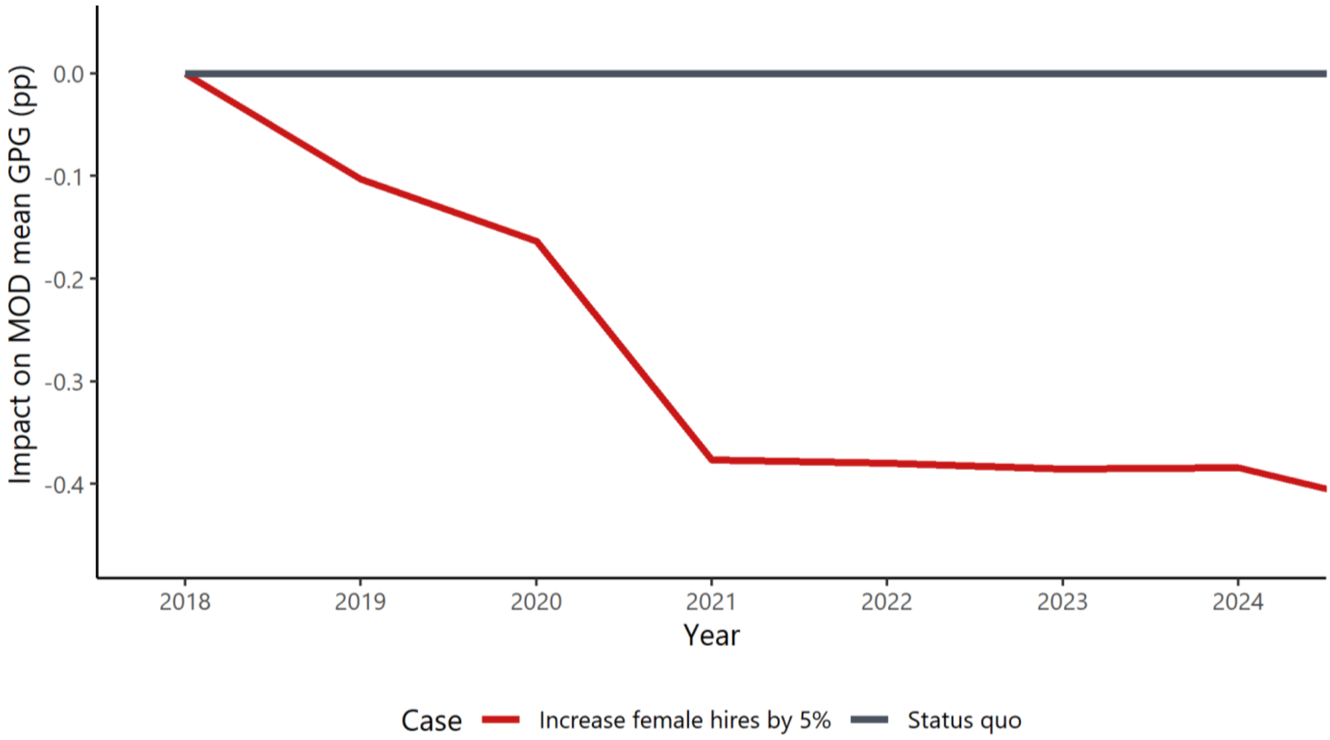
- Part-time workers experienced penalties in performance ratings, bonuses, promotions, and were more likely to resign. Women made up the majority (74%) of part-time workers.

- The uptake of Shared Parental Leave was low despite enhanced pay and while men experienced a 5-15pp boost in promotions in the years after taking leave, women did not.

We ran a simulation analysis to estimate the impact of a range of interventions on MOD's GPG compared to no intervention. We projected forwards the full dataset using annual changes in outcomes for gender, grade and full-time/part-time groups. While the GPG would decrease without intervention over time, this analysis indicated that increasing women hires by 5% at every grade would reduce the GPG several years faster than status quo (Figure 1).

[23] The availability of data for this time period varied by HR process
[24] Apart from the relatively lower paid and male-dominated 'industrial' grade

Figure 1: Impact on MOD's GPG relative to status quo (no intervention)



Qualitative research and findings

We interviewed five managers and employees at MOD to understand their experiences at MOD. These interviews were focused on the performance and bonus process and differing experiences for part-time and full-time employees. We learned that MOD employees take pride in their work, and that monetary incentives, in the form of bonuses, do not strongly motivate them to perform in their daily work. There was an ethos of doing work well for its own sake out of professional pride. However, bonuses were appreciated when given and made employees feel valued. This professional pride suggests that employees would be effective champions of MOD and that an incentives-based referrals scheme could be counterproductive.[25] Positive feedback for those who made referrals could communicate that those actions are valued.

Intervention development

Based on our analysis and discussions with MOD, we decided that increasing applications from women to MOD was the most promising opportunity for reducing the GPG. We designed an intervention to encourage employees to share open vacancies at MOD with women they knew and encourage them to apply. Hiring managers for civilian external and cross-government vacancies were randomly allocated to either the control or the intervention group as summarised in Table 1 below.

[25] It is worth noting that an incentive-based referral scheme was not under consideration at the time.

Table 1: Condition summary

Group	Description
Control group	Hiring managers did not receive an invitation to join the ‘Referrals Challenge’.
Intervention group	Hiring managers received an email inviting them to ask their team to share the role with women in their networks as part of the ‘Referrals Challenge’.

The intervention included the following behaviourally-informed components:

1. **Timely moment:** Managers received this email shortly after they created a new vacancy (within the week after) and were most likely to be concerned with reaching applicants. Messages sent at the ‘right moment’ are often effective at prompting action.[26]

2. **Personalised content:** The email was addressed directly to the hiring manager and referenced the vacancy they had created by name. Personalising communication can help attract attention.[27]

3. **Achievable target:** The email asked the manager to challenge their team to reach out to five women. This provided a reference point and a motivational target (‘anchor’) that would be sufficiently high but still attainable based on a typical team size of 5-7 people. Research shows that goals can be achieved by setting personal targets, receiving feedback on performance, and recognition when goals have been achieved.[28]

4. **Small actionable steps:** The email explained what managers needed to do and provided specific tips. When we want people to take an action, it can help to divide information into easy-to-digest chunks and straightforward, specific next steps.[29]

5. **Transparency and feedback:** The email included a leaderboard where business areas were ranked by the number of women applicants. This created a non-financial incentive for employees to participate, but also signalled that their participation was being tracked.[30] Providing positive and relative feedback can also improve performance.[31]

[26] For example, Hoff, G., & Bretthauer, M. (2008). Appointments timed in proximity to annual milestones and compliance with screening: randomised controlled trial. *Bmj*, *337*.
[27] Shapiro, K. L., Caldwell, J., & Sorensen, R. E. (1997). Personal names and the attentional blink: A visual" cocktail party" effect. *Journal of Experimental Psychology: Human Perception and Performance*, 23(2), 504.
[28] Gosnell, G. K., List, J. A., & Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, 128(4), 1195-1233.
[29] Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *TRENDS in Cognitive Sciences*, 5(6), 236–243.
[30] The Behavioural Insights Team (2017) Update Report 2016-17. Previous BIT work found that providing specific team performance in terms of ranking and average number of steps increased activity in a workplace trial using Fitbits.
[31] Murthy, U. S., & Schafer, B. A. (2011). The effects of relative performance information and framed information systems feedback on performance in a production task. *Journal of Information Systems*, 25(1), 159–184.

6. **Respected messenger:** The email was signed off by the HR Director. The impact of information tends to be higher if it comes from someone with authority in the field, someone like us or someone we like.[32]

We ran a co-design workshop with four HR Business Partners from different TLBs to develop the content of the emails. We then conducted user-testing with managers to ensure the content was easy to understand and appealing, and incorporated their feedback.

Figure 2: Email Content[33]

Dear Jane,

I see that you are seeking candidates for Civilian Driver. We are piloting an initiative called the **Referrals Challenge** to create a more diverse workforce, starting by encouraging more women to apply for our roles. The lockdown period has highlighted pre-existing gender inequalities across society, which has reinforced the MOD's commitment to closing its gender pay gap. The **Referrals Challenge** is supported by the Executive Committee as part of the MOD's Diversity and Inclusion plan to improve gender equality. If it is successful, we intend to expand it to improve diversity in other areas.

What is the Referrals Challenge?

The **Referrals Challenge** invites employees to encourage women they know to apply to roles in areas of the MOD that struggle to attract women. Can you challenge your team to reach out to **five women** they either know directly or through other contacts they may have in their networks? Help Air TLB to reach the top of the leader board:

Rank	Business area	Business area referrals
1	Head Office	76
2	DIO	28
3	Navy	13
4	DNO	10
5	Army	8
6	Air	3

Referrals are based on applicant data from closed campaigns where they have indicated a) they heard about the role from an MOD employee and b) they are a woman. As referrals are indicated in a free-text box, we may not pick up on all of them. We will analyse the responses in detail at the end of the initiative for final numbers and will update you on the final count.

[32] Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, 33(1), 264-277.
[33] The content of the emails was designed following a workshop with HR Business Partners from a number of TLBs and was user-tested with employee managers. The addressee, role and leader board displayed here are fictional to protect confidentiality.

Your team may know women who would be great for the MOD, but they may not have considered a career in Defence. Please be aware that while we are piloting the intervention, not everyone will be invited to take part in the **Referrals Challenge**.

How can I encourage my team to take part in the Referrals Challenge?

Take 10 minutes in a meeting with your team to get them thinking about their networks and how to share the role with potential candidates. Top tips for your team:

- Reach out to a **range of your networks**: school friends, university classmates, previous co-workers
- Share through **social media**: LinkedIn, Twitter and Facebook, if you have an account
    - Before sharing online, consider and security constraints that may be necessary for certain vacancies
- **Understand the role** and think about who is appropriate in your network
- Have a **one-to-one conversation** with the person in your network and **share your experiences** at MOD

You can further encourage your team to engage in the **Referrals Challenge** by utilising your existing team updates, sites and newsletters. These actions can help you to get a great field of candidates that enable you to get the best person for the role.

As part of this pilot and to ensure its integrity please do not discuss this with anyone outside of your team.

Remember, all recruitment in the MOD is subject to the Civil Service Recruitment Principles. We are keen to hear your experiences and may contact you in the future for feedback on the challenge.

Kind regards,

Siobhan Sheridan CBE, FCIPD | Director Civilian HR, Diversity and Inclusion |

Chief of Defence People

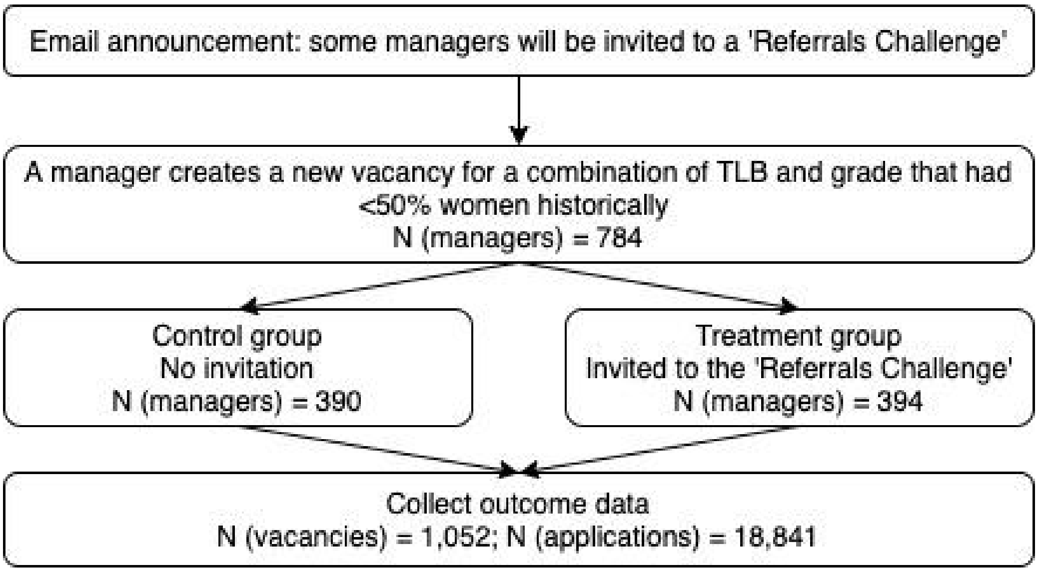
Please be aware that responses to this email will not be monitored

Trial design

We ran a two-armed randomised controlled trial (RCT) to test whether encouraging employees to share open vacancies with women they know and encourage them to apply would increase the referrals, applications and hires of women. Randomisation was clustered at the hiring manager level. This was done to eliminate spillovers between vacancies handled by the same manager. The analytical strategy is covered in Appendix C.

Applicant journey
The applicant journey is summarised in Figure 3.

Figure 3: Applicant journey



Before the start of the intervention, Civilian Workforce Advisors (senior HR leads) in each TLB shared a high-level email or intranet announcement with their TLB informing them of the 'Referrals Challenge' and encouraging them to take part if they receive an invitation.

Once the trial started, if a hiring manager was assigned to the intervention group after creating a job vacancy, they received an invitation email. This email informed them of the 'Referrals Challenge' and provided the team a target to refer five women.

Description of data and sample

We ran the trial for five months, from August 2020 to December 2020. The final sample consisted of 784 hiring managers for 1,052 civilian external and cross-government vacancies created in the trial period. We restricted the sample to only include vacancies where the combination of TLB and grade had fewer than 50% women applicants in our pre-trial dataset.[34] This meant 75.3% of TLB and grade combinations were in scope for the trial. This was done so we could target the intervention at areas of MOD where women are underrepresented in the applicant pool.

When calculating the share of applications from women for each vacancy, we only include applicants who identified as male or female in their application form (disclosure rates are high, with 97.3% of MOD applicants disclosing their gender). Table 2 below summarises the sample for the control and intervention groups and Table 3 shows referrals, applications and offers split by gender. There were originally 816 managers and 1,158 vacancies in the trial, but some had to be excluded due to errors in condition assignment and because some vacancies had been withdrawn

[34] This dataset contains all applications to MOD vacancies submitted between July 2016 and June 2018.

while being advertised. Further details of the sample descriptive statistics and balance checks are in Appendix B.

Table 2: Summary statistics

Variable	Control	Intervention	Total
Managers	390	394	784
Vacancies	549	503	1,052
Applications	8,899	9,117	18,016
Offers	438	501	939
Hires	251	290	541

Table 3: Statistics by gender

Overall	Control		Intervention	
	Women	Men	Women	Men
Referrals	78 (39.6%)	119 (60.4%)	125 (54.6%)	104 (45.4%)
Applications	3,581 (41.5%)	5,041 (58.5%)	3,611 (40.7%)	5,272 (59.3%)
Offers	175 (41.6%)	246 (58.4%)	189 (38.2%)	306 (61.8%)
Hires	111 (46.4%)	128 (53.6%)	124 (43.2%)	163 (56.8%)

Outcome measures

Table 4 summarises the outcome measures in the trial.

Table 4. Summary of outcome measures.

Outcome	Outcome measure
Referrals	Secondary: number of applications from women indicating they were referred per vacancy
	Exploratory: number of applications from men referred per vacancy
	Exploratory: share of women among referrals
Applications	Primary: share of applications from women per vacancy
	Exploratory: number of applications from women per vacancy
	Exploratory: share of applications from ethnic minority candidates per vacancy
	Exploratory: share of applications from lesbian, gay, bisexual candidates or other minority sexualities (LGB+) per vacancy
	Exploratory: share of applications from candidates with a disability per vacancy
Recruitment outcomes	Exploratory: share of women among offers
	Exploratory: number of offers made to women per vacancy
	Exploratory: number of women hires (accepted offers) per vacancy
	Exploratory: proportion of vacancies that made at least one offer

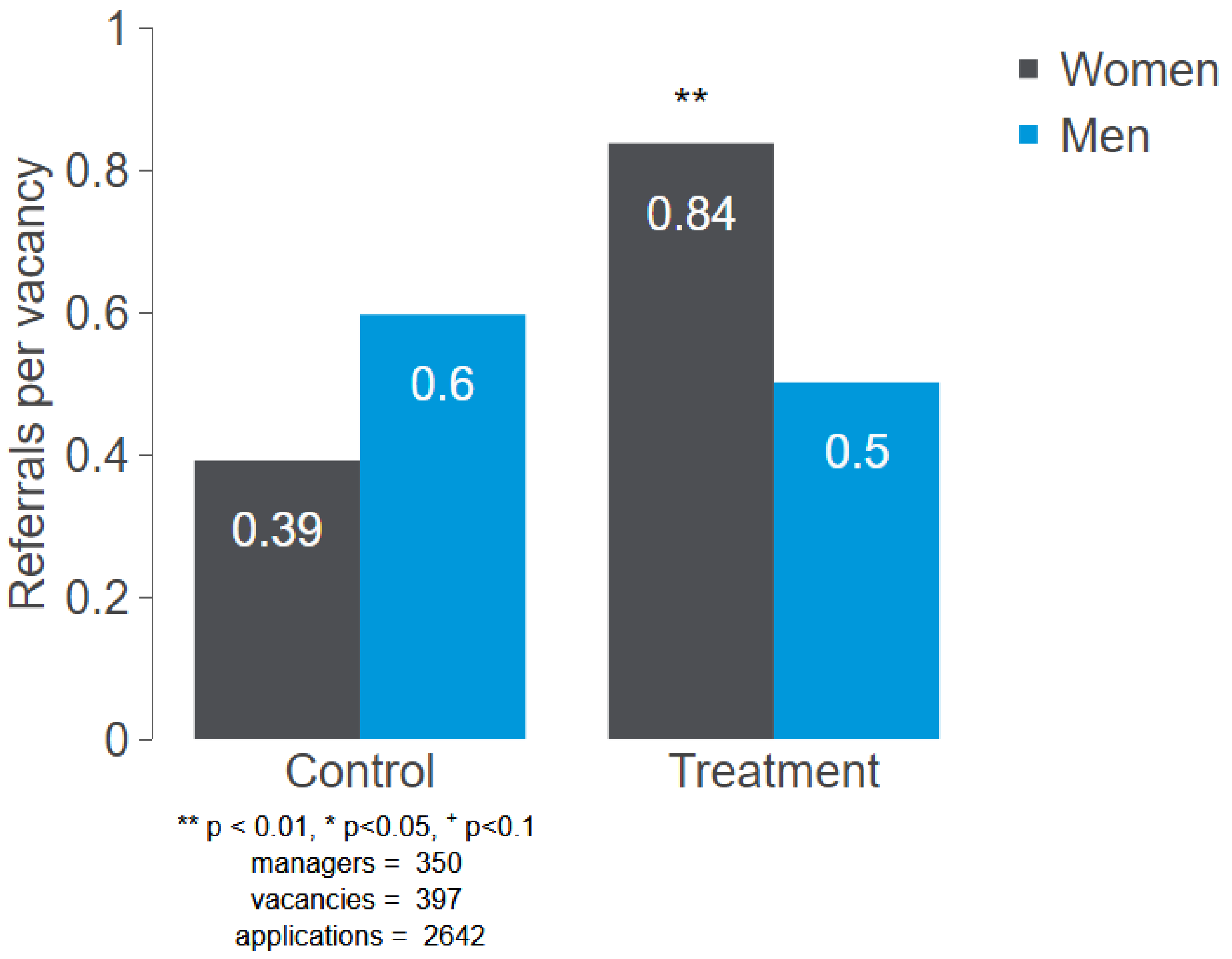
Trial results

Trial results

Referrals

Number of referred women
The Intervention had a positive effect on the average number of referred women per vacancy which was 0.39 in the control group compared to 0.84 in the Intervention group. This result was significant at the 1% level and is supported by the results of our robustness checks (see Appendix D.1).

Figure 4: Number of referred applications per vacancy by gender



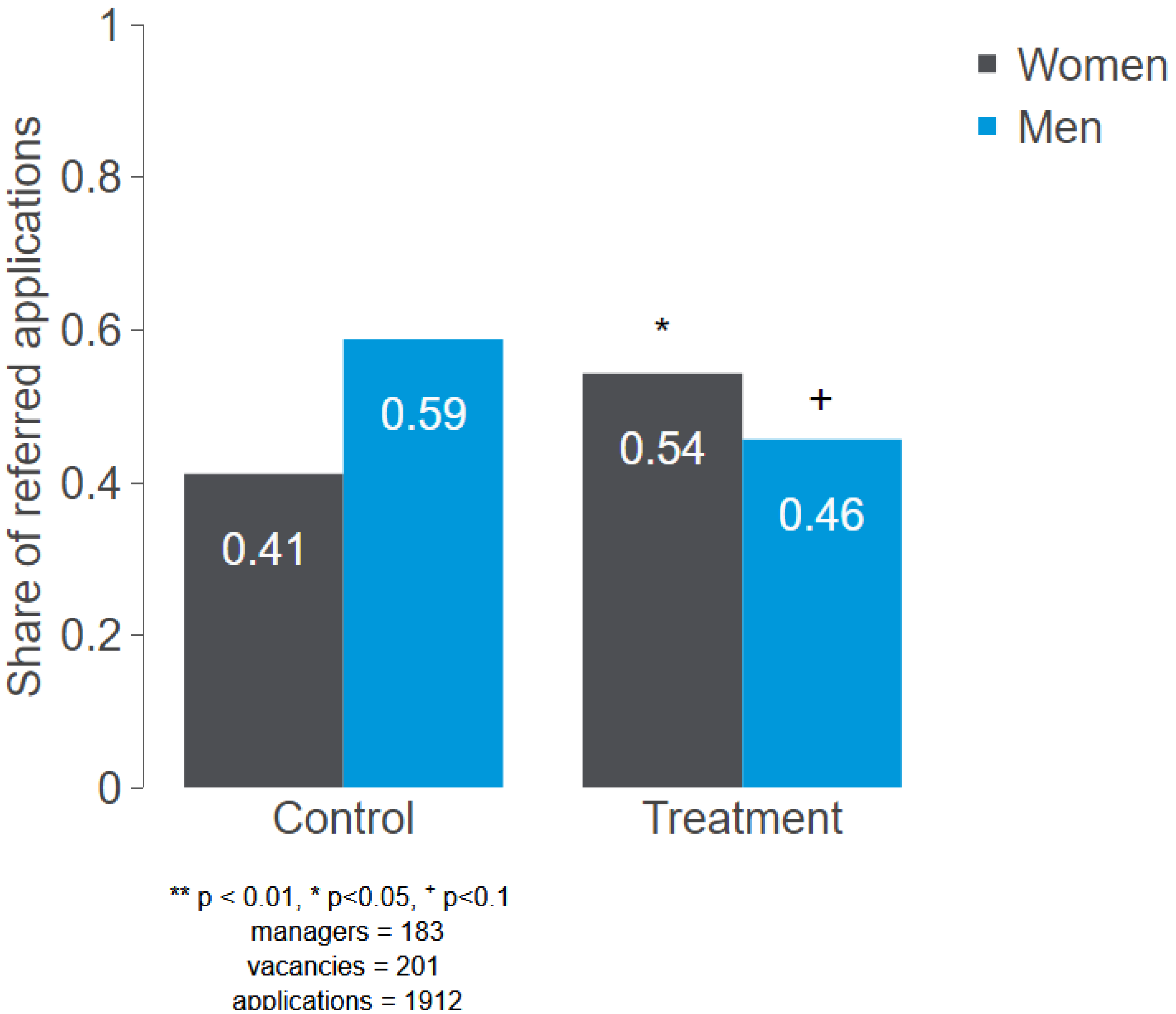
Number of referred men
To investigate whether the increase in the number of referred women led to fewer men being referred, we repeated the analysis for the number of referred men.

There was no significant effect at the 10% level of the intervention on the number of referred men, suggesting that our intervention did not decrease the number of referred men. However, alternative model specifications run during robustness checks were inconclusive, with one model suggesting a marginal increase and another suggesting a marginal decrease in the number of referred men (see Appendix D.2).

Share of women among referrals

The intervention had a positive effect on the share of women among referrals which increased from 41% in the control group to 54% in the intervention group (+13 percentage points). This result was significant at the 5% level and remains significant at the 10% level in alternative model specifications (see Appendix D.3).

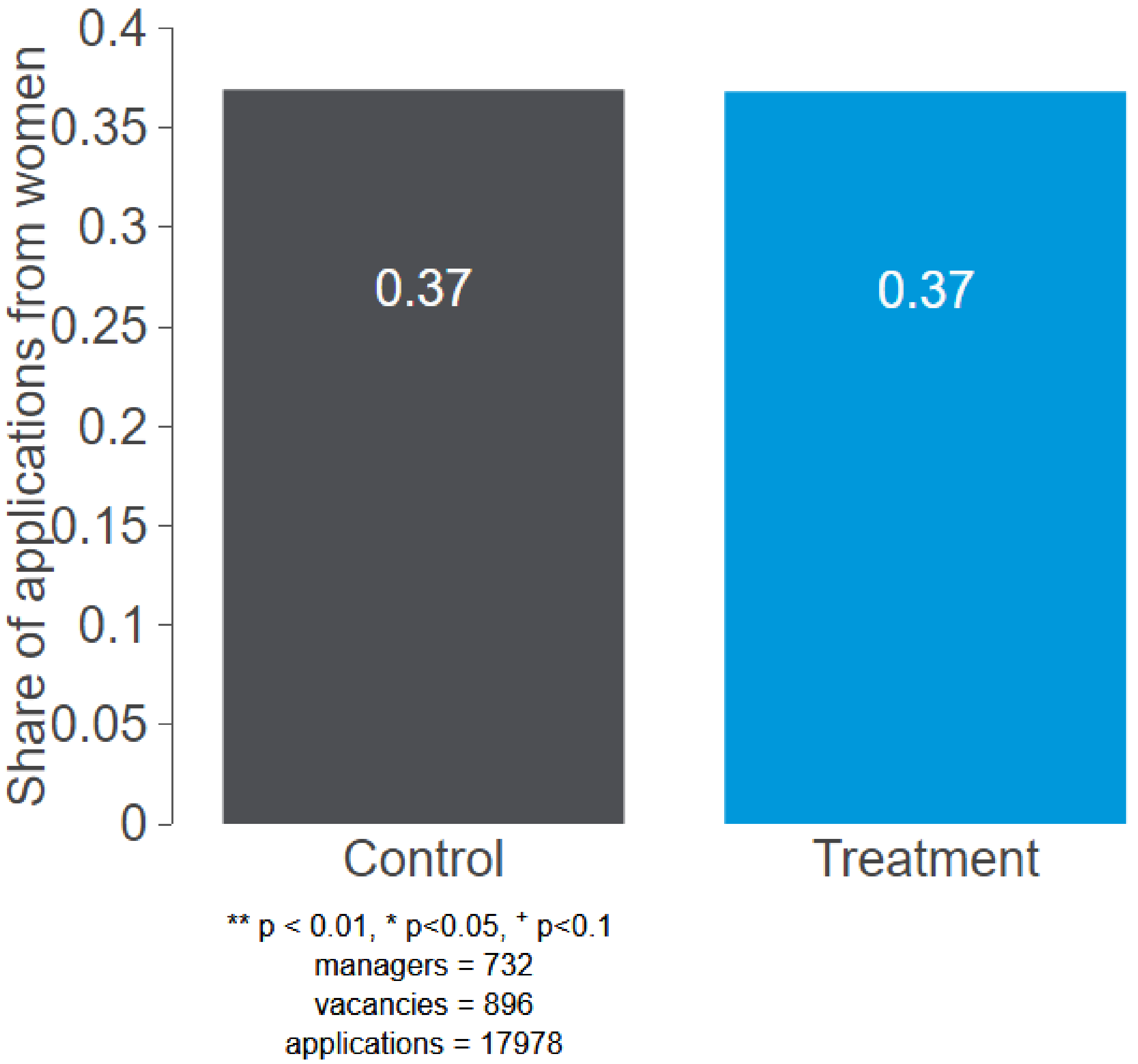
Figure 5: Share of referred applications by gender



Applications

Share of applications from women
There was no significant effect at the 10% level of the intervention on the share of applications from women (see Appendix D.4).

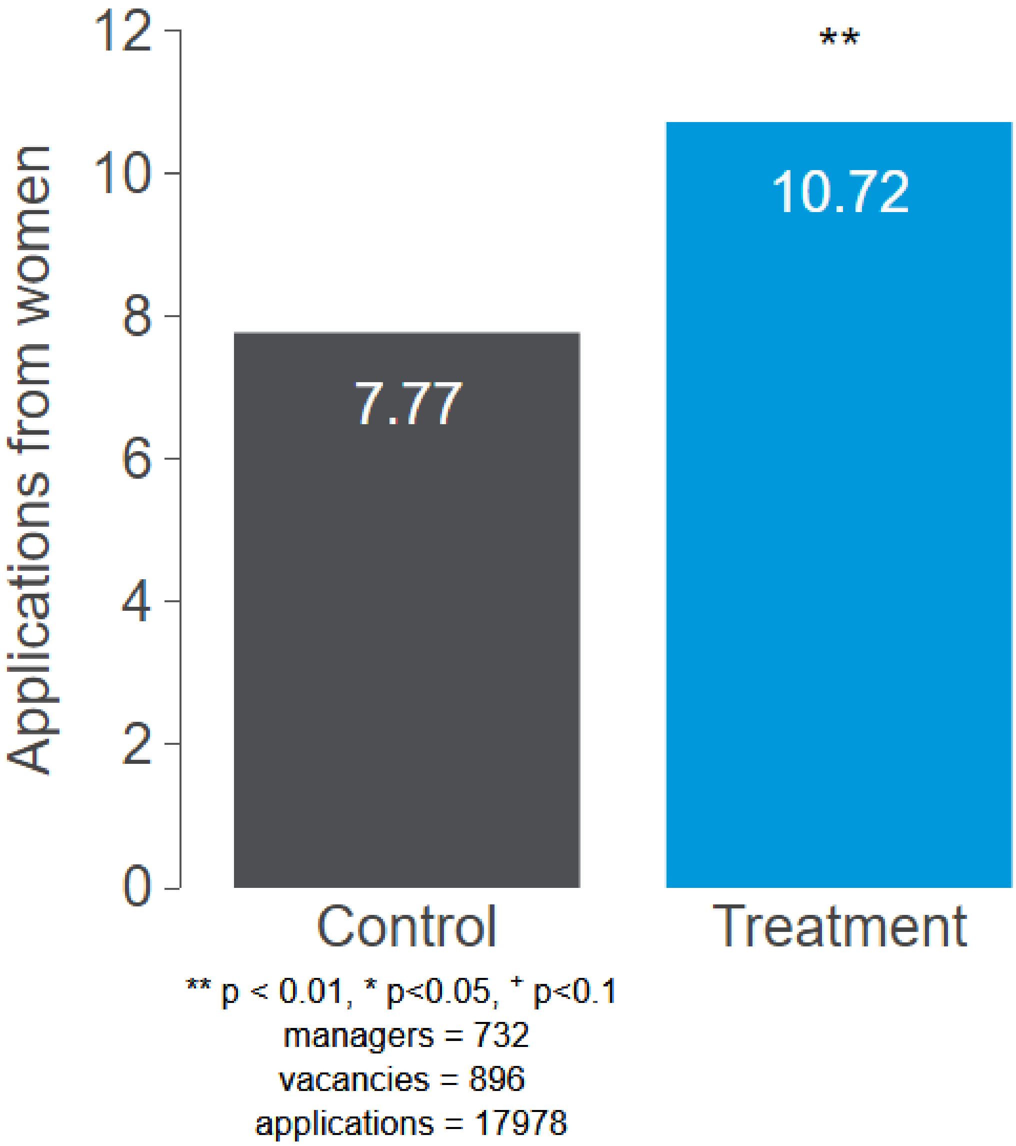
Figure 6: Share of applications from women per vacancy



Note that the analysis excluded any vacancies without applications.

Number of applications from women
The intervention had a significantly positive effect on the average number of applications from women which increased from 7.8 women applicants per vacancy in the control group to 10.7 in the intervention group. This result was significant at the 1% level in all model specifications (see Appendix D.5).

Figure 7: Number of applications from women per vacancy



Note that the analysis excluded any vacancies without applications.

Share of applications from minority applicants
We wanted to understand the impact of the intervention on applicants from an ethnic minority, applicants with a disability or LGB+ applicants.

The intervention did not have a significant effect at the 10% level on the share of applications from minority applicants from these groups (see Appendix D.6).

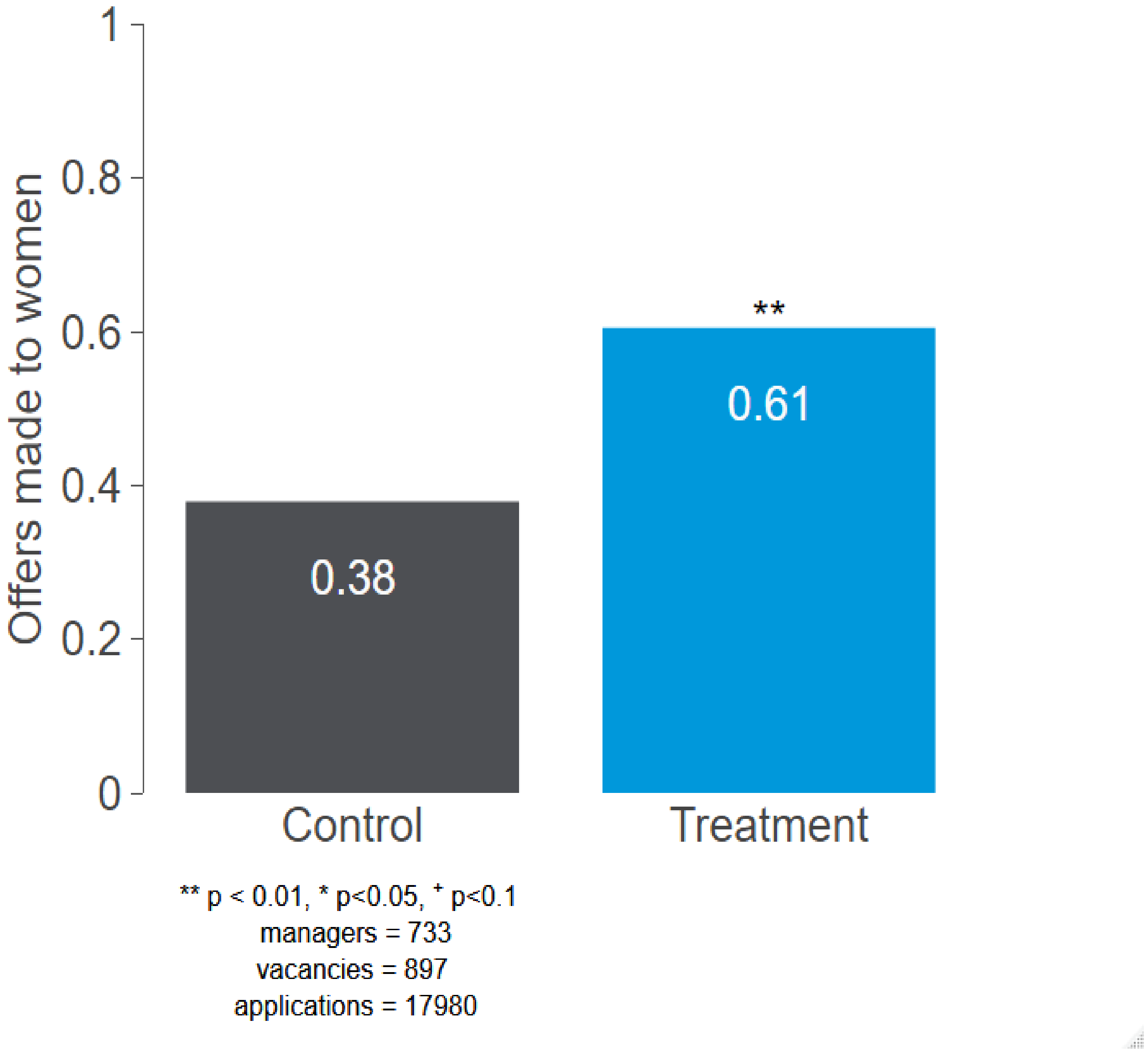
Hires

Share of women among applicants who were offered a job
The intervention did not have a significant effect at the 10% level on the share of women among applicants who were offered the job and among those who accepted the job (see Appendix D.7).

Number of women offered a job
We performed additional analysis on the number of women who were offered the job for each vacancy. We only included vacancies where at least one person was offered the job.

The intervention had a significant positive effect on the number of women who were offered the job, which increased from an average of 0.38 women per vacancy in the control group to 0.61 in the intervention group (see Appendix D.8).

Figure 8: Number of offers made to women per vacancy



Number of women hires

The intervention had a significant positive effect on the number of women who accepted offers for the job, which increased from an average of 0.24 women per vacancy in the control group to 0.43 women in the intervention group (see Appendix D.8).

Quality of applications

Only around half of the vacancies in the trial had found a suitable candidate by the time of taking the final extract of the data, which was about 2 months after the last week of the trial (16 February 2021). A significantly higher proportion of vacancies in the intervention group (49.8%) had made at least one offer (0.52, $p < 0.01$) at this point compared with the control group (43.1%).

Discussion and conclusion

Discussion and conclusion

We ran a two-armed randomised controlled trial to test whether encouraging employees to share open vacancies with women they know and encourage them to apply would increase the referrals, applications and hires of women. Hiring managers in the intervention group who advertised a new vacancy received an email inviting them to ask their team to share this vacancy with five women they know.

Compared to the control group, the intervention increased the number of referred women, the share of women among referrals, the number of applications from women and the number of women among applicants who were offered a job. At the same time, it did not lead to any changes in the share of applications from women, the share of offers to women and the share of applications from minority candidates.

Personal recommendations are one of the key sources of applicants for many organisations.[35] However, these may limit the diversity of the applicant pool and hires because employees are more likely to refer people who are similar to them.[36] Our results suggest that targeted referrals where employees are encouraged to refer people who are currently underrepresented can mitigate this inequality. In particular, when women are underrepresented in an organisation, asking employees to refer women for newly advertised roles can increase the referrals of women.

Our intervention meant twice as many women were referred for a vacancy compared to the control group.[37] Importantly, without intervention, employees referred men nearly 60% of the time, which is close to the representation of men at MOD (57%). With intervention, the proportion of women (54%) compared with men among referrals was more balanced. This demonstrates the value of targeted action in the context of existing inequality.

The intervention also increased the number of applications from women and offers made to women. This is important as it shows that the impact of the intervention followed through to later stages in the hiring process. Over time the increased numbers of women going into the MOD may lead to more women in senior roles at MOD, given that women have higher promotion rates than men.

However, the intervention also appeared to increase the number of applications from men and offers made to men, so that there was no significant change in the share of applications from women or offers made to women. In our pre-trial dataset, we could see that men were much more likely to apply multiple times to MOD. Due to a change in MOD's recruitment IT system, we were no longer able to uniquely identify applicants. We cannot know how many of the additional applications from men represent the same men. Regardless, there was no difference in the share of offers made to men between conditions. This is surprising because the intervention did not increase the number of referrals for men. It may be that the intervention encouraged managers to

[35] LinkedIn. (2017). Global Recruiting Trends 2017.
[36] Beaman, L., Keleher, N., & Magruder, J. (2018). Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi. *Journal of Labor Economics*, 36(1), 121–157.
[37] When controlling for the vacancy's region, business area, grade and manager type (could be the line manager or the person who created the vacancy).

take a more active role in their recruitment process more generally. Further analysis is needed to investigate this.

Our findings support wider research that finds informal referrals are more likely to be hired and stay in the organisation for longer than recruits who are not referred.[38] We gave managers several prompts to think deeply about the role and their range of networks and it is likely that this slower and more thoughtful process meant they were more likely to refer candidates who would prove to be a better match. We chose to implement and evaluate this intervention partly because we felt that MOD employees could provide a more personal portrayal of the organisation and potential applicants would be able to ask questions and discuss the role. Having an existing connection at the organisation is also likely to make it easier to build a network once in the company. Furthermore, an intervention like this engages managers in the process of reducing the organisation's gender pay gap, which has been found to be more effective than imposing a top-down policy that managers feel they are forced to comply with.[39]

We had concerns that targeting referrals at women could negatively affect diversity in other characteristics. However, the intervention was neither harmful nor beneficial for the representation of applicants from an ethnic minority background, with a disability, or identifying as lesbian, gay, bisexual or other minority sexualities, nor did it seem to harm male candidates. Future research should test the impact of targeted referrals for other underrepresented groups.

For technical reasons, our intervention did not include sending personal feedback on the number of referrals or completed leaderboards comparing TLBs throughout most of the trial. These accountability and transparency elements could have increased the impact of the intervention. We recommend further research that includes these elements to understand whether it increases the impact.

Overall, our findings suggest that a targeted referrals approach is promising for improving diversity. Further research should seek to understand whether targeted referrals could be successful at increasing the representation of other underrepresented groups and whether in the longer term it can have a significant impact on the share of women in senior positions.

[38] Brown, M., Setren, E., & Topa, G. (2016). Do informal referrals lead to better matches? Evidence from a firm's employee referral system. *Journal of Labor Economics*, 34(1), 161–209.
[39] Dobbin, F., & Kalev, A. (2017). Are Diversity Programs Merely Ceremonial? Evidence-Free Institutionalization. In *The SAGE Handbook of Organizational Institutionalism* (pp. 808–828).

Appendices

Appendices

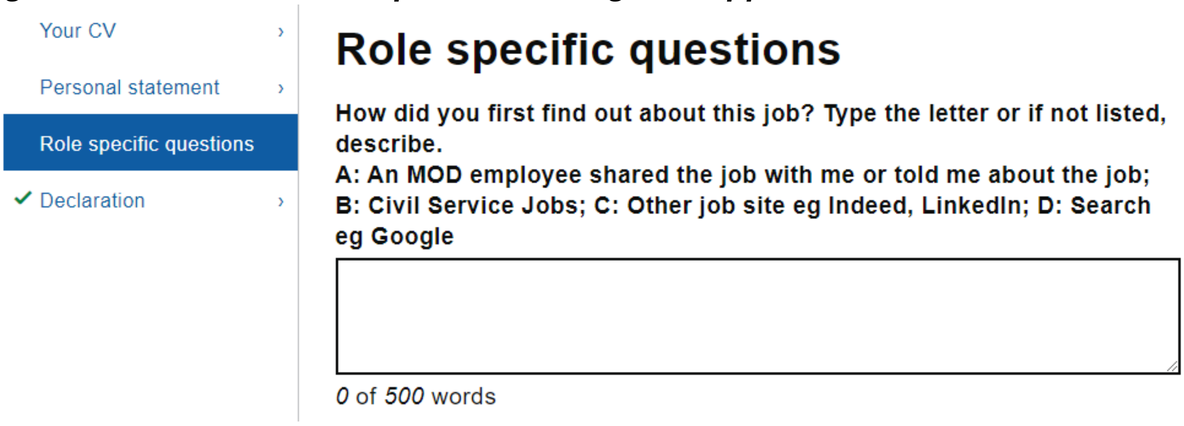
Appendix A: Implementation challenges

Incomplete leaderboard in the invitation email
In the initial invitation emails we included a leaderboard comparing the TLBs in terms of women referrals and applications with placeholder question marks to illustrate what it would look like. We planned to start adding data to the leaderboard one month into the trial to allow time for the first vacancies to receive applications. However, due to technical challenges implementing the Excel macro required to semi-automate collecting and combining this data during the trial, we only sent completed leaderboards in invitation emails in the last five weeks of the 19 weeks of the trial.

No live personalised feedback
We planned that after each intervention vacancy had closed for applications, the hiring manager would receive a personal feedback email with the latest leaderboard and the number of referrals and applications from women their team achieved. Unfortunately, due to technical challenges with building the macro required in order to semi-automate this process, we were not able to provide this feedback during the trial.

Missing data for how applicants came across the role
We implemented a new question into MOD's application form in order to measure how applicants came across the role and whether they were referred. Due to many constraints with the civil service jobs platform, the question was implemented with multiple issues that reduced response rates. Managers were able to remove the question from the application form, it was not mandatory for applicants to answer it and the question was free-text. The question itself had to be written within a limited number of characters and we could not use formatting to improve comprehension. This resulted in high rates of missing data for this question in 85.2% of applications. The level of missing data was balanced across control and intervention, so we do not think it is a threat to our interpretation. A member of the team previously uninvolved in the trial coded the free-text responses into the categories provided in the question. Below is how the question appeared in the application form.

Figure A1: Screenshot the question asking how applicants came across the role



Among applications that provided data for this question, 16.1% overall first found the job through a referral from an employee.

Table A1: data for how applicants came across the role by condition

Data type	Control	Intervention
Applications with data for the question 'how you found out about the job'	1228	1414
Applications that were referred	197 (16.0%)	229 (16.2%)

Applicants could not be uniquely identified
In the pre-trial dataset, we were able to uniquely identify applicants. From this data we found that men reapplied to MOD at much higher rates than women. In the trial dataset we were not able to identify unique applicants because of changes to MOD's recruitment IT systems. This means that we may underestimate the effect on unique women applicants for any cross-vacancy analysis.

Appendix B: Summary statistics and balance checks

Descriptive statistics

Table B1: Averages per vacancy

Per vacancy	Control		Intervention	
	Women	Men	Women	Men
Referrals*	0.39	0.60	0.63	0.53
Applications	7.77	10.93	8.22	12.01
Offers overall	0.38	0.53	0.43	0.70

**Calculated out of vacancies that had at least one application with data for 'how did you first find out about this job?'*

Table B2: Breakdown of vacancies

Overall	Control	Intervention
Vacancies with at least 1 application	462 (84.2%)	440 (87.5%)
Vacancies with at least 1 offer	306 (66.2%)	315 (71.6%)
Vacancies with at least 1 hire	199 (43.1%)	219 (49.8%)

Balance checks

We observed balance between intervention and control groups across regions (Chi-square[11] = 15.57, $p = 0.16$), business areas (Chi-square[6] = 9.06, $p = 0.17$). However, we observed an imbalance between intervention and control groups across grades (Chi-square[7] = 20.10, $p < 0.01$) and the number of missing data about referrals (Chi-square[1] = 10.60, $p < 0.01$). It seems that control had more vacancies from grades C2, but less from Industrial grades, and slightly higher number of missing data about referrals (85.7% Intervention vs. 87.4%). High percentage of missing data overall for referrals limited our ability to impute it, and created an important caveat in interpreting the results about referrals.

Table B3: Frequency of different vacancy grades in control and intervention groups.

Grade	Intervention	Control
B1	22	20
B2	42	47
C1	120	138
C2	90	138
D	70	50
E1	15	15
E2	15	11
Industrial/Other	65	42

Appendix C: Analytical strategy

Primary outcome: share of applications from women per vacancy
We used a quasi-binomial model. We did not use a linear model because many predicted shares of women applicants from such a model are likely to lie outside the feasible [0,1] range. In our pre-trial dataset, 18.4% of within-scope vacancies only received applications from men and 4.6% only attracted applications from women. A quasi-binomial model also accounts for heteroskedasticity by making the variance of the outcome variable a function of the covariates. The specification was as follows:

$$sharewomen_j \sim quasibinomial(N_j, p_j, \varphi)/N_j;$$

$$logit(p_j) = \alpha + \beta T_j + region_j + businessarea_j + grade_j + managertype_j$$

$$var(sharewomen_j) = p_j(1-p_j)\varphi/N_j$$

Here $sharewomen_j$ is the share of women applicants for vacancy j. T_j is an indicator for the vacancy being posted by a hiring manager allocated to the intervention group, and so β represents the average effect of the intervention. $region_j$, $businessarea_j$ and $grade_j$ are fixed effects for the vacancy's region, TLB and grade respectively. N_j is the number of applicants. We used bootstrapped standard errors (with clustering at the hiring manager level)[40] [41] and weight observations by the number of applicants.

We also controlled for $managertype_j$. There were two sources of identifying the manager: one from the HR system that specified the intended line manager for the vacancy, and one from the recruitment system in terms of the person who created the vacancy. If the line manager email address was available we used this, but where this was not in the system, we used the email address that created the vacancy. MOD felt that the latter was more likely to be an administrative member of the team rather than a manager, which is why we used the line manager email address where available. 88% of the managers in the trial were line managers and 12% were the person who created the vacancy.

As a robustness check, we also repeat the above analysis using an OLS regression model.

Secondary outcome: number of applications from women indicating they were referred per vacancy
We used the following specification:

$$log(1 + numreferrals_j) = \alpha + \beta T_j + region_j + businessarea_j + grade_j + managertype_j + \varepsilon_j$$

Here $numreferrals_j$ is the number of referrals. The indexing, covariates and level of clustering for standard errors are the same as in the primary analysis. We favoured a log transformation of the outcome in this case because it reduced the positive skew which was likely in this count variable.

---

[40] Broström, G., & Holmberg, H. (2011). Generalized linear models with clustered data: Fixed and random effects models. *Computational Statistics and Data Analysis*, 55(12), 3123–3134.
[41] Package 'glmmML' May 28, 2020

As robustness checks, we perform an additional analysis with Poisson regression using untransformed referral counts and logistic regression, using a binary variable indicating if a vacancy had more than 0 referred women. In both of these robustness checks, we removed 'region' covariate, due to problems with estimating correctly the coefficients.

Exploratory outcome: number of applications from men indicating they were referred per vacancy

We repeat the above analysis, but for men.

Exploratory outcome: number of applications from women/men per vacancy

The impact of the intervention on the number of women and men applicants for each vacancy was evaluated in the same way as its impact on the number of referrals.

Exploratory: share of women among referrals

We estimate the following model, using only the sample of those who were referred:

$$sharewomen_j = \alpha + \beta T_j + region_j + businessarea_j + grade_j + \ managertype_j$$

We assumed quasibinomial and gaussian distributions of the share.

Due to estimation problems for the region covariate, we also repeat the analysis without the region covariate.

Exploratory: share of women among offers

The trial protocol specified the share of positions filled by women applicants as the outcome. To account for the fact that more than one person can be hired for each vacancy, we modify the pre-specified analysis as follows (matching exactly the primary analysis).

$$sharewomen_j \sim quasibinomial(N_j, p_j, \varphi)/N_j;$$

$$logit(p_j) = \alpha + \beta T_j + region_j + businessarea_j + grade_j + \ managertype_j$$

Where $sharewomen_j$ is proportion of women who were offered the advertised position (out of everybody who was offered the position).

As a robustness check, we perform the same analysis as above but only for women who accepted the offers.

Exploratory: number of offers made to women per vacancy

We perform an additional analysis looking at the number of women who were offered the job for each vacancy (not pre-specified). We use the following model

$$log(1 + numwomen_j) = \alpha + \beta T_j + region_j + businessarea_j + grade_j + managertype_j + \varepsilon_j$$

where $numwomen_j$ is the number of women offered the job for each vacancy. The indexing, covariates and level of clustering for standard errors are the same as in the primary analysis. As a robustness check, we repeat the above analysis using Poisson regression, and untransformed dependent variable $numwomen_j$

Exploratory outcome: number of women hires (accepted offers) per vacancy
We repeat the above analysis, but this time counting the number of women who were offered an accepted the offer.

Exploratory outcome: number of applications from women for each vacancy
The impact of the intervention on the number of applications from women for each vacancy was evaluated in the same way as its impact on the number of referrals.

Exploratory outcomes: share of applications from minority applicants for each vacancy
We estimated a quasi-binomial model as in the primary outcome, but with the share of white / heterosexual / without a disability applicants as the outcome variable and the sample of applicants from which vacancy-level shares are calculated containing women only.

For all analysis where we use quasibinomial regression, we also repeat the analysis using OLS regression as a robustness check.

Likelihood of being offered a position, based on gender and condition
To check if our intervention increased the pool of quality applicants and if it increased the chance of a women being hired, we estimate the following logistic regression models:

Model 1 Without gender interaction

$$offer_j = \alpha + \beta_1 T_j + \beta_2 gender_j + region_j + businessarea_j + grade_j + managertype_j + \varepsilon_j$$

Model 2 With gender interaction

$$offer_j = \alpha + \beta_1 T_j + \beta_2 gender_j + \beta_3 T_j * gender_j + region_j + businessarea_j + grade_j + managertype_j + \varepsilon_j$$

Where $offer_j$is a binary variable indicating if a person received an offer, T_j indicates if an application was in the intervention or control, $gender_j$ indicates the gender of the applicant.

Appendix D: Detailed results

D.1 Number of applications from women indicating they were referred per vacancy

Table D1 provides the results of the secondary analysis for the number of referred women. As robustness checks, we performed an additional analysis with poisson distribution using untransformed referral counts (Column 2 in Table D1) and Binomial distribution (Column 3 in Table D1), using a binary variable indicating if a vacancy had more than 0 referred women. In both of these robustness checks, we removed ‘region’ covariate, due to problems with estimating correctly the coefficients (as indicated by extreme S.E. values). Both of these robustness checks supported the main conclusion.

Table D1: Effects of intervention on number of referred women. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table.

	(1)	(2)	(3)
	Log-transform	Poisson	Binomial
Intervention	0.24**	0.76**	0.63**
	(0.06)	(0.10)	(0.20)
Constant	-0.09	-2.06**	-2.38**
	(0.20)	(0.35)	(0.59)
Observations	397	397	397

* $p < 0.05$, ** $p < 0.01$

D.2 Number of applications from men referred per vacancy
Table D2 provides the results of the analysis as specified in the 'Analytical strategy' section.

Table D2: Effects of intervention on number of referred men. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table.

	(1) Log-transform	(2) Poisson	(3) Binomial
Intervention	0.01	-0.18+	0.35+
	(0.06)	(0.09)	(0.20)
Constant	0.42	-0.00	-2.32*
	(0.19)	(0.30)	(0.59)
Observations	397	397	397

$^{+}$ p < 0.10, * p < 0.05, ** p < 0.01

D.3 Share of women among referrals
Table D3 provides the results of the exploratory analysis as specified in the 'Analytical strategy' section. We assumed quasibinomial (Column 1) and gaussian (Column 2) distributions of the share. Both analyses suggested a significant positive effect of intervention on the share of referred women (Columns 1 and 2). The quasibinomial model could not properly estimate the coefficients for regions (judging based on extreme SE values, e.g. see the intercept in the table below). We therefore repeated the analysis without the region covariate (Columns 3 and 4)

Table D3. Effects of intervention on share of applications from women. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table

	(1) Quasi-binomial	(2) OLS	(3) Quasi-binomial Without region	(4) OLS Without region
Intervention	0.64*	0.13*	0.51+	0.11+
	(0.31)	(0.06)	(0.27)	(0.06)
Constant	-17.63	-0.20	-2.10*	0.02
	(1938.03)	(0.36)	(0.81)	(0.17)

Observations	201	201	201	201

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

D.4 Share of applications from women per vacancy

Table D4 provides the results of the analysis as specified in the 'Analytical strategy' section.

Table D4: Effects of the intervention on share of applications from women. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table.

	(1) Quasibinomial	(2) OLS
Intervention	0.03	-0.00
	(0.08)	(0.02)
Constant	-1.52**	0.15*
	(0.30)	(0.06)
Observations	896	896

* $p < 0.05$, ** $p < 0.01$

D.5 Number of applications from women per vacancy

Table D5 provides the results of the exploratory analysis as specified in the 'Analytical strategy' section. We observed a significant positive effect of the Intervention on the number of applications from women. A robustness check using a Poisson model with untransformed counts suggested the same conclusion.

Table D5. Effects of intervention on number of applications from women. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table.

	(1) Log-transformed	(2) Poisson
Intervention	0.16*	0.32**
	(0.08)	(0.02)
Constant	0.81**	0.70**
	(0.28)	(0.07)

Observations	897	897

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

D.6 Share of applications from minority applicants

Table D6 provides the results of the exploratory analysis as specified in the 'Analytical strategy'

Table D6. Effects of intervention on share of applications from minority applicants. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table.

	(1)	(2)	(3)	(4)	(5)	(6)
	Quasibinomial	OLS	Quasibinomial	OLS	Quasibinomial	OLS
	Ethnic minority	Ethnic minority	Disability	Disability	LGB+	LGB+
Intervention	0.01	0.00	-0.16	-0.01	0.05	0.00
	(0.09)	(0.01)	(0.10)	(0.01)	(0.66)	(0.01)
Constant	-2.94**	-0.01	-2.96	0.05	-3.47**	0.02
	(0.37)	(0.05)	(0.42)	(0.03)	(0.46)	(0.02)
Observations	891	891	894	894	889	889

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

D.7 Share of women among offers

Table D7 provides the results of the exploratory analysis as specified in the 'Analytical strategy' section.

Table D7. Effects of intervention on share of women who were offered the job. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table.

	(1)	(2)	(3)	(4)
	Quasibinomial	OLS	Quasibinomial	OLS
	Offered	Offered	Accepted	Accepted
Intervention	-0.06	-0.01	0.05	0.01

	(0.16)	(0.03)	(0.21)	(0.04)
Constant	-0.57	0.34**	-1.52	0.14
	(0.56)	(0.12)	(0.76)	(0.16)
Observations	605	605	406	406

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

D.8 Number of offers made to women per vacancy and number of hires
Table D8 provides the results of the exploratory analysis as specified in the 'Analytical strategy'

Table D8. Effects of intervention on the number of women who were offered the job and accepted the job. The regression additionally included covariates for manager type, business area, region and grade, but were omitted from the table.

	(1) Log-transformed offered	(2) Poisson offered	(3) Log-transformed Accepted	(4) Poisson accepted
Intervention	0.11**	0.47**	0.11**	0.57**
	(0.03)	(0.08)	(0.03)	(0.10)
Constant	0.21+	-1.33**	0.10	-2.02**
	(0.12)	(0.29)	(0.10)	(0.38)
Observations	897	897	897	897

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$