Predicting In-State Workforce Retention After Graduate Medical Education Training

Tracy J. Koehler, PhD
Jaclyn Goodfellow, MM
Alan T. Davis, PhD
Jessaca Spybrook, PhD
John E. vanSchagen, MD, FAAFP
Lori Schuh, MD, FAAN

ABSTRACT

Background There is a paucity of literature when it comes to identifying predictors of in-state retention of graduate medical education (GME) graduates, such as the demographic and educational characteristics of these physicians.

Objective The purpose was to use demographic and educational predictors to identify graduates from a single Michigan GME sponsoring institution, who are also likely to practice medicine in Michigan post-GME training.

Methods We included all residents and fellows who graduated between 2000 and 2014 from 1 of 18 GME programs at a Michigan-based sponsoring institution. Predictor variables identified by logistic regression with cross-validation were used to create a scoring tool to determine the likelihood of a GME graduate to practice medicine in the same state post-GME training.

Results A 6-variable model, which included 714 observations, was identified. The predictor variables were birth state, program type (primary care versus non–primary care), undergraduate degree location, medical school location, state in which GME training was completed, and marital status. The positive likelihood ratio (+LR) for the scoring tool was 5.31, while the negative likelihood ratio (–LR) was 0.46, with an accuracy of 74%.

Conclusions The +LR indicates that the scoring tool was useful in predicting whether graduates who trained in a Michigan-based GME sponsoring institution were likely to practice medicine in Michigan following training. Other institutions could use these techniques to identify key information that could help pinpoint matriculating residents/fellows likely to practice medicine within the state in which they completed their training.

Introduction

One of the major decisions in physicians' lives as they approach the end of their graduate training is where to practice medicine. Federal and state governments, along with other sources, devote substantial financial resources to the training and development of physicians who may leave the state in which they trained to practice elsewhere. In this time of concern over an impending shortage of physicians and the cost of graduate medical education (GME), there is a benefit to identifying GME graduates who are most likely to practice medicine in the state in which they trained. Therefore, it is important to learn more about the factors that influence in-state practice location decisions.

Over the years, in-state retention has been examined in different ways, including the use of logistic regression to examine variables related to in-state retention, and studies that report summary data related to percentages of graduates who practice in the state in which they completed their GME training.[1–12] While published literature provides useful information on retention, the majority of these studies are based on data from the same historical source, the American Medical Association Physician Masterfile.[4–12] What seems to be lacking is a current look at predictors of in-state retention, specifically demographic and educational characteristics of physicians. The purpose of this study was to use demographic and educational predictors to identify graduates from a single Michigan-based GME sponsoring institution who are likely to practice medicine in Michigan after completing training.

DOI: http://dx.doi.org/10.4300/JGME-D-16-00278.1

Editor's Note: The online version of this article contains a data sources table, best subsets model selection description, and the scoring tool.

Methods

Study Sample

All individuals who graduated from the 18 GME programs offered by Grand Rapids Medical Education Partners (GRMEP) in Michigan between 2000 and 2014 were included in the initial review. Residents and fellows who graduated from a GRMEP training program and were still in training (such as fellowship) at the time of data collection were excluded from the review, as were transitional year and preliminary residents who left GRMEP after 1 year of GME to enroll in another residency program. Data collected included birth state, undergraduate degree location, medical school location, time in program, state GME training completed, marital

status, sex, visa status, type of program, and whether the graduate ever practiced in Michigan posttraining. Data sources included the New Innovations database (New Innovations, Uniontown, OH), GRMEP GME records, Google, the Michigan Department of Licensing and Regulatory Affairs website, and GRMEP program directors and coordinators (provided as online supplemental material).

This study was approved by the Spectrum Health Institutional Review Board prior to data collection.

Data Analysis

Data were analyzed using Stata version 13.0 (StataCorp, College Station, TX). The approach to building the model included cross-validation, which randomly splits a data set into 2 portions, a training sample that includes 80% of the original data set, and a validation sample that is made up of the other 20%.[13] A best subsets logistic regression approach was used on the training sample, which included, as the outcome variable, having practiced in Michigan posttraining and, as predictor variables, birth state (Michigan versus not Michigan), undergraduate degree location (Michigan versus not Michigan), medical school location (Michigan versus not Michigan), time in program (years), state where GME training was completed (Michigan versus not Michigan), marital status (ever married versus never married), sex (male versus female), visa status (visa versus no visa), and type of program (primary care versus non–primary care). Primary care was defined as residency in family medicine, internal medicine, internal medicine–pediatrics, and pediatrics. Criteria used to assess the best subsets analysis included Mallow's Cp, adjusted R^2, Akaike information criterion, and Bayesian information criterion. Significance was assessed at $P < .05$. Logistic regression assumptions were checked and met prior to running the analyses. Also, missing data were assessed and considered to be missing at random. The justification for this decision is described by Frieswyk.[14] The model was cross-validated and checked for overfitting with the bootstrap procedure, using the area under the curve (AUC) as the criterion.[15]

Scores for the variables included in the model were then derived using a method by Sullivan et al.[16] The β coefficients obtained from the independent variables were used to create the scores. The β coefficients were compared, with the lowest value representing the referent value, from which the remaining scores were determined. Each β coefficient was divided by the referent value to determine the score for that variable. Products from this calculation were rounded to the nearest 0.5. A receiver operating characteristic (ROC) curve analysis using the scores derived from the model was performed. Youden's J was used to determine the optimal cut point for the scoring tool.[17]

What was known and gap
Teaching institutions and state governments are interested in the in-state retention of graduates of their physician training programs.

What is new
A 6-variable model using birth state, program type, undergraduate degree location, medical school location, state of graduate medical education (GME) training, and marital status produced a scoring tool with acceptable predictive accuracy.

Limitations
Single state, single institution study may reduce generalizability.

Bottom line
The scoring tool is useful in predicting which graduates are likely to practice medicine within the state of their GME training.

The utility of the scoring tool was evaluated by comparing the accuracy of predicting the outcome variable (having practiced in Michigan) between the training data set and the validation data set. Further subanalyses were performed to assess whether the results held true for primary care and non–primary care programs.

Results

The entire data set included 988 graduates. Summary data for the sample are shown in TABLE 1. Just over half of graduates practiced in Michigan at some point after completing GME. The sample consisted of 58% men, and nearly one-third had attended medical school in Michigan, attended an undergraduate institution in Michigan, or were born in Michigan. Just under half of the graduates were from a primary

TABLE 1
Summary Data for the Sample

Variable	n/N (%)
Ever practiced medicine in Michigan	504/988 (51.0)
Primary care program graduate	475/988 (48.1)
Completed medical school in Michigan	302/988 (30.6)
Male	573/988 (58.0)
Completed GME in Michigan	793/988 (80.3)
Born in Michigan	266/945 (28.1)
Ever married	661/925 (71.5)
Visa status	126/970 (13.0)
Completed undergraduate education in Michigan	336/988 (34.0)
Time in program, y (mean ± SD)	3.5 ± 1.0

Abbreviation: GME, graduate medical education.

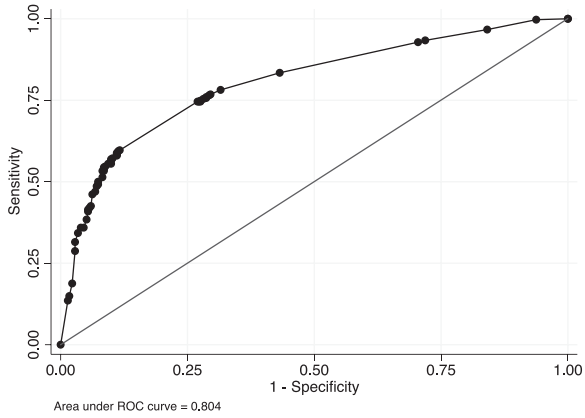

FIGURE
Receiver Operating Characteristic (ROC) Curve for Final Model

care program. Approximately 80% completed all of their GME training in Michigan.

The results of the best subsets regression analysis produced the 9 best variable combinations for the model. The process used to select the final model is described in the online supplemental material. Variables in the model for the training sample and validation sample are described in TABLE 2. A ROC analysis found the AUC to be 0.804 (FIGURE). This final model met the criteria established by Hosmer et al,[18] deeming it an excellent model for discriminating between graduates who have practiced in Michigan and those who have not.

The value for the AUC was compared with the AUC from the bootstrap cross-validation procedure for consistency. The original AUC and corrected AUC were 0.804 and 0.800, respectively. The results suggest that overfitting was not a concern.

TABLE 3 shows how scores were calculated for each variable included in the predictor model. The scoring system ranged from 0 to 8.5 points. The next step included assigning scores associated with whether or not the graduate ever practiced medicine in the state of Michigan to all graduates in the training and validation data sets.

The results of the ROC analysis produced the optimal cut point for the scoring tool (provided as online supplemental material). Youden's J was determined to be 0.478, which is associated with a cut point of 4. This cut point has a sensitivity of 58.8%, a specificity of 88.9%, and a correct classification rate of 73.7%. The positive likelihood ratio (+LR) is 5.31, while the negative likelihood ratio (–LR) is 0.46. The interpretation for this cut point means that individuals with a score ≥ 4 were more likely to practice in Michigan than those who receive a score < 4.

When using the validation data ($n = 175$), the accuracy of predicting whether or not the individual would ever practice in Michigan at some point after graduation was 72.0%, which was not significantly different from the accuracy of the training data set ($P = 0.65$). Using the validation data set, subanalyses for primary care ($n = 85$) and non–primary care ($n = 90$) programs also showed similar values for accuracy, compared to the training data set (68.2%, $P = 0.29$ and 75.6%, $P = 0.70$, respectively).

Discussion

This study examined the likelihood that graduates from a Michigan GME training institution would practice in Michigan after graduation, predictive variables included being born in Michigan, attending medical school in Michigan, obtaining an undergraduate degree in Michigan, graduating from a primary care residency, completing GME training in Michigan, and having been married. The resulting scoring tool had similar predictive ability for the derivation and validation cohorts, and supports the theory that GME graduates with some tie to Michigan may be more likely to practice in the state.[1,3–9,19]

Completion of GME in Michigan was predictive of practicing in Michigan. This is consistent with much of the literature, suggesting that location at the

TABLE 2
Training Sample and Validation Sample Summary Data

Variable	Training Sample, n (%); (N = 714)	Validation Sample, n (%); (N = 175)
Ever practiced medicine in Michigan	362 (50.7)	96 (54.9)
Primary care program graduate	340 (47.6)	85 (48.6)
Completed medical school in Michigan	206 (28.9)	62 (35.4)
Born in Michigan	197 (27.6)	56 (32.0)
Male	412 (57.7)	95 (54.3)
Ever married	516 (72.3)	115 (65.7)
Completed undergraduate education in Michigan	227 (31.8)	66 (37.7)
Completed GME training in Michigan	574 (80.4)	141 (80.6)

Abbreviation: GME, graduate medical education.

TABLE 3
Scoring Tool Calculations

Variable	β Coefficient	Score Computation	Rounded Score
Born in Michigan	0.889		
Yes = 1	1 * 0.889 = 0.889	0.889/0.593 = 1.499	1.5
No = 0[a]	0 * 0.889 = 0	0/0.593 = 0	0
Primary care program graduate	0.768		
Yes = 1	1 * 0.768 = 0.768	0.768/0.593 = 1.295	1.5
No = 0[a]	0 * 0.768 = 0	0/0.593 = 0	0
Completed undergraduate education in Michigan	1.135		
Yes = 1	1 * 1.135 = 1.135	1.135/0.593 = 1.914	2
No = 0[a]	0 * 1.135 = 0	0/0.593 = 0	0
Ever married	0.698		
Yes = 1	1 * 0.698 = 0.698	0.698/0.593 = 1.177	1
No = 0[a]	0 * 0.698 = 0	0/0.593 = 0	0
Completed medical school in Michigan	0.901		
Yes = 1	1 * 0.901 = 0.901	0.901/0.593 = 1.519	1.5
No = 0[a]	0 * 0.901 = 0	0/0.593 = 0	0
Completed GME training in Michigan	0.593		
Yes = 1	1 * 0.593 = 0.593	0.593/0.593 = 1.000	1
No = 0[a]	0 * 0.593 = 0	0/0.593 = 0	0

Abbreviation: GME, graduate medical education.
[a] Denotes reference category.

end of GME is associated with practice within the state.[1,6–8,20]

Training in a primary care specialty also was included in the final model, with individuals who graduated from a primary care program more likely to practice in Michigan than those who did not. This result is consistent with the study by Seifer et al,[5] who reported that general practitioners (defined as family medicine, internal medicine, and pediatrics physicians) were 1.4 times more likely to practice medicine within the state of GME than non–primary care physicians, which was a significant predictor in their regression model.

The decision to practice in Michigan was not influenced by sex. This is in contrast to the studies of Burfield et al,[4] which showed that 60% of female graduates practice in the state in which they trained compared to 50% of male graduates, and Seifer et al[5] that demonstrated that female GME graduates were significantly more likely to practice medicine in the state in which they completed GME.

The ROC analysis showed that the model had excellent discrimination between those who practiced in Michigan and those who did not, based on the criteria established by Hosmer et al.[18] Overfitting the model to the data is often a challenge in regression analyses; however, the results of the bootstrap procedure showed that overfitting was not an issue. The accuracy of predicting the outcome variable was similar between the training sample and the validation sample. This also extended to the subanalyses of the validation sample, indicating that the scoring tool should prove equally useful for primary and non–primary care program graduates.

Limitations of the study include missing data. However, the variables in the data set were at least 90% complete. Another limitation is the use of data from a single institution in Michigan, which limits the generalizability of the scoring tool to other GME institutions or states. The extended time frame of the study (15 years) introduces the potential for historical factors that could influence practice location decisions. For example, Michigan went through a recession, which could have negatively influenced decisions to practice in the state, and physician employment opportunities are unlikely to be the same from 1 year to the next.

This tool could be used to assess residents and fellows at any time during their training as identification for potential recruitment by hospitals, local offices/clinics, and/or state-based physician recruiters. An instrument like this could potentially aid in increasing in-state retention rates, and could be used as a performance measure for local and state funding sources. However, this application will require future research into the attributes of the tool for this purpose. While some elements of the tool (eg, medical school location, birth state) may be useful consider-

ations when reviewing candidates for residency, the use of the tool in this context is not recommended due to ethical concerns regarding the rationale for ranking candidates.

Conclusion

We produced a scoring tool that is useful in predicting whether graduates who trained in a Michigan-based GME sponsoring institution would practice medicine in Michigan after completing training. Other institutions could use the analytical techniques we describe to identify key data for determining graduates who are likely to practice medicine within the state of their GME training.

References

1. Koehler TJ, Goodfellow J, Davis AT, et al. Physician retention in the same state as residency training: data from 1 Michigan GME institution. *J Grad Med Educ*. 2016;8(4):518–522.
2. Veldhouse J, Switzer R, Coe C, et al. GRMEP residents after graduation: who is staying to practice primary care in West Michigan? Poster presented at: GRMEP Annual Research Day, April 25, 2013; Grand Rapids, MI.
3. Georgia Statewide Area Health Care Network. 2013 Primary Care Summit summary. http://www.gru.edu/ahec/documents/fy15gme_request.pdf. Accessed November 7, 2016.
4. Burfield WB, Hough DE, Marder WD. Location of medical education and choice of location of practice. *J Med Educ*. 1986;61(7):545–559.
5. Seifer SD, Vranizan K, Grumbach K. Graduate medical education and physician practice location: implications for physician workforce policy. *JAMA*. 1995;274(9):685–691.
6. Owen JA, Hayden GF, Bowman RC. Influence of places of birth, medical education, and residency training on the eventual practice locations of family physicians: recent experience in Virginia. *South Med J*. 2005;98(6):674–675.
7. Fagan EB, Gibbons C, Finnegan SC, et al. Family medicine graduate proximity to their site of training: policy options for improving the distribution of primary care. *Access Fam Med*. 2015;47(2):124–130.
8. Office for Healthcare Workforce Analysis and Planning. Data brief, March 2011: retaining physicians educated in South Carolina. http://www.officeforhealthcareworkforce.org/docs/Retaining%20Physicians%20Educated%20in%20South%20Carolina.pdf. Accessed November 7, 2016.
9. Armstrong D, Chung R, Forte GJ. 2014 New York residency training outcomes: a summary of responses to the 2014 New York resident exit survey. Rensselaer, NY: Center for Health Workforce Studies, School of Public Health, SUNY Albany; 2015. http://chws.albany.edu/archive/uploads/2015/04/2014_NY_Residency_Training_Outcomes.pdf. Accessed November 7, 2016.
10. Rabinowitz HK, Diamond JJ, Markham FW, et al. Critical factors for designing programs to increase the supply and retention of rural primary care physicians. *JAMA*. 2001;286(9):1041–1048.
11. Robert Graham Center. Specialty and geographic distribution of the physician workforce: what influences medical student and resident choices? http://www.graham-center.org/dam/rgc/documents/publications-reports/monographs-books/Specialty-geography-compressed.pdf. Accessed November 7, 2016.
12. Skillman SM, Stover B; Center for Health Workforce Studies. Washington State's physician workforce in 2014. http://depts.washington.edu/uwrhrc/uploads/CHWS_WA_Phys_Workforce_2014.pdf. Accessed November 7, 2016.
13. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 6th ed. Upper Saddle River, NJ: Pearson Education Inc; 2013.
14. Frieswyk TJ. GME graduate retention rates: a single institution study [dissertation]. Kalamazoo, MI: Western Michigan University; 2015. http://scholarworks.wmich.edu/dissertations/1173. Accessed November 7, 2016.
15. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and measuring and reducing error. *Stat Med*. 1996;15(4):361–387.
16. Sullivan LM, Massaro MM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med*. 2004;23(10):1631–1660.
17. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006;163(7):670–675.
18. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: John Wiley & Sons Inc; 2013.
19. Rosenthal TC, Rosenthal GL, Lucas CA. Factors in the physician practice location puzzle: a survey of New York State residency-trained family physicians. *J Am Board Fam Pract*. 1992;5(3):265–273.
20. French R. Michigan gets med-school boom, doctor bust. *Bridge Magazine*. June 12, 2012. http://www.bridgemi.com/quality-life/michigan-gets-med-school-boom-doctor-bust. Accessed November 7, 2016.

Tracy J. Koehler, PhD, is Assistant Director of Research, Grand Rapids Medical Education Partners; **Jaclyn Goodfellow, MM,** is Director of Accreditation and Compliance, Grand Rapids Medical

Education Partners; **Alan T. Davis, PhD,** is Director of Research, Grand Rapids Medical Education Partners, and Associate Professor, Department of Surgery, Michigan State University; **Jessaca Spybrook, PhD,** is Associate Professor, Western Michigan University; **John E. vanSchagen, MD, FAAFP,** is Director of Medical Education and Designated Institutional Official, Mercy Health Saint Mary's, and Associate Professor, Department of Family Medicine, Michigan State University; and **Lori Schuh, MD, FAAN,** is Designated Institutional Official, Grand Rapids Medical Education Partners, and Vice President of Academic Affairs, Spectrum Health.

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

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

The authors would like to thank the many Grand Rapids Medical Education Partners staff, program coordinators, and program directors who assisted with data collection and/or verification of information.

This article is the continuation of a study introduced in a previous *JGME* article published in the October 1, 2016, issue: Koehler TJ, Goodfellow J, Davis AT, et al. Physician retention in the same state as residency training: data from 1 Michigan GME institution. *J Grad Med Educ*. 2016;8(4):518–522.

Corresponding author: Tracy J. Koehler, PhD, Grand Rapids Medical Education Partners, 945 Ottawa Avenue, Grand Rapids, MI 49503, 616.732.6223, fax 616.732.6275, tracy.koehler@grmep.org

Received May 9, 2016; revision received September 9, 2016; accepted October 4, 2016.