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)

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.

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...
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. 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. Compared with the United States Medical Licensing Exam (USMLE), an assessment required to consider granting an unrestricted medical license to a physician, 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. 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. Residency program directors want to predict which students will be most successful in their program as efficiently and cost-effectively as possible; however, there is a paucity of consistent predictors of residency performance that have been well documented in the literature. 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); 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; however his methodology was aimed to evaluate residents performance no medical schools. A study by Go and colleagues 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 un-paired 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. 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. 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. Another study in 2015 showed that private medical school alumni had higher attrition rates than public medical school alumni for the neurosurgery residency. 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. In 2018, a Z-Score System for Normalizing Residency Evaluations variance components was estimated using analysis of variance (Anova). 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. 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). 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. 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 ap-

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

Palabras clave: formación académica; educación; preguntas de examen; escuelas de medicina; especialidad
proach 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; 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 México 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. 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). 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. 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 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.25
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 repre-
sented 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); statisti-
cal 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 socio-
economic 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 mod-
erate positive correlation of Selected \( \leftrightarrow \) Applicants
\( R_s = 0.564, p<0.001 \); and a poor but positive correlation
of Selected \( \leftrightarrow \) Total Score \( R_s = 0.149, p<0.001 \). There
were three statistically significant and negative correla-
tions between Percentage of Selected \( \leftrightarrow \) Selected
\( R_s = -0.383, p<0.001 \); Percentage of Selected \( \leftrightarrow \) total
score \( R_s = -0.403, p<0.001 \); and Percentage of Selected \( \leftrightarrow \) Applicants
\( R_s = -0.061, p= 0.025 \). Figure 1 depicts the correlation ma-
trices 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 sig-
nificant 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 num-
ber 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 geo-
graphic region and socioeconomic.

Table II shows the summary measures and statisti-
cal significances for the percentage of selected applicants
and total score compared by type of school, geographic
**Table 1**

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

<table>
<thead>
<tr>
<th>Variables</th>
<th>Applicants</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Selected applicants</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median</td>
<td>Minimum</td>
<td>Maximum</td>
<td>AUC</td>
<td>Rate of change</td>
<td>Median</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Type of school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 60 (53.10)</td>
<td>93.0</td>
<td>16.0</td>
<td>175.0</td>
<td>18 584.00</td>
<td>5.57</td>
<td></td>
<td>27.0</td>
<td>3.0</td>
<td>41.5</td>
</tr>
<tr>
<td>Private 53 (46.90)</td>
<td>92.0</td>
<td>13.0</td>
<td>224.0</td>
<td>30 369.50</td>
<td>10.59</td>
<td></td>
<td>23.0</td>
<td>2.0</td>
<td>45.0</td>
</tr>
<tr>
<td>p-value</td>
<td>0.698</td>
<td>0.156</td>
<td>0.736</td>
<td>0.247</td>
<td>0.245</td>
<td></td>
<td>0.945</td>
<td>0.498</td>
<td>0.778</td>
</tr>
<tr>
<td>Geographic regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northwest 18 (15.93)</td>
<td>169.50</td>
<td>18.00</td>
<td>303.00</td>
<td>39 294.00</td>
<td>7.03</td>
<td></td>
<td>64.0</td>
<td>4.0</td>
<td>90.5</td>
</tr>
<tr>
<td>Northeast 17 (15.04)</td>
<td>125.00</td>
<td>12.00</td>
<td>266.00</td>
<td>47 010.50</td>
<td>11.06</td>
<td></td>
<td>27.0</td>
<td>2.0</td>
<td>45.0</td>
</tr>
<tr>
<td>West 17 (15.04)</td>
<td>137.00</td>
<td>18.00</td>
<td>256.00</td>
<td>11 796.50</td>
<td>7.66</td>
<td></td>
<td>27.0</td>
<td>3.0</td>
<td>47.0</td>
</tr>
<tr>
<td>East 14 (12.39)</td>
<td>36.75</td>
<td>12.50</td>
<td>78.50</td>
<td>10 195.00</td>
<td>9.45</td>
<td></td>
<td>7.00</td>
<td>1.0</td>
<td>15.0</td>
</tr>
<tr>
<td>North-central 10 (8.86)</td>
<td>119.00</td>
<td>13.00</td>
<td>263.50</td>
<td>64 780.00</td>
<td>10.18</td>
<td></td>
<td>49.75</td>
<td>4.0</td>
<td>93.0</td>
</tr>
<tr>
<td>South-central 19 (16.81)</td>
<td>154.00</td>
<td>16.00</td>
<td>194.00</td>
<td>12 601.50</td>
<td>5.35</td>
<td></td>
<td>47.0</td>
<td>5.0</td>
<td>71.0</td>
</tr>
<tr>
<td>Southeast 9 (7.96)</td>
<td>58.00</td>
<td>11.00</td>
<td>87.00</td>
<td>310.75</td>
<td>18.00</td>
<td></td>
<td>7.00</td>
<td>2.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Southwest 9 (7.96)</td>
<td>26.00</td>
<td>15.00</td>
<td>35.00</td>
<td>7 084.00</td>
<td>1.91</td>
<td></td>
<td>4.00</td>
<td>3.0</td>
<td>4.0</td>
</tr>
<tr>
<td>p-value</td>
<td>0.010‡</td>
<td>0.629</td>
<td>0.008‡</td>
<td>0.406</td>
<td>0.661</td>
<td></td>
<td>0.001‡</td>
<td>0.040‡</td>
<td>0.011‡</td>
</tr>
<tr>
<td>Socioeconomic level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1 9 (7.96)</td>
<td>26.00</td>
<td>15.00</td>
<td>35.00</td>
<td>7 084.00</td>
<td>1.92</td>
<td></td>
<td>4.00</td>
<td>3.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Level 2 20 (17.70)</td>
<td>51.25</td>
<td>14.50</td>
<td>104.50</td>
<td>10 792.75</td>
<td>9.47</td>
<td></td>
<td>9.00</td>
<td>1.5</td>
<td>17.5</td>
</tr>
<tr>
<td>Level 3 15 (13.27)</td>
<td>146.00</td>
<td>13.00</td>
<td>283.00</td>
<td>64 780.00</td>
<td>6.34</td>
<td></td>
<td>53.00</td>
<td>3.0</td>
<td>91.0</td>
</tr>
<tr>
<td>Level 4 15 (13.27)</td>
<td>58.00</td>
<td>11.00</td>
<td>72.00</td>
<td>-1 886.00</td>
<td>13.01</td>
<td></td>
<td>7.00</td>
<td>1.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Level 5 21 (18.5)</td>
<td>99.00</td>
<td>16.00</td>
<td>231.00</td>
<td>23 787.50</td>
<td>9.47</td>
<td></td>
<td>28.00</td>
<td>3.0</td>
<td>56.0</td>
</tr>
<tr>
<td>Level 6 20 (17.70)</td>
<td>170.50</td>
<td>13.50</td>
<td>289.00</td>
<td>26 082.25</td>
<td>8.54</td>
<td></td>
<td>31.50</td>
<td>3.0</td>
<td>67.5</td>
</tr>
<tr>
<td>Level 7 13 (11.50)</td>
<td>272.00</td>
<td>21.00</td>
<td>373.00</td>
<td>13 395.75</td>
<td>4.67</td>
<td></td>
<td>66.00</td>
<td>7.0</td>
<td>96.0</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt; 0.01‡</td>
<td>0.591</td>
<td>&lt; 0.01‡</td>
<td>0.433</td>
<td>0.354</td>
<td></td>
<td>&lt; .001‡</td>
<td>0.008‡</td>
<td>&lt; .001‡</td>
</tr>
</tbody>
</table>

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 Licego 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. In a recent report about ENARM published in 2017, 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.
# Table II

**SUMMARY MEASURES AND STATISTICAL SIGNIFICANCES FOR THE NUMBER OF PERCENTAGE OF SELECTED APPLICANTS AND TOTAL SCORE**

<table>
<thead>
<tr>
<th>type of school</th>
<th>N (% )</th>
<th>Percentage of selected applicants</th>
<th>Total score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Public 60 (53.10)</td>
<td>5.26</td>
<td>3.647</td>
<td>9.83</td>
<td>-0.07</td>
</tr>
<tr>
<td>Private 53 (46.90)</td>
<td>4.55</td>
<td>3.05</td>
<td>8.66</td>
<td>-0.07</td>
</tr>
<tr>
<td>Type of school</td>
<td></td>
<td>Median</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>p-value</td>
<td>0.066</td>
<td>0.076</td>
<td>0.635</td>
<td>0.593</td>
</tr>
<tr>
<td>Northwest 18 (15.93)</td>
<td>4.21</td>
<td>2.80</td>
<td>7.39</td>
<td>-0.07</td>
</tr>
<tr>
<td>Northeast 17 (15.04)</td>
<td>4.51</td>
<td>3.00</td>
<td>8.39</td>
<td>-0.01</td>
</tr>
<tr>
<td>West 17 (15.04)</td>
<td>5.40</td>
<td>3.64</td>
<td>11.75</td>
<td>-0.30</td>
</tr>
<tr>
<td>East 14 (12.39)</td>
<td>7.10</td>
<td>4.37</td>
<td>14.00</td>
<td>-0.28</td>
</tr>
<tr>
<td>Geographic regions</td>
<td></td>
<td>Median</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>p-value</td>
<td>0.005</td>
<td>0.036</td>
<td>0.073</td>
<td>0.026</td>
</tr>
<tr>
<td>North-central 10 (8.86)</td>
<td>4.26</td>
<td>3.08</td>
<td>8.38</td>
<td>-0.08</td>
</tr>
<tr>
<td>South-central 19 (16.81)</td>
<td>4.57</td>
<td>3.22</td>
<td>8.00</td>
<td>-0.07</td>
</tr>
<tr>
<td>Southeast 9 (7.96)</td>
<td>6.57</td>
<td>4.04</td>
<td>13.50</td>
<td>0.13</td>
</tr>
<tr>
<td>Southwest 9 (7.96)</td>
<td>3.40</td>
<td>2.88</td>
<td>13.07</td>
<td>0.71</td>
</tr>
<tr>
<td>Socioeconomic level</td>
<td></td>
<td>Median</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>p-value</td>
<td>0.016</td>
<td>0.075</td>
<td>0.151</td>
<td>0.051</td>
</tr>
<tr>
<td>Level 1 9 (7.96)</td>
<td>3.40</td>
<td>2.88</td>
<td>13.07</td>
<td>0.71</td>
</tr>
<tr>
<td>Level 2 20 (17.70)</td>
<td>6.55</td>
<td>4.21</td>
<td>10.83</td>
<td>0.05</td>
</tr>
<tr>
<td>Level 3 15 (13.27)</td>
<td>4.41</td>
<td>3.28</td>
<td>8.94</td>
<td>-0.18</td>
</tr>
<tr>
<td>Level 4 15 (13.27)</td>
<td>5.86</td>
<td>4.00</td>
<td>12.50</td>
<td>0.03</td>
</tr>
<tr>
<td>Level 5 21 (18.5)</td>
<td>4.25</td>
<td>2.87</td>
<td>9.36</td>
<td>-0.01</td>
</tr>
<tr>
<td>Level 6 20 (17.70)</td>
<td>4.71</td>
<td>3.13</td>
<td>10.12</td>
<td>-0.10</td>
</tr>
<tr>
<td>Level 7 13 (11.50)</td>
<td>3.76</td>
<td>2.66</td>
<td>7.91</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

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 <0.05).
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

**Figure 2. Times series graph of the selected variables grouped by type of school (public vs. private)**
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. The median summary measure was the variable able to identify significant differences between groups; our explanation of why the comparison of slopes

*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**
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, 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; readers can find a brief review on this topic in the article published by Schober and colleagues.

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. 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. 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.

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.

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 Osegua, 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

11. Go PH, Klaassen Z, Chamberlain RS. An ERAS-based survey evaluating academic performance, United States Medical Licensing Examina...