Mexican ENARM: performance comparison of public vs. private medical schools, geographic and socioeconomic regions

Daniela Catherine Hernández-Gálvez, MD,(1) Ernesto Roldán-Valadez, MSc, DSc.(1,2)

Hernández-Gálvez DC, Roldán-Valadez E. Mexican ENARM: performance comparison of public vs. private medical schools, geographic and socioeconomic regions. Salud Publica Mex. 2019;61:637-647.
https://doi.org/10.21149/10078

Abstract

Objectives. This study aimed to compare the performance in the National Assessment for Applicants for Medical Residency (ENARM in spanish) of private versus public medical schools, geographic regions and socioeconomic levels by using three different statistical methods (summary measurements, the rate of change and the area under the receiver operator characteristics [AUROC]). These methods have not been previously used for the ENARM; however, some variations of the summary measurements have been reported in some USA assessments of medical school graduates. **Materials and methods.** Cross-sectional study based on historical data (2001-2017). We use summary measures and colour-filled map. The statistical analysis included Mann-Whitney U, Kruskal-Wallis, Spearman correlation coefficient (Rs), and linear regression. **Results.** A total of 113 medical schools were included in our analysis; 60 were public and 53 private. We found difference in the median of total scores for type of schools, MD= 54.07 vs. MD= 57.36, p= 0.011. There were also significant differences among geographic and socioeconomic regions (p<0.05). **Conclusions.** Differences exist in the total scores and percentage of selected test-takers between type of schools, geographic and socioeconomic regions. Higher scores are prevalent in the Northeast and Norwest

Hernández-Gálvez DC, Roldán-Valadez E. ENARM: comparación del desempeño de escuelas de medicina públicas y privadas, regiones geográficas y socioeconómicas. Salud Publica Mex. 2019;61:637-647.
https://doi.org/10.21149/10078

Resumen

Objetivo. Comparar el desempeño en el Examen Nacional de Aspirantes a Residencias Médicas (ENARM) de escuelas de medicina privadas y públicas, regiones geográficas y niveles socioeconómicos mediante el uso de tres métodos estadísticos diferentes (medidas de resumen, tasa de cambio y el área bajo las características del operador receptor [AUROC en inglés]). Estos métodos no han sido utilizados previamente para el ENARM; sin embargo, se han informado algunas variaciones de las mediciones de resumen en algunas evaluaciones de graduados de medicina de Estados Unidos. **Material y métodos.** Estudio transversal basado en datos históricos (2001-2017). Se usaron medidas de resumen y un mapa lleno de color. El análisis estadístico incluyó Mann Whitney U, Kruskal-Wallis y coeficiente de correlación de Spearman (Rs). **Resultados.** Se incluyeron 113 escuelas de medicina en el análisis; 60 eran públicas y 53 privadas. Se encontraron diferencias en la mediana de las puntuaciones totales para el tipo de escuelas, MD= 54.07 vs. MD= 57.36, p= 0.011. También hubo diferencias significativas entre las regiones geográficas y socioeconómicas (p<0.05). **Conclusiones.** Existen diferencias en los puntajes totales y el porcentaje de examinados seleccionados entre el tipo de escuelas, regiones geográficas y socioeconómicas. Las puntuaciones más altas

(1) Dirección de Investigación, Hospital General de México Dr. Eduardo Liceaga. Mexico City, Mexico.
(2) Department of Radiology, IM Sechenov First Moscow State Medical Univeristy, Sechenov University. Moscow, Russia.

Received on: September 29, 2018 • **Accepted on:** April 10, 2019
Corresponding author: DSc. Ernesto Roldán-Valadez. Hospital General de México Dr. Eduardo Liceaga. Dr. Balmis 148, col. Doctores, Del. Cuauhtemoc. 06726 Mexico City, Mexico.
E-mail: ernest.roldan@usa.net

regions. Additional research is required to identify factors that contribute to these differences. Unsuspected differences in examination scores can be unveiled using summary measures.

Keywords: academic training; education; examination questions; medical schools; medical speciality

prevalecen en las regiones noreste y noroeste. Se requieren investigaciones adicionales para identificar los factores que contribuyen a estas diferencias. Las diferencias insospechadas en los puntajes de los exámenes se pueden revelar usando medidas de resumen.

Palabras clave: formación académica; educación; preguntas de examen; escuelas de medicina; especialidad

In the current competency-based medical education model, the comparison of student's performance versus their peer's with grades is the rule, not only for persons but also for institutions.[1] In Mexico, the score that a general practitioner (GP) obtains in the National Evaluation for Medical Residency Applicants (*Examen Nacional de Aspirantes a Residencias Médicas*, ENARM) is the entrance door to a specialization course endorsed by a Mexican University.[2,3] Compared with the United States Medical Licensing Exam (USMLE), an assessment required to consider granting an unrestricted medical license to a physician,[4,5] the ENARM is a one-day-only exam, that uses multiple-choice questions and computerized patient cases to assess examinee's knowledge related to foundational science concepts applicable to medical and scientific theories to clinical medicine; details concerning the logistics' of the exam has been published previously.[6] The ENARM and the percentages of selection of their graduates are indicators of efficiency and reason of prestige and even of propaganda among the aspirants to study medicine.[7] Residency program directors want to predict which students will be most successful in their program as efficiently and cost-effectively as possible;[8] however, there is a paucity of consistent predictors of residency performance that have been well documented in the literature.[9] Although there are tools able to perform a standardised, comprehensive, summative assessment of a medical student's performance such as the Medical Student Performance Evaluation (MSPE);[8] to the best of our knowledge, there is no consensus of the best statistical methods to compare medical schools performance worldwide. Some examples of the different methodology used to compared medical schools are presented below.

In 2011, Baker at Harvard Medical School described a summative score system that normalised resident evaluations by using Z-score transformations,[10] however his methodology was aimed to evaluate residents performance no medical schools. A study by Go and colleagues[11] in 2012, compared the USMLE scores, demographics, medical school type and location, among American medical graduates and USA international medical graduates; but because it was a cross-sectional study, the authors only compared groups using unpaired Student t-test for mean data. The same year, the use of maps to represent Medical schools of origin of 64 preliminary residents recruited to the UC-Denver Department of Surgery was presented by Montero and colleagues.[12] Some medical schools from USA reported in 2016, the use of linear regression, with estimation of the β coefficient and 95% confidence intervals (CI) to compare grades in the field of obstetrics and gynecology.[13] A study from the USA in 2016 reported specific evidence regarding the association between public and private medical school curricular settings and the proportion of medical students matching into family medicine careers.[14] Another study in 2015 showed that private medical school alumni had higher attrition rates than public medical school alumni for the neurosurgery residence.[15] In 2017, the type of medical school attended, public vs. private, was considered relevant in countries like the USA where tuition loans of over 200 000 dollars are not uncommon for students because medical schools are often private institutions, and the topic is also essential for universities where the vast majority of medical schools are tax-funded and, for this reason, free of tuition like Germany.[16] In 2018, a Z-Score System for Normalizing Residency Evaluations variance components was estimated using analysis of variance (Anova).[17] In Europe, a recent study in 2019 of the differences in subjective well-being over 20 years at two Norwegian medical schools analysed the continuous variables of scores using t-tests and one-way Anova.[18] A four year comparative study published in 2019 between students that learned anatomy at the special master's program in physiology at the University of Cincinnati compared the average scores of written examination questions between groups for each year using analysis of covariance (Ancova).[19] In Mexico, a recent study published in 2017 about the ENARM described the number of Mexican test-takers and accepted GPs belonging to each Mexican medical school registered in the ENARM.[2] However, the authors did not include a method section in their study, they reported only descriptive statistic without calculations of statistical significance among groups. These authors did not comment about their approach for the time-series graphs they presented; their study included more than 100 medical schools during the period 2001-2016.

Because there have not been studies that compare the performance of private versus public schools; neither they have explored the significant differences in the performance of medical schools based on their geographic regions and socioeconomic levels in the Mexican states;[2] we aimed to compare the performance in the ENARM of private versus public medical schools, geographic regions of Mexican states, and socioeconomic levels by using three different statistical methods (summary measurements, rate of change, and the area under the receiver operator characteristics [AUROC]). Although the proposed methods have not been previously reported for the ENARM assessment, we believe they can unveil unsuspected differences in comparing the scores of medical schools. We hypothesised that private medical schools reach higher scores than public ones and that Mexican states in the north and central parts of the country achieve better performance than states from central and south regions.

Materials and methods

Study design and data acquisition

This study was conducted in July 2018 at the Directorate of Research of the *Hospital General de Mexico Dr. Eduardo Liceaga* in Mexico City. It is a cross-sectional study and used public, historical data that did not require the approval of an Institutional Review Board. We based our analyses in the annual public report of the ENARM in 17 years from 2001 to 2017. The reports were issued by the Interinstitutional Commission for Human Resources Training for Health (*Comisión Interinstitucional para la Formación de Recursos Humanos para la Salud,* CIFRHS), a department of the Undersecretariat of Innovation and Quality of the Mexican Ministry of Health.[20] The reports contain quantitative information of the academic performance of medical schools whose graduate students took the ENARM, but it does not contain information of individual test-takers; a PDF file containing the annual reports is available as a Supplemental Digital Content (the datasets analyzed during the current study are available from the corresponding author on reasonable request).

Reported variables

For each year we reported the three variables that coincidentally appear in the CIFRHS report: number of applicants (GP) who took the exam, the number who were selected, and the total score (as a percentage) in the exam. Results were recorded from 113 medical schools. We calculated a fourth variable named the percentage of selected for each medical school dividing number who were chosen by the number of applicants and multiplying by 100.

We classified medical schools as public or private. The location of each medical schools in each Mexican state was coded by socioeconomic regions from one to seven. We used a classification system proposed by the Mexican National Institute of Statistics and Geography (INEGI, by its acronym in spanish).[21] In ascending magnitude, the numbers show the last available information from the Mexican National Population Census encompassing education, employment, occupation, housing, and health; higher numbers represent the Mexican states with the best levels of well-being in Mexico.

We also sorted the medical schools by their geographic regions (Mexican states where the Universities endorsing Medical schools were located) using the InterPlanetary File System (IPFS), a protocol and network designed to create a content-addressable, peer-to-peer method of storing and sharing hypermedia in a distributed file system.[22] This system classifies the Mexican states in eight regions: 1) Northwest (Baja California, Baja California Sur, Chihuahua, Durango, Sinaloa, Sonora); 2) Northeast (Coahuila, Nuevo León, Tamaulipas); 3) North-central (Aguascalientes, Guanajuato, Querétaro, San Luis Potosí, Zacatecas); 4) South-central (Mexico City, Estado de México, Morelos); 5) West (Colima, Jalisco, Michoacán, Nayarit); 6) East (Hidalgo, Puebla, Tlaxcala, Veracruz); 7) Southeast (Campeche, Quintana Roo, Tabasco, Yucatán); 8) Southwest (Chiapas, Guerrero, Oaxaca). In ascending order, one and two represent regions of Mexican states in the north and east west regions; three to six represent states of the central regions; highest numbers, seven and eight, depict states in the south and west east regions.

Statistical analysis

The independent variables were the four variables derived from the ENARM (number of applicants, the number of selected, total score, and percentage of selected applicants). Most of the schools had up to 17 measurements representing the years of examinations (2001-2017), we performed an analysis of serial measurements following the recommendations by Mathews and colleagues[23] to obtain one measurement representing each of the selected variables from our universe of 113 medical schools.

We performed Kolmogorov-Smirnoff and Shapiro-Wilk tests to know if the variables had a normal or non-normal distribution, and because we found that all (variables depicted non-normal distributions) we decided to use distribution-free non-parametric tests. We obtained the median, minimum and maximum of each variable. Also, we calculated the slopes using a regression model to test the hypothesis of a difference in the rate of change and the area under the curve (AUC) for each school during the 17 annual examinations. We applied the summary measures method to collapse repeated measures into a single observation per group so that we could avoid complex repeated measured analyses; a detailed description of the foundations of this methods and formula used to calculate the AUC is explained in the article by Mathews and colleagues.[23]

We compared the differences between private and public schools using the Mann-Whitney U test, and the differences among geographic regions and socioeconomic levels were analysed using the Kruskal-Wallis test. The associations between categorical variables, type of school vs. geographic, and kind of school vs. socioeconomic level were evaluated with the Spearman correlation coefficient (R_S). We also used R_S and scatterplots with regression analysis to assess the correlation between the continuous variables: number of applicants, number of accepted, the percentage of accepted, and total score.

Data visualisation techniques

We generated maps of the Mexican states that represented the geographic regions or socioeconomic levels coloured by number of applicants, number of selected, percentage of selected, and total score; it was used Tableau software (version 10.5.3, Seattle, Washington, USA); scatter plots were designed with JMP Pro from SAS Institute Inc. (version 13.2.1, Cary, NC, USA); statistical analyses were carried out using and the IBM SPSS Statistics (software version 25.0.0.1 IBM Corporation; Armonk, NY, USA). Statistical significance considered a p-value <0.05 (two-tailed).

Results

Demographic data and correlations between variables

A total of 113 medical school were included in our analysis. There were not differences in the percentages of participant schools sorted by each variable: type of school c (1)= 0.434, p= 0.510; geographic region c (7)= 8.841, p= 0.264; and socioeconomic level c (6)= 7.239, p= 0.299; samples sizes for each category are presented in table I.

We found a poor not significant correlation between the type of school with the geographic region R_S= - 0.042, p= 0.657; and between the type of school with the socioeconomic level R_S= 0.015, p= 0.876. There was a significant negative correlation between geographic region and the socioeconomic level R_S= - 0.325, p<0.001.

The selected continuos variables revealed a moderate positive correlation of Selected«Applicants R_S= 0.564, p<0.001; and a poor but positive correlation of Selected↔Total Score R_S= 0.149, p<0.001. There were three statistically significant and negative correlations between Percentage of Selected↔Selected R_S= -0.383, p<0.001; Percentage of Selected↔total acore R_S= - 0.403, p<0.001; and Percentage of Selected«Applicants R_S= -0.061, p= 0.025. Figure 1 depicts the correlation matrices and histograms between selected variables of ENARM.

Number of applicants and number of selected applicants

The socioeconomic level showed differences in the median of applicants and the maximum number of applicants. Geographics regions presented significant differences in the median, minimum, and maximum numbers of selected applicants; also there were significant differences in the socioeconomic level evinced by the median, minimum, and maximum number of selected applicants.

Table I shows the summary measures and statistical significances for the number of applicants and the number of selected applicants compared by type of school, geographic regions and socioeconomic level. Figure 2 shows the times series of the number of applicants and the number of applicants chosen grouped by type of school (public vs. private). Figure 3 shows the colour maps indicating the median values for the four selected variables grouped by regions (3A) and socioeconomic level (3B).

Percentage of selected applicants and the total score

We found that the median value for geographic regions was significantly different. There were also differences in the minimum and rate of change for the percentage of selected applicants. The median of the socioeconomic level was also different.

We did find a significant difference for the total score in the median between public and private schools, and the minimum and maximum scores for the geographic region and socioeconomic.

Table II shows the summary measures and statistical significances for the percentage of selected applicants and total score compared by type of school, geographic

Table I
Summary measures and statistical significances for the number of applicants and selected applicants*

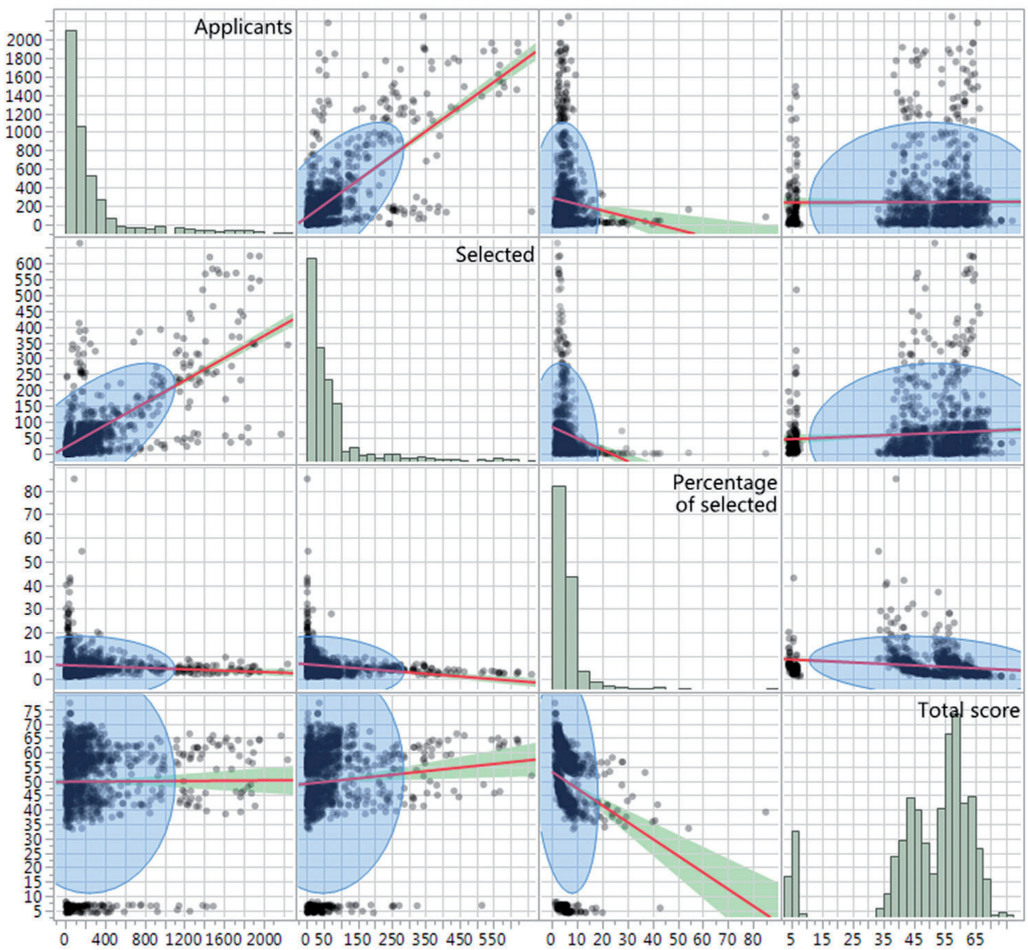
Variables	N (%)	Applicants					Selected applicants				
		Median	*Minimum*	*Maximum*	*AUC*	*Rate of change*	*Median*	*Minimum*	*Maximum*	*AUC*	*Rate of change*
Type of school	Public 60 (53.10)	93.0	16.0	175.0	18 584.00	5.57	27.00	3.00	41.50	2 991.75	2.00
	Private 53 (46.90)	92.0	13.0	224.0	30 369.50	10.59	23.00	2.00	45.00	2 464.25	1.89
	p-value	0.698	0.156	0.736	0.247	0.245	0.945	0.498	0.778	0.297	0.725
Geographic regions	Northwest 18 (15.93)	169.50	18.00	303.00	39 294.00	7.03	64.00	4.00	90.50	3 240.00	3.81
	Northeast 17 (15.04)	125.00	12.00	266.00	47 010.50	11.06	27.00	2.00	45.00	1 120.00	1.71
	West 17 (15.04)	137.00	18.00	256.00	11 796.50	7.66	27.00	3.00	47.00	1 375.75	1.56
	East 14 (12.39)	36.75	12.50	78.50	10 195.00	9.45	7.00	1.00	15.00	3 250.50	1.75
	North-central 10 (8.86)	119.00	13.00	263.50	64 780.00	10.18	49.75	4.00	93.00	2 915.50	3.20
	South-central 19 (16.81)	154.00	16.00	194.00	12 601.50	5.35	47.00	5.00	71.00	3 080.00	1.72
	Southeast 9 (7.96)	58.00	11.00	87.00	1 310.75	18.00	7.00	2.00	15.00	303.00	1.00
	Southwest 9 (7.96)	26.00	15.00	35.00	7 084.00	1.91	4.00	3.00	4.00	2 508.00	1.53
	p-value	0.010‡	0.629	0.008‡	0.406	0.661	0.001‡	0.040‡	0.001‡	0.260	0.359
Socioeconomic level	Level 1 9 (7.96)	26.00	15.00	35.00	7 084.00	1.92	4.00	3.00	4.00	2 508.00	1.53
	Level 2 20 (17.70)	51.25	14.50	104.50	10 792.75	9.47	9.00	1.50	17.50	3 250.50	1.40
	Level 3 15 (13.27)	146.00	13.00	283.00	64 780.00	6.34	53.00	3.00	91.00	3 240.00	3.49
	Level 4 15 (13.27)	58.00	11.00	72.00	-1 886.00	13.01	7.00	1.00	17.00	877.00	1.90
	Level 5 21 (18.5)	99.00	16.00	231.00	23 787.50	9.47	28.00	3.00	56.00	2 470.00	3.11
	Level 6 20 (17.70)	170.50	13.50	289.00	26 082.25	8.54	31.50	3.00	67.50	1 736.00	1.84
	Level 7 13 (11.50)	272.00	21.00	373.00	13 395.75	4.67	66.00	7.00	96.00	5 008.00	1.64
	p-value	< 0.01‡	0.591	< 0.01‡	0.433	0.354	< .001‡	0.008‡	< .001‡	0.499	0.634

AUC: area under the curve.
* At the time this article was written in July 2018, both authors had affiliation at the Directorate of Research of the *Hospital General de México Dr. Eduardo Liceaga* in Mexico City
‡ statistical significance (*p*-value <0.050).

regions and socioeconomic level. Figure 2 shows the times series of the portion of selected applicants and total score grouped by type of school (public vs. private). Figure 3 shows the colour maps indicating the median values for the four chosen variables grouped by regions (3A) and socioeconomic level (3B).

Discussion

In this study, we were able to prove the hypothesis posed in the introduction that private medical schools reach significantly higher scores than those of public schools and that Mexican states in the north parts of the country showed better performance than central and south regions. The clinical relevance of our study is supported by the comprehensive statistical analysis we performed and the tables containing summary measures that we share with our readers. The logistics and transparency of the ENARM exam have been previously described.[6] In a recent report about ENARM published in 2017,[2] the authors compared the performance of public versus private school. However, they presented only bar graphs with an independent comparison for each Mexican state without a regional or global analysis. In the same study, the percentage of selected students was presented for 31 out of 32 Mexican states, but there were no mentions of the statistical significances at a global level or region subgroups. The study used mean values and our analysis revealed that, except for the total score variable, all selected variables showed a non-normal distribution, which requires the report of results use medians and parametric-free tests or transformed variables.



* At the time this article was written in July 2018, both authors had affiliation at the Directorate of Research of the *Hospital General de México Dr. Eduardo Liceaga* in Mexico City.
ENARM: *Examen Nacional de Aspirantes a Residencias Médicas.*

FIGURE 1. CORRELATION MATRICES AND HISTOGRAMS BETWEEN CONTINUOUS VARIABLES OF ENARM*

Table II

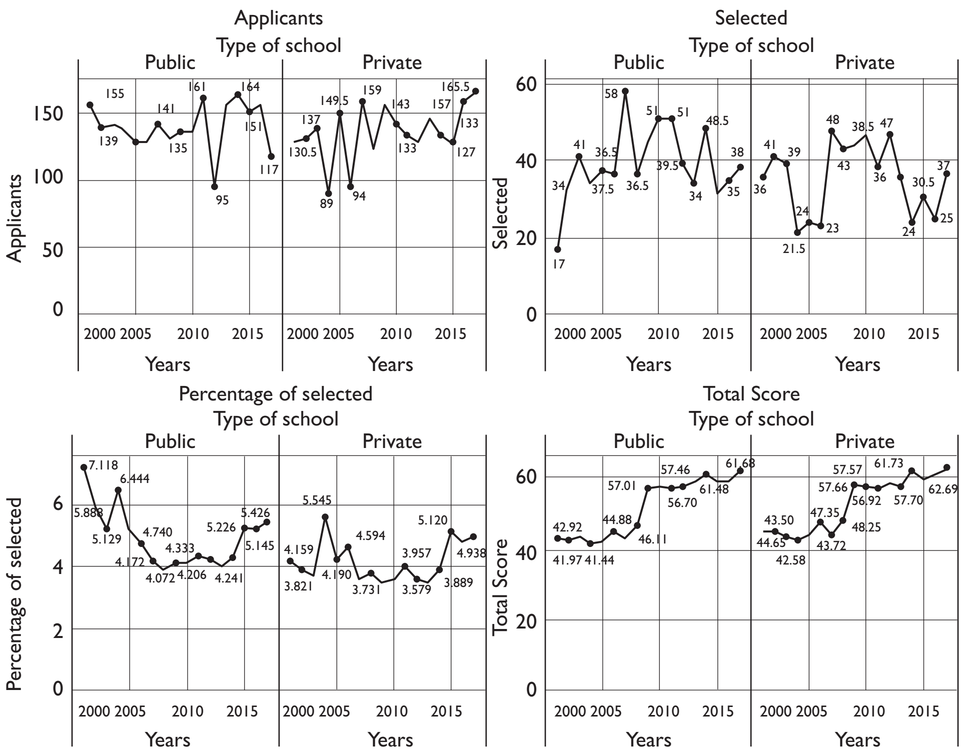
Summary measures and statistical significances for the number of percentage of selected applicants and total score*

	N (%)	Percentage of selected applicants					Total Score				
		Median	*Minimum*	*Maximum*	*AUC*	*Rate of change*	*Median*	*Minimum*	*Maximum*	*AUC*	*Rate of change*
Type of school	Public 60 (53.10)	5.26	3.647	9.83	-0.07	-0.07	54.07	6.60	61.89	1.40	1.40
	Private 53 (46.90)	4.55	3.05	8.66	-0.07	-0.07	57.36	35.63	63.16	1.45	1.45
	p-value	0.066	0.076	0.635	0.593	0.593	0.011‡	0.363	0.050‡	0.868	0.868
Geographic regions	Northwest 18 (15.93)	4.21	2.80	7.39	-0.07	-0.07	58.51	41.79	64.16	1.40	1.40
	Northeast 17 (15.04)	4.51	3.00	8.39	-0.01	-0.01	57.21	6.70	63.48	1.42	1.42
	West 17 (15.04)	5.40	3.64	11.75	-0.30	-0.30	54.14	4.75	61.71	1.59	1.59
	East 14 (12.39)	7.10	4.37	14.00	-0.28	-0.28	54.33	42.91	58.65	1.77	1.77
	North-central 10 (8.86)	4.26	3.08	8.38	-0.08	-0.08	58.47	21.78	63.82	1.46	1.46
	South-central 19 (16.81)	4.57	3.22	8.00	-0.07	-0.07	54.65	5.77	62.91	1.31	1.31
	Southeast 9 (7.96)	6.57	4.04	13.50	0.13	0.13	55.28	52.99	59.35	1.50	1.50
	Southwest 9 (7.96)	3.40	2.88	13.07	0.71	0.71	56.43	51.76	62.84	1.21	1.21
	p-value	0.005‡	0.036‡	0.073	0.026‡	0.026‡	0.406	0.044‡	0.006‡	0.910	0.910
Socioeconomic level	Level 1 9 (7.96)	3.40	2.88	13.07	0.71	0.71	56.43	51.76	62.84	1.21	1.21
	Level 2 20 (17.70)	6.55	4.21	10.83	0.05	0.05	54.33	42.91	58.89	1.77	1.77
	Level 3 15 (13.27)	4.41	3.28	8.94	-0.18	-0.18	52.27	6.60	62.48	1.43	1.43
	Level 4 15 (13.27)	5.86	4.00	12.50	0.03	0.03	55.28	39.80	59.73	1.18	1.18
	Level 5 21 (18.5)	4.25	2.87	9.36	-0.01	-0.01	57.21	6.85	63.48	1.39	1.39
	Level 6 20 (17.70)	4.71	3.13	10.12	-0.10	-0.10	55.34	6.36	62.55	1.56	1.56
	Level 7 13 (11.50)	3.76	2.66	7.91	-0.08	-0.08	52.45	5.72	63.70	880.97	1.48
	p-value	0.016‡	0.075	0.151	0.051	0.051	0.539	0.044‡	0.012‡	0.794	0.664

AUC: area under the curve.

* At the time this article was written in July 2018, both authors had affiliation at the Directorate of Research of the *Hospital General de México Dr. Eduardo Liceaga* in Mexico City.

‡ Statistical significance (p-value <0.050).



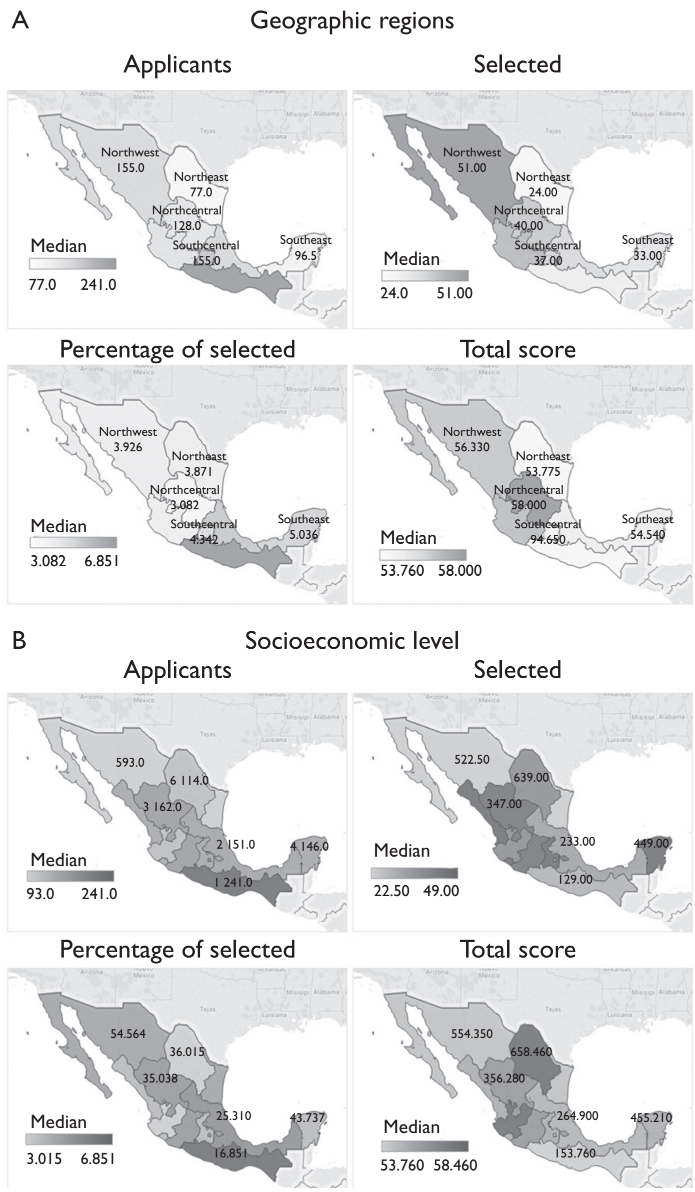
*At the time this article was written in July 2018, both authors had affiliation at the Directorate of Research of the *Hospital General de México Dr. Eduardo Liceaga* in Mexico City

Figure 2. Times series graph of the selected variables grouped by type of school (public vs. private)*

We found significant differences in the academic performance of medical schools when grouped by geographic regions and socioeconomic levels; our results mirror the heterogeneity in the quality of education and unequal distribution of economic resources at each Mexican state. Regarding socioeconomic levels, the interpretation of a significant negative correlation between the numeric code assigned to geographic regions and the code assigned to socioeconomic levels meant that the states in the Northwest; and Northeast present the highest socioeconomic levels. Our findings relating better scores to specific socioeconomic levels might reflect the social determinants of the Mexican Health system. For example, the Total health expenditure of Mexico represented 6.2% of the gross domestic product (GDP) in 2012; however, in 2015 it declined to 5.6%.[24]

The main strength of our study is the statistical comparisons using different methods between public vs. private medical schools. We found a proportion of selected students from public vs. private school of approximately 4.9:1.[2] Our findings proved a significantly higher score over 17 years of a private school in the total score measure. Medical schools with the best academic performance registered percentages of total score around 70%; this number is below the score reported by the USMLE of 94%.[25] The observation of low scores at the ENARM had been previously addressed in a study by de la Garza-Aguilar.[3] However, his study did not compare the performance of public vs. private medical schools.

To the best of our knowledge, no other study about the Mexican ENARM exam has performed a comparison of results using the currently accepted approaches for the analysis of longitudinal data as our study did. When there are only two measurements per subject, the data may be analysed using a fair summary value (difference, ratio, etc.) of paired measurements or use regression such as linear or logistic regression adjusting for baseline variable. Nonetheless, the analysis becomes more complicated when more than two measurements are involved (in our study we have up to 17 measures per medical school). The comparison of multiple means might have a detrimental effect of loss of power by estimating many parameters, besides the implication of the multiplicity effect. An option is the use of a sophisticated analysis named mixed-effect models to account for within-patient correlation.[26] But failing to consider within-subject correlation can impact



*At the time this article was written in July 2018, both authors had affiliation at the Directorate of Research of the *Hospital General de México Dr. Eduardo Liceaga* in Mexico City

Figures 3. Map distributions for applicants, selected applicants, the percentage of selected applicants, and total scores sorted by geographic regions and socioeconomic level*

on the standard error of the parameter estimates resulting in affecting *p*-value or power of analysis. The summary measure method we used allowed us to condense medical schools' longitudinal observations (repeated measures) for each selected variable (mean, median, minimum and maximum value, estimated slope and AUC) to a single number per institution and to compare these univariate measures across groups (type of school, geographic regions and socioeconomic levels). This method eliminated the within-subject repeated measurements and allowed for a straightforward comparison of groups using standard statistical hypothesis tests (Mann-Whitney U and Kruskal-Wallis in our study) that assume independent measurements.[27] The median summary measure was the variable able to identify significant differences between groups; our explanation of why the comparison of slopes of growth for the four selected variables did not reveal significant differences among medical schools is that medical schools' scores behave with a gradual, constant growth rate, although each institution was depicting different performance levels. Similar behaviour can be inferred of the AUC values, they did not show a rise to a peak and then return to baseline (as observed for example in pharmacokinetic data), then AUC comparisons were not statistically different. We did not use the historically fundamental repeated-measures analysis of variance,[28] because we identified the majority of our variables depicted non-normal distributions. Also, we wanted to present a more straightforward comparison of groups that could be replicated by other authors. The methodology to perform summary statistics to condense the repeated measures has been used since the last three decades; it is gaining trend in clinical medicine, for example in the analysis of clinical trials; [29] readers can find a brief review on this topic in the article published by Schober and colleagues. [27]

Several limitations need to be acknowledged for this study. We acknowledge that for this article we focused the discussion in the statistical methods used with a second place for some of the findings of the ENARM. We intentionally adopted this approach considering the scarce public information available would not allow us to present a solid discussion of the real causes of differences between public and private schools. Also, we did not have access to reliable individual information from each medical school. With the ENARM, the Mexican Secretariat of Health get to select the best candidates each year with reasonable confidence, but a number much higher than the accepted is left without entering a medical speciality; we did not analyse those numbers as this topic was out of the scope of our study. Also, we did not comment the context regarding the offer and demand of Mexican physicians per number of inhabitants; in 2015, Mexico had 2.2 physicians per 1 000 population, including professionals in the private sector, these numbers represent a significant disparity in the distribution of human health resources in the country. The same year, the USA reported a ratio of 3.1 physicians per 1 000 inhabitants.[24] Although the number and needs of the medical specialists are not found fully identified, the number of existing doctors and possible training at the current rate will be insufficient for the needs of the country.[2] We did not get deep in the analysis of some schools per city, but it is clear that there is no homogeneous distribution, there is at least one medical school per Mexican state, but some of them shows the highest density. For example, in the metropolitan area of Mexico City, there are 18 medical schools, 12 in the State of Jalisco, and 10 in the State of Tamaulipas. Until 2014, only 84 Mexican schools were included in the International Medical Education Directory.[30]

Our assessment was not focused on examinee's performance, but the performance of 113 medical schools over 17 years. Similarly, subgroup performance differences considering age, gender, a race of test takers, English as a second language, and the influence of self-selected specialities was not performed because this information was not publicly available; neither we include international medical graduates because no information regarding our selected variables was available for this group. Detractors of the summary measures method could argue that a significant drawback is that a substantial amount of data is lost when all individual measurements are condensed into a single number. The main advantage of this approach is that it is straightforward but yet provides *valid results*, and clinicians with a limited statistical background easily understand the technique and results.[27]

In conclusion, there must be a balance between the training of general practitioners and the opportunities they have to receive education as medical specialists. Medical schools have different ways of supporting students through their curricula; evaluation of accrued scores at ENARM will help medical schools improve their understanding of how they contribute to the examination performance of their graduates. Identification of insufficiently trained examinees will avoid them from practising medicine, at least until appropriate remediation is undertaken. A self-assessment of their academic performance at ENARM will allow medical schools to identify candidate examinees who may need additional help with ENARM preparation. Wise use of ENARM scores to screen applicants for interviews could reduce potential bias against certain socioeconomic and ethnic minorities. The existing heterogeneity in academic performance across demographic groups could be a sign that additional work is required to identify factors contributing to geographic and socioeconomic differences among medical schools. The implications of this study deserve a sequel; future studies should explore if ENARM scores can be predictive of performance on subsequent assessments of speciality in-training and certification examinations.

Acknowledgments

The authors thank Ma. Carolina Silva Oseguera, President of *Organización del Voluntariado del Hospital General de Mexico Dr. Eduardo Liceaga* for her donation of a workstation that allowed us to produce the data visualization and statistical analysis for this article.

Authors thank Sergio Nigenda and Michelle Cahue Vargas, for their help extracting data and assembling the database of ENARM.

Declaration of conflict of interests. The authors declare that they have no conflict of interests.

References

1. Pereira AG, Woods M, Olson APJ, van den Hoogenhof S, Duffy BL, Englander R. Criterion-based assessment in a norm-based world: How can we move past grades? Acad Med. 2018;93(4):560-4. https://doi.org/10.1097/ACM.0000000000001939
2. Ramiro M, Cruz E, Zerón-Gutiérrez L, Arévalo-Vargas A. The ENARM and the schools and faculties of medicine. An analysis that nobody will like. Rev Med Inst Mex Seguro Soc. 2017;55(4):498-511.
3. de la Garza-Aguilar J. Reflections on the quality of medical school curricula in Mexico. Gac Med Mex. 2005;141(2):129-41.
4. Gauer JL, Jackson JB. Relationships of demographic variables to USMLE physician licensing exam scores: a statistical analysis on five years of medical student data. Adv Med Educ Pract. 2018;9:39-44. https://doi.org/10.2147/AMEP.S152684
5. Rubright JD, Jodoin M, Barone MA. Examining demographics, prior academic performance, and United States Medical Licensing Examination scores. Acad Med. 2018;94(3):364-70. https://doi.org/10.1097/ACM.0000000000002366
6. Barajas-Ochoa A, Ramos-Remus C. Equity, validity and reliability of the National Examination for Candidates to Medical Residences (ENARM): opportunities for improvement. Salud Publica Mex. 2017;59(5):501-2. https://doi.org/10.21149/8769
7. Blog ENARM. Mejores Escuelas y Facultades de Medicina - Ranking 2018. Mexico City: Foro ENARM, 2018 [cited May 2 2018]. Available from: https://comunidadenarm.blogspot.com
8. Andolsek KM. Improving the medical student performance evaluation to facilitate resident selection. Acad Med. 2016;91(11):1475-79. https://doi.org/10.1097/ACM.0000000000001386
9. Borowitz SM, Saulsbury FT, Wilson WG. Information collected during the residency match process does not predict clinical performance. Arch Pediatr Adolesc Med. 2000;154(3):256-60.
10. Baker K. Determining resident clinical performance: getting beyond the noise. Anesthesiology. 2011;115(4):862-78. https://doi.org/10.1097/ALN.0b013e318229a27d
11. Go PH, Klaassen Z, Chamberlain RS. An ERAS-based survey evaluating demographics, United States Medical Licensing Examination Performance, and research experience between American medical graduates and United States citizen international medical graduates: is the bar higher on the continent? J Surg Educ. 2012;69(2):143-8. https://doi.org/10.1016/j.jsurg.2011.07.013
12. Montero P, Powell R, Travis CM, Nehler MR. Selection, mentorship, and subsequent placement of preliminary residents without a designated categorical position in an academic general surgery residency program. J Surg Educ. 2012;69(6):785-91. https://doi.org/10.1016/j.jsurg.2012.04.009
13. Fay EE, Schiff MA, Mendiratta V, Benedetti TJ, Debiec K. Beyond the Ivory Tower: a comparison of grades across academic and community OB/GYN clerkship sites. Teach Learn Med. 2016;28(2):146-51. https://doi.org:/10.1080/10401334.2016.1146603
14. Wimsatt LA, Cooke JM, Biggs WS, Heidelbaugh JJ. Institution-Specific Factors Associated With Family Medicine Residency Match Rates. Teach Learn Med. 2016;28(3):269-78. https://doi.org/10.1080/10401334.2016.1159565
15. Lynch G, Nieto K, Puthenveettil S, Reyes M, Jureller M, Huang JG, *et al.* Attrition rates in neurosurgery residency: analysis of 1361 consecutive residents matched from 1990 to 1999. J Neurosurg. 2015;122(2):240-9. https://doi.org/10.3171/2014.10.JNS132436
16. Zavlin D, Jubbal KT, Noe JG, Gansbacher B. A comparison of medical education in Germany and the United States: from applying to medical school to the beginnings of residency. Ger Med Sci. 2017;15:Doc15. https://doi.org/10.3205/000256
17. Wanderer JP, de Oliveira Filho GR, Rothman BS, Sandberg WS, McEvoy MD. Implementation and Evaluation of the Z-Score System for Normalizing Residency Evaluations. Anesthesiology. 2018;128(1):144-58. https://doi.org/10.1097/ALN.0000000000001919
18. Sletta C, Tyssen R, Lovseth LT. Change in subjective well-being over 20 years at two Norwegian medical schools and factors linked to well-being today: a survey. BMC Med Educ. 2019;19(1):45. https://doi.org/10.1186/s12909-019-1476-3
19. Thompson AR, Marshall AM. Participation in dissection affects student performance on gross anatomy practical and written examinations: results of a four-year comparative study. Anat Sci Educ. 2019. https://doi.org/10.1002/ase.1859
20. Comisión Interinstitucional para la Formación de Recursos Humanos para la Salud. Reportes académicos. Médicos inscritos, sustentantes, seleccionados y resultados por escuela o facultad de medicina. Mexico City: CIFRHS, 2017 [cited 2018 Jul 22]. Available from: http://www.cifrhs.salud.gob.mx/site1/enarm/reportes_academicos.html
21. Instituto Nacional de Estadística y Geografía. Clasificación de Entidades Federativas. Mexico City: INEGI, 2018 [cited 2018 Jul 2018]. Available from: http://sc.inegi.gob.mx/niveles/index.jsp
22. Institucion Protocol Labs Inc. List of regions of México. San Francisco: IPFS InterPlanetary File System, 2016 [cited 2018 Jul 22]. Available from: https://ipfs.io/ipfs/QmXoypizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uco/wiki/List_of_regions_of_Mexico.html
23. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. BMJ. 1990;300(6719):230-5. https://doi.org/10.1136/bmj.300.6719.230
24. Pan American Health Organization. Health in the Americas, 2017 Edititon. Summary: Regional Outlook and Country Profiles. Washington DC: Pan American Health Organization, 2017.
25. Lehman EP, Guercio JR. The Step 2 Clinical Skills exam-a poor value proposition. N Engl J Med. 2013;368(10):889-91. https://doi.org/10.1056/NEJMp1213760
26. Roldan-Valadez E, Orbe-Arteaga U, Rios C. Eigenfactor score and alternative bibliometrics surpass the impact factor in a 2-years ahead annual-citation calculation: a linear mixed design model analysis of Radiology, Nuclear Medicine and Medical Imaging journals. Radiol Med. 2018;123(7):524-34. https://doi.org/10.1007/s11547-018-0870-y
27. Schober P, Vetter TR. Repeated measures designs and analysis of longitudinal data: If at first you do not succeed-try, try again. Anesth Analg. 2018;127(2):569-75. https://doi.org/10.1213/ANE.0000000000003511
28. Roldan-Valadez E, Rios-Piedra E, Favila R, Alcauter S, Rios C. Diffusion tensor imaging-derived measures of fractional anisotropy across the pyramidal tract are influenced by the cerebral hemisphere but not by gender in young healthy volunteers: a split-plot factorial analysis of variance. Chin Med J (Engl). 2012;125(12):2180-7.
29. Albert PS. Longitudinal data analysis (repeated measures) in clinical trials. Stat Med. 1999;18(13):1707-32.
30. Duvivier RJ, Boulet JR, Opalek A, van Zanten M, Norcini J. Overview of the world's medical schools: an update. Med Educ. 2014;48(9):860-9. https://doi.org/10.1111/medu.12499