Taylor & Francis
Taylor & Francis Group

The effect of demographic characteristics, Country of birth and country of medical training on the peer evaluations of internal medicine resident physicians

G.D. Everett[a], L. Albadin[b] and Y. Du[c]

[a]Residency Program, Florida Hospital Medical Center, Orlando, FL, USA; [b]Department of Physicians Assistant Studies, Adventist University of Health Sciences, Orlando, FL, USA; [c]Statistics, faculty member of Research Analytics and Academics Institute, Graduate Medical Education, Florida Hospital Medical Center, Orlando, FL, USA

ABSTRACT

Background: Peer review by resident physicians, a standard evaluation technique, has rarely been studied for potential biases related to demographic and cultural characteristics of trainees.

Objective: The study sought to determine whether peer evaluations were favorably biased toward trainees of similar background.

Methods: This observational study was conducted in the Internal Medicine residency of a large, metropolitan, community hospital, and included all 91 Internal Medicine residents who had entered the program from 1 July 2009 thru 30 June 2017. Of 3,445 Peer Evaluation Forms (PEF)s offered, 2,922 (84%) were completed and studied. Multivariate statistical analysis was completed. The primary dependent variable was the Peer Evaluation Score (PES). Independent variables included age, gender, race, birth country and country of medical school training. Confounding variables included United States Medical Licensing Examination (USMLE) and In-Training Examination (ITE) scores, and the American Board of Internal Medicine (ABIM) yearly assessment.

Results: Confounding factors accounted for most of the variation. Among the independent variables, only age difference and medical school country were statistically associated with PES. Race and Gender were not significant.

Conclusions: Peer evaluations were not significantly biased by race or gender similarities and only minimally biased by age and medical school country similarities.

1. Introduction

It is now common practice to obtain assessments of resident physicians from multiple sources, including faculty, nurses, ancillary staff, patients, and peers. The rationale for garnering multisource assessments is to improve the breadth and accuracy of subjective elements of evaluation through the use of multiple observers. Peer evaluations, most often studied in medical students, have been shown to have reliability in narrowly defined settings [1–8]. All subjective evaluations, including clinical evaluations of resident physicians, are potentially prone to bias. Gender bias has been documented in American Board of Internal Medicine (ABIM) evaluation of female residents by male attending physicians [9]. Gender bias or disparity also has been described in research study applications in the United States and in the Netherlands [10,11]. Race, ethnicity, and country are interrelated potential biases that appear to influence evaluations of residency applicants and research quality reviews [12–14]. Age also may be a factor in decisions related to resident selection or other healthcare assessments [12,15].

With the increasing diversity of resident physicians, some biases may be decreased through diversity exposure, whereas others may be revealed or even amplified. Internal Medicine residents have become a diverse group of trainees. Nationally, nearly half are female, nearly 60% are not US seniors from Liaison Committee on Medical Education (LCME) accredited schools, and nearly 1/3 were neither born in the US nor attended medical school in the US [16]. This complex environment could affect the way that peers evaluate each other. Therefore, we sought to determine if any of the common demographic factors (race/ethnicity, gender, age, country of birth, and country of medical school) affected peer evaluations. Because peer evaluations should be influenced by the skill and knowledge of the trainee, we controlled for the effects of standard, objective measures of knowledge and clinical skills.

CONTACT G.D. Everett ✉ George.Everett.MD@FLHOSP.org 🖃 Residency Program, Florida Hospital Medical Center, 2501 North Orange Avenue, Suite 235, Orlando, FL 32804; L. Albadin ✉ LillieAlbadin@gmail.com 🖃 Adventist University of Health Sciences; Y. Du ✉ Yuan.Du@FLHOSP.org 🖃 Statistics, faculty member of Research Analytics and Academics Institute, Graduate Medical Education, Florida Hospital Medical Center, 2501 North Orange Ave, Orlando, FL 32804

2. Methods

Hypothesis: We hypothesized that resident physicians give more favorable evaluations to peer resident physicians who have greater demographic similarity to themselves than to other resident physicians, after controlling for objective measures expected to influence evaluation scores.

2.1. Subjects

All 91 Internal Medicine residents, from 1 July 2009 to 30 June 2017, were included in the study. These dates correspond to the initial residency program start-up date (1 July 2009) to the end of the most recent academic year of this single institutional study. Data were taken from the standard Peer Evaluation Form (PEF) used by the residents for peer evaluation (see appendix Table A1) and from portfolio information on all residents. The PEF was the same one used throughout the study period. The PEF was given to every resident working on a patient care team. Basic instructions on the use and purpose of the form were provided. Team assignments arose 8–9 months in the PGY-1 year, 5–6 months in the PGY-2 year, and 4–5 months in the PGY-3 year. Resident teams mostly consisted of one upper-level resident (PGY-2 or 3) and two PGY-1 residents. Every team member was evaluated by all other team members. Residents were assigned to teams randomly and were not permitted to select a team or team member they preferred. The PEFs were confidential. The evaluated resident could not ascertain the evaluating person's identity. The Florida Hospital Institutional Review Board approved the study.

2.2. Design: the study design was an observational study

2.2.1. Setting

The study took place within the Internal Medicine residency program at a 1400 bed quaternary, community hospital in Florida. The ambulatory and administrative facilities for the residency program were directly attached to the hospital inpatient facilities.

2.2.2. Demographic variables

The classification of the birth and medical school demography (country, sub-region, region) of each resident was taken from the Statistics Division of the United Nations [17]. The race and ethnicity characterization was obtained from the United States Census Bureau, Office of Management and Budget standards [18].

2.2.3. Statistical analysis

Multivariate analysis was completed. The primary dependent variable was the Peer Evaluation Score (PES). The PES on each peer evaluation was the total score of all 27 items on the PEF divided by the number of items answered by the evaluator (not all items were answered on every evaluation form). Demographic Independent variables of interest were age, gender, race, country of birth, and country of medical school training. Age was compared as within/outside 5 years difference and by age difference from 26 at PGY-1 (age 26 is the approximate expected age of a PGY-1 resident entering residency from an LCME-accredited medical school). The Gender and Race/Ethnicity were analyzed as binary variables (same or different). Country of birth and of medical school were classified as the same country, sub-region, or region of the world. Confounding variables were United States Medical License Examination (USMLE) Step 1 and 2 scores, In-Training Examination (ITE) percentile scores, and Program Director's yearly American Board of Internal Medicine (ABIM) assessment, which was classified as Unsatisfactory, Marginal, Satisfactory, or Superior. For use in the multivariate analysis (General Linear Modeling of Statistical Analysis System), the USMLE score was the average of Step 1 and 2, the ITE score was the average percentile of each year completed, and the ABIM was the ratio of evaluations received divided by maximum possible evaluation (unsatisfactory = 0, marginal = 1, satisfactory = 2, superior = 3). Confounding variables were expected to affect the PES score. The confounding variables (USMLE scores, ITE scores, and ABIM assessment) associated with the person being evaluated were not known by the evaluating resident. Thus, these three confounding variables were surrogate quality markers of the evaluated resident. Demographic variables, by contrast, would not be expected to affect the PES score unless a bias was present.

3. Results

The basic demographic and academic characteristics of the 91 residents in the study is presented in Table 1. The average age of the residents was 28 years, and a majority were male (55.0%) and of Asian ethnicity (69.3%). The average USMLE and ITE scores were significantly higher than the mean of all persons taking the examination across the US.

Table 2 reveals the details of the birth and medical school countries of the resident physicians by Region, Sub-region and Country. The birth and medical school country were usually, but not always, the same. Pakistan, China, United States, and India were the most common countries represented.

Table 1. Demographic and Academic Characteristics of Resident Physicians.

Characteristics	Result
Age (years), mean ± SD	28 ± 5
Gender, No. (% male)	91 (55.0)
Race/Ethnicity, No. (%)	
Asian – India	34 (37.4)
Asian – Other	29 (31.9)
Caucasian	14 (15.4)
Hispanic or Latin	11 (12.1)
Middle East	3 (3.3)
USMLE Step 1 score, mean ± SD	237.4 ± 22.6
USMLE Step 2 score, mean ± SD	243.1 ± 21.3
PGY-1 ITE percentile score, mean ± SD	70.5 ± 28.0
PGY-2 ITE percentile score, mean ± SD	69.0 ± 27.6
PGY-3 ITE percentile score, mean ± SD	71.2 ± 26.8
ABIM Evaluation by Program Director – Year 1, No. (%)	
Superior	48 (56.5)
Satisfactory	34 (40.0)
Marginal	1 (1.2)
Unsatisfactory	2 (2.4)
ABIM Evaluation by Program Director – Year 2, No. (%)	
Superior	41 (56.2)
Satisfactory	32 (43.8)
ABIM Evaluation by Program Director – Year 3, No. (%)	
Superior	41 (66.1)
Satisfactory	21 (33.9)

Abbreviation: SD, standard deviation; USMLE, United States Medical Licensing Examination; PGY, postgraduate year; ITE, In-training Examination; ABIM, American Board of Internal Medicine

Table 3 displays the univariate analysis table of demographic characteristics of residents in relation to the PES. The numbers in the Frequency column are the number of PEFs analyzed. The total number of expected PEFs would have been 3,445 if all forms had been completed. A total of 2,922 evaluations were completed for a completion rate of 84.8%. Because of skewness, the Kruskal Wallis test was used to assess the significance of the PES differences for the individual variables. Ethnicity/race, birth country, and medical school country were statistically significant at traditional p-values.

Table 4 depicts the multivariate analysis. USMLE scores strongly correlated with ITE scores, and thus ITE score was dropped from the multivariate analysis. Likewise, birth country and medical school country were highly correlated; thus, birth country was dropped from the analysis. The ABIM Evaluation by Program Director was the most predictive factor followed by medical school country, age difference, and age at PGY-1 above 26. After accounting for ABIM evaluation, USMLE score was not significant. Gender was not statistically significant. The final multivariate analysis had an R-square of 6% indicating that the components of the analysis accounted for only a small proportion of all variance.

4. Discussion

Our study represents one of only a very few studies evaluating peer assessments among resident physicians and the first to assess multiple factors that pertain to peer assessments. We have found that the most important explanatory factors in peer assessments are factors that would be expected to relate to the quality of resident capabilities: ABIM Evaluation by Program Director, which strongly correlates to ITE and USMLE scores. These objective assessments were not available to the peer residents doing the peer evaluations and thus did not influence the residents' assessments of each other. We found evidence of a very low level of bias favoring residents with similar background and demography. The quantitation of this bias was less than 0.1 on a scale of 1–3. We also found that age variation bias, while statistically significant, was quantitatively trivial. We did not find gender bias.

The study has several strengths in that our residents come from many areas of the world, thus allowing for assessment of a wide range of birthplaces and training. There was also a nearly equal distribution of gender and enough age variation to allow for a robust study of these factors as well. The size and length of the study permitted the evaluation of even very small levels of bias. The high proportion of completed evaluations and the use of the same instrument over the entire study helped to assure that the study was comprehensive and comparable over time.

The study also has several notable limitations. It was a single institution study; therefore, generalization must be applied cautiously. The demographic characteristics of the internal medicine residents in our program are representative of community hospital programs but not all programs. The PES had only a limited range of 1–3, and there was significant skewing of scores toward the higher range, thus limiting the range of statistical analysis. It is possible that a wider range of PES scores would have permitted better discrimination of small differences in scores. Finally, the confounding factors used in the multivariable equation as surrogates of resident quality are limited to measurements of knowledge (USMLE examinations and ITE examinations) and of global clinical assessment (Program Director's yearly American Board of Internal Medicine (ABIM) assessment). This combination of assessments may not optimally gauge the effectiveness of residents in the clinical setting.

5. Conclusions

Peer evaluations by Internal Medicine resident physicians revealed statistically significant, but very modest evidence of bias favoring similar country of origin and training, ethnicity, and age. There was no evidence of gender bias. Objective measures of resident quality strongly predicted peer evaluations, as expected.

Table 2. Resident physician birth and medical school location by region, Sub-region, and Country.

Location	Region		Sub-region		Country	
	Birth	Med [a]	Birth	Med [a]	Birth	Med [a]
Americas	23(25.3) [b]	26(28.6)				
North America			16(17.6)	17(18.7)		
USA					16(17.6)	17(18.7)
Central America			1(1.1)	1(1.1)		
Mexico					1(1.1)	1(1.1)
South America			2(2.2)	2(2.2)		
Colombia					1(1.1)	2(2.2)
Venezuela					1(1.1)	
Latin America/Caribbean			4(4.4)	6(6.6)		
Puerto Rico					4(4.4)	4(4.4)
Dominica						1(1.1)
Grenada						1(1.1)
Asia	59(64.8)	57(62.64)				
Eastern Asia			17(18.7)	16(17.6)		
China					17(18.7)	16(17.6)
Southeast Asia			8(8.8)	6(6.6)		
Myanmar					5(5.5)	5(5.5)
Thailand					1(1.1)	
Indonesia					1(1.1)	
Philippines					1(1.1)	1(1.1)
Southern Asia			33(36.3)	33(36.3)		
India					13(14.3)	12(13.2)
Pakistan					19(20.9)	20(22.0)
Nepal					1(1.1)	1(1.1)
Western Asia			1(1.1)	2(2.2)		
Iraq					1(1.1)	
Syria						1(1.1)
Jordan						1(1.1)
Europe	8(8.8)	6(6.6)				
Eastern Europe			5(5.5)	5(5.5)		
Russia					3(3.3)	2(2.2)
Slovakia					1(1.1)	1(1.1)
Czech Republic					1(1.1)	1(1.1)
Hungary						1(1.1)
Northern Europe			1(1.1)	1(1.1)		
UK					1(1.1)	1(1.1)
Southern Europe			1(1.1)			
Albania					1(1.1)	
Western Europe			1(1.1)			
France					1(1.1)	
Africa	1(1.1)	2(2.2)				
Northern Africa			1(1.1)	2(2.2)		
Egypt					1(1.1)	2(2.2)

a denotes Medical School
b denotes N (%). Each column totals 91 and 100%

Table 3. Univariate analysis of demographic characteristics on PES[a].

Characteristic	Frequency N (%)	PES mean (± SD)	PES Median	P-Value [b]
Age Difference				0.902
≤5 years	1,785 (61.09)	2.79 (0.34)	3.00	
≥5 years	1,137 (38.91)	2.79 (0.33)	3.00	
Ethnicity/Race				0.039
Same	1,549 (53.19)	2.83 (3.30)	3.00	
Different	1,363 (46.81)	2.76 (0.37)	3.00	
Gender				0.487
Same	1,506 (51.54)	2.79 (0.34)	3.00	
Different	1,416 (48.46)	2.79 (0.34)	3.00	
Birth Country				0.001
Same Country	43.5 (14.94)	2.80 (0.34)	3.00	
Same Sub-Region Only	192 (6.59)	2.88 (0.24)	2.96	
Same Region Only	843 (28.95)	2.79 (0.32)	3.00	
All Different	1,442 (49.52)	2.78 (0.36)	3.00	
Medical School Country				<0.001
Same Country	439 (15.08)	2.83 (0.31)	3.00	
Same Sub-Region Only	171 (5.87)	2.88 (0.24)	3.00	
Same Region Only	787 (27.03)	2.80 (0.32)	3.00	
All Different	1,515 (52.03)	2.77 (0.36)	3.00	

[a] = Peer Evaluation Score
[b] = Kruskal Wallis Test

Table 4. Multivariable analysis of demographic and academic characteristics of residents on Peer Evaluation Score (PES).

Characteristic	Estimate	Standard Error	T Value	P Value
ABIM Evaluation by Program Director	0.400	0.0366	10.92	<0.0001
USMLE Score	−0.0001	0.0003	−0.38	0.7019
PGY 1 Age Above 26	−0.0046	0.0014	−3.30	0.0010
Age Difference > 5	0.0315	0.0142	2.22	0.0265
Gender Difference	0.0004	0.0122	0.03	0.9738
Medical School Country				
Same Country vs. Diff Region	0.0718	0.0180	3.98	<0.001
Same Sub-region vs. Diff Region	0.0767	0.0270	2.84	0.0045
Same Region vs. Diff Region	0.0283	0.0148	1.91	0.0563

Acknowledgments

Authors Contributions

a. Author who designed study, assisted in data retrieval, assisted in analysis and prepared the manuscript. Corresponding author.
b. Author who assisted in data retrieval, assisted in analysis and assisted in manuscript preparation
c. Author who assisted in data retrieval, performed statistical analysis and reviewed the manuscript.

Disclosure statement

Funding

Prior Publications/Abstracts/Presentations

References

[1] Basehore PM, Pomerantz SC, Gentile M. Reliability and benefits of medical student peers in rating complex clinical skills. Med Teach. 2014 May;36 (5):409–414. PubMed PMID: 24597711

[2] Beckman TJ, Lee MC, Mandrekar JN. A comparison of clinical teaching evaluations by resident and peer physicians. Med Teach. 2004 Jun;26(4):321–325. PubMed PMID: 15203844

[3] Bentley BS, Hill RV. Objective and subjective assessment of reciprocal peer teaching in medical gross anatomy laboratory. Anat Sci Educ. 2009 Jul-Aug;2 (4):143–149. PubMed PMID: 19637291

[4] Burgess A, Clark T, Chapman R, et al. Senior medical students as peer examiners in an OSCE. Med Teach. 2013;35(1):58–62. PubMed PMID: 23102164

[5] Kovach RA, Resch DS, Verhulst SJ. Peer assessment of professionalism: a five-year experience in medical clerkship. J Gen Intern Med. 2009 Jun;24(6):742–746. PubMed PMID: 19390903; PubMed Central PMCID: PMCPMC2686767

[6] Levine RE, Kelly PA, Karakoc T, et al. Peer evaluation in a clinical clerkship: students' attitudes, experiences, and correlations with traditional assessments. Acad Psychiatry. 2007 Jan–Feb;31(1):19–24. PubMed PMID: 17242048.

[7] Spandorfer J, Puklus T, Rose V, et al. Peer assessment among first-year medical students in anatomy. Anat Sci Educ. 2014 Mar–Apr;7(2):144–152. PubMed PMID: 23959790.

[8] Speyer R, Pilz W, Van Der Kruis J, et al. Reliability and validity of student peer assessment in medical education: a systematic review. Med Teach. 2011;33 (11):e572–e585. PubMed PMID: 22022910

[9] Rand VE, Hudes ES, Browner WS, et al. Effect of evaluator and resident gender on the American board of internal medicine evaluation scores. J Gen Intern Med. 1998 Oct;13(10):670–674. PubMed PMID: 9798813; PubMed Central PMCID: PMC1500895

[10] Kaatz A, Lee YG, Potvien A, et al. Analysis of National Institutes of health R01 application critiques, impact, and criteria scores: does the sex of the principal investigator make a difference? Acad Med. 2016 Aug;91 (8):1080–1088. PubMed PMID: 27276003; PubMed Central PMCID: PMCPMC4965296.

[11] van der Lee R, Ellemers N. Gender contributes to personal research funding success in The Netherlands. Proc Natl Acad Sci U S A. 2015 Oct 6;112(40):12349–12353. PubMed PMID: 26392544; PubMed Central PMCID: PMCPMC4603485.

[12] de Oliveira GS Jr., Akikwala T, Kendall MC, et al. Factors affecting admission to anesthesiology residency in the United States: choosing the future of our specialty. Anesthesiology. 2012 Aug;117 (2):243–251. PubMed PMID: 22739761.

[13] Harris M, Macinko J, Jimenez G, et al. Does a research article's country of origin affect perception of its quality and relevance? A national trial of US public health researchers. BMJ Open. 2015 Dec 30;5(12):e008993. PubMed PMID: 26719313; PubMed Central PMCID: PMCPMC4710821.

[14] Harris M, Macinko J, Jimenez G, et al. Measuring the bias against low-income country research: an implicit association test. Global Health. 2017 Nov 6;13(1):80. PubMed PMID: 29110668; PubMed Central PMCID: PMCPMC5674740

[15] FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. BMC Med Ethics. 2017 Mar 1;18(1):19. PubMed PMID: 28249596; PubMed Central PMCID: PMCPMC5333436.

[16] Annual Report on Residents: American Association of Medical Colleges. 2017. [cited 2018 Apr 26]. Available from: https://www.aamc.org/data/448474/residentsreport.html

[17] United Nations Statistical Division: Standard Country or Area Codes for Statistical Use. 2018. [cited 2018 Apr 26]. Available from: https://unstats.un.org/unsd/methodology/m49

[18] United States Census Bureau 1997 Office of Management and Budget Standards 1997. Available from: https://www.census.gov/topics/population/race/about.html

Table A1. Resident Peer Evaluation.

Resident Photo	Subject Name	Evaluator
	Subject Employer	
	Evaluation Dates	Evaluator Name
	Subject Rotation	Evaluator Employer

- *Please be as complete and honest as possible in assigning values to each component*
- *Try to identify areas that should be improved and give suggestions of how improvement could be achieve*
- *Comments expected regarding any 'Improvement Needed' areas and encouraged for 'Exceptional Performance' areas*

	Room for Improvement 1	Meets Standards 2	Exceptional Performance 3	Not Applicable
Patient Care & Professionalism				
(1) Shows compassion in patient care				
(2) Communicates effectively with patients' family				
(3) Culturally sensitive when delivering care				
(4) Maintains patient confidentiality				
(5) He/she arrives to work on time and is well prepared for daily rounds				
(6) Follows up on patient issues after rounds on tests/serial exams/consultants' input				
Interpersonal and Communication Skills				
(7) Communicates effectively with the rounding team and is a team player				
(8) Coordinates care effectively for patients with other health professionals, physicians and health care resource centers				
Medical Knowledge				
(9) Case presentations are appropriate, clear and concise				
(10) Considers appropriate DDX				
(11) Selects and assesses diagnostic tests appropriately				
(12) Manages patients with complex medical problems well				
(13) Consults in an appropriate manner				
Practice-Based Learning				
(14) Use of Internet as source of information				
(15) Applies pertinent evidence-based medicine to clinical care				
Systems-Based Practice				
(16) Maintains quality written progress notes				
(17) Accepts constructive criticism				
(18) Understands cost effective care and costs to patients				
Senior Resident as Teacher (Use this section only if peer is supervising)				
(19) Was a good role model of a caring doctor				
(20) Encouraged teamwork				
(21) Demonstrated a broad working fund of knowledge				
(22) Modeled self-directed learning				
(23) Raised critical teaching points on rounds				
(24) Encouraged team members to pursue their own learning issues				
(25) Encouraged critical appraisal of the literature/evidenced based learning				
(26) Participated actively in teaching/review of current literature				
(27) Demonstrated good judgement in patient management				
Overall Comments:				