The American Journal of Surgery 217 (2019) 335–345

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

A 15-year residency program report card: Differences between the crème of the crop and the bottom of the barrel on the American Board of Surgery examinations

John L. Falcone [a, b, *]

[a] Owensboro Health, Owensboro Health Surgical Specialists, Department of Surgery, Owensboro, KY, USA
[b] University of Louisville, Department of Surgery, Louisville, KY, USA

ARTICLE INFO

Article history:
Received 9 April 2018
Received in revised form
18 June 2018
Accepted 1 September 2018

Keywords:
American board of surgery
Qualifying examination
Certifying examination
Residency program

ABSTRACT

Background: American Board of Surgery examination performance represents an important residency metric. The hypothesis is that demographic differences exist between the most and least successful programs.

Methods: This was a retrospective fifteen-year study. Data focused on program Examination Index (EI). The first and tenth decile programs were compared across demographics, using an $\alpha = 0.05$.

Results: The first decile had a higher EI than the tenth decile (91.0% ± 2 .6% vs 51.4% ± 5.4% [$p < 0.001$]). The first decile programs were larger ($p = 0.001$). The first decile had more military and academic programs, with fewer community programs ($p = 0.01$). More first decile programs were in the West with fewer in the Northeast ($p = 0.02$).

Conclusion: There are clear differences in ABS examination performance based on program size, type, and location. These results essentially perform a national needs-assessment, and may evoke a spirit of competition and collaboration.

Introduction

In order to become a board-certified general surgeon, one must pass the computerized multiple choice Qualifying Examination (QE)[1] and the oral Certifying Examination (CE)[2] offered by the American Board of Surgery (ABS). At the individual level, general surgery graduates view ABS certification as a major career milestone in the transition to independent practice.[3,4] Credentialing bodies, in turn, view ABS certification as an imprimatur of competence.[3,4] As a result, the primary measure of residency program success in the educational domain is measured by the performance of its graduates in the form of first-attempt pass rate on ABS examinations.[3–6] As a consequence of poor performance, programs can be subject to disciplinary action such as probationary periods and lack of accreditation by the Accreditation Council of Graduate Medical Education (ACGME) and by the Residency Review Committee (RRC).[3,5,7] The current RRC guidelines highlight the ACGME minimum standard as a first-attempt examinee pass rate of at least 65% on both the QE and CE during the last five year period.[8] Program pass rates on the ABS examinations are also some metrics considered by medical students in the application and interview process to general surgery programs.[4,7,9–11] Rolling five year program pass rates are electronically published and available on the ABS website.[9]

There have been several studies specifically evaluating program performance metrics associated with ABS examination outcomes. Program size is positively correlated with program performance on the QE and on the CE.[7,12] Larger programs likely have more resources and monies available for educational materials and curricula.[7] Program size is also positively correlated with the in-training examination equivalent for Osteopathic residency programs.[13]

Program location is also associated with program performance.[11–15] On the CE, programs on the West coast of the United States outperform other regions of the United States.[11] Regionally, states that are members of the Southwestern Surgical Congress outperform other areas on both the QE and CE.[14] In contrast, states and the territory of Puerto Rico that are members of the Southeastern Surgical Congress do not do as well on the QE or CE compared to other areas in the United States.[15]

* Ridgecrest Medical Park, 2801 New Hartford Road, Owensboro, KY, 42303, USA.
E-mail address: John.FalconeMD@owensborohealth.org.

https://doi.org/10.1016/j.amjsurg.2018.09.002
0002-9610/

There are a few other variables associated with residency program demographics and ABS examination outcomes. There have been some studies to suggest that military programs outperform their civilian counterparts on standardized testing.[4,16] One recent study showed that the programs subscribed to the Surgical Council on Resident Education (SCORE) Curriculum (n = 200) had higher mean pass rates than programs without subscriptions (n = 33) (86,4% vs 82.7%, $p > 0.05$).[12] Other curricula that programs offer, such as mock oral board examinations have conferred benefit on the ABS CE.[6,17,18] A final known published program characteristic associated with ABS examination performance include an inverse relationship in pass rates and the percentage of international medical graduates (IMGs) in a residency program.[12]

At the residency program level, ABS performance is indeed a point of pride for some residencies. Several general surgery residency programs list specific examination performance on websites and other electronic media; these programs use ABS examination performance as a positive selling point and a marketing tool for attracting high-quality applicants and persuading applicants to match at their specific residency program.[19–21] Examples of this include: "All of our graduates from the past five years have passed both the ABS Qualifying and Certifying Examinations on their first attempt."[19] Another program states that it has "In-Service Training Examination (ABSITE) scores well above the national average and a near 100% first-time pass rate on the Qualifying Examination (written boards)."[20] Still another program highlights its performance as a distinct regional selling and marketing point:

High Board of Surgery Passage Rates

> The dedication of our faculty, in part, has helped our residents achieve exemplary American Board of Surgery passage rates. From 2006 to 2010 our residents had a 93% pass rate on their first attempt at the qualifying exam (national average = 86%), a 92% pass rate on their first attempt at the certifying exam (national average = 83%) and an 87% overall first time taker success rate (national average = 73%). Our 87% QE/CE combined success rate is higher than almost all of the other residency programs in Boston.[21]

In sum, general surgery residency program performance on the ABS QE and ABS CE is very important. With the importance and the pressure for success on these high-stakes examinations, the primary purpose of this study is really to compare the highest-performing established residency programs and the lowest-performing established residency programs while presenting a very explicit honor roll format of program performance. The hypotheses, based on previous studies, are that larger programs, military programs, programs in the Western United States, and programs with a lower IMG percentage are more commonly found in the highest-performing residency programs on the ABS examinations.

Materials and methods

This was a retrospective study from the 2001–2002 academic year to the 2015–2016 academic year. This study was evaluated by the Institutional Review Board from the University of Louisville, and the application was accepted as an exempt study. The data sets were obtained through previously-published residency program performance documents by the ABS.[22–24] The data sets represented three consecutive and non-overlapping five year time frames: early,[22] middle,[23] and late.[24]

All residency programs were included for potential comparison if the program had data available for all three of the time frames and had at least one examinee take the QE and the CE in each time frame. Double data entry was performed to ensure extrapolation accuracy. The data were collated across time frames using program codes. For each program, the data were comprised of the number of examinees that took the QE and the number that passed and failed on the first attempt, the number of examinees that took the CE and the number that passed and failed on the first attempt, and the combined index, or examination index (EI), which represents the proportion of examinees per program that passed both the QE and the CE on the first attempt.[22–24]

To compare the programs with the most favorable ABS outcomes to the least favorable ABS outcomes, the residency programs were arbitrarily split into deciles, based on overall ABS EI. The first decile and the tenth decile programs were then compared by QE pass percentage, CE pass percentage, and EI with t-tests. Cohorts were also compared with univariate analyses by program size, type, location, and IMG percentage. Program size was estimated by the number of examinees that attempted the QE during the study period, as has been performed in previous study of ABS examinations.[7] As this distribution was normal, t-tests were used for univariate comparisons of program size between deciles.

Programs were classified by type, in that they were defined as either a military program, an academic program, or a community program. The same algorithm used in two previous studies was used in this study.[4,26] The distribution of program types between the first and tenth deciles was compared with a 2×3 contingency table Fisher's exact test.

Programs were classified by location according to their United States Census Bureau Region, with defined regions the Northeast, Midwest, South, and West.[27] This sorting algorithm had also been used in previous studies.[11,13] The distribution of program Census Bureau Regions between the first and tenth deciles was compared with a 2×4 contingency table Fisher's exact test.

The percentage of IMGs was found through program-specific and reported data available through the Fellowship and Residency Electronic Interactive Database Access (FREIDA).[25] As this distribution was not normal, Mann-Whitney U tests were used for univariate comparisons of IMG percentage between deciles. A FREIDA search engine data point available also includes the program's range of the United States Medical Licensure Examination (USMLE) Step 1 by its residents. The FREIDA data points of IMG percentage and USMLE Step 1 performance were incomplete.

Multiple regression was also performed with the EI as the dependent variable. As comparing just the extremes in EI performance introduces statistical bias across a continuous outcome variable, the multiple regression included all programs. Independent explanatory variables included program size, program type (with the community designation as the dummy/comparative variable), and program location (with the Northeast as the dummy/comparative variable). As the IMG percentage data were incomplete, the multiple regression was performed twice. The first multiple regression included programs with IMG data. The second multiple regression included all programs, but omitted IMG percentage as an explanatory variable because of its incomplete state. All statistics were performed with Stata 11.1 statistical software (StataCorp, College Station, TX), using an $\alpha = 0.05$.

Results

From the 2001–2002 academic year to the 2015–2016 academic year, there were 231 programs that satisfied inclusion criteria. The mean program pass rate on the QE was 85.4% ± 8.5%. (The total pass rate by all examinees was 12,568/14,543 = 86.4%.) The mean program pass rate on the CE was 83.0% ± 8.2%. (The total pass rate by all examinees was 10,787/12,817 = 84.2%.) The mean program EI was 72.9% ± 11.6%. (The total examination index off all examinees was

10,800/14,543 = 74.3%.) Paired *t*-test showed no trends in overall program EI/rank between the early and late time periods ($p = 0.38$). A scatter plot of the included programs by QE and CE performance is shown in Fig. 1. There are three programs that are "at risk" for violating the minimum RRC/ACGME standards of 65% first-attempt pass rates.[8] All of the programs that satisfied inclusion criteria ($n = 231$) are given by rank and overall EI and number of examinees in Table 1.

The first decile and tenth decile programs are, by default, also listed by overall EI in Table 1. The first decile ($n = 23$) represents rank #1 – rank #23. The tenth decile ($n = 23$) represents rank #209 – rank #231. Over time, between the early and late time periods, the ABS EI increased globally in the bottom decile programs (57%), and decreased globally in the top decile programs (70%). Paired t-tests did not show any trends in overall EI/rank between the early and late time periods in either decile (both $p > 0.25$). The first decile programs had a higher pass percentage on the QE (96.3% ± 2.2% vs 71.4% ± 7.2% [$p < 0.001$]), CE (93.9% ± 2.6% vs 70.0% ± 8.9% [$p < 0.001$]), and had higher examination indices (91.0% ± 2.6% vs 51.4% ± 5.4% [$p < 0.001$]). With regard to program size, the first decile programs were larger than the tenth decile programs (77 ± 36 vs 47 ± 23 [$p = 0.001$]).

The differences in program types between first and tenth decile programs are shown in Table 2. First decile programs had more military and academic programs, whereas a majority of tenth decile programs were community programs ($p = 0.01$).

The differences in geographic locations between first and tenth decile programs are shown in Table 3. First decile programs had more presence in the West and lesser of a presence in the Northeast ($p = 0.01$).

With regard to IMG percentage, data were available for 13/23 (57%) of the first decile and 13/23 (57%) of the tenth decile programs. Univariate comparisons showed that first decile programs had a median lower IMG percentage than tenth decile programs (4.4% (Interquartile Range (1.0%–6.5%)) vs 34.6% (Interquartile Range (IQR) [25.2%–42.8%]) [$p < 0.001$]).

The multiple regression analyses included all 231 programs, using program size, program type, program location, and IMG percentage as the independent variables. The mean program size during the study was 63 examinees ±27 examinees; the scatterplot of program size to EI is shown in Fig. 2, and the simple linear regression showed a positive slope ($p < 0.001$).

By program type, there were 8 (3.4%) military programs, 136 (58.9%) academic programs, and 87 (37.7%) community programs. A Kruskal-Wallis test by program type showed EI differences based on military program (87.0% (IQR [81.5%–89.8%])), academic

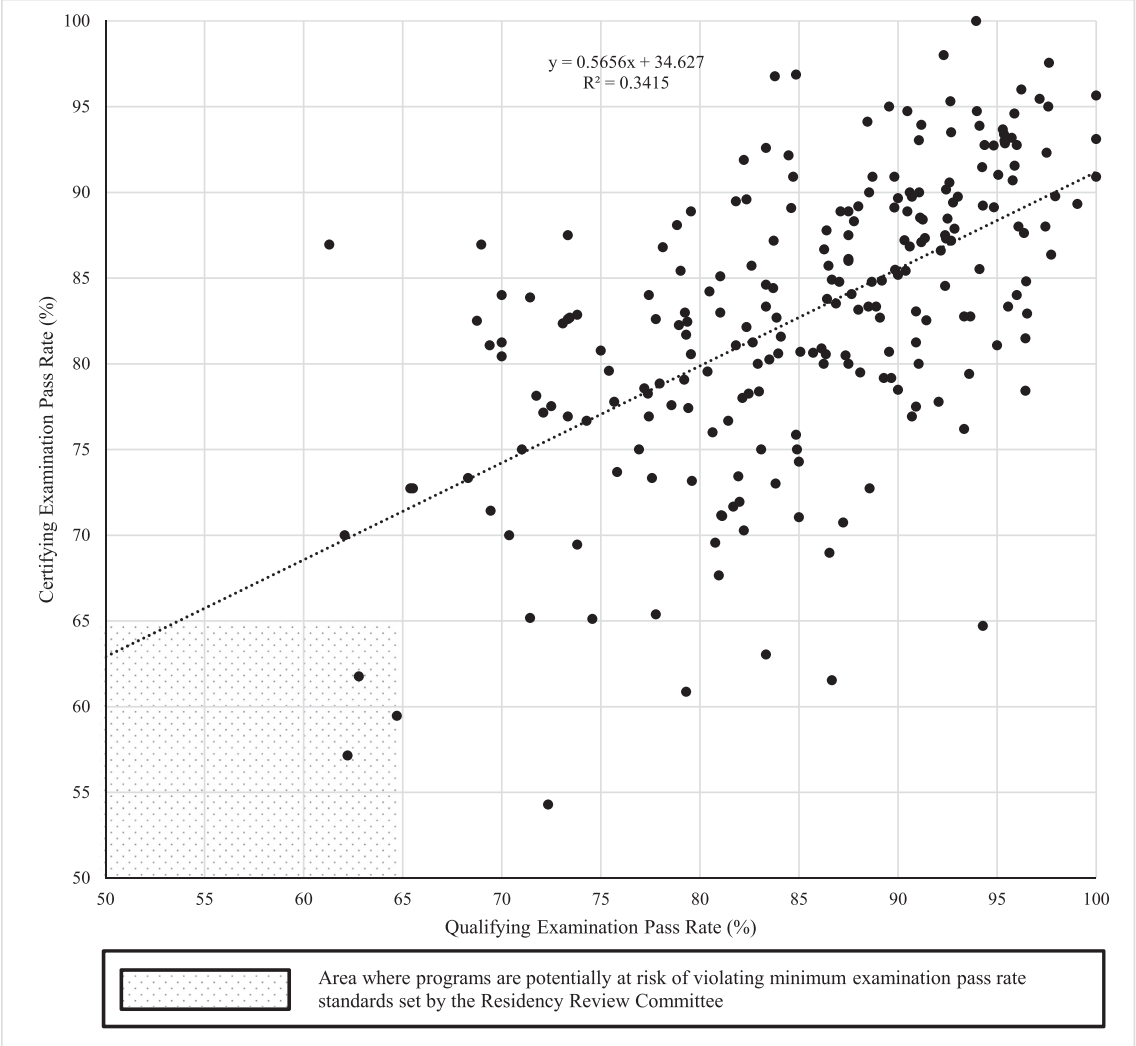

Fig. 1. Overall performance on the American Board of Surgery Qualifying Examination and Certifying Examination over a 15-year time frame by 231 residency programs.

Table 1
Program rank by American Board of Surgery examination index over 15 years.

Rank	Program Name	Examination Index	Number of Examinees (n)
1	Madigan Army MC	97.6%	(42)
2	Methodist - Dallas	95.8%	(24)
3	S.A.Unif.Svc.-Brooke Army	94.3%	(53)
4	Riverside Methodist	93.9%	(33)
5	Gundersen	93.1%	(29)
6	Northwestern Univ.	92.9%	(70)
7	U. Texas Southwest.	92.8%	(166)
8	Mass. General Hosp.	91.7%	(121)
9	Iowa Methodist	91.5%	(47)
10	Vanderbilt Univ.	91.3%	(104)
11	North Carolina	90.6%	(85)
12	Mayo Clinic Arizona	90.0%	(40)
13	Carolinas Medical Center	89.4%	(47)
14	Brown University	89.3%	(75)
15	Univ. of Michigan	89.2%	(83)
16	Mayo Clinic	89.1%	(129)
17	NY & Presbyt. - Cornell	88.9%	(108)
18	University of Utah	88.7%	(71)
19	Univ. of Washington	88.7%	(97)
20	UC - Davis	88.6%	(105)
21	UCLA	88.5%	(87)
22	U. Tenn.- Memphis	88.4%	(95)
23	Natl. Capital Consortium	88.2%	(68)
24	Naval Hospital-San Diego	88.2%	(51)
25	TX A&M (Scott & White)	87.9%	(58)
26	Univ. of Kentucky	87.7%	(73)
27	Univ. of Wisconsin	87.2%	(78)
28	U. of Kansas - Wichita	86.6%	(82)
29	Univ. of Chicago	86.6%	(67)
30	Stanford University	86.4%	(81)
31	Bronx-Lebanon	86.4%	(44)
32	Johns Hopkins	86.2%	(87)
33	Good Samaritan	85.7%	(56)
34	Univ. of Virginia	85.7%	(70)
35	Tripler Army MC	85.7%	(42)
36	Univ. Pennsylvania	85.6%	(97)
37	Mount Carmel Health	85.3%	(34)
38	Univ. of Florida	84.8%	(66)
39	Emory University	84.6%	(123)
40	Baylor Univ. - Dallas	84.5%	(110)
41	Washington Univ.	84.5%	(97)
42	St.Jos. Mercy-Ann Arbor	84.3%	(51)
43	Mayo - Jacksonville	84.0%	(25)
44	Baptist Health System	83.8%	(37)
45	Tufts Medical Center	83.7%	(43)
46	Yale - New Haven	83.3%	(78)
47	Greenville Hospital	83.3%	(54)
48	New Hanover	83.3%	(30)
49	Brigham & Women's	83.3%	(108)
50	LAC - Harbor UCLA	82.9%	(70)
51	Univ. of Pittsburgh	82.5%	(80)
52	U Texas - Houston	82.5%	(80)
53	Good Samaritan (Phoenix)	82.4%	(85)
54	Univ. of Cincinnati	82.4%	(85)
55	Med Col of Wisconsin	82.4%	(85)
56	Swedish Medical Center	82.4%	(34)
57	Virginia Mason MC	82.3%	(62)
58	Kaiser - Los Angeles	82.2%	(45)
59	Univ. of Colorado	81.9%	(105)
60	William Beaumont	81.8%	(66)
61	Naval MC - Portsmouth	81.8%	(33)
62	Loma Linda	81.5%	(81)
63	Mt. Sinai	81.4%	(86)
64	UC - San Francisco	81.4%	(102)
65	Huntington Memorial	81.0%	(21)
66	Indiana University	80.7%	(135)
67	Rush-Presb.-St.Luke's	80.6%	(108)
68	Wm. Beaumont Army MC	80.5%	(41)
69	Grand Rapids MSU	80.2%	(86)
70	Univ. of Alabama	80.2%	(96)
71	Georgetown Univ.	80.0%	(65)
72	U. Texas San Antonio	80.0%	(125)
73	Spartanburg General	80.0%	(30)
74	Univ. Miami-Jackson	80.0%	(90)

Table 1 (*continued*)

Rank	Program Name	Examination Index	Number of Examinees (n)
75	LAC - USC	79.6%	(93)
76	Hennepin County	79.4%	(63)
77	Univ. of Toledo	78.8%	(52)
78	Case Western Reserve	78.8%	(85)
79	New York Univ.	78.8%	(118)
80	San Joaquin General	78.6%	(28)
81	U. Tenn. - Knoxville	78.6%	(56)
82	Providence - Southfield	78.6%	(42)
83	Dartmouth-Hitchcock	78.4%	(51)
84	Duke University	78.1%	(96)
85	Baylor - Houston	77.9%	(104)
86	Exempla/St.Joseph (CO)	77.8%	(45)
87	Loyola University	77.6%	(85)
88	U. Tenn.-Chattanooga	77.6%	(58)
89	Guthrie Healthcare (PA)	77.5%	(40)
90	Univ.Missouri-Columbia	77.5%	(40)
91	Atlanta Medical Center	77.4%	(31)
92	Boston University	76.8%	(69)
93	St.Jos. Mercy - Pontiac	76.7%	(30)
94	Easton Hospital	76.7%	(30)
95	UC-SF (Fresno)	76.5%	(51)
96	George Washington U.	76.4%	(55)
97	Oregon Health Sc.Un.	76.2%	(147)
98	Wake Forest University	75.8%	(66)
99	Med. Univ. South Carolina	75.8%	(66)
100	Ochsner	75.7%	(70)
101	Beth Israel Deaconess	75.7%	(111)
102	Eisenhower Army MC	75.7%	(37)
103	Univ. of Rochester	75.6%	(78)
104	Abington Memorial	75.5%	(53)
105	Univ. Massachusetts	75.3%	(81)
106	Louisiana State	75.0%	(100)
107	U. of New Mexico	75.0%	(60)
108	St. Barnabas-NJ	75.0%	(56)
109	Univ. of Louisville	74.7%	(99)
110	Washington Hosp.Cntr	74.6%	(67)
111	U. of Mississippi	74.6%	(67)
112	Memorial - Savannah	74.4%	(43)
113	Conemaugh	74.2%	(31)
114	Lehigh Valley Hosp.	74.1%	(58)
115	East Carolina Univ.	74.1%	(54)
116	Thomas Jefferson U.	74.1%	(81)
117	Cedars-Sinai	74.1%	(54)
118	Univ. of Minnesota	73.6%	(87)
119	NY/Presbyt.-Columbia	73.6%	(87)
120	Harlem Hospital	73.3%	(45)
121	Western Michigan	73.3%	(30)
122	Albany Med. Cntr.	73.2%	(56)
123	U. of Ill. - Peoria	73.2%	(41)
124	U.of Ct.	72.8%	(92)
125	U. Illinois Metro	72.7%	(88)
126	Bassett Healthcare	72.7%	(33)
127	University of Vermont	72.7%	(44)
128	Univ. of Maryland	72.1%	(68)
129	Univ. of Arkansas	71.9%	(64)
130	Cleveland Clinic	71.1%	(90)
131	Ohio State Univ.	71.0%	(69)
132	Christiana Care (DE)	71.0%	(62)
133	Med. College Georgia	71.0%	(62)
134	UC - San Diego	70.5%	(78)
135	Summa - Akron City	70.5%	(44)
136	Orlando Regional	70.5%	(44)
137	St. Agnes	70.5%	(44)
138	Wright State Univ.	70.3%	(101)
139	West Virginia U. Chrlstn	69.8%	(43)
140	Lahey Clinic	69.7%	(33)
141	U. Florida - Jacksonville	69.2%	(52)
142	Henry Ford Hosp.	69.1%	(81)
143	Univ. of Oklahoma	69.0%	(58)
144	Eastern Virginia	69.0%	(58)
145	University of Iowa	68.8%	(80)
146	SUNY - Syracuse	68.8%	(64)
147	VCU - MCV	68.7%	(67)
148	U. of North Dakota	68.6%	(35)

(*continued on next page*)

Table 1 (*continued*)

Rank	Program Name	Examination Index	Number of Examinees (n)
149	UPMC Mercy	68.3%	(41)
150	Mercer University	68.2%	(44)
151	Maine Medical Cntr.	68.1%	(47)
152	U. of Illinois Aff.	68.0%	(97)
153	Michigan State Univ.	67.9%	(53)
154	Univ. South Florida	67.1%	(76)
155	UC - Irvine	66.7%	(75)
156	West Virginia U. Mrgntown	66.7%	(42)
157	Jewish Hosp. - Cincinnati	66.7%	(27)
158	Univ. of Kansas	66.7%	(57)
159	St. John Hospital	66.1%	(56)
160	Monmouth Med. Cntr.	65.9%	(44)
161	Univ. of Hawaii	65.7%	(35)
162	Rutgers- New Brunswick	65.5%	(87)
163	U. of Mo.	65.1%	(63)
164	Palmetto Health Alliance	65.0%	(40)
165	Marshfield-St.Joseph's	64.5%	(31)
166	Rutgers - Newark	64.4%	(104)
167	LSU - Shreveport	64.3%	(70)
168	Univ. of Puerto Rico	64.3%	(70)
169	Maricopa County	64.2%	(53)
170	The Methodist Hospital	63.8%	(47)
171	Allegheny Genl.Hosp.	63.4%	(71)
172	Lenox Hill Hosp.	63.3%	(30)
173	Einstein - Montefiore	63.3%	(90)
174	Akron General	63.3%	(30)
175	Univ. of South Alabama	63.2%	(57)
176	Kern Medical Center	63.0%	(27)
177	Univ. of Nebraska	62.7%	(51)
178	St. Louis University	62.7%	(59)
179	Southern Illinois U.	62.5%	(40)
180	St.Luke's-Roosevelt	62.5%	(72)
181	Wayne State	62.4%	(101)
182	York Hospital	62.2%	(45)
183	Maimonides Med. Cntr.	62.2%	(74)
184	Texas Tech - El Paso	62.2%	(37)
185	NY Hosp. - Queens	62.1%	(58)
186	SUNY - Stony Brook	62.0%	(71)
187	East Tennessee St.	61.8%	(68)
188	West. Reserve-Youngstown	61.5%	(26)
189	St. Mary's Hospital	61.5%	(26)
190	Lankenau Medical Center	61.3%	(31)
191	North Shore Univ. Hosp.	61.0%	(100)
192	Nassau University	61.0%	(41)
193	UC - SF, East Bay	60.7%	(84)
194	Penn State Univ.	60.7%	(61)
195	NY Med.Col-Westchester	60.0%	(60)
196	U. of Ok. - Tulsa	60.0%	(35)
197	Geisinger Med. Cntr.	60.0%	(35)
198	Texas Tech (Lubbock)	60.0%	(35)
199	Cooper - Camden	60.0%	(45)
200	Baystate Med. Cntr.	59.4%	(64)
201	St. Luke's - Bethlehem	59.4%	(32)
202	Univ. of Arizona	59.3%	(86)
203	Univ. of Nevada	59.2%	(49)
204	Carilion Health Sys.	59.2%	(49)
205	Univ. of Buffalo	59.2%	(120)
206	Albert Einstein (PA)	59.0%	(39)
207	Santa Barbara Cottage	58.8%	(34)
208	Mount Sinai-Beth Israel	58.7%	(46)
209	Stamford Hospital	58.6%	(29)
210	U. Texas - Galveston	58.5%	(53)
211	Creighton University	57.4%	(54)
212	New York Methodist	57.1%	(42)
213	Temple University	57.1%	(91)
214	Mt. Sinai	56.3%	(48)
215	Drexel University	56.3%	(80)
216	Howard University	55.1%	(69)
217	St. Elizabeth's (OH)	54.8%	(42)
218	Presence Saint Joseph Hsp	53.3%	(15)
219	Mercy Catholic MC	51.7%	(29)
220	Pinnacle Hlth-Harrisbgh	51.6%	(31)
221	SUNY - Brooklyn	50.9%	(112)
222	Atlantic - Morristown	50.8%	(59)
223	Waterbury Hospital	50.0%	(26)

Table 1 (*continued*)

Rank	Program Name	Examination Index	Number of Examinees (n)
224	St. Elizabeth's (MA)	50.0%	(36)
225	Union Memorial	48.3%	(29)
226	Mt.Sinai - Miami	48.1%	(27)
227	Morehouse - Atlanta	44.8%	(29)
228	Brookdale Hospital	44.7%	(47)
229	Marshall University	44.2%	(43)
230	Brooklyn Hospital Ctr.	43.1%	(51)
231	Staten Island Hosp.	40.0%	(45)

program (74.9% (IQR [66.4%–82.0%])), and community program (69.7% (IQR [61.1%–79.0%])) ($p < 0.001$). Mann-Whitney U tests showed that military programs had higher EIs than academic ($p = 0.001$) and community ($p < 0.001$) programs. Academic programs had higher EIs than community programs ($p < 0.01$).

By program location, there were 70 (30.3%) Northeast programs, 54 (23.4%) Midwest programs, 71 (30.7%) Southern programs and 36 (15.6%) Western programs. A Kruskal-Wallis test by program location showed EI differences by program location in Northeast (65.0% (IQR [59.5%–75.4%])), Midwest (73.4% (IQR [66.7%–82.4%])), South (75.7% (IQR [67.6%–82.9%])), and West (81.2% (IQR [70.5%–84.4%])) ($p < 0.001$). Mann-Whitney U tests showed differences between groups (all $p < 0.001$) except comparisons between Western and Southern programs ($p = 0.06$), Western and Midwestern programs ($p = 0.08$) and Southern and Midwestern programs ($p = 0.96$).

There were IMG data available for 135/231 (58.4%) of the included study programs. The median program IMG percentage mean program size during the study was 10.1% (IQR [3.3%–24.7%]); the scatterplot of IMG percentage to EI is shown in Fig. 3, and the simple linear regression showed a negative slope ($p = 0.001$).

The multiple regression analyses including the programs with IMG data ($n = 135$) showed that 46.1% of the variance was explained by the model and that the regression was statistically-significant overall ($p < 0.001$). Statistically significant regression variable coefficients included program size ($p < 0.001$) and academic designation to community designation ($p = 0.02$). There were no military programs included. A location in the West ($p = 0.03$), Midwest ($p = 0.01$), and South ($p = 0.01$) were significant compared to the Northeast. The IMG coefficient was also statistically-significant ($p = 0.03$).

The multiple regression including all programs ($n = 231$) showed that 50.2% of the variance was explained by the model and that the regression was statistically-significant overall ($p < 0.001$). Statistically significant regression variable coefficients included program size ($p < 0.001$) and military designation to community designation ($p < 0.001$). No differences were seen in the coefficient comparing academic and community designations ($p = 0.12$). A location in the West ($p < 0.001$), Midwest ($p < 0.001$), and South ($p < 0.001$) were significant compared to the northeast.

There were 154 (66.7%) of programs with FREIDA data showing resident performance on USMLE Step 1. The range of performance for programs was 221–240 ($n = 16$ (10.4%)), 241–260 ($n = 126$ (81.8%)) or >260 ($n = 12$ (7.8%)). A Kruskal-Wallis test showed differences between EI of programs in the low range (64.8% (IQR [62.0%–80.0%])), middle range (72.8% (IQR[62.4%–79.5%])), and upper range (86.7% (IQR[82.1%–88.7%])) ($p < 0.001$). Mann-Whitney U tests showed differences in EIs between the upper range and the middle/lower range programs (both $p < 0.001$), but no differences in EIs between the middle and lower range programs ($p = 0.16$).

Discussion

The main objective of this study was to define and rank program performance on the ABS examinations, comparing the most and least successful established residency programs on the ABS examinations. Overall, there is a wide spectrum of performance on the ABS examinations, as the first and tenth decile programs differed statistically in overall QE performance, CE performance, and EI over the study period (all $p < 0.001$). The hypotheses of this study were that larger programs, military programs, programs in the Western United States, and programs with a lower IMG percentage were more commonly found in the highest-performing residency programs on the ABS examinations. The study results do support the study hypotheses. The univariate comparisons showed that the cohort of first decile programs were larger than the cohort of tenth decile programs. There were more military programs in the first decile cohort and none in the tenth decile cohort. There were more programs in the West in the first decile cohort and none in the tenth decile cohort. The percentage of IMGs was also lower in the first decile programs. All of the comparisons/distributions were statistically-significant.

Ultimately, the multiple regression is the strongest evidence of true differences across the spectrum of residency performance, with significant regression coefficients for program size, between types, between locations, and plus or minus IMG percentage while controlling for the other variables. This included all of the programs that satisfied initial study inclusion criteria, and is much more generalizable. The multiple regression analyses explained a moderate amount of variance, and were statistically-significant overall. The multiple regression with all 231 programs (omitting the IMG percentage as an explanatory variable) had every explanatory variable coefficient as statistically-significant. There were no significant differences in terms of overall EI rank changes over time, suggesting relative stability of program performance. The multiple

Table 2
Comparison of residency program type and decile of the American Board of Surgery examination index.

	Program Type			Total
	Military	Academic	Community	
First Decile	3 (13.0%)	14 (60.9%)	6 (26.1%)	23 (100.0%)
Tenth Decile	0 (0.0%)	8 (34.8%)	15 (65.2%)	23 (100.0%)
				$p = 0.01$

p: p-value of a 2×3 contingency table Fisher's Exact Test.

Table 3
Comparison of residency program location and decile of the American Board of Surgery examination index.

	United States Census Bureau Region				Total
	Northeast	South	Midwest	West	
First Decile	4 (17.4%)	9 (39.1%)	6 (26.1%)	4 (17.4%)	23 (100.0%)
Tenth Decile	13 (56.5%)	7 (30.4%)	3 (13.0%)	0 (0.0%)	23 (100.0%)
					$p = 0.01$

p: p-value of a 2×4 contingency table Fisher's Exact Test.

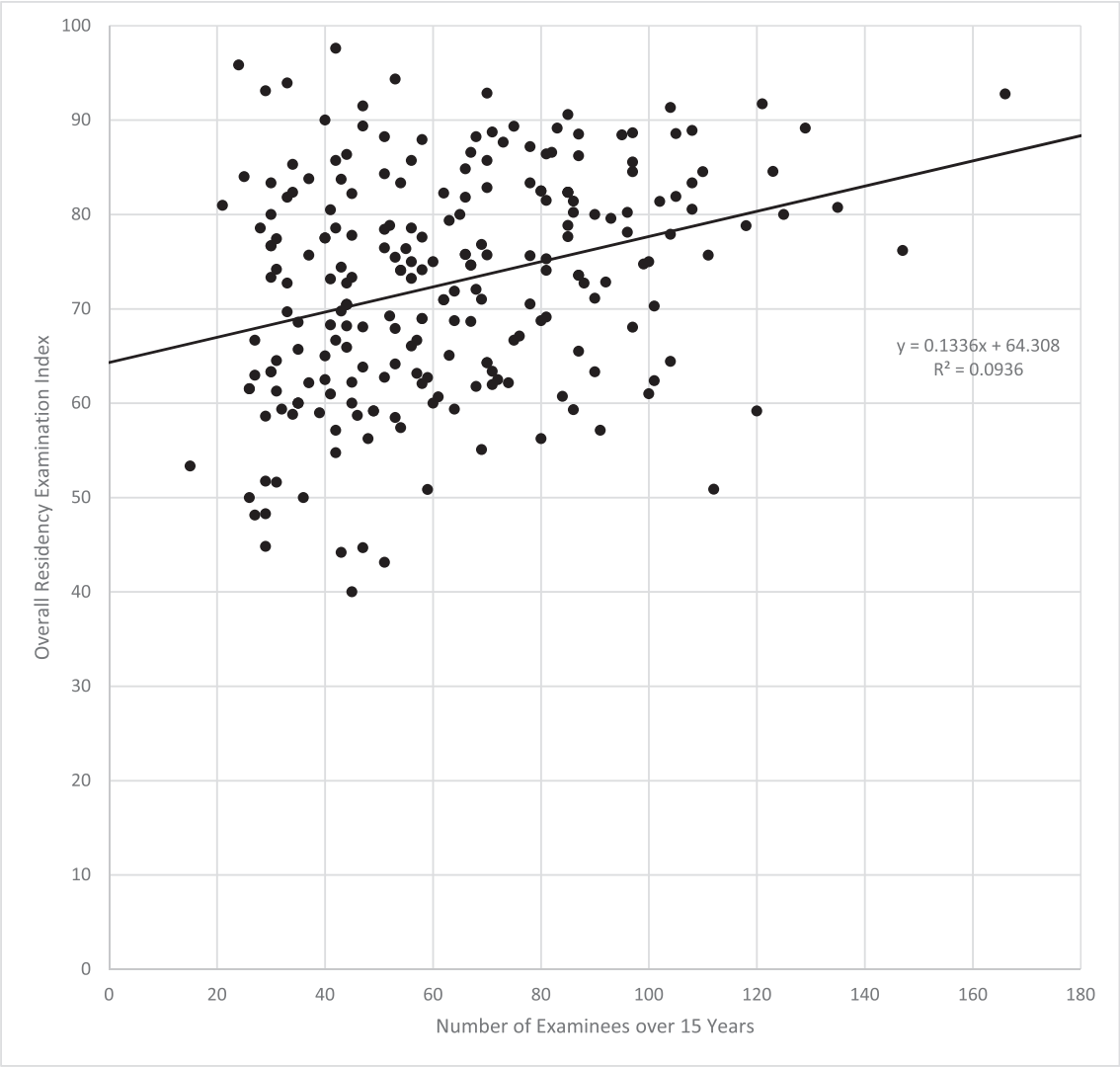

Fig. 2. Overall Examination Index of 231 programs by the number of examinees during the study period.

regression findings confirm and validate the findings and distributions seen at the extremes of EI performance occur across the spectrum of programs in The United States.

Table 1 of this study is a very practical and meaningful listing. It is analogous to a report card or a leaderboard. This table shows the long-term performance of every established residency program over the last fifteen years. Now, the ABS does publish rolling five-year residency performance on its website.[9] But, there are several limitations of that listing, in that it does not compare programs in any ranked outcome order, but gives an alphabetical listing of programs. That data set with regard to pass rates and EI can also be more greatly affected by smaller numbers, where one failing examinee affects the overall pass rate percentage for smaller programs much more than one failing examinee for larger programs. Table 1, however, reflects program numbers over 15 years, and has a much higher number of examinees per program. It should be highlighted, that ABS examination performance does not in any way to reflect on the quality of care provided to patients by the parent institution.[9] Table 1 does have some interesting findings in that there are some smaller programs in the first few deciles and some larger programs in the last deciles. Similarly, there are some community programs in the first few deciles and academic programs in the last few deciles. The same phenomenon is true with regard to program location, and IMG percentage. So, while there are indeed some real differences in ABS performance by demographics, these are not without overlap.

Interestingly, there were eight military programs that satisfied initial inclusion criteria. It is very interesting that a rather high number, 3/8 (38%), are in the first decile. The range of military designation EI rank was from 1 to 102. There could indeed be some personality characteristics of those with a military background such as more disciplined study habits; this could make for a very interesting future study in the domain of ABS examination performance.

Providing Table 1 has a few potential ramifications. First, it can be a useful tool for medical students that are applying to categorical general surgery positions,[4,7,9–11] as program performance in the past is likely an indicator of program performance in the future. So, it is a useful resource for applicants and could potentially affect application and interview patterns. The author of this manuscript personally looked at the list to see where his previous and current affiliated programs were located. The suspicion is that all of the readers of this manuscript will do the same.

This table could also serve as a resource for residency programs, especially ones that like to use performance as a means to recruit and market their programs.[19–21] Table 1 could provide an incentive

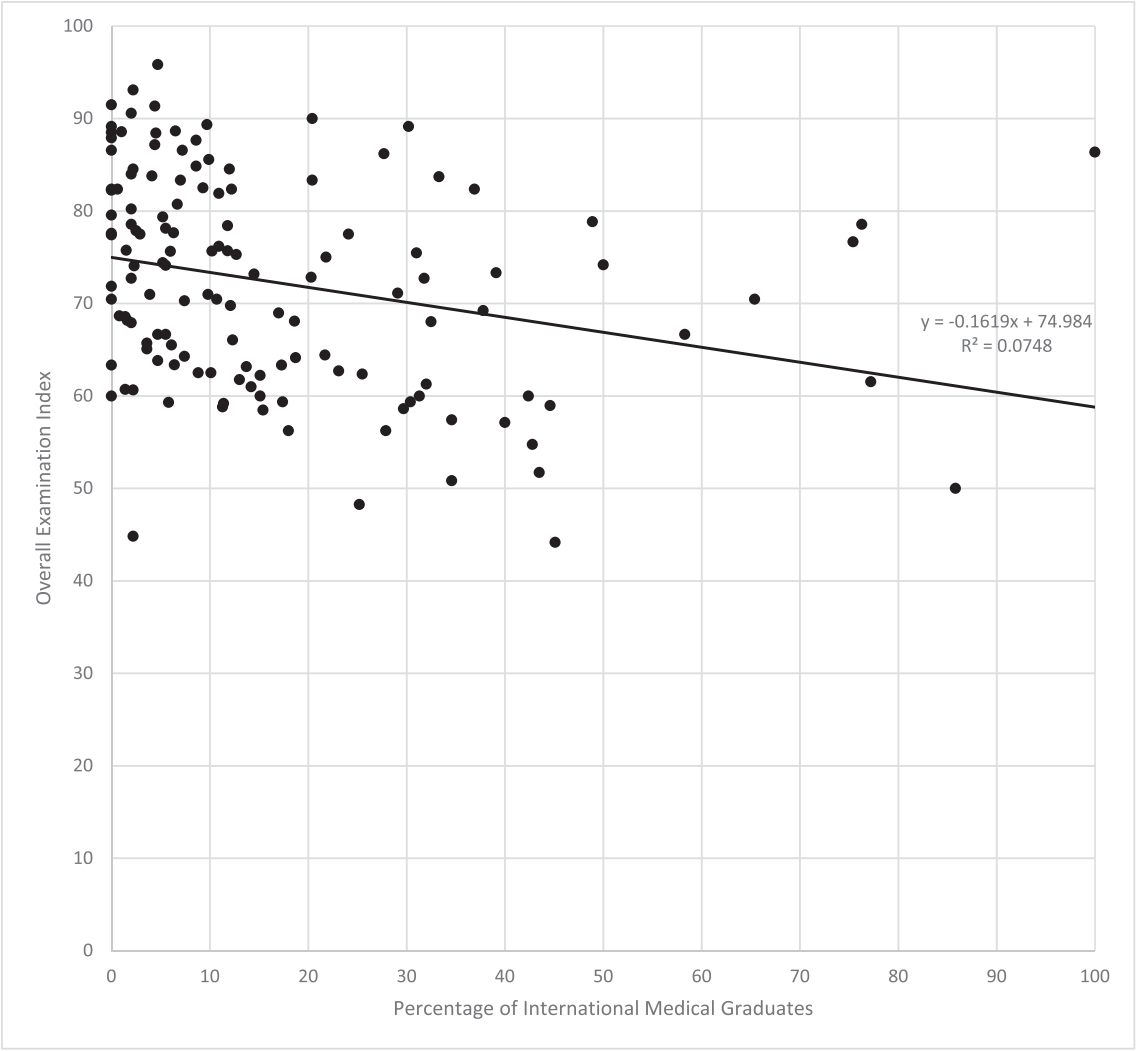

Fig. 3. Overall Examination Index of 135 programs by the percentage of international medical graduates during the study period.

and evoke a competitive spirt for residency programs to improve performance, as surgeons do have competitive personalities with regard to outcomes.[28,29] This presentation essentially performs a national needs-assessment, and promotes a sort of "gamification" of ABS examination performance by providing a report card/leaderboard, which certainly contribute to motivation for improvement.[29] While this study certainly highlights the programs with the most success on the ABS examinations, it also highlights the programs that are the least successful. Some of these programs could be "at risk" for potential disciplinary action by the RRC,[3,5,7] as the minimum standard is currently defined by at least 65% on both the QE and CE during the last five year period.[8] That said, a majority of the tenth decile programs had a global increase in overall EI over the study period.

Finally, some programs hoping for improved performance may be geographically located nearby more successful programs with which they could potentially collaborate. Regional collaboration is a very important element with regard to ABS performance, as it has been an important revelation in other regional studies of ABS performance in the Southwestern Surgical Congress and the Southeastern Surgical Congress.[14,15]

There are certainly many limitations of this study. First, it is a retrospective program performance review, and it subject to all the limitations of retrospective studies, such as the inability to establish a causal relationship between explanatory variables and ABS examination performance as well as the presence of confounding variables. Second, there are additional program-level variables that have previously been associated with ABS examination performance that were not evaluated in this study or had incomplete data. These would include SCORE subscription status,[12] offering mock oral board curricula,[6,17,18] and a complete data set of the percentage of international medical graduates.[12] The lack of complete data is a limitation of the data sets that were used in this study.

The SCORE portal is a very common surgical program resource. In 2012, up to 96% of all of the programs carried subscriptions.[12,30] That said, if subscription is a ubiquitous variable, it would be unlikely to be associated with differences between decile performance. This author would contend that the SCORE portal is more important at the individual level than at the program level, as there is not a statistical advantage of the SCORE portal at the program level.[12] Programs with SCORE portal access also have users/residents with zero logins and zero self-assessment question attempts,[12] suggesting that access alone is not always effective.

Mock oral board data were also not available for the studied residency programs. The suspicion is that most programs have some element of mock oral board preparation given the importance of the CE, but the nature and standardization of those curricula are

unknown. All programs offer the ABSITE annually as a mock written examination for QE preparation. Again, if variables are ubiquitous, it is not possible to show differences between cohorts.

The percentage of international medical graduates and residency USMLE performance are variables available for general surgery applicants through the FREIDA search engine; but these data were incomplete for the programs.[25] The validity of these self-reported program metrics are unknown. The IMG percentage was a statistically-significant variable in univariate comparisons, had a negative slope in the simple linear regression involving EI performance, and was also statistically-significant in the multiple regression analysis.

This study also did not look at individual examinee characteristics associated with ABS examination performance, such as medical school rank, USMLE performance, ABSITE performance, and the reception of awards, and having a mandatory research year.[3,5] Examinees that delay taking the QE also do not do as well as those who take the examination immediately after residency completion.[31] This said, individual characteristics were well outside the scope of this study, as this study focused on performance at the residency program level. And while outside of the scope of the study, interestingly, a majority of programs had USMLE performance available. The programs with residents with higher USMLE performance ranges also were programs with higher examination indices. This phenomenon makes a lot of sense with the previous work done at the individual level.[3,5] This could represent the chicken-and-the-egg debate, whether it is unknown if programs with academically-stronger residents (as defined by USMLE performance) do better on the ASE examinations, or if programs that do better on the ASE examinations attract stronger residents. Examining this perpetual cycle with future study would be very interesting in this sector of surgical education research.

On the other hand, there are also several strengths of this study. First, it has a very straightforward study design. The comparisons between the first and tenth decile programs are neatly done with straightforward statistical methods. Second, this study represents an evaluation of the population of established residency programs in the United States over a substantial fifteen year time frame. This breadth and depth is far superior to short-term studies of single institutions or regional institutions. Moreover, the fidelity of the outcome data used is very high, as the data were obtained directly from the archives of the esteemed American Board of Surgery. The study could easily be repeated. The sorting methods used in this study were similar to those used in other studies. This is the first-known long-term study of established programs presented in a ranked format. The multiple regression analysis is methodologically much more generalizable. This study has essentially performed a needs-assessment of program performance on the crucial ABS examinations. This study, as aforementioned, can certainly elicit a competitive spirit, as well as a collaborative spirit.

This study does lay a framework from which futures studies can be done. A very interesting next step could potentially be an interview or survey-based study of the first decile programs and tenth decile programs, to see if there are any institution-specific curricula, incentives, or programs that are utilized for examinee motivation and associated with ABS examination outcomes. It also could potentially be very interesting to find other programmatic variables that could be important with regard to this very high-stakes set of ABS examinations.

Conclusions

Overall, there is a wide variation of performance with regard to the ABS EI by established residency programs. Programs in the first decile were statistically larger programs, with a higher percentage of military designated institutions, were more likely to be in the West, and had a lower percentage of IMGs. These findings are also evident across the entire spectrum of program performance. A valuable resource in the form of a report card/leaderboard for general surgery applicants as well as programs has been developed as a result of this study. This study essentially performed a national needs-assessment, and the results could potentially evoke a competitive spirt of the “gamification” of ABS examination performance, as well as a collaborative spirit with regard to the high-stakes ABS examinations.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

None.

References

1. General Surgery Qualifying Examination. *American Board of Surgery*; 2018. Available through: http://www.absurgery.org/default.jsp?certgsqe. Accessed June 19, 2017.
2. General Surgery Certifying Examination. *American Board of Surgery*; 2018. Available through: http://www.absurgery.org/default.jsp?certcehome. Accessed June 19, 2017.
3. Shellito JL, Osland JS, Helmer SD, Chang FC. American Board of Surgery examinations: can we identify surgery residency applicants and residents who will pass the examinations on the first attempt? *Am J Surg*. 2010;199:216–222.
4. Falcone JL, Charles AG. Military and academic programs outperform community programs on the American Board of Surgery Examinations. *J Surg Educ*. 2013;70(5):613–617.
5. de Virgilio C, Yaghoubian A, Kaji A, et al. Predicting performance on the American Board of Surgery qualifying and certifying examinations: a multi-institutional study. *Arch Surg*. 2010;145(9):852–856.
6. Guzman E, Babakhani A, Maker VK. Improving outcomes on the ABS Certifying Examination: can monthly mock orals do it? *J Surg Educ*. 2008;65(6):441–444.
7. Falcone JL, Hamad GG. The state of performance on the American Board of Surgery Qualifying Examination and Certifying Examination and the effect of residency program size on program pass rates. *Surgery*. 2012;151(5):639–642.
8. ACGME Program Requirements for Graduate Medical Education in General Surgery. *Accreditation Council for Graduate Medical Education (ACGME)*; 2017. Available through http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/440_general_surgery_2017-07-01.pdf?ver=2017-05-25-084853-043. Accessed November 11, 2017.
9. General Surgery Residency Pass Rates. American Board of Surgery. Available through http://www.absurgery.org/default.jsp?prog_passreport. Accessed June 19, 2017.
10. Falcone JL. A longitudinal study of the American Board of Surgery examination index: can residents pass both the qualifying examination and certifying examination the first time? *J Surg Educ*. 2014;71(4):455–458.
11. Falcone JL, Hamad GG. The American Board of Surgery Certifying Examination: a retrospective study of the decreasing pass rates and performance for first-time examinees. *J Surg Educ*. 2012;69(2):231–235.
12. Klingensmith ME, Jones AT, Smiley W, et al. Subscription to the Surgical Council on Resident Education web portal and qualifying examination performance. *J Am Coll Surg*. 2014;218(4):566–570.
13. Falcone JL, Rosen ME. The importance of residency program size and location on American Osteopathic Board of Surgery In-Training Examination outcomes. *J Surg Res*. 2013;184(1):61–65.
14. Al Fayyadh MJ, Rawlings JA, Willis RE, et al. The American Board of Surgery examinations: how are the Southwestern Surgical Congress programs performing compared to the rest of the United States? *Am J Surg*. 2016;212(6):1243–1247.
15. Falcone JL. Performance of southeastern surgical congress examinees on the American board of surgery examinations: an educational opportunity for regional collaboration. *Am Surg*. 2017;83(7):249–252.
16. Sullivan MC, Sue G, Bucholz E, et al. Effect of program type on the training experiences of 248 university, community, and US military-based general surgery residencies. *J Am Coll Surg*. 2012;214(1):53–60.
17. Fischer LE, Snyder M, Sullivan SA, et al. Evaluating the effectiveness of a mock oral educational program. *J Surg Res*. 2016;205(2):305–311.
18. Fingeret AL, Arnell T, McNelis J, et al. Sequential participation in a multi-institutional mock oral examination is associated with improved American board of surgery certifying examination first-time pass rate. *J Surg Educ*. 2016;73(6):e95–e103.

19. General Surgery Residency. University of Missouri health School of Medicine. Available through: https://medicine.missouri.edu/download/MizzouSurgery_Brochure.pdf Accessed 18 June 2017.
20. USA General Surgery Residency Program. University of South Alabama School of Medicine. Available through: http://www.usahealthsystem.com/surgeryresidency.Accessed 18 June 2017.
21. General Surgery Residency overview. Tufts Medical center. Available through: https://www.tuftsmedicalcenter.org/patient-care-services/Departments-and-Services/Surgery/Training-Education/General-Surgery-Residency/Overview.aspx Accessed 18 June 2017.
22. *Program Summary of Performance on ABS Examinations: 2001-2006*. Philadelphia, PA: American Board of Surgery; 2007.
23. Summary of Performance on ABS Examinations. *2006-2011. First-taker Examination Pass Rates General Surgery Qualifying and Certifying Examinations (QE & CE)*. Philadelphia, PA: American Board of Surgery; 2012.
24. Program Summary of Performance on ABS Examinations. *2011-2016. First-taker Examination Pass Rates General Surgery Qualifying and Certifying Examinations (QE & CE)*. Philadelphia, PA: American Board of Surgery; 2017.
25. *Fellowship and Residency Electronic Interactive Database Access FREIDA Online, the AMA Residency & Fellowship Database*. American Medical Association; 2018. Available through https://www.ama-assn.org/life-career/search-freida-online-ama-residency-fellowship-database. Accessed June 19, 2017.
26. Charles AG, Walker EG, Poley ST, et al. Increasing the number of trainees in general surgery residencies: is there capacity? *Acad Med*. 2011;86(5):599–604.
27. Census Regions and divisions of the United States. United States Census Bureau. Available through: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf Accessed 18 June 2017.
28. Enter DH, Lee R, Fann JI, et al. "Top Gun" competition: motivation and practice narrows the technical skill gap among new cardiothoracic surgery residents. *Ann Thorac Surg*. 2015;99(3):870–875.
29. Mokadam NA, Lee R, Vaporciyan AA, et al. Gamification in thoracic surgical education: using competition to fuel performance. *J Thorac Cardiovasc Surg*. 2015;150(5):1052–1058.
30. ABS News. *American Board of Surgery*; 2013. Available through http://www.absurgery.org/xfer/newsletter2013.pdf. Accessed December 5, 2017.
31. Malangoni MA, Jones AT, Rubright J, et al. Delay in taking the American Board of Surgery qualifying examination affects examination performance. *Surgery*. 2012;152(4):738–743.