Interventions That Affect Gender Bias in Hiring: A Systematic Review
Carol Isaac, PhD, PT, Barbara Lee, PhD, and Molly Carnes, MD, MS

Abstract

Purpose
To systematically review experimental evidence for interventions mitigating gender bias in employment. Unconscious endorsement of gender stereotypes can undermine academic medicine’s commitment to gender equity.

Method
The authors performed electronic and hand searches for randomized controlled studies since 1973 of interventions that affect gender differences in evaluation of job applicants. Twenty-seven studies met all inclusion criteria. Interventions fell into three categories: application information, applicant features, and rating conditions.

Results
The studies identified gender bias as the difference in ratings or perceptions of men and women with identical qualifications. Studies reaffirmed negative bias against women being evaluated for positions traditionally or predominantly held by men (male sex-typed jobs). The assessments of male and female raters rarely differed. Interventions that provided raters with clear evidence of job-relevant competencies were effective. However, clearly competent women were rated lower than equivalent men for male sex-typed jobs unless evidence of communal qualities was also provided. A commitment to the value of credentials before review of applicants and women’s presence at above 25% of the applicant pool eliminated bias against women. Two studies found unconscious resistance to “antibias” training, which could be overcome with distraction or an intervening task. Explicit employment equity policies and an attractive appearance benefited men more than women, whereas repeated employment gaps were more detrimental to men. Masculine-scented perfume favored the hiring of both sexes. Negative bias occurred against women who expressed anger or who were perceived as self-promoting.

Conclusions
High-level evidence exists for strategies to mitigate gender bias in hiring.


The success of female physicians is recognized and celebrated both in popular television series such as “ER,” “Providence,” and “Strong Medicine” and by the National Library of Medicine. Despite explicit support for gender equity in academic medicine, however, female physicians advance more slowly toward seniority than do male physicians, earn less than male physicians in similar positions, and have not entered the ranks of leadership at rates predicted by their proportional presence in academic medicine for the past 30 years.

Physicians are committed to evidence-based practice. Studies with random assignment of participants to an intervention or control group, in particular, provide high levels of evidence in informing physician decision making. Decades of social cognitive research exists on how gender stereotypes lead to assumptions—both implicit (unconscious) and explicit (conscious)—that consistently impede women’s advancement in historically male-dominant fields. The success of a job applicant in obtaining a position is a major determinant of that person’s ability to advance in any career. To facilitate the adoption of evidence-based employment practices in academic medicine, we performed a systematic review of studies with randomized controlled designs that investigated the impact of an intervention on the activation and application of gender bias in hiring settings.

Method
Study selection
The studies we selected met the following inclusion criteria: random assignment of participants to the intervention or control group, assumption by participants of the role of personnel decision makers evaluating applicants for employment, publication after 1972 (the year that Congress passed the Title IX Amendment to the Civil Rights Act), blinding of participants to the intervention, the presence of both men and women in the contrived applicant pool and the participant (rater) groups, and comparison of the impact of an intervention on ratings of male and female applicants with identical qualifications. We excluded studies that

Dr. Isaac is a postdoctoral fellow, Center for Women’s Health Research, and lecturer, Department of Orthopedics and Rehabilitation, University of Wisconsin–Madison, Madison, Wisconsin.

Dr. Lee is a consultant in statistics and evaluation, Tarpon Springs, Florida.

Dr. Carnes is director, Center for Women’s Health Research, University of Wisconsin–Madison and Meriter Hospital, Madison, Wisconsin; professor, Departments of Medicine and Psychiatry, School of Medicine and Public Health; professor, Department of Industrial & Systems Engineering, and codirector, Women in Science and Engineering Leadership Institute (WISELI), College of Engineering, University of Wisconsin–Madison, Madison, Wisconsin; and director, Women Veterans Health Program, William S. Middleton Memorial Veterans Hospital, Madison, Wisconsin.

Correspondence should be addressed to Dr. Carnes, Center for Women’s Health Research, University of Wisconsin–Madison, 700 Regent Street, Suite 301, Madison, WI 53715; telephone: (608) 263-9770; fax: (608) 265-6425; e-mail: (mlcarnes@wisc.edu).
assessed bias only by reaction time or accuracy in matching gender-linked stereotypic words or pictures, studies in which the participants were stated to be less than 18 years old, studies with only women in the applicant pool (e.g., pregnant and nonpregnant participants), and studies that did not specifically indicate random assignment of the intervention. We also excluded dissertations, letters, and abstracts. Although searches had no language restriction, all studies identified were in English. When the presence of an inclusion criterion was in doubt, the authors achieved resolution through consensus. This effort usually involved distinguishing between an intervention that had an impact on gender bias and one that simply documented gender bias in a different hiring setting (e.g., jobs supervising men or women9).

Data sources and search strategy
The authors electronically searched the following sites from 1973 (when possible) to June 2008: PubMed, PsychINFO, Web of Science (including Social Science Citation Index), Cochrane Library, CINAHL, ProQuest, ABI/INFORM (U.S. and international articles on business and management), ERIC, and SocINDEX. Terms entered from the Medical Subject Headings (MeSH) of the National Library of Medicine were Human, Female, Prejudice, and Stereotype(s). Other terms entered individually or in combination were Gender, Women, Hire/hiring, Bias, Sex roles, Sex, Discrimination, and Research. The authors narrowed database searches using the term Experimental to identify studies with randomized controlled designs. Professional librarians performed supplemental searches of ProQuest, PubMed, and Women’s Studies International. Additional reference mining included selected author searches, hand searches of bibliographies of retrieved studies and meta-analyses, and review of files of senior faculty who study gender and leadership. The search was considered saturated when relevant articles reappeared in multiple searches. The authors identified and reviewed abstracts from citations through each of the above searches (N = 1,920) and retrieved and examined articles that seemed to meet inclusion criteria (Figure 1). Because of the heterogeneity in interventions and outcomes, the data were not pooled.

![Figure 1](http://links.lww.com/ACADMED/A1) Search strategy and final selection of studies for inclusion in systematic review. UW, University of Wisconsin; WISELI, Women in Science and Engineering Leadership Institute.

Data extraction
We three authors independently reviewed in detail 130 studies. One of us (B.L.), a statistician, evaluated articles for quality and effectiveness of controls, validity checks on interventions, and appropriateness of statistical tests. We scored articles for quality by using a modified Jadad numerical system of one to four points (a point was allowed for single blinding).10 Inclusion required a score of at least 2. After verification of inclusion criteria, we extracted the following information: author, year, and country in which study was performed; intervention; outcome; study design; demographic information on study participants (i.e., gender and race–ethnicity); the construct measured; results; and the P values of statistical procedures.

When an article described more than one experiment, we included only those substudies that met our inclusion criteria. If more than one of the substudies in a given paper met the criteria, we reviewed each one but still counted the citation as one study. Twenty-seven studies met all inclusion criteria. See the Appendix (http://links.lww.com/ACADMED/A1). The Jadad score was 3 for 4 of the studies11–14 and 2 for the other 23 studies.

Results
Overview of selected studies
Participants in 18 of the 27 studies were college students. Other studies used business (MBA) or graduate students (3 studies and 1 substudy),15–18 managers,12,19,20 adult workers,21,22 and members of human resource associations.14 Twenty-three studies were conducted in the United States: 3 at specified universities,11,13,17 7 in identified regions,12,14,19–21,23,24 and 13 at unspecified locations. Two studies were conducted in the Netherlands25,26 and 2 in Germany.27,28 Participants in all studies were categorized by gender; 11 had descriptors of age (means or ranges),13,14,17,19,21,23,27–31 and 2 provided some description of race and ethnicity.21,23 Whites made up 72% to 90% of participants in these two studies. Studies established applicant gender visually by photograph19,28,29,32–36 or video,13,21,37 designation of sex on the application,18,24 in-person interview,27,37 and/or the use of gendered names and pronouns (modifications of the Goldberg paradigm38).11,12,14–18,22,32,34 Twenty-four studies11–13,15,16,18–30,32,34–37,39 examined
gender bias in decision making with regard to applicants for “male sex-typed jobs,” the term applied in much of this research to positions historically or predominantly occupied by men and/or assumed to require stereotypically male traits. Such positions included mechanical engineer,11,14 assistant vice president for financial affairs,14 chair of a district’s association of physicians,25,26 sales manager for a heavy-machinery company,12 high-ranking chief executive officer,21 and police officer.22,39 Twelve studies12–14,17,21,24,27,30,33,34,36,39 examined outcomes for female sex-typed jobs (e.g., nurse,39 dental receptionist,13 and day care worker22) or gender-neutral jobs (e.g., copy editor,24 assistant trainee,21 and compensation analyst14). One study13 manipulated the sex-typing of a neutral job (computer lab manager) by emphasizing the requirement of either stereotypic male traits (i.e., technically skilled and able to work under pressure) or stereotypic female traits (i.e., helpfulness and sensitivity to coworkers). Studies confirmed job sex-typing with pretested scales11,22–24,26–29,33 or previous studies12–17,19–22,25,27,28,30,34,36,39 that used, for example, job sex-typing inventories.40–44 Twenty-three studies used MANOVA,11–17,27-29,32,33,34-36,39 or ANCOVA28,36 to compare main effects of the intervention and other independent variables and to test for interactions with gender on the dependent variables of interest. These comparisons were followed by individual comparisons of findings for male and female applicants with previously planned contrasts or appropriate post hoc tests. The remaining study used the chi-square test.31

All but one study24 confirmed that male applicants are evaluated more positively than female applicants for employment in male sex-typed jobs. See the Appendix (http://links.lww.com/ACADMED/A1). It was easier for men than for women with identical qualifications to be assessed the impact on bias against female applicants (e.g., having a biology/political science degree rather than a business/economics degree when applying for a lower management position) resulted in lower ratings than did the absence of such information. Glick and colleagues12 provided individuating information that...
established gender-counterstereotypic personality traits (e.g., men working in retail sales at a jewelry store and women working in grounds maintenance) but that was job-irrelevant; they found higher employability ratings for both male and female applicants with stereotypic masculine traits, although the preference of raters for a match between sex-type and applicant gender remained. To measure the degree of gender stereotyping, the participants in the study by Heilman
assessed applicants by using five adjectival scales associated with gender-related work attributes (e.g., emotional–rational, ambitious–unambitious, tough–soft). Providing a high degree of job-relevant information about a female applicant eliminated the difference in gender stereotyping between male and female applicants seen with low job-relevant information or no information. Furthermore, when composites of these adjectival scores were covaried with applicant ratings, the perceptions of gender-related attributes rather than the applicant’s actual gender accounted for assessments of hireability and of potential for advancement. Rudman and Glick
found that highly competent female applicants benefited from applications that included a written “life philosophy” endorsing communal (stereotypically female) rather than agentic (stereotypically male) values, particularly when they were applying for female sex-typed jobs.

Two studies examined the impact of including information on parental status in the application.
Male and female applicants without children received comparable ratings on all employment-relevant measures. Parenthood resulted in lower ratings for both male and female applicants, but women whose applications indicated that they had children were more disadvantaged. Although both female and male parents were rated as less committed and less dependable than nonparents, only female applicants with children were rated lower on measures of hiring and promotion.

One study included both marital and parental status information in the applications. Marital status had little effect on applicant ratings, although married men with children and single women were ranked as the most suitable applicants for two neutral sex-typed positions. One study examined the impact of applications that contained discontinuities in employment and found that men were generally judged more harshly than women in such cases.

One study compared the effect of gender ambiguity in the application.
When an applicant’s gender was apparent from the application, women were disadvantaged; however, when applicants had gender-ambiguous names (e.g., Pat or Chris), job suitability was based solely on the applicants’ qualifications (even if the inferred gender was female).

**Applicant behavior, scent, or appearance**

Three studies assessed the impact of interview behavior on gender bias.
All found negative reactions to women who exhibited stereotypic male behaviors. Rudman
found that, when applicants of either gender violated behavioral norms—men by being self-effacing and women by being self-promoting—both were rated lower than applicants who behaved in a more gender-congruent manner. In one of the few differences by participant gender, female raters judged self-promoting women more harshly than did male raters. Rudman and Glick
found that women who exhibited an agentic interview style were rated lower on social skills than men, although this difference was eliminated when women’s applications included a communal life philosophy statement. Brescoll and Uhlmann
found that the expression of anger by an applicant enhanced the evaluation of men and lowered the evaluation of women, particularly women applying for a high-status position. The existence of a specific external cause for anger mitigated but did not eliminate the negative bias toward women; external attribution for anger improved the status and salary ratings for women who expressed anger but had no impact on the lower rating of competence.

Sczesny and Stahlberg
and Sczesny and Kühnen
found that visual and olfactory cues can activate gender stereotypes independent of the actual biological sex of the applicant. Male and female applicants wearing a masculine-scented perfume or submitting paper applications to which such a scent was applied received more positive ratings than did identically qualified applicants who used a feminine scent.

This group also found that both men and women who looked more stereotypically masculine in photographs were favored for hiring into a leadership position.

Five studies examined the impact of physical attractiveness and found that overall attractiveness is advantageous, but more so for men than women.

Highly attractive women can be disadvantaged in applying for male sex-typed jobs, and less attractive women can be disadvantaged in applying for female sex-typed and neutral jobs. Heilman and Saruwatari
found that attractiveness predicted ratings of stereotypic male or female traits among applicants and that, when these ratings were factored out, the impact of attractiveness was eliminated.

**Conditions under which raters assessed applicants**

Five studies sought to manipulate automatic gender bias in hiring by informing raters of employment equity directives or by prior training of raters with an exercise to decrease the response time to gender-counterstereotypic word associations. In response to employment equity directives, Ng and Wiesner
found that men who were less qualified than women for a female sex-typed job (i.e., nurse) were more likely to be hired, but this positive bias for the underrepresented candidate did not hold true for women who were less qualified than male applicants for a male sex-typed job (i.e., police officer). In the study by Biernat and Fuegen,
raters with the expectancy of accountability for their hiring decisions were less likely to hire a female applicant. Rosen and Mericle
found that, even under strong employment equity directives, female applicants were recommended for lower salaries than were men with identical qualifications. Kawakami and colleagues
engaged raters in “antibias” training that successfully reduced response time in matching gender-counterstereotypic words that were displayed sequentially on a computer screen. However, this training did not reduce gender bias in a subsequent mock-hiring situation unless an intervening task or concurrent cognitive distraction prevented subjects from correcting against the perceived coercion of training.

If participants were able to correct for perceived coercion on an initial task, the preference for male over female job candidates and the attribution of gender-stereotypic traits were eliminated.

Two studies varied the order in which aspects of the hiring process occurred.
Uhlmann and Cohen found that requiring raters to commit to the value of credentials before reviewing any applicants eliminated gender bias in hiring a police chief. Cann et al. found better correlation between applicant ratings and recommendations to hire when raters were forced to rate applicants’ qualifications separately before, rather than after, providing summary employability judgments.

Heilman found that, when women composed 25% or less (i.e., no more than two) of the applicants in a pool of eight, they were viewed as less qualified than male applicants for a managerial job and as being more stereotypically female on gender-related adjectival scales than when women made up at least 37.5% of the pool (three of eight applicants). Covariance analysis of gender-stereotypic and hireability ratings indicated that the impact of gender proportion in the applicant pool could be completely accounted for by the stronger attribution of female gender stereotypes to women when they were made up 25% or less of the pool. In a study by Heilman and Martell, priming raters with data that women are succeeding in a relevant male-dominated field eliminated bias against female applicants, although priming with information about a single successful woman did not.

Sczesny and Kühnen found that rating applicants in the presence of a competing cognitive demand (i.e., memorizing a nine-digit number) enhanced the evaluation of male applicants for leadership competence and certainty of hiring. This effect was most pronounced in female raters.

**Discussion**

This systematic review reaffirmed the ubiquity of unconscious stereotypes regarding the behaviors and traits associated with being male or female, the ease with which these stereotypes are activated, and the consequent negative bias against women applicants for jobs historically occupied by men. More important, however, this review documents the capability for mitigating the automatic activation and subsequent application of these biases.

Taken together, these studies indicate that, when ambiguity exists in an individual’s qualifications or competence, evaluators will fill the void with assumptions drawn from gendered stereotypes. Providing individuating proof of competence and past performance excellence that are relevant to the employment opportunity seems to be effective in mitigating gender bias, provided that raters do not feel coerced, conditions enable raters to fully attend to the information provided, and raters commit to the value of specific credentials both before the review and before giving an overall rating. Informing raters about research confirming women’s competence in sex-typed male tasks is also effective.

Given the large number of competent women physicians and scientists, this approach would seem to be a fairly straightforward way to ensure gender equity. The studies reviewed also indicate, however, that the issue is more complex than expected. Women who are clearly competent in male sex-typed roles may engender negative reactions and lower ratings simply because their competence violates the prescriptive norms for female behavior. This outcome seems particularly likely for women who exhibit anger (a “male” emotion) and for women who use self-promoting, powerful verbal and nonverbal status cues. At the same time, men are penalized in evaluations for exhibiting communal or stereotypic female behaviors (e.g., parenthood or self-effacing speech). Providing evidence that agentic, competent women also behave in gender-congruent communal ways helps mitigate this negative bias, however, women must be careful not to seem overly communal by bringing attention to the fact that they are parents or by seeming too feminine in appearance or scent. The potential benefit to a woman who is applying for a male sex-typed job of having a gender-ambiguous name is worth noting.

Diversity training and employment equity policies would seem logical institutional initiatives to promote gender equity. Evidence from our review suggests, however, that these directives do not ensure gender equity in hiring. Furthermore, if such directives result in women’s presence as a small proportion of an applicant pool, individuating from the stereotypes of the social group that women occupy becomes more difficult, and they may be less likely to be hired.

Counterstereotype training was effective only under certain circumstances. This review covered more than 30 years of publications. More recent studies often built on previous work and tended to employ more sophisticated interventions and analyses, but there was no clear diminution of gender bias in the findings between earlier and more recent studies. Several studies did not meet all inclusion criteria but are worth mentioning. Bragger and colleagues found that structured interviews with standardized, sequential questions that were relevant to the position eliminated the hiring bias against pregnant applicants found when the same information was obtained through haphazard conversation. Glick and colleagues found “sexy” attire was a particular disadvantage, as compared with neutral dress, for women applying for a managerial position. Wiley and Eskilson found that applicants with tentive speech patterns, regardless of gender, received lower ratings. The benefit of gender ambiguity was striking in a study comparing employer response to identical resumes with female names or initials. Davies and colleagues found that affirmation that both men and women are equally capable prevented female-stereotype priming from undermining women’s subsequent leadership aspirations. McConnell and Fazio found that use of the title “chairman” primed raters to give a position more stereotypic masculine ratings than did the use of “chairperson” or “chair.” Martell found that gender bias in rating police officers was eliminated by the reducing time pressure and cognitive distraction during evaluation. Heilman and Okimoto confirmed the importance for highly agentic women of providing evidence of communality, to prevent negative ratings. Hugenberg and colleagues found less gender bias in selection when raters decided whom to include rather than whom to exclude from a list of individuals in a male sex-typed job.

This study had some limitations. Evidence-based recommendations are limited by the predominant use of college students as participants, although gender bias in evaluation was also found in the six studies with adult nonstudent participants. Furthermore, Marlowe and colleagues found gender biases even in the evaluations of experienced managers. The absence of any study in an academic medicine setting is a limitation in the capacity to generalize our findings to academic medicine. We also...
have little information on the ethnic–racial diversity of the participants, but, given the populations from which these studies drew participants, it is likely that nearly all were white. Finally, although the randomized controlled design of these studies is important for establishing a causal relationship between the intervention and the outcome, the success of these interventions in actual employment settings is unknown.

List 2
Evidence-based recommendations to reduce the application of bias that could disadvantage women applicants in hiring settings*

Recommendations for Institutions

- Design process to allow applicants to provide individuating evidence of job-relevant competency
- Visibly display research evidence that men and women are equivalently successful in male sex-typed roles
- Work hard to ensure that women comprise at least 25% of an applicant pool
- Insist that raters commit to the value of specific credentials before seeing actual applicants
- Rate specific qualifications before making summary judgments about applicant
design and antibias training so that raters do not feel coerced during evaluation
- Do not ask about parenthood status in the application
- Encourage raters to spend adequate time and avoid cognitive distractions during evaluation
- Use structured rather than unstructured interviews
- Do not use man-suffix in job titles (e.g., use “chair” or “chairperson” as opposed to “chairman”)
- Implement training workshops for personnel decision makers that include examples of common hiring biases and group problem solving for overcoming such biases
- Encourage raters to use an inclusion rather than an exclusion selection strategy in constructing a final list of applicants

Recommendations for female applicants

- Provide some evidence of communal job-relevant behaviors (e.g., being helpful and sensitive to the needs of subordinates)
- Indicate clear evidence of competency (e.g., resume, third-party endorsements) but avoid appearing self-promoting in an interview
- Do not show anger or discuss previous job-related situations that made you angry
- Best to avoid feminine-scented perfume, but wearing masculine-scented perfume may be beneficial (although you would need to pretest the scent to ensure that it is considered “masculine”)
- Avoid revealing parenthood status until job and salary are secured
- In your initial application, if you have a female-gendered first name, consider using initials instead of the full name
- Strive for an “attractive” but neutral appearance for interviews or presentation photographs. Avoid interviewing in overly feminine clothing (more masculine clothing and facial features may be beneficial)
- If you are visibly pregnant, it might be wise to obscure it with your clothing
- Avoid tentative speech patterns (e.g., use of intensifiers such as “really” and “definitely,” “hedges such as “I guess” or “sort of,” and hesitations such as “well” or “let’s see”)

* All studies cited in this table, except those with superscript symbols, met inclusion criteria for the systematic review. The studies with superscripts are all experimental, controlled studies, but they were excluded from the systematic review for the reasons listed below. We include the citations in this table because the studies support the recommendation.

†† Intervention appeared to be randomly assigned, but this was not specifically stated.

Conclusions
This review identifies several institutional interventions with a high level of evidence promising the possibility of promoting gender equity in hiring (List 2). The limitations of the studies, in combination with the continual and rapid evolution of social norms, make us reluctant to dictate to individual female applicants behaviors that may enhance their hireability. Whereas we are mindful of these caveats, we also provide recommendations for individual applicants that are supported by the existing research evidence (List 2).

The National Institutes of Health Roadmap calls for scientists to move beyond the limits of their own discipline and explore new organizational models for interdisciplinary science. Evidence-based practice has become a core value of academic medicine. With this systematic review, we encourage those within the institution of academic medicine to apply evidence from social science research to the practice of personnel decision making.

Acknowledgments
The authors would like to acknowledge the assistance from the following individuals: Sandra Phelps and Heidi Markeu (Ebling Library librarians), Phyllis Holman Weisbard (Women’s Studies librarian), and Jennifer Sheridan (chief librarian) at Meriter Hospital.

This research was funded by the University of Wisconsin School of Medicine and Public Health and Meriter Hospital. Dr. Isaac was supported by grant no. T32 AG00265 from the National Institute on Aging. Dr. Carnes is employed part time by the William S. Middleton Veterans Hospital. This report is GRECC manuscript no. 2000-30.

References
Muchinsky PM, Harris SL. The effect of


Academic Medicine, Vol. 84, No. 10 / October 2009


References Cited in List 2 Only.


Academic Medicine, Vol. 84, No. 10 / October 2009