



Gender bias in the evaluation of interns in different medical specialties: An archival study

Roy Adar, Rotem Kahalon, Johannes Ullrich, Arnon Afek & Vered H. Eisenberg

To cite this article: Roy Adar, Rotem Kahalon, Johannes Ullrich, Arnon Afek & Vered H. Eisenberg (2022) Gender bias in the evaluation of interns in different medical specialties: An archival study, *Medical Teacher*, 44:8, 893-899, DOI: [10.1080/0142159X.2022.2046715](https://doi.org/10.1080/0142159X.2022.2046715)

To link to this article: <https://doi.org/10.1080/0142159X.2022.2046715>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 22 Mar 2022.



[Submit your article to this journal](#)



Article views: 2437



[View related articles](#)



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)

Gender bias in the evaluation of interns in different medical specialties: An archival study

Roy Adar^{a*}, Rotem Kahalon^{b,c*}, Johannes Ullrich^b, Arnon Afek^d and Vered H. Eisenberg^d

^aDepartment of Urology, Sheba Medical Center, Ramat Gan, Israel; ^bDepartment of Psychology, University of Zurich, Zurich, Switzerland; ^cDepartment of Psychology, Simon Fraser University, Canada; ^dManagement & research unit, Sheba Medical Center, Tel Hashomer, Israel

ABSTRACT

Introduction: The field of medicine is characterized by within-field gender segregation: Gender ratios vary systematically by subdisciplines. This segregation might be, in part, due to gender bias in the assessment of women and men medical doctors.

Methods: We examined whether the assessments, i.e. overall score, department scores and skills scores, interns receive by their superiors during their internship year, vary as a function of their gender and the representation of women in the field. We analyzed an archival data set from a large hospital in Israel which included 3326 assessments that were given to all interns who completed their internship year between 2015 and 2019.

Results: Women received lower department scores and skills scores in fields with a low (versus high) representation of women. Men received higher scores in fields with a high (versus low) representation of men, yet there was no difference in their skills scores.

Conclusions: Women are evaluated more negatively in fields with a low representation of women doctors. Similarly, men are evaluated more negatively in fields with a low representation of men, yet this cannot be explained by their skills. This pattern of results might point to a gender bias in assessments. A better understanding of these differences is important as assessments affect interns' career choices and options.

KEYWORDS



Gender bias; resident selection; trainee selection; career choice; women in surgery; Gender differences

In the last decades women's labor-force participation patterns in most Western countries have shifted (U.S. Bureau of Labor Statistics 2018). Initially, the job-market was characterized by a between-field gender segregation—men were overrepresented in occupations requiring physical strength, and analytical/technical skills, while women were mostly represented in occupations requiring social and nurturing skills. Today, more and more women are entering formerly male dominated fields. This has led to the new phenomenon of within-field gender segregation, with women dominated subfields emerging in formerly male dominated occupations (Levanon and Grusky 2016). Within-field gender segregation is well documented in medicine. Women mostly specialize in residencies that are considered as requiring more social skills, such as Family Medicine, Internal Medicine, and Pediatrics, but are underrepresented in surgical fields, which are mostly occupied by men (Kilminster et al. 2007; Alers et al. 2014; Burgos and Josephson 2014; Abelson et al. 2016). This segregation, which exists across countries (Buddeberg-Fischer et al. 2006; Fukuda and Harada 2010; Abelson et al. 2016; Pelley and Carnes 2020), can be problematic both for society—when women's residential choices do not match real systemic needs (i.e. lack of surgical trainees in Japan and Canada; Scott et al. 2008; Fukuda and Harada 2010)—and for women's career options and future income (Buddeberg-

Practice points

- The field of medicine is characterized by within-field gender segregation—women mostly specialize in residencies that are considered as requiring more social skills, but are underrepresented in surgical fields, which are mostly occupied by men.
- Gender bias in the evaluation of interns in different medical specialties might contribute to this segregation.
- Women receive lower department scores and skills scores in fields with a low (versus high) representation of women, while men receive higher scores in fields with a high (versus low) representation of men.
- Assessments received by superiors might affect interns' career choices and options.
- Awareness of subtle gender biases should be encouraged, and more objective measures for intern assessments are required.

Fischer et al. 2006; Jena et al. 2016), as surgeons are usually better compensated than non-surgical physicians (Weaver et al. 2015).

CONTACT Roy Adar, MD  Roy.Adar@sheba.health.gov.il  Department of Urology, Sheba Medical Center, Tel Hashomer, Derech Sheba 2, Ramat Gan, Israel

*The first and second authors contributed equally to the manuscript. Their names are presented in alphabetical order.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Previous research, which aimed to identify the reasons for this disparity mostly sought to understand and mitigate the barriers that block women's interest in male dominated fields. Using self-report measures and interviews, this literature suggests that career barriers for women in surgery, as identified by women, are complex. They include personal preferences and social influences, such as quality of life, social values conflict, and concerns regarding issues of work-life balance (Gjerberg 2003; Lambert et al. 2003; Wendel et al. 2003; Goldacre et al. 2012; Alers et al. 2014; Giantini Larsen et al. 2021; Zdravkovic et al. 2021). They also include barriers within surgery such as ineffective mentorship, lack of belonging, organizational culture and institutional policies that negatively affect opportunities for advancement and unclear expectations for advancement (Cochran et al. 2019; Thompson-Burdine et al. 2019; Zdravkovic et al. 2020).

Importantly, not all reasons can be easily identified by self-report measures. Subtle factors can also affect the sense of fit to a specific environment (Schmader and Sedikides 2018), eventually affecting residential decisions. In the present research we focused on assessment scores as a factor that might contribute to the gender segregation in medicine. Specifically, we tested whether the numerical representation of women and men in a given specialty creates a difference in how women and men interns are assessed by their superiors. We focused on both women and men interns, because although the lack of women in surgery is a well-known problem, understanding the reasons for men's underrepresentation in some fields in medicine (e.g. family medicine) is also important for reducing gender segregation in medicine and for promoting equal opportunities for both women and men (Croft et al. 2015).

Our prediction that the numerical representation of men and women in a given specialty will affect how interns are assessed by their superiors was based on research from the field of social psychology. This research suggests that stereotypes about the suitability of men/women to a given occupation are strongly influenced by the numerical representation of men and women in the field. Specifically, people tend to assume that in occupations that are dominated by men, stereotypically masculine attributes (e.g. competitiveness, assertiveness) are necessary for success, whereas in occupations that are dominated by women, stereotypically feminine attributes (e.g. empathy, warmth, sociability) are necessary instead (Cejka and Eagly 1999). In the case of Medicine, the low proportion of women in surgical specialties and the relatively low proportion of men in more social specialties (Shemesh et al. 2012; AAMC 2019), may result in perceiving surgical roles as more suitable for men, and some non-surgical roles as more suitable for women.

The present research was conducted in Israel in which medical students are required to complete a one-year internship in an accredited hospital upon completion of their medical studies, as a prerequisite to receiving licensure. Rotatory internship is customary in many countries around the world, including Australia, Ecuador, Indonesia, Ireland, Peru, Slovenia and others. In Israel, since 2014, interns are assigned to hospitals based on an algorithm that takes into account students' preferences and the average number of hospital beds in each hospital (Bronfman

et al. 2017). Importantly, interns are not assigned to hospitals on the basis of merit.

As in many other countries, the internship period in Israel is a critical point in a doctor's medical career. This is because during this time interns are expected to choose where they would like to specialize. Most doctors in Israel have a specialization (80% of those who studied in Israel and 57% of those who studied abroad; Tor-Sinai et al. 2020). While some students show an inclination to a specific specialty during medical school, it is common in Israel (as well as in other countries) that interns use their internship year to 'feel' the type of work in different medical fields before they choose a specialty and a medical facility (Houlden et al. 2004; Mihalynuk et al. 2006; Lumb and Murdoch-Eaton 2014; Katz et al. 2018).

As a part of the internship year, interns work at five mandatory departments (General Surgery, Internal Medicine, Emergency Medicine, Pediatrics and Anesthesia) and one to two elective departments—which are chosen by the interns individually based on their interest. In each of these departments the interns receive assessments on their skills and performance by a senior doctor in the department. These assessments are recorded in interns' personal files and can be used to help differentiate between residency applicants in the resident selection process, as no uniform selection criteria for residency acceptance exist. The assessments given to young interns at this critical junction may affect superiors' decisions to accept an intern to a specific residency, and they can also affect the interns' own perceptions about their suitability to different medical specialties. Both can influence career options, choices and goals (Bing-You et al. 2017; Hu et al. 2017). We examined whether the assessments men and women interns receive by their superiors in different departments vary as a function of the proportion of women in the respective medical specialty in Israel.

Method

Procedure and materials

The study was approved by the Institutional Review Board at the Sheba Medical Center (SMC-19-5963). We analyzed the hospitals' archival data set which included interns' assessment scores from the years 2015–2019. A total of 3326 review reports were collected from 421 interns at Sheba Medical Center, the largest hospital in Israel (which therefore receives the largest number of interns per year). The data set contains the assessments of all interns that enrolled in the hospital during these years. Each intern received between six to ten assessments in different departments, therefore the assessments were nested within participants. Based on Arend and Schäfer (2019), our sample size affords 80% power for detecting a medium sized effect.

Primary outcomes included the overall *department score* participants received in each department, which ranged from 1(*poor*) to 3(*excellent*). In addition, in each department, participants were assessed on five different skills: Theoretical knowledge, clinical judgment, practical skills, curiosity and motivation and care for patients. Each of these ranged from 1(*poor*) to 5(*excellent*), yet some assessments included higher numbers such as 5.5 or 6. A factor analysis with a varimax rotation was conducted to examine

Table 1. Descriptive statistics for assessment scores.

	Range	Overall mean M (SD)	Men (n = 246) M (SD)	Women (n = 175) M (SD)
Internship scores	1–4	2.49 (1.05)	2.31 (1.07)	2.72 (.98)
Department scores	1–3	2.57 (.50)	2.51 (.51)	2.66 (.47)
Specific skills scores	1–6	4.46 (.59)	4.38 (.59)	4.58 (.51)

Note. All means are presented while controlling for place of study, which was entered to the ANCOVA model as a covariate. M = mean, SD = Standard deviation. Higher numbers represent better scores.

the number of dimensions which are assessed by the specific skills. The results suggested a one-factor solution (i.e. all five skills had a similar response pattern, eigenvalue = 4.52, explaining 75.29% of the variance), therefore the specific skills were averaged to create one score representing interns' *specific skills scores* (internal consistency reliability was high, $\alpha = .93$), which serves as another primary outcome. Department scores and specific skills scores were given by the same evaluator. Finally, the last primary outcome was the *internship score*, which was given to interns based on their performance across all departments and ranged from 1(*passed*) to 4 (*excelled*).

The archival data set also included basic demographic information for each intern: gender (man/woman)¹ and internship year (2015–2019). It also included information regarding place of study (coded Israel/Abroad). Because students who study abroad usually do so because they were not accepted to Israeli institutions, this served as an indicator for pre-existing ability (Note that in Israel most students study abroad due to the inability of the domestic medical training system to meet demand, therefore the acceptance requirement are extremely high (Gamzu et al. 2016). Descriptive statistics for internship scores, department scores and specific skills scores, while controlling for place of study, are available in Table 1. The dataset is available upon request from the first author.

For each medical field we classified the representation of women as relatively high or relatively low. This classification was based on the proportion of women who are board-certified (i.e. licensed) in each specialty obtained from reports of the Israeli Ministry of Health (average from the relevant years). In Israel, physicians are deemed specialists (i.e. board-certified) if they had fulfilled their respected specialty's requirements (for instance passed several tests, performed a specific number of a given procedure) dictated by the Scientific Council in the Israeli Medical Association. Specifically, specialties in which the proportion of board-certified women was lower than the median (39.75%) were considered as departments with a low representation of women (e.g. Orthopedic Surgery, General Surgery), while specialties in which the proportion of board-certified women was higher than the median were considered as departments with a relatively high representation of women (e.g. Pediatrics, Family Medicine).² This dichotomous categorization was chosen due to the low number of women who choose elective departments in specialization with a relatively low percentage of women.

Statistical analyses

Data were analyzed with SPSS version 25.0 (Chicago, IL, USA). Missing data analysis on the main variables of interest revealed that only 0.5% of all values were missing.

Considering individual cases, 98.5% of the participants had no missing data. Listwise deletion was used to deal with missing data.

Two hierarchical linear models (HLM) were conducted with department scores and specific skills scores as the dependent variables. All variables were standardized before entering them to the model. Hierarchical linear models are an extension of simple linear models that allow both fixed and random effects and are particularly used when there is non-independence in the data due to clustering of observations within persons or groups ($ICC > .10$).³ In this case, 6–10 assessments were available for each intern. These department level observations are not independent, as within a given intern the assessments of different departments are more similar, therefore a hierarchical model, with grouping by intern, is more suitable (Raudenbush and Bryk 2002). In both models, department type (relatively high/low representation of women) was the level one variable while gender (man/woman) was the level two variable. Study place (Israel/abroad) was included as a level two covariate to control for differences in perceived abilities that stem from the place of study, considering that doctors who study abroad tend to receive lower scores than those who study in Israel (Shemesh et al. 2012). Finally, the cross-level interaction between gender and department type was entered in the model. The AIC⁴ values (i.e. indicators for the relative quality of the statistical model) in both models were smaller than the null models (that take into account only the subject-level variability), indicating a better model fit. Specifically, including the predictors mentioned above improved the models; for department scores from 5812.15 to 5493.20, and for specific skills from 6901.30 to 6635.33. To analyze the overall internship scores, we used an analysis of variance (ANOVA) with gender as a predictor and study place as a covariate.

Results

Characteristics of study participants. Of the 421 interns, 246 (58.4%) were men and 175 (41.6%) women, representing the gender distribution between recently licensed men and women medical doctors in Israel (Ministry of Health 2019). Of the women, 68.6% studied in Israel while 38.3% of the men studied in Israel, representing a trend among Israeli men, especially Israeli-Arab men, to study abroad (Tor-Sinai et al. 2020). In line with the Ministry of Health reports, women mostly chose elective departments in fields with a relatively high representation of women (72.3%), while the distribution for men was more balanced – 50.9% chose elective in fields with a low representation of women, and 49.1% in fields with a high representation of women. Figure 1 presents the proportion of board-certified women in each specialty according to the Ministry of Health reports, which reflects the specialty selection in Israel.

Primary outcomes

Department scores

The results indicated that study place was associated with department scores ($b = .33$, $t = 15.56$, $p < .001$, 95% CI [.29 to .37]), such that those who had studied abroad received lower assessment scores than those who had studied in

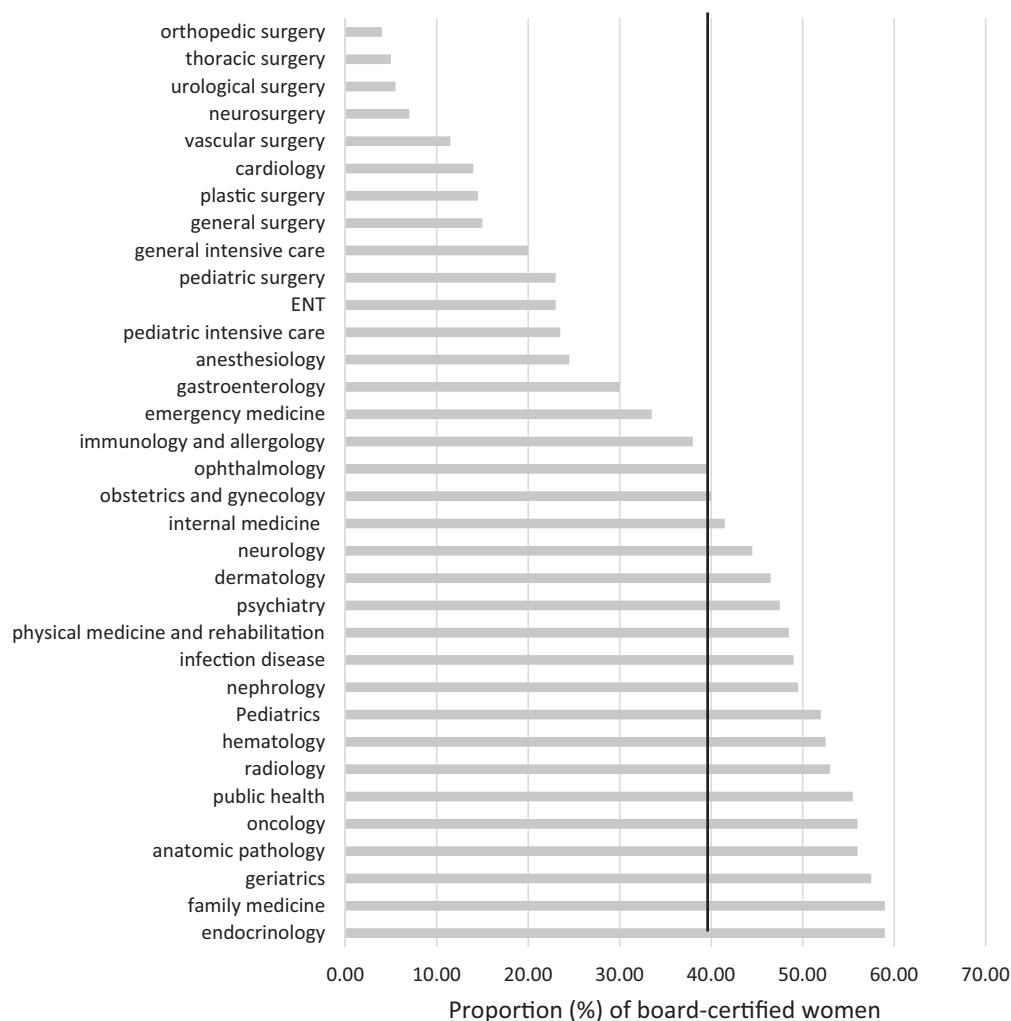


Figure 1. The proportion (%) of board-certified women in each specialty based on the ministry of health reports (average for the years 2015 and 2018). Note. The line represents the median of the proportion of board-certified women in all specialties in Israel.

Israel. Department type was not related to department scores⁵ ($b = -.01$, $t = -.04$, $p = .965$, 95% CI $[-.03$ to $.03]$). At the participant level, gender was found to be associated with department scores ($b = -.05$, $t = -2.39$, $p = .017$, 95% CI $[-.09$ to $-.01]$), such that women received better assessments in comparison to men, yet this effect was qualified by a significant cross-level interaction between gender and department type ($b = -.11$, $t = -3.43$, $p = .001$, 95% CI $[-.17$ to $-.05]$). Probing the interaction using Preachers' HLM online calculator (Preacher et al., 2006), showed that women received higher scores in departments with a relatively high, in comparison to low, representation of women, *simple slope* = .05 (.02), $Z = 2.21$, $p = .027$. For men, the opposite pattern occurred, such that they received higher scores in departments with a low, in comparison to high representation of women, *simple slope* = $-.05$ (.02), $Z = -2.69$, $p = .007$. Figure 2 presents the pattern of results.

Specific skills scores. The results indicated that study place was also associated with specific skills assessment scores ($b = .40$, $t = 15.16$, $p < .001$, 95% CI $[.35$ to $.45]$), such that those who had studied abroad received lower assessment scores than those who had studied in Israel. Department type was associated with specific skills assessment scores ($b = .04$, $t = 2.08$, $p = .038$, 95% CI $[.01$ to $.08]$), such that interns received higher scores in departments with a

relatively high, in comparison to low representation of women. At the participant level, gender was found to be associated with specific skills assessment scores ($b = -.08$, $t = 2.08$, $p = .004$, 95% CI $[-.13$ to $-.02]$), such that women tended to receive better assessments in comparison to men. These main effects were qualified by a significant cross-level interaction between gender and department type ($b = -.10$, $t = -2.64$, $p = .008$, 95% CI $[-.17$ to $-.03]$), such that women received higher assessment scores in departments with a relatively high representation of women in comparison to a low representation of women, *simple slope* = .08 (.03), $Z = 3.07$, $p = .002$. For men, no difference was found between the departments, *simple slope* = $-.01$ (.02), $Z = -.44$, $p = .662$. Figure 3 presents the pattern of results.

Overall internship score

No difference was found between women and men, $F(1,387) = .08$, $p = .784$, $\eta^2_p < .001$.

Discussion

This archival study revealed that assessment of interns might be biased by the percentage of men and women in the field. Women received lower department scores and specific skills scores in fields with a low, versus high,

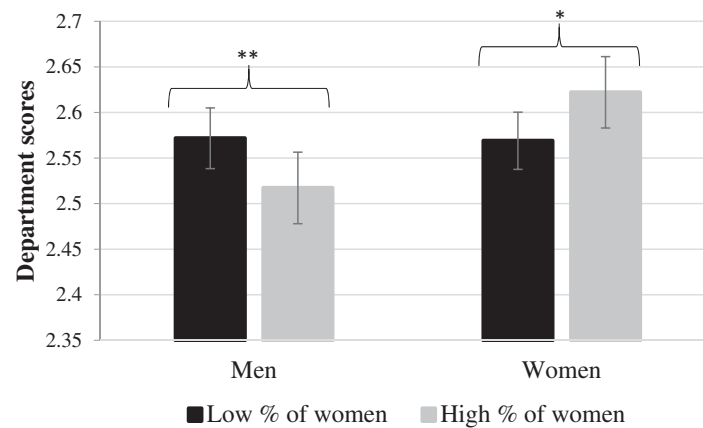


Figure 2. Interns' department scores in medical fields with a low and high representation of women.

Note. Lines represent 95% confidence intervals. Means scores are presented, higher numbers represent better scores. * $p < 0.05$, ** $p < 0.01$.

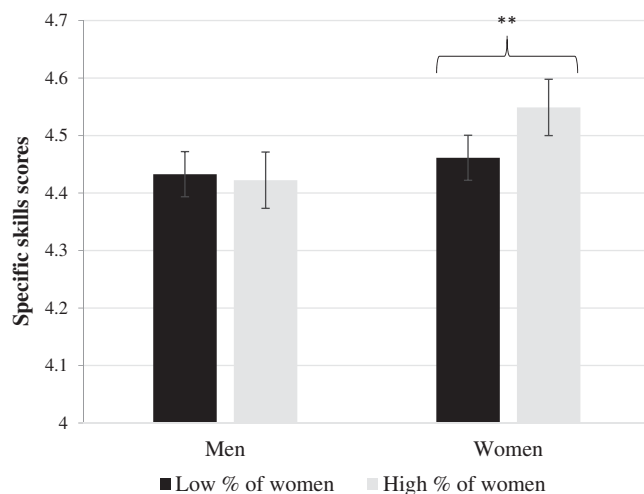


Figure 3. Interns' scores on specific skills in medical fields with a relatively low and a relatively high representation of women.

Note. Lines represent 95% confidence intervals. Means scores are presented, higher numbers represent better scores. * $p < 0.05$, ** $p < 0.01$.

representation of women. These results suggest that the representation of women in a field is related to how women are assessed in a given department, such that women are evaluated more negatively in fields with a low representation of women doctors. Men received lower department scores in fields with a high representation of women in comparison to those with a low representation of women, yet no difference was found in how men were assessed on specific skills. This suggests that skills alone cannot explain the lower department scores that men received in fields with a high representation of women.

Although the actual differences between scores were relatively small, even small differences in assessment scores can amplify to large differences later on and have a large cumulative impact (Greenwald et al. 2014). This can, therefore, serve as a barrier to the inclusion of women and men in gender-incongruent fields. To illustrate, in a study conducted in the United States, it was found that small differences in assessed performance between underrepresented in medicine (UIM) and not-UIM students, led to larger differences in grades and selection for awards. The authors concluded that this can result in differences in opportunities for competitive residency programs and entrance to academic careers (Teherani et al. 2018). Moreover, even small signals of lack of fit could affect the motivation of

young people from underprivileged groups (e.g. women, ethnic minorities) from further pursuing careers in which they feel that their group is socially devalued (Schmader and Sedikides 2018).

These differences could also affect residency acceptance chances. The Israeli medical system does not have a uniform selection process or criteria for residency acceptance. Therefore, when applying for residency programs, scores from departments that share similar characteristics (i.e. surgical fields), might be deemed as more relevant and be used to help differentiate amongst residency candidates.

As for the theoretical explanation, the observed pattern of intern's assessments might imply a subtle gender bias held by physicians. This might stem from the fact that gender beliefs are associated with anticipated performance levels, even when gender is actually unrelated to relevant outcomes (Ridgeway and Erickson 2000). The results of the present study are in line with previous findings showing that men and women are evaluated differently in different specialties within medicine (Brown et al. 2020; Samuriwo et al. 2020). Specifically, health care professionals were found to hold implicit and explicit biases associating men with career and women with family, and surgeons tended to associate men with surgery and women with Family Medicine (Salles et al. 2019). In the same line, female General Surgery residents were given less autonomy than their male peers during laparoscopic surgery, even after controlling for technical performance, post graduate year level, and case factors (Hoops et al. 2019). These results together suggest that women are judged as less suitable for surgery than their male colleagues. Yet, our results also suggest that there might also be a bias regarding the judgment of men's suitability to fields with a relatively high representation of women, which might direct men away from these departments.

Our research provides an important contribution to the literature on gender segregation within the field of medicine, as it points to a possible subtle mechanism which might affect women's and men's career choices—gender differences in assessments which might stem from gender bias. Importantly, even people who want to be fair and objective can be affected by biases because they 'catch them off guard' (Uhlmann and Cohen 2007). Senior doctors who teach and evaluate young physicians or even medical students, are, therefore, at risk for such biases, and this can affect how their interns view themselves and their medical abilities (Lai

et al. 2009). As such, gender differences in interns' assessments can direct both women and men to gender congruent fields, resulting in loss of talent in all medical fields and economic disadvantage for women.

Limitations

The natural setting in which these findings were observed is a clear advantage of this study, as assessment scores allow to examine how interns are judged in different departments without the risk of demand characteristic effects (i.e. awareness or suspicion of the purpose of the study). On the other hand, the retrospective nature of our study also holds limitations, particularly with regard to the availability of control variables. One of such factors is the gender of the doctor making the assessment. Although we could not control for this measure in our analysis, a report from HR suggests that the majority of heads of departments in the examined medical center are men (above 85%). In addition, previous research found that although faculty's gender can affect evaluations (e.g. gender-congruent resident faculty pairings were more beneficial to male than to female residents), both female and male residents hold gender biases which can affect their assessments (Klein et al. 2020).

Another limitation is that the nature of the study does not allow to establish causal relations. Therefore, we cannot rule out completely that the results were due to actual differences in women's and men's abilities in different departments. This is less likely for men, as the pattern of the results for their skills score was inconsistent with their department scores. Importantly, empirical findings on gender differences in abilities are mixed (Kilminster et al. 2007). Some found that female medical students received higher scores in basic surgical skills training and in theoretical examination than male students (Lou et al. 2016). Yet, others found no differences between women's and men's microsurgical skills (Flyckt et al. 2017; Sudario-Lumague et al. 2018), or found that differences occurred in the learning process but not necessarily in skills (Ali et al. 2015). Future research should take into account that the evidence required to rule out differences in abilities might suffer from the same bias observed here, suggesting that more objective measures for assessments are required.

Conclusions

Women are evaluated more negatively in fields with a low representation of women doctors. Similarly, men are evaluated more negatively in fields with a low representation of men, yet this cannot be explained by their skills. This pattern of results might point to a gender bias in assessments. Raising awareness of subtle gender biases and their possible effects is an important step toward minimizing their effects. We hope that the data presented here are a useful step in that direction.

Ethical approval

The study was approved by the Institutional Review Board at the Sheba Medical Center (SMC-19-5963), March 19th, 2019.

Disclosure statement

We hereby declare that we have no potential conflicts of interest with respect to the authorship or the publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research of this article: Swiss Government Excellence Scholarship awarded to Rotem Kahalon (Grant No. 2020.0085).

Notes

1. The interns' gender in the present study was recorded as man or a woman. We had no data regarding other genders.
2. Note that women became more represented in Obstetrics and Gynecology from 2015 (37% women) to 2018 (43% women). No changes were found between the examined years in other departments.
3. ICC- Intraclass Correlation.
4. AIC- Akaike's information criterion. The lower the AIC value is, the better fit a model is.
5. We report *b* - beta coefficients (i.e., the degree of change in the outcome variable for every 1-unit of change in the predictor variable) and *t* - test statistics (i.e., the ratio of the estimated coefficient to its standard error).

Notes on contributors

Roy Adar, MD, is a urology resident at the department of Urology at Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel.

Rotem Kahalon, PhD, is a postdoctoral fellow at the department of psychology at the University of Zurich, Zurich Switzerland, and at the department of psychology at Simon Fraser University, Canada.

Johannes Ullrich, PhD, is a professor of social psychology at the department of psychology, University of Zurich, Zurich Switzerland.

Arnon Afek, MD, MHA, is a deputy director-general at Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel.

Vered H Eisenberg, MD, is the chief quality officer and a gynecologist at Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel.

References

- Abelson JS, Chartrand G, Moo TA, Moore M, Yeo H. 2016. The climb to break the glass ceiling in surgery: trends in women progressing from medical school to surgical training and academic leadership from 1994 to 2015. *Am J Surg*. 212(4):566–572.e1.
- Alers M, Pepping T, Bor H, Verdonk P, Hamberg K, Lagro-Janssen A. 2014. Speciality preferences in Dutch medical students influenced by their anticipation on family responsibilities. *Perspect Med Educ*. 3(6):443–454.
- Alers M, van Leerdam L, Dielissen P, Lagro-Janssen A. 2014. Gendered specialities during medical education: a literature review. *Perspect Med Educ*. 3(3):163–178.
- Ali A, Subhi Y, Ringsted C, Konge L. 2015. Gender differences in the acquisition of surgical skills: a systematic review. *Surg Endosc*. 29(11):3065–3073.
- Arend MG, Schäfer T. 2019. Statistical power in two-level models: a tutorial based on monte carlo simulation. *Psychol Methods*. 24(1): 1–19.
- Association of American Medical Colleges (AAMC). 2019. Physician workforce data report. <https://www.aamc.org/what-we-do/mission-areas/patient-care/workforce-studies/reports>.
- Bing-You R, Hayes V, Varaklis K, Trowbridge R, Kemp H, McKelvy D. 2017. Feedback for learners in medical education: What is known? A scoping review. *Acad Med*. 92(9):1346–1354.
- Bronfman S, Hassidim A, Kalif G, Romm A. 2017. Matching practices for entry-labor markets – The Israeli Medical Internship Match, MiP

- Country Profile 25. <https://www.matching-in-practice.eu/the-israeli-medical-internship-match/>.
- Brown MEL, Hunt GEG, Hughes F, Finn GM. 2020. Too male, too pale, too stale: A qualitative exploration of student experiences of gender bias within medical education. *BMJ Open*. 10(8):e039092.
- Buddeberg-Fischer B, Klaghofer R, Abel T, Buddeberg C. 2006. Swiss residents' speciality choices-impact of gender, personality traits, career motivation and life goals. *BMC Health Serv Res*. 6(1):137.
- Burgos CM, Josephson A. 2014. Gender differences in the learning and teaching of surgery: a literature review. *Int J Med Educ*. 5:110-124.
- Cejka MA, Eagly AH. 1999. Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Pers Soc Psychol Bull*. 25(4):413-423.
- Cochran A, Neumayer LA, Elder WB. 2019. Barriers to careers identified by women in academic surgery: a grounded theory model. *Am J Surg*. 218(4):780-785.
- Croft A, Schmader T, Block K. 2015. An underexamined inequality: cultural and psychological barriers to men's engagement with communal roles. *Pers Soc Psychol Rev*. 19(4):343-370.
- Flyckt RL, White EE, Goodman LR, Mohr C, Dutta S, Zanotti KM. 2017. The use of laparoscopy simulation to explore gender differences in resident surgical confidence. *Obstet Gynecol Int*. 2017:1-7.
- Fukuda Y, Harada T. 2010. Gender differences in specialty preference and mismatch with real needs in Japanese medical students. *BMC Med Educ*. 10(1):15.
- Gamzu R, Kaidar N, Afek A, Horev T. 2016. Physician density planning in a public healthcare system: complexities, threats and opportunities-The case of the Israeli healthcare system. *Health Policy*. 120(8):920-927.
- Giantini Larsen AM, Pories S, Parangi S, Robertson FC. 2021. Barriers to pursuing a career in surgery: an institutional survey of harvard medical school students. *Ann Surg*. 273(6):1120-1126.
- Gjerberg E. 2003. Women doctors in Norway: the challenging balance between career and family life. *Soc Sci Med*. 57(7):1327-1341.
- Goldacre MJ, Goldacre R, Lambert TW. 2012. Doctors who considered but did not pursue specific clinical specialties as careers: questionnaire surveys. *J R Soc Med*. 105(4):166-176.
- Greenwald AG, Banaji MR, Nosek BA. 2014. Statistically small effects of the implicit association test can have societally large effects. *J Pers Soc Psychol*. 108(4):553-561.
- Hoops H, Heston A, Dewey E, Spight D, Brasel K, Kiraly L. 2019. Resident autonomy in the operating room: does gender matter? *Am J Surg*. 217(2):301-305.
- Houlden RL, Raja JB, Collier CP, Clark AF, Waugh JM. 2004. Medical students' perceptions of an undergraduate research elective. *Med Teach*. 26(7):659-661.
- Hu S, Creed PA, Hood M. 2017. Career goal revision in response to negative feedback: testing a longitudinal cross-lagged model. *J Couns Psychol*. 64(3):335-345.
- Jena AB, Olenski AR, Blumenthal DM. 2016. Sex differences in physician salary in US public medical schools. *JAMA Intern Med*. 176(9):1294-1304.
- Katz DA, Shnek O, Rave-Braver B, Yinon A, Monter G. 2018. Choice of internship among interns in a changing era. *Ha-Refuah*. 157(6):356-360.
- Kilminster S, Downes J, Gough B, Murdoch-Eaton D, Roberts T. 2007. Women in medicine-is there a problem? A literature review of the changing gender composition, structures and occupational cultures in medicine. *Med Educ*. 41(1):39-49.
- Klein R, Ufere NN, Rao SR, Koch J, Volerman A, Snyder ED, Schaeffer S, Thompson V, Warner AS, Julian KA, et al. 2020. Association of gender with learner assessment in graduate medical education. *JAMA Netw Open*. 3(7):e2010888.
- Lai NM, Nalliah S, Jutti RC, Hla YY, Lim VKE. 2009. The educational environment and self-perceived clinical competence of senior medical students in a Malaysian medical school. *Educ Heal Chang Learn Pract*. 22(2):148.
- Lambert TW, Davidson JM, Evans J, Goldacre MJ. 2003. Doctors' reasons for rejecting initial choices of specialties as long-term careers. *Med Educ*. 37(4):312-318.
- Levanon A, Grusky DB. 2016. The persistence of extreme gender segregation in the twenty-first century. *Am J Sociol*. 122(2):573-619.
- Lou Z, Yan F. h, Zhao Z. q, Zhang W, Shui X. q, Liu J, Zhuo D. I, Li L, Yu E. d. 2016. The sex difference in basic surgical skills learning: a comparative study. *J Surg Educ*. 73(5):902-905.
- Lumb A, Murdoch-Eaton D. 2014. Electives in undergraduate medical education: AMEE Guide No. 88. *Med Teach*. 36(7):557-572.
- Mihalynuk T, Leung G, Fraser J, Bates J, Snadden D. 2006. Free choice and career choice: clerkship electives in medical education. *Med Educ*. 40(11):1065-1071.
- Ministry of Health. 2019. Medical professions human resource report 2019.
- Pelley E, Carnes M. 2020. When a specialty becomes "women's Work": trends in and implications of specialty gender segregation in medicine. *Acad Med*. 95(10):1499-1506.
- Preacher KJ, Curran PJ, Bauer DJ. 2006. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *J Educ Behav Stat*. 31(4):437-448.
- Raudenbush SW, Bryk AS. 2002. Hierarchical linear models: applications and data analysis Methods. Newbury Park (CA): Sage.
- Ridgeway CL, Erickson KG. 2000. Creating and spreading status beliefs. *Am J Sociol*. 106(3):579-615.
- Salles A, Awad M, Goldin L, Krus K, Lee JV, Schwabe MT, Lai CK. 2019. Estimating implicit and explicit gender bias among health care professionals and surgeons. *JAMA Netw Open*. 2(7):e196545.
- Samuriwo R, Patel Y, Webb K, Bullock A. 2020. 'Man up': medical students' perceptions of gender and learning in clinical practice: a qualitative study. *Med Educ*. 54(2):150-161.
- Schmader T, Sedikides C. 2018. State authenticity as fit to environment: the implications of social identity for fit, authenticity, and self-segregation. *Pers Soc Psychol Rev*. 22(3):228-259.
- Scott IM, Matejcek AN, Gowans MC, Wright BJ, Brenneis FR. 2008. Choosing a career in surgery: factors that influence Canadian medical students' interest in pursuing a surgical career. *Can J Surg*. 51(5):371-377.
- Shemesh A, Rotem N, Haklahi Z, Georgi M, Horev T. 2012. Occupational characteristics of doctors in Israel. *J Occup Environ*. 39(9):874-881.
- Sudario-Lumague R, Chiang YC, Lin TS. 2018. Gender comparison of medical student microsurgical skills in a laboratory model. *J Reconstr Microsurg*. 34(5):359-362.
- Teherani A, Hauer KE, Fernandez A, King TE, Lucey C. 2018. How small differences in assessed clinical performance amplify to large differences in grades and awards: a cascade with serious consequences for students underrepresented in medicine. *Acad Med*. 93(9):1286-1292.
- Thompson-Burdine JA, Telem DA, Waljee JF, Newman EA, Coleman DM, Stoll HI, Sandhu G. 2019. Defining barriers and facilitators to advancement for women in academic surgery. *JAMA Netw Open*. 2(8):e1910228.
- Tor-Sinai A, Zontag N, Blondheim O, Weinreb A, Tchernichovsky D. 2020. Doctors in Israel and trends in their training and characteristics. www.taubcenter.org.il.
- U.S. Bureau of Labor Statistics 2018. The economics daily:Labor force participation rates by age and sex, seasonally adjusted. <https://www.bls.gov/>.
- Uhlmann EL, Cohen GL. 2007. I think it, therefore it's true: effects of self-perceived objectivity on hiring discrimination. *Organ Behav Hum Decis Process*. 104(2):207-223.
- Weaver AC, Wetterneck TB, Whelan CT, Hinami K. 2015. A matter of priorities? Exploring the persistent gender pay gap in hospital medicine. *J Hosp Med*. 10(8):486-490.
- Wendel TM, Godellas CV, Prinz RA, Scott-Conner C, Estes NC, Pollak R. 2003. Are there gender differences in choosing a surgical career? *Surgery*. 134(4):591-596.
- Zdravkovic M, Neskovic V, Berger-Estilita J. 2021. Surveys on gender issues among anaesthesiologists: where do we go from here? *Journal of Gender Studies*. 30(7):868-871.
- Zdravkovic M, Osinova D, Brull SJ, Prielipp RC, Simões CM, Berger-Estilita J, Matas M, Santos S, Stroo K, Bouzia A, et al. 2020. Perceptions of gender equity in departmental leadership, research opportunities, and clinical work attitudes: an international survey of 11 781 anaesthesiologists. *Br J Anaesth*. 124(3):e160-e170.