

Diversity, Equity, Inclusion, and Justice

The Impact of Trainee and Standardized Patient Race and Gender on Internal Medicine Resident Communication Assessment Scores

Janae K. Heath, MD, MS
 C. Jessica Dine, MD, MS
 Denise LaMarra, MS, CHSE
 Serena Cardillo, MD

ABSTRACT

Background Standardized patient (SP) encounters are commonly used to assess communication skills in medical training. The impact of SP and resident demographics on the standardized communication ratings in residents has not been evaluated.

Objective To examine the impact of gender and race on SP assessments of internal medicine (IM) residents' communication skills during postgraduate year (PGY) 1.

Methods We performed a retrospective cohort study of all SP assessments of IM PGY-1 residents for a standardized communication exercise from 2012 to 2018. We performed descriptive analyses of numeric communication SP ratings by gender, race, and age (for residents and SPs). A generalized estimating equation model, clustered on individual SP, was used to determine the association of gender (among SP and residents) with communication ratings. A secondary analysis was performed to determine the impact of residents and SP racial concordance in communication scores.

Results There were 1356 SP assessments of 379 IM residents (199 male residents [53%] and 178 female residents [47%]). There were significant differences in average numeric communication rating (mean 3.40 vs 3.34, $P = .009$) by gender of resident, with higher scores in female residents. There were no significant interactions between SP and resident gender across the communication domains. There were no significant interactions noted with racial concordance between interns and SPs.

Conclusions Our data demonstrate an association of resident gender on ratings in standardized communication exercises, across multiple communication skills. There was not an interaction impact for gender or racial concordance between SPs and interns.

Introduction

Implicit biases, unconscious mental attitudes toward a person, thing, or group, are pervasive through society, including across medical academia and medical training.^{1,2} These biases influence perception of physician competencies across domains of race, ethnicity, sexual orientation, and gender.³⁻¹² Similarly, these biases impact competency-based assessments of medical trainees.¹³⁻¹⁵ Competency-based medical education relies on accurate assessments of trainees as they progress. Understanding the impact of implicit biases on these assessments is critical.

One commonly used assessment strategy in medical education is standardized patient (SP) encounters. These are commonly operationalized to assess communication and clinical skills¹⁶⁻²⁰; yet, the full impact

of implicit biases within this assessment strategy remains unknown.

Several studies assessed the role of gender and ethnicity of both trainee and SP, finding an association with gender and racial concordance with medical student assessments.^{21,22} However, these studies were limited in focus (solely concentrating on empathy ratings rather than the spectrum of communication and interpersonal skills often assessed by SP raters), or analyzed the impact on medical student assessment, but not postgraduate trainees. A subsequent single center study of emergency medicine residents focused on case-based simulations did not find any educationally significant rating differences based on residents or demographics, but did not evaluate communication skills (which may be prone to more implicit biases than checklist assessments).²³ Yet, similar to medical school curricula, residency programs frequently use SP programs to assess communication skills. As SP assessments aid in identifying residents for communication remediation, it is imperative to evaluate the impact of gender and race on SP assessments of these trainees.

DOI: <http://dx.doi.org/10.4300/JGME-D-21-00106.1>

Editor's Note: The online version of this article contains a description of communication domains.

Furthermore, effective communication requires skills beyond empathy. According to the Accreditation Council for Graduate Medical Education (ACGME) Milestones, interpersonal and communication skills deemed aspirational include “role models effective communication and development of therapeutic relationships,” “models cross-cultural communication,” and “establishes a therapeutic relationship with . . . persons of different socioeconomic and cultural backgrounds.”²⁴ Aligning with this expectation, assessment tools focus on discrete abilities such as eliciting and sharing relevant information, listening, respectfulness, and professionalism, as well as empathy. Differential ratings of trainees along these other communication skills is worth analyzing to fully understand the role of SP and provider traits on SP assessments.

The primary purpose of this study is to examine the effects of ethnicity and gender on SP assessments of residents’ communication skills. Our primary objective was to examine the interaction of SP and internal medicine (IM) resident gender on ultimate communication rating across several communication domains. A secondary objective was to examine the interaction of race (with both SP and resident) and the association with communication ratings. We hypothesized that implicit biases (across gender and race) exist within SP assessment of medical resident communication beyond empathy ratings. Specifically, we hypothesized that communication ratings would be affected by resident and SP gender, race, ethnicity, and the interaction between these demographics. Overall, understanding the influence of rater and resident demographics on these assessments is paramount in order to optimally interpret SP ratings for use in competency-based communication assessments.

Methods

Setting and Participants

We performed a retrospective cohort study of all SP assessments of University of Pennsylvania (UPenn) IM residents for a standardized communication exercise from 2012 to 2018. Each resident completed 4 distinct SP encounters during the communication assessments, which occurred in a single time point for each resident from October to November of postgraduate year (PGY) 1 over the study time line.

The program includes 4 discrete cases that assess skills in behavioral change counseling, assessment of health literacy, difficult conversations (delivery of bad news and navigating an emotionally charged encounter), and obtaining informed consent. Following each encounter, SPs rate resident communication skills on a 4-point Likert scale (ranging from “almost never” to “almost always”) in 6 domains: Eliciting Information, Listening, Giving Information, Respectfulness,

Objectives

To examine the impact of gender and race on standardized patient (SP) assessments of internal medicine residents’ communication skills during postgraduate year (PGY) 1.

Findings

There were significant differences in average numeric communication ratings by gender of resident (with higher scores in female residents), although no significant impact of concordance between SP and resident gender or race.

Limitations

The study was limited by generalizability, as well as limited racial diversity in the included residents.

Bottom Line

Differences in SP scoring by demographics could have an important implication for residents if communication ratings are used for high-stakes assessment purposes, and more work is needed to understand the impact of race on these assessments.

Empathy, and Professionalism. A seventh domain—Likelihood of Referring a Family Member—was removed in the 2018 academic year (so analysis along this domain included only 2012–2017). This scale had previously demonstrated validity evidence in PGY-1 residents at UPenn, using the correlation of SP ratings to faculty ratings. Specifically, we conducted an initial screen of SP ratings through a pilot in 2012. During this pilot, each encounter was scored by SPs and 2 faculty members (for each PGY-1 resident in the program). Based on the high correlation between SP and faculty ratings, future assessments consisted of SPs’ ratings alone. The remainder of the rating scales and the SP clinical encounters remained unchanged over the duration of the study (see online supplementary data). Following the encounter, the complete communication assessments were used to screen individuals in need of additional remediation and communication training.

SPs were recruited from the UPenn SP Program, which was established in 1997 and serves more than 2000 learners annually across multiple professions and at all levels of training. SPs are rigorously trained in a standardized manner to record the events of encounters, assess communication and interpersonal skills, and provide verbal feedback to trainees. For the resident communication assessment, SP training takes 5 hours and involves memorizing case facts and practicing the scenarios. After initial home study of case facts, SPs attend training together to ensure standardized, consistent portrayal and scoring.

The institutional review board at UPenn approved the study as exempt from full review.

Data Analysis

We performed descriptive analyses of numeric ratings by gender, race, and age for both trainees and SPs.

Resident demographic information was based on self-identification through a database derived from the Electronic Residency Application System (ERAS), and SP demographic information was based on self-identification through an SP database (maintained by the UPenn SP program). Residents and/or SPs without complete demographic data (gender, race, and/or ethnicity) were excluded from analysis due to our objective of assessing the gender and racial impact on assessments. We also performed summary statistics across the 7 communication domains (provided as online supplementary data), as well as over a single SP global communication score (which consisted of the mean rating across the 7 domains).

To assess the association between IM resident and SP gender with communication ratings, we performed Mann–Whitney testing (given the leftward skewed communication ratings), using an outcome of SP communication scores.

We used a mixed effects linear regression model to obtain a model associating the final communication scores with individual resident and SP demographic factors. This model was chosen to adjust for the non-independent nature of SP raters throughout the dataset (ie, individual SPs perform as raters for multiple residents across multiple years). We proceeded with a mixed effects linear regression model using the global communication score (as a continuous variable) as the outcome. We hypothesized that resident gender and race would be associated with the outcome of the overall communication score. The independent variables included gender, age, and race for both residents and SPs, year of testing, and case number. The year of testing was intentionally included in the model to account for secular trends leading to alternations in SP individual rater assessments. The model incorporated the interaction between SP gender and resident gender to determine the impact of gender discordance on ultimate ratings. Gender was provided as a binary variable in the dataset (either “male” or “female”).

A secondary analysis was performed to determine the effect of racial concordance or discordance (as an interaction effect) between SP and trainees, and the impact on ultimate communication ratings. A similar mixed effects linear regression model was implemented, as outlined above, using the global communication rating as the outcome. For this dataset, participants self-identified in 1 of 5 choices for a combined race and ethnicity identifier based on a residency program database derived from the ERAS application, and were only able to choose one option: African American, Asian, Caucasian, Hispanic, Indian, or Other. Further classification into additional subgroups was not possible (ie, the ERAS application

does not allow for separate race and ethnicity identification). However, in order to identify any perceived biases based on external appearance and to address potential misclassification bias in the dataset, an additional sensitivity analysis was performed using “Caucasian” versus “non-Caucasian.”

Statistical analysis was completed using STATA version 15.1 (StataCorp, College Station, TX).

Results

The cohort consisted of communication assessment data for 379 unique residents over the 2012 to 2018 academic years, 375 of which had complete demographic data available (TABLE 1). The complete resident cohort consisted of 199 male (53%) and 176 female residents (47%). Of the 1500 assessments on residents with complete demographic data, complete SP demographic data was available on a total of 1425 assessments.

The mean SP ratings across the 7 communication domains is shown in TABLE 2, stratified by resident gender. There were statistically significant differences in average numeric global communication rating (mean 3.40 vs 3.34, $P = .009$) by gender of resident, with higher scores in female trainees. There were significant differences in average numeric communication rating along the domains of Listening, Giving Information, Empathy, and overall Likelihood of Referring a Family Member, all favoring female residents.

There were no significant differences in average numeric communication rating (mean 3.39 vs 3.35, $P = .10$) by SP gender, nor were there differences along the remainder of the 7 communication domains by SP gender.

TABLE 3 presents the multivariable mixed effects linear regression model of communication ratings adjusted for year of assessment, SP gender and race, and resident gender and race. We did not observe significant interactions between SP gender and resident gender across the 7 communication domains or the global SP communication score.

Our secondary analysis analyzed the role of race and ethnicity on ultimate SP assessment, which is outlined in TABLE 4. There were no differences in ratings of residents by race or ethnicity of the SP. There were also no significant interactions between resident ethnicity and race and SP ethnicity and race.

A sensitivity analysis examining Caucasian vs non-Caucasian status of residents revealed no significant difference and no significant interaction effect in the global communication assessment rating by SPs ($\beta - 0.024$, $P = .34$). There was a significant interaction noted in the Eliciting Information domain ($\beta 0.26$, $P = .037$) in race-concordant pairs, but otherwise no

TABLE 1
Demographic Information of Cohort

| Demographic | Residents (n = 379) | Standardized Patients (SPs; n = 58) ^a | | | |
|----------------------------|------------------------|--|-----------------------|-----------------------|-----------------------|
| | | SP Case 1 (n = 16) | SP Case 2 (n = 15) | SP Case 3 (n = 14) | SP Case 4 (n = 13) |
| Age (y) | 27.9 (±2.2) | 33.3 (±4.1) | 33.3 (±4.1) | 44.2 (±7.5) | 61.8 (±5.8) |
| Gender, n (%) | | | | | |
| Male | 199 (53) | 16 (100) | 0 (0) | 14 (100) | 0 (0) |
| Female | 178 (47) | 0 (0) | 15 (100) | 0 (0) | 13 (100) |
| Not Specified ^b | 2 (0.5) | ... | ... | ... | ... |
| Race/Ethnicity, n (%) | | | | | |
| African American | 14 (4) | 2 (13) | 2 (13) | 4 (29) | 3 (23) |
| Asian | 102 (27) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| Caucasian | 223 (59) | 14 (88) | 12 (80) | 10 (71) | 10 (77) |
| Hispanic | 28 (8) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| Indian | 8 (2) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| Other | 1 (0.2) | 0 (0) | 1 (7) | 0 (0) | 0 (0) |
| Not Specified ^b | 3 (0.8) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |

^a Standardized patient (SP) numbers include those with available SP IDs (otherwise the assessments were excluded due to lack of demographic data available).

^b Residents without either gender or race/ethnicity data were excluded from analysis, consisting of 4 learners.

significant interactions were noted throughout the remainder of the domains.

Discussion

Our data demonstrated an association of resident gender on ratings in standardized communication exercises across 7 communication domains. We did not note any communication rating differences of residents based on race and ethnicity concordance with the SPs across the communication domains. This study was the first to evaluate communication ratings along 7 distinct domains (expanding beyond previous studies that focused on empathy ratings) based on demographics of the SPs and residents.

The gender findings in our study are consistent with prior knowledge of gender discrepancies in SP assessment ratings and communication scoring. Previous evidence suggests that communication style and expression of empathy is manifested differently according to gender.^{25–28} It is certainly possible that gender stereotypes promote certain communication styles in one gender, which may align with the way SPs are trained to assess individuals. However, in contrast to prior literature,²¹ we did not find significant interaction between SP gender on these ratings after adjustment for SP-related individual factors. While gender of the resident is associated with differential ratings, this does not appear to be based on SP gender, suggesting there is not an added bias in the ratings based on gender concordance or discordance.

TABLE 2
Mean Standardized Patient Ratings Across 7 Communication Domains and Global Communication Assessment, by Resident Gender

| Communication Domain | Mean Rating (SD) | | P Value |
|--|------------------|--------------|---------|
| | Male PGY-1 | Female PGY-1 | |
| Global Assessment ^a | 3.33 (±0.47) | 3.40 (±0.44) | .009 |
| Eliciting Information | 3.30 (±0.65) | 3.31 (±0.64) | .76 |
| Listening | 3.34 (±0.66) | 3.41 (±0.63) | .019 |
| Giving Information | 3.25 (±0.69) | 3.33 (±0.66) | .031 |
| Respectfulness | 3.47 (±0.61) | 3.53 (±0.58) | .05 |
| Empathy | 3.23 (±0.69) | 3.32 (±0.67) | .015 |
| Professionalism | 3.54 (±0.58) | 3.57 (±0.55) | .34 |
| Likelihood of Referring a Family Member ^b | 3.15 (±0.74) | 3.24 (±0.70) | .042 |

Abbreviation: PGY, postgraduate year.

^a Global Assessment reflects the mean of all 7 standardized patient assessments across the 4 standardized patient cases.

^b Question was dropped from the 2018 assessment (so analysis includes 2012–2017).

TABLE 3

Associations Between Resident Gender, Standardized Patient (SP) Gender, and Gender Interaction With Ultimate Global Communication Assessment Scores^a

| Effects | Mean Global Communication Assessment Score ^b | Standard Error | P Value |
|------------------------------------|---|----------------|---------|
| Main Gender Effects | | | |
| Female Intern | 3.37 | 0.03 | .016 |
| Male Intern | 3.24 | 0.03 | .22 |
| Female SP | 3.37 | 0.10 | .28 |
| Male SP | 3.24 | 0.10 | .48 |
| Interaction Effects | | | |
| Female Intern <i>and</i> Female SP | 3.24 | 0.06 | .21 |
| Female Intern <i>and</i> Male SP | 3.27 | 0.04 | .39 |
| Male Intern <i>and</i> Female SP | 3.24 | 0.04 | .39 |
| Male Intern <i>and</i> Male SP | 3.41 | 0.04 | .39 |

Abbreviation: SP, standardized patient.

^a Global assessment reflects the mean of all 7 standardized patient assessments.

^b Adjusted for trainee gender, race, age, standardized patient gender, race, age, case number (1, 2, 3, or 4), and year.

Importantly, our findings showed a rating difference of 0.06, which may not represent an educationally significant finding. However, if such differences contribute to a resident falling below a designated cutoff, these findings could imply educationally significant differences if communication ratings are used for high-stakes assessment purposes. In contrast to the gender findings, the lack of race or ethnicity impact on communication assessment diverges from prior research regarding SP ratings,²¹ which showed

TABLE 4

Associations Between Resident Race/Ethnicity, Standardized Patient (SP) Race/Ethnicity, and Race/Ethnicity Interaction With Ultimate Global Communication Assessment Scores^a

| Effects | Mean Global Communication Assessment Score ^b | Standard Error | P Value |
|--|---|----------------|---------|
| Intern-level Race/Ethnicity Effects | | | |
| African American | 3.31 | 0.06 | .07 |
| Caucasian | 3.32 | 0.05 | .49 |
| Other ^c | 3.39 | 0.03 | .59 |
| SP-Level Race/Ethnicity Effects | | | |
| African American | 3.30 | 0.09 | .31 |
| Caucasian | 3.35 | 0.09 | .48 |
| Other ^c | 3.59 | 0.20 | .32 |
| Interaction Effects | | | |
| African American Intern <i>and</i> African American SP | 3.70 | 0.19 | .11 |
| African American Intern <i>and</i> Caucasian SP | 3.02 | 0.15 | .08 |
| African American Intern <i>and</i> Other ^c SP | 3.59 | 0.22 | .39 |
| Caucasian Intern <i>and</i> African American SP | 3.40 | 0.06 | .63 |
| Caucasian Intern <i>and</i> Caucasian SP | 3.29 | 0.05 | .88 |
| Caucasian Intern <i>and</i> Other ^c SP | 3.30 | 0.10 | .61 |
| Other ^c Intern <i>and</i> African American SP | 3.36 | 0.06 | .42 |
| Other ^c Intern <i>and</i> Caucasian SP | 3.37 | 0.05 | .41 |
| Other ^c Intern <i>and</i> Other ^c SP | 3.43 | 0.10 | .79 |

^a Global assessment reflects the mean of all 7 SP assessments.

^b Adjusted for trainee gender, race, age, SP gender, race, age, case number (1, 2, 3, or 4), and year.

^c Other includes Asian, Hispanic, Indian, and Other.

a significant interaction effect in the rating of empathy based on race and ethnicity. This also contrasts with expectations from prior research highlighting different patient preferences for provider communication style depending on race and ethnicity.^{8,9,11} Importantly, the limited diversity within our resident cohort may limit the power of our findings. In our secondary analysis, we did note a positive interaction (leading to a more lenient rating) of African American residents by African American SPs within the domain of Eliciting Information. This aligns with a recent study using scripted video vignettes which found that African American patients viewed the doctor more positively with an African American rather than a Caucasian physician, which was mitigated (but not eliminated) with patient-centered communication.²⁹ This aligns with real-life observations where race-ethnicity concordance increased likelihood of seeking preventative care³⁰ and highlights the importance of provider diversity to meet the needs of their populations and optimize effective and patient-centered communication.

In real-world practice, recent evidence suggests an impact of provider demographics, namely gender and race, on patient perception of communication. While prior studies showed a significant difference based on concordance (or discordance) in gender as well as race (between the learner and the SP rater), we found no interaction between SP and resident demographics on ultimate ratings. However, our study analysis specifically adjusted for individual SP variation, which may account for these differences. Regardless, we did note some overarching trends according to resident gender, suggesting demographic factors may influence SP ratings. These factors will need to be considered if they are used for high-stakes assessments (including for standardized examination purposes). Ultimately, potential impacts of implicit bias mitigation strategies for residency educators, SP trainers, or SP medical directors serves as an area of interesting future work.

Our study is limited by generalizability, as this was a single center study in an IM residency program. However, our cohort spanned 8 years and included SPs and residents from diverse backgrounds. Our study is also limited by a small number of African American residents included in the cohort, which limits the effect size observed in our study, and warrants further work in larger studies of assessments in individuals who are underrepresented in medicine. Additionally, while time-based trends in bias awareness are an important consideration, the year of the assessment was included within our model to account for these secular trends. Within our database, there is potential for misclassification of race (as the database combined both race and ethnicity in a single variable). However, a sensitivity analysis did not reveal significant changes to our original findings. Finally, while this study is

specifically evaluating SP ratings, we do not have direct evidence that this represents implicit biases as opposed to different communication skillsets. Furthermore, we do not know how the SP assessments translate to individual patient experience along these communication domains, or the role of the community demographic mix on these findings, but these would be important areas of future research.

Conclusions

Our study revealed demographic differences in SP ratings, mainly associated with gender of residents. There was no interaction between demographic concordance of the residents and SPs on communication assessments.

References

1. Capers Q. How clinicians and educators can mitigate implicit bias in patient care and candidate selection in medical education. *ATS Scholar*. 2020;1(2). doi:10.34197/ats-scholar.2020-0024PS
2. Kirwan Institute for the Study of Race and Ethnicity. Understanding Implicit Bias. <https://kirwaninstitute.osu.edu/implicit-bias-training>. Accessed July 13, 2021.
3. Nosek BA, Smyth FL, Hansen JJ, et al. Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur Rev Soc Psychol*. 2007;18(1):36–88. doi:10.1080/10463280701489053
4. Braun HJ, O'Sullivan PS, Dusch MN, McGrath MH, Ascher NL. The roles of gender and demeanor in perceptions of female surgeons. *Arch Psychol*. 2017;1(2):1–9.
5. Boring A. Gender biases in student evaluations of teaching. *J Public Econ*. 2017;145:27–41. doi:10.1016/j.jpubeco.2016.11.006
6. Heilman ME. Gender stereotypes and workplace bias. *Res Organ Behav*. 2012;32:113–135. doi:10.1016/j.riob.2012.11.003
7. Blanch-Hartigan D, Hall JA, Roter DL, Frankel RM. Gender bias in patients' perceptions of patient-centered behaviors. *Patient Educ Couns*. 2010;80(3):315–320. doi:10.1016/j.pec.2010.06.014
8. LaVeist TA, Nuru-Jeter A. Is doctor-patient race concordance associated with greater satisfaction with care? *J Health Soc Behav*. 2002;43(3):296–306. doi:10.2307/3090205
9. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. *Ann Intern Med*. 2003;139(11):907–915. doi:10.7326/0003-4819-139-11-200312020-00009
10. Aruguete MS, Roberts CA. Participants' ratings of male physicians who vary in race and communication style.

- Psychol Rep.* 2002;91(3):793–806. doi:10.2466/pr0.2002.91.3.793
11. Johnson RL, Roter D, Powe NR, Cooper LA. Patient race/ethnicity and quality of patient–physician communication during medical visits. *Am J Public Health.* 2004;94(12):2084–2090. doi:10.2105/AJPH.94.12.2084
 12. Risdon C, Cook D, Willms D. Gay and lesbian physicians in training: a qualitative study. *CMAJ.* 2000;162(3):331–334.
 13. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med.* 2017;177(5):651–657. doi:10.1001/jamainternmed.2016.9616
 14. Mueller AS, Jenkins TM, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender differences in attending physicians' feedback to residents: a qualitative analysis. *J Grad Med Educ.* 2017;9(5):577–585. doi:10.4300/JGME-D-17-00126.1
 15. Capers Q, Clinchot D, McDougle L, Greenwald AG. Implicit racial bias in medical school admissions. *Acad Med.* 2017;92(3):365–369. doi:10.1097/ACM.0000000000001388
 16. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992–2003. *Med Teach.* 2003;25(3):262–270. doi:10.1080/0142159031000100300
 17. Howley LD. Standardized Patients. In: Levine AI, DeMaria S, Schwartz AD, Sim AJ, eds. *The Comprehensive Textbook of Healthcare Simulation.* Springer; 2013:173–190. doi:10.1007/978-1-4614-5993-4_13
 18. Anderson MB, Stillman PL, Wang Y. Growing use of standardized patients in teaching and evaluation in medical education. *Teach Learn Med.* 1994;6(1):15–22. doi:10.1080/10401339409539637
 19. Cohen DS, Colliver JA, Marcy MS, Fried ED, Swartz MH. Psychometric properties of a standardized-patient checklist and rating-scale form used to assess interpersonal and communication skills. *Acad Med.* 1996;71(suppl 1):87–89. doi:10.1097/00001888-199601000-00052
 20. Stillman PL, Swanson DB. Ensuring the clinical competence of medical school graduates through standardized patients. *Arch Intern Med.* 1987;147(6):1049–1052. doi:10.1001/archinte.1987.00370060045009
 21. Berg K, Blatt B, Lopreiato J, et al. Standardized patient assessment of medical student empathy: ethnicity and gender effects in a multi-institutional study. *Acad Med.* 2015;90(1):105–111. doi:10.1097/ACM.0000000000000529
 22. Berg K, Majdan JF, Berg D, Veloski J, Hojat M. Medical students' self-reported empathy and simulated patients' assessments of student empathy: an analysis by gender and ethnicity. *Acad Med.* 2011;86(8):984–988. doi:10.1097/ACM.0b013e3182224f1f
 23. Siegelman JN, Lall M, Lee L, Moran TP, Wallenstein J, Shah B. Gender bias in simulation-based assessments of emergency medicine residents. *J Grad Med Educ.* 2018;10(4):411–415. doi:10.4300/JGME-D-18-00059.1
 24. Accreditation Council for Graduate Medical Education. Internal Medicine Milestones. <https://www.acgme.org/Portals/0/PDFs/Milestones/InternalMedicineMilestones.pdf>. Accessed July 1, 2020.
 25. Hall JA, Irish JT, Roter DL, Ehrlich CM, Miller LH. Gender in medical encounters: an analysis of physician and patient communication in a primary care setting. *Health Psychol.* 1994;13(5):384–392. doi:10.1037/0278-6133.13.5.384
 26. Roter DL, Hall JA, Aoki Y. Physician gender effects in medical communication: a meta-analytic review. *JAMA.* 2002;288(6):756–764. doi:10.1001/jama.288.6.756
 27. Bertakis KD. The influence of gender on the doctor–patient interaction. *Patient Educ Couns.* 2009;76(3):356–360. doi:10.1016/j.pec.2009.07.022
 28. Franks P, Bertakis KD. Physician gender, patient gender, and primary care. *J Womens Health.* 2003;12(1):73–80. doi:10.1089/154099903321154167
 29. Saha S, Beach MC. Impact of physician race on patient decision-making and ratings of physicians: a randomized experiment using video vignettes. *J Gen Intern Med.* 2020;35(4):1084–1091. doi:10.1007/s11606-020-05646-z
 30. Ma A, Sanchez A, Ma M. The impact of patient-provider race/ethnicity concordance on provider visits: updated evidence from the medical expenditure panel survey. *J Racial Ethn Health Disparities.* 2019;6(5):1011–1020. doi:10.1007/s40615-019-00602-y



All authors are with the Perelman School of Medicine at the University of Pennsylvania. **Janae K. Heath, MD, MS**, is Assistant Professor of Medicine, Department of Medicine; **C. Jessica Dine, MD, MS**, is Associate Professor of Medicine, Department of Medicine; **Denise LaMarra, MS, CHSE**, is Director, Standardized Patient Program; and **Serena Cardillo, MD**, is Professor of Medicine, Department of Medicine.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

This work was presented at the University of Pennsylvania Department of Medicine Medical Education Research Day, Philadelphia, PA, June 11, 2019, and the Alliance for Academic Internal Medicine APDIM Online 2020, October 9, 2020.

Corresponding author: Janae K. Heath, MD, MS, Perelman School of Medicine, janae.heath@penmedicine.upenn.edu, Twitter @JanaeHeath1

Received January 26, 2021; revision received April 12, 2021; accepted June 3, 2021.