RESEARCH

From admissions to licensure: education data associations from a multi-centre undergraduate medical education collaboration

S. Chahine[1] · I. Bartman[2] · K. Kulasegaram[3,4] · D Archibald[5,6] · P. Wang[7] · C. Wilson[7] ·
B. Ross[8] · E. Cameron[8,9] · J. Hogenbirk[10,11] · C. Barber[12,13] · R. Burgess[12,14] ·
E. Katsoulas[1] · C. Touchie[15] · L Grierson[12,16]

Abstract

This paper reports the findings of a Canada based multi-institutional study designed to investigate the relationships between admissions criteria, in-program assessments, and performance on licensing exams. The study's objective is to provide valuable insights for improving educational practices across different institutions. Data were gathered from six medical schools: McMaster University, the Northern Ontario School of Medicine University, Queen's University, University of Ottawa, University of Toronto, and Western University. The dataset includes graduates who undertook the Medical Council of Canada Qualifying Examination Part 1 (MCCQE1) between 2015 and 2017. The data were categorized into five distinct sections: demographic information as well as four matrices: admissions, course performance, objective structured clinical examination (OSCE), and clerkship performance. Common and unique variables were identified through an extensive consensus-building process. Hierarchical linear regression and a manual stepwise variable selection approach were used for analysis. Analyses were performed on data set encompassing graduates of all six medical schools as well as on individual data sets from each school. For the combined data set the final model estimated 32% of the variance in performance on licensing exams, highlighting variables such as Age at Admission, Sex, Biomedical Knowledge, the first post-clerkship OSCE, and a clerkship theta score. Individual school analysis explained 41–60% of the variance in MCCQE1 outcomes, with comparable variables to the analysis from of the combined data set identified as significant independent variables. Therefore, strongly emphasising the need for variety of high-quality assessment on the educational continuum. This study underscores the importance of sharing data to enable educational insights. This study also had its challenges when it came to the access and aggregation of data. As such we advocate for the establishment of a common framework for multi-institutional educational research, facilitating studies and evaluations across diverse institutions. This study demonstrates the scientific potential of collaborative data analysis in enhancing educational outcomes. It offers a deeper understanding of the factors influencing performance on licensure exams and emphasizes

Extended author information available on the last page of the article

the need for addressing data gaps to advance multi-institutional research for educational improvements.

Introduction

It is essential that our physicians are trained to a high level of competence. Yet, there are numerous approaches to achieving this goal. Indeed, each medical school in Canada takes pride in its distinctive qualities that distinguish it from other institutions. Even within the context of rigorous accreditation constraints, the admissions processes, education policies, curricula, and assessment practices of each school blend together to form a unique training culture. This culture shapes how medical learners come to conceptualize their future professional responsibilities and, ultimately, how they deliver care to patients and communities. Single institution studies have demonstrated meaningful associations between admissions criteria, procedures, and systems, and downstream performance during medical training (Barber et al., 2022; Donnon et al., 2007; Dore et al., 2017; Eva et al., 2009; Reiter et al., 2007). Similar studies have also established links between performance in medical training and performance on licensing exams (Casey et al., 2016; Deng et al., 2015; Gohara et al., 2011; Kimbrough et al., 2018). The value of these types of studies becomes apparent when considered in light of complementary work that shows meaningful associations between training and performance on licensing exams and measures of practice effectiveness (e.g., Asch et al., 2009; De Champlain et al., 2020; Tamblyn et al., 2007). However, only a few studies have linked and analyzed education outcomes data that was procured and aggregated from multiple training institutions (e.g., Grierson et al., 2017; Santen et al., 2021). This gap has left leaders of medical education with uncertainty about the degree to which reported relationships have generalizable relevance within unique learning contexts. To fill this gap, we conducted a multi-institutional study involving the aggregation and linkage of retrospective learner-level administrative education data to identify key associations between generalizable features of admissions and in-program assessment systems and performance on licensing exams. The ultimate goal was to develop evidence that enables educational improvements that support student success.

Background

There have been mounting calls to leverage educational data to generate fresh insights into educational practices and advance conceptual models of training (Chahine et al., 2018; Phillips et al., 2022; Janssen et al., 2022). This is especially salient within the context of Canadian medical education, wherein country-wide curricular reforms (Tannenbaum et al., 2011; College of Family Physicians of Canada, 2009; Frank et al., 2010; Fowler et al., 2022) are prompting increased scrutiny on the value of educational activities for clinical performance. Previous stand-alone investigations using administratively collected data for the purposes of admissions, curricular assessment, and licensing have shown to be effective at evaluating

educational efficacy in relation to educational, professional, and patient outcomes (Grierson et al., 2017; Tamblyn et al., 2007). Further, these studies have enabled theory driven investigations on the social and psychological nature of expertise (Asch et al., 2014). Accordingly, several academic and professional organizations have called for the creation of data infrastructure to make this type of research more programmatic, continuous, feasible, and sustainable (Grierson et al., 2022).

The central idea is that the quality of medical school admissions policies, curricula, and assessment practices influence the overall quality of medical education, which, in turn, influences the quality of trainee performance on licensing and certification examinations, and ultimately the quality of healthcare provided by medical school graduates. For example, the educational shift to competency-based medical education (CBME) is taking place with the assumption that this curricular improvement in medical education will improve the provision of healthcare and patient outcomes (Frank et al., 2010; Holmboe & Batalden, 2015). Many question the assumption that CBME will meet its professed goals (Klamen et al., 2016; Whitehead & Kuper, 2017). Some are more optimistic and recommend that evidence be gathered before and after such reforms are implemented (Salim & White, 2018) to determine whether changes are leading us in the intended direction. However, the chain of evidence *proving* that a change in educational principles leads to actual improvement of healthcare is challenging to evaluate empirically.

This study represents a key step in addressing this challenge. Our project is a collaborative endeavour between medical education researchers from the six Ontario medical schools (McMaster University, Northern Ontario School of Medicine University, Queen's University, University of Ottawa, University of Toronto, and Western University), and the Medical Council of Canada (MCC). Together, we aim to illuminate generalizable data associations early in the medical education trajectory; from admission into medical school through undergraduate training and to the first stage of licensure. Until 2020, to practice medicine in Canada, a physician was required successfully complete two licensing examinations. As of 2021, new physicians must only complete the first MCC qualifying examination (MCCQE1), which focuses on medical knowledge and clinical decision-making, in order to qualify for a medical license (Medical Council of Canada, 2023). Notably, numerous single institution studies have documented meaningful relationships between training data and licensing outcomes (Violato & Donnon, 2005; Eva et al., 2012; Barber et al. 2018; Dore et al. 2017; Gullo et al., 2015; Coumarbatch et al., 2010). Given that the literature also highlights relationships between licensing examination performance and future prescribing competence, proactive health advocacy behaviour, and rates of diagnostic error (Tamblyn et al., 2007; Wenghofer et al., 2009; Kawasumi et al., 2011), it is clearly worthwhile to develop a more robust understanding of this part of the training continuum. Accordingly, our first objective was to identify the admissions and assessment variables commonly used across all six Ontario medical schools that are significantly associated with stronger performance on the MCCQE1. Our second objective was to identify the admissions and assessment variables unique to each medical school that are also significantly associated with success on the MCCQE1. Importantly, in both cases, we consider the associated value of any one admission and assessment variable in light of all relevant admission and assessment data.

Methods

Study sample

We extracted comprehensive sets of defined variables for learners who graduated from the medical schools located at McMaster University, the Northern Ontario School of Medicine University, Queen's University, University of Ottawa, University of Toronto, and Western University in 2015, 2016, and 2017, and who had challenged the MCCQE1 prior to 2018. This inclusion criteria ensured all MCCQE1 data were generated on the same test blueprint.

Data harmonization

Given that each of the six schools operated independent admissions processes and assessments, interoperable expressions of data needed to be developed prior to data collation. We began this process by organizing the data from the six schools into five categories: demographic information, admissions metrics, course performance metrics, objective structured clinical examination metrics, and clerkship performance metrics. Within each category, we defined relevant variables via an iterative process of team-wide consensus building. Ultimately, we sought agreement that the data brought forward from each school were appropriately reflective of the measurement construct that the variable was intended to represent and, as such, analogous within a collated dataset. Where necessary, we then transformed the data that comprised each defined variable at each school so that they were expressed on comparable scales. We concluded with a second process of categorization, which involved labelling each variable as either *common* or *unique*. A variable was denoted as common when it was present in each school's independent dataset. A variable was denoted as unique when it was present in five or fewer of the independent datasets. Variables that were not realized at all schools were deemed unique even if they were relevant to most participating institutions in order to protect against institutional identification in the larger collated datasets. For instance, while the Medical College Admissions Test (MCAT) is used as an admission tool at most participating institutions, it is not used at the University of Ottawa. As such, if it were included in the inter-institutional collation of common variables, then it would have been easy to determine data specifically associated with the University of Ottawa in the larger dataset by determining those observations characterized by an absence of MCAT variables.

Data definitions

The common variables defined in the demographic information category include age (Age), sex (Sex), national neighbourhood income quintile per single person equivalent after-tax (QNATIPPE), and geographic status (Geographic Status). Age was coded in years (*continuous* variable) and reflected age at the beginning of medical school. Sex was coded as either male or female (*categorical* variable) and reflected self-reported answers provided to admissions prompts regarding gender. It was not possible to determine whether responses reflected declarations of legal sex, sex assigned at birth, or gender. QNATIPPE was built upon the residential postal codes at the time of secondary school reported in application materials (*categorical*) and was derived via Statistics Canada's Postal Code Conversion

File (Statistics Canada). The neighbourhood quintile is understood as a proxy measure of socioeconomic status. Geographic Status reflected the region from which the successful medical school application was made, either in Ontario, elsewhere in Canada, or outside of Canada (*categorical*). There were no unique variables defined within the demographic information category.

The common variables defined in the admissions metrics category included undergraduate grade point average (GPA), graduate degree status (Graduate Degree), admissions interview score (Interview), and admission into a traditional or conjoint MD-PhD stream of study (Study Stream). GPA reflected a four-point expression of the undergraduate grade point average appraised during the admissions cycle at each school, as collected by each school (*continuous*). Graduate Degree reflected whether matriculants held or did not hold a masters or doctorate degree at the time of admissions (*categorical*). The Interview variable reflected a transformation of the final scores on multiple mini-interviews, panel interviews, or combined interview approaches applied at each school (*continuous*). The unique variables within the admissions metrics category included final and sub-section scores derived via the Medical College Admissions Test (MCAT; *continuous*), scores on the Computer-based Assessment for Sampling Personal Characteristics situational judgement test (CASPer; *continuous*), and language stream in English or French (*categorical*).

The only common variable defined in the course performance metrics category was derived from performance ratings pertaining to biomedical knowledge (Biomedical Knowledge). These reflected cumulative and transformed assessments of anatomy and biomedical systems learning from each of the six institutions (*continuous*). The only unique variable defined in the course performance metrics category was derived from assessment of professional competence (Professional Knowledge), which included assessments of patient and interprofessional communication, collaboration, and professional behaviors (*continuous*). While the curricula at all six schools contemplate the development and assessment of professional competence, the nature of its integration into the wider programming made it difficult for some of the collaborating institutions to accurately separate out an assessment that was specifically and exclusively reflective of the Professional Knowledge construct; hence its inclusion in the unique category.

The common variables defined in the objective structured clinical examination (OSCE) metrics category included performance on the final pre-clerkship OSCE (OSCE1; *continuous*) and on the first post-clerkship OSCE (OSCE2; *continuous*). While all participating schools conduct two OSCEs, some hold more. As such, unique variables in the OSCE category included scores from OSCEs that were supernumerary to those defined as common within this category (*continuous*).

The common clerkship performance score (Clerkship), were derived from assessments from the six core clerkship rotations at each school: Family Medicine, Medicine, Obstetrics and Gynecology (OBGYN), Pediatrics, Psychiatry, and Surgery. Two methods were employed to generate a common clerkship score.

The first method involved converting clerkship scores to a 3-point categorical variable to establish a consistent metric across schools. For schools with continuous data, clerkship scores were converted to z-scores, with the lowest decile within each clerkship coded as "-1", scores in the second through ninth deciles coded as "0", and scores in the top decile coded as "1". In cases where categorical data were used, the labels "fail," "pass," and

"excellent" were respectively coded as "-1," "0," and "1." As a result, each school utilized a common categorical scale for each of the six clerkships.

The second method was needed due to blocks of missing data in certain course specialties. This was problematic in utilizing clerkship scores for each specialty. As a result, we decided to generate a comprehensive metric for each learner, reflecting their scores across the different clerkships. First, the psychometric properties were assessed by using Cronbach's alpha and exploratory factor analyses (EFA). The reliability analysis yielded a satisfactory alpha value of 0.67, and an EFA with maximum likelihood estimation resulted in a one-factor solution. Subsequently, a partial-credit Rasch model (Mair, & Hatzinger, 2007) was employed to generate a single theta score that represents clerkship performance. This approach allowed for the retention of variation in difficulty across the core specialties. Similar approaches for data combination and aggregation were recently used when working with varied data sources (Norcini et al., 2023). The theta scores can be interpreted similarly to z-scores, where 0 represents average performance, lower values indicate lower performance, and higher values indicate higher performance on a continuous scale. The scores were generated using the eRM R package (Mair, & Hatzinger, 2007; R Core Team, 2022).

The only dependent variable defined for this study was MCCQE1 final performance (*continuous*). These data were generated during the 2015–2017 examination administrations. The examination was scored on a scale ranging from 50 to 950. The minimum score for passing the exam was set at 427.

Table 1 provides full details of all common and unique variables including their coding structure, data transformations performed, and the distribution of unique variables across collaborating schools.

Research data management

After agreeing on the common variables to include, we established a standard organizational structure for the datasets (e.g. standard variable names, order). The datasets from each school were organized according this structure and then forwarded to the Ontario Physician Research Centre (OPRC; formerly the Ontario Physician Human Resource Data Centre (OPHRDC)), which managed the safe and ethical amalgamation of all data into research-ready datasets. The OPRC operates with state-of-the-art data management technology and under longstanding data sharing agreements that mediate regular exchange of data between the medical schools and the Province of Ontario's Ministry of Health and Long-Term Care (MOHLTC) in support of provincial health human resource planning. The OPRC received approval to facilitate this project from its Steering Committee, which is responsive to the strategic mandates of both the MOHTLC, the Council of Ontario Faculties of Medicine (COFM), and the Council of Ontario Universities (COU). For these reasons, the OPRC was deemed ideally suited to serve as a trusted data facility for this research. Oversight of the research data management, as well as the enactment of the research study, was provided by a Research and Data Governance Committee, which was developed specifically for the purposes of this study. This committee was comprised of representatives of COFM and the Ontario Medical Student Association (OMSA).

The OPRC linked the data from the six medical schools and the MCC to create seven analysis-ready datasets. The primary dataset included common variables from all six schools, matching common variables with the MCCQE1 data at the level of each student

Table 1 Characteristics of common and unique variables and coding system.[a]

Variable	Brief definition	Type of data	Scale and coding	Common or unique variable designation
Demographic Variables:				
Age at admission	The learner's age at the time of entry to medical school.	Continuous	Years	Common
Sex	The individual's self-identified sex or gender.	Categorical (Nominal)	Man (0), Woman (1)	Common
Neighbourhood income quintile (national) after tax (QNATIPPE)	Household size-adjusted measure of household income (after tax).	Categorical (Ordinal)	Lowest quintile (1); Medium-low quintile (2); Middle quintile (3); Medium-high quintile (4); Highest quintile (5); Missing (9)	Common
Geographic status	The region in which the learner's home residence during secondary school was located.	Categorical (Nominal)	Ontario (0), other province or territory in Canada (1), outside Canada (2), or missing/unknown (3).	Common
Admissions Variables:				
Undergraduate GPA	The unweighted, overall cumulative undergraduate grade point average of the applicant.	Continuous	0.00 to 4.00; 2 decimal points included	Common
Graduate degree status	Whether or not the applicant had a graduate degree (Master's or PhD) at the time of application.	Categorical (Nominal)	No graduate degree (0); Graduate degree (PhD and/or Master's) (1)	Common
Interview Score	Score received for performance in admissions interview.	Continuous	Z-score, calculated within-school and within-cohort	Common
MD/PhD Status	An indication of whether the learner is enrolled in the MD program or the MD/PhD program.	Categorical (Nominal)	MD program (0); MD/PhD program (1)	Common
Medical College Admissions Test (MCAT)	Final and sub-section MCAT scores received by the learner.	Continuous	Score	Unique
Computer Based Assessment for Sampling Personal Characteristics situational judgement test (CASPer)	The score on the CASPer test (situational judgement test) received by the learner.	Continuous	Score	Unique
Course Performance Metrics:				
Biomedical knowledge	Encompasses activities related to the learner's acquisition of biomedical knowledge during their medical degree.	Continuous	Cohort-mean centered scores	Common
Professional Knowledge	An assessment of professional competence.	Continuous	Cohort-mean centered scores	Unique
Objective Structured Clinical Examinations (OSCE) metrics:				

Table 1 (continued)

Variable	Brief definition	Type of data	Scale and coding	Common or unique variable designation
Final Pre-clerkship OSCE (OSCE 1)	Performance on the OSCE that occurred most recently to a learner's clerkship period. The OSCE is designed to test clinical skill performance and other competencies.	Continuous	Z-score, calculated within-school and within-cohort	Common
First Post-Clerkship OSCE (OSCE 2)	Performance on the first OSCE that the learner completed following clerkship.	Continuous	Z-score, calculated within school and within cohort	Common
Additional OSCEs	Performance on additional OSCEs held by some participating medical schools.	Continuous	Z-score, calculated within cohort	Unique
Clerkship Evaluations:				
Overall Clerkship Performance Score	A score derived from assessments in six core clerkship rotations at each school: Family Medicine, Medicine, Obstetrics and Gynecology (OBGYN), Pediatrics, Psychiatry, and Surgery.	Continuous	Theta score, based on partial-credit Rasch model	Common
Individual Clerkship Performance Scores:	This variable encompasses evaluations of the learner's performance on their individual clerkships (General Surgery; Psychiatry; Pediatrics; Family Medicine; Obstetrics & Gynecology; Internal Medicine)	Categorical (Ordinal)	Excellent Performance (Highest decile performers) (coded as 2); Pass (2nd through 9th decile) (coded as 1); Fail/borderline pass (Lowest decile) (coded as 0)	Unique
Dependent Variable:				
MCCQE Part I Final Performance[b]	Defined as the total score the learner received on their first attempt of the MCCQE Part I.	Continuous	Possible scores range from 50 to 950. A passing score is 427 or higher.	Common

[a]See the Supplementary Material for more detailed information about the variables, at each medical respective school

[b]For more information about how the MCCQE Part I is scored, please see the technical reports available here: https://mcc.ca/research-and-development/technical-reports/

observation. This dataset was then de-identified at the level of both the student and the school before being returned to the research team. This dataset was analyzed in support of our first objective (i.e., to determine the admissions and assessment variables commonly used by our six Ontario medical schools that are significantly associated with success on the MCCQE1). Six secondary datasets were also created. These datasets linked MCCQE1 data with the individual datasets provided by each participating school. As such these datasets were inclusive of each school's common variables but also their school-specific unique variables. These datasets were de-identified at the level of the student before being returned to the relevant participating institution. These datasets were analyzed in support of our second objective (i.e., to determine the admissions and assessment variables unique to each of our medical schools that are also significantly associated with success on the MCCQE1).

Analytic strategy

Hierarchical linear regression was used in support of both our first and second objectives. In all cases, a manual stepwise approach, in which variables are retained or removed according to their significance at different steps in the model building process. The final models for both objectives were retained to explain the most amount of variance using the minimal number of variables. Statistical significance was set to an alpha-level of 0.05 for all models.

The final model for the first objective was progressively constructed through the development of steps that reflected the introduction of data in a manner consistent with the trajectory of undergraduate medical students over time. The first step was comprised exclusively of the common variables from the demographic information category. Step 2 retained variables that were statistically significant in from step 1 while adding in common variables from the admissions metrics category. Similarly, step 3 retained variables that were statistically significant from step 2, while adding in the common variable from the course performance metrics category. Step 4 introduced common OSCE metrics variables to the model that retained from step 3. Lastly, step 5 retained the significant variables from step 4 and added in the common variable from the clerkship performance category. The final models are presented in the results. Analyses were performed by IB and verified by SC.

In order to meet our second objective, a common model building syntax was developed by one investigator (SC) and verified by another investigator (KK). The syntax was shared with the relevant investigators at each participating institution for application to their school-specific dataset of common and unique variables. In this way, six independent final models were constructed in support of the second objective. These were built in the same process as described above; however, were also considerate of potential associations between the unique variables for each school at the 5 different steps and MCCQE1 outcomes. To support consistency in analyzing the independent datasets, the investigators affiliated with each institution oversaw all analysis during iterative, videoconference "data parties", which are commonly used in program evaluation to facilitate review and discussion of data among key stakeholders (Adams & Neville, 2020). Analyses across the various sites were conducted using the SPSS (IBM, United States) software either versions 26 or 28.

Ethics approval

Ethics clearance was received by the relevant Research Ethics Board at each participating institution (Hamilton Integrated Research Ethics Board Project ID # 4652; Laurentian REB # 6,017,122, Lakehead REB # 1,467,011, Queen's University Ethics Board # 6,024,231, Ottawa Health Science Network Research Ethics Board ID # 20,180,669–01 H and Bruyère Research Ethics Board Protocol # M16-18-058; University of Toronto REB #00039173, Western University REB # 110,837). The MCC received ethical clearance from Advarra Canada (IRB Pro0030825, December 5, 2018).

Results

The overall descriptive statistics for the amalgamated dataset are presented in Table 2. The dataset contained 2727 rows representing individual medical students over the four years of medical school, the mean or percentage, and the standard deviations are in Table 2.

Objective 1: common variable analysis

As described earlier the model building process encompasses five main steps that represent the different stages off medical school from application to completion, at each step variables are added and removed based on their significance. The final model developed involved a sample size of 2637. In the first step, Age at Admission (Age) and Sex were significantly associated with the MCCQE1 scores ($F(2,2634)=26.86$, $p<0.001$) and explained 2% of the variance. In the second step there were no significant variables retained. The third step ($F(3,2633)=227.60$, $p<0.001$), which added a Biomedical Knowledge, showed a small but significant improvement from the first step ($\Delta F(1,2633)=616.52$, $p=<0.001$) and overall explained 21% of the variance. The fourth step ($F(4,2632)=181.05$, $p<0.001$), retained OSCE2 performance and showed significant improvement from the third step ($\Delta F(1,2632)=33.08$, $p<0.001$) and added 1% to the explained variance. The fifth step ($F(5,2631)=244.79$, $p<0.001$) included the Clerkship theta score It showed significant improvement from the third step ($\Delta F(1,2631)=392.14$, $p<0.001$) and explained 10% more variance. Overall, this final model explained 32% of the variance. The retained variables at each step for the final model are presented in Table 3. The final list of variables associated with the performance on the MCCQE1includes: Age, and Sex, Biomedical Knowledge, OSCE2 and clerkship theta scores.

Table 2 Descriptive statistics of amalgamated data set

Variable	No.	Metric (m/%)	s.d.
Demographic Variables			
Age	2719	23.36	2.73
QNATIPPE	2580	3.67	1.37
Sex (Female)	2727	55.70%	
Geographic Status (Ontario)	2629	87.88%	
Admissions Variables			
GPA (accumulative)	2674	3.83	0.16
Interview scores (Within School Z-Scores)	2702	0.02	1.17
Graduate Degree (Yes)	2711	20.21%	
MdPhD (Yes)	2727	2.02%	
Undergraduate Knowledge Assessments			
Biomedical Knowledge	2692	0.10	5.57
Undergraduate Performance Assessments			
OSCE 1	2683	−0.03	1.28
OSCE 2	2683	−0.01	1.10
Clerkship Assessments			
Clerkship Theta Scores	2727	0.45	1.55
Outcome			
MCCQE1	2705	545.76	70.37

Table 3 Hierarchical regression estimates of final model

Step 1			
Age	-3.39	-0.13	-6.70***
Sex	7.92	0.06	2.91**
Step 2			
None	-	-	-
Step 3			
Age	−2.56	−0.10	−5.59***
Sex	10.85	0.08	4.43***
Biomedical Knowledge	5.54	0.43	24.83***
Step 4			
Age	−2.50	−0.10	−5.51***
Sex	9.60	0.07	3.93***
Biomedical Knowledge	5.36	0.42	23.93***
OSCE2	6.40	0.10	5.75***
Step 5			
Age	−2.06	−0.08	−4.84***
Sex	6.35	0.05	2.77**
Biomedical Knowledge	3.84	0.30	17.24***
OSCE2	4.20	0.07	4.02***
Clerkship Theta Scores	16.11	0.35	19.81***

Note $n=2637$; $R^2=0.02$ for step 1, $p<0.001$; $R^2\triangle=0.19$ for step 3, $p<0.001$; Total $R^2=0.21$, $p<0.001$; $R^2\triangle=0.01$ for step 4, $p<0.001$; Total $R^2=0.22$, $p<0.001$; $R^2\triangle=0.10$ for step 5, $p<0.001$; Total $R^2=0.32$, $p<0.001$

$*p<0.05$, $**p<0.01$, $***p<0.001$

Objective 2: school-specific analyses

The model building process used by each school mirrored the 5 main steps used with the amalgamated dataset, with the addition of unique variables. The variables retained in the final steps across the 6 schools and the combined data set are presented in Table 4. The results of the final models (i.e., step 5) are described below and presented in Appendix 1. The descriptive for each school are provided in Appendix 1.

McMaster university

The final model developed with the McMaster University data involved a sample size of 270. There were no significant variables in the first or second step. The third step explained 25% of the variance and included Biomedical Knowledge which was significantly associated with MCCQE1 scores ($F(1,268)=90.64$, $p<0.001$). In step 4 ($F(2,267)=57.66$, $p<0.001$), a common variable (OSCE 2) was added, which showed significant improvement from the model at step 3 ($\Delta F(1,267)=18.67$, $p<0.001$), and explained 5% more variance. Finally, in the fifth step ($F(5,264)=36.08$, $p<0.001$), clerkship variables (Surgery, OBGYN, Medicine) were added, which showed significant improvement from the model at step 4 ($\Delta F(3,264)=15.45$, $p<0.001$), and explained 11% more variance. Overall, this final model explained 41% of the variance. The final list of variables associated with the performance on the MCCQE1 at McMaster University includes: Biomedical Knowledge, OSCE2 and three clerkship scores (Surgery, OBGYN, and Internal Medicine).

Table 4 Retained variables of final model

Variables	Amalgamated	McMaster	NOSM	Queen's	UofO	UofT	Western
Demographic Variables:							
Age at admission	1						
Sex	1				1		
Neighbourhood income quintile (national) after tax (QNATIPPE)							
Geographic status							
Admissions Variables:							
Undergraduate GPA			1				
Graduate degree status							
Interview Score							
MD/PhD Status							
Medical College Admissions Test (MCAT) Total				1			
MCAT Verbal Reasoning						1	1
Computer Based Assessment for Sampling Personal Characteristics situational judgement test (CASPer)							
Course Performance Metrics:							
Biomedical knowledge	1	1	1	1	1	1	1
Unique Institution Specific Course			1			1	
Professional Knowledge							
Objective Structured Clinical Examinations (OSCE) metrics:							
Final Pre-clerkship OSCE (OSCE 1)					1		1
First Post-Clerkship OSCE (OSCE 2)	1	1	1		1		1
Additional OSCEs							
Clerkship Evaluations:							
Overall Clerkship Performance Score	1						
Individual Clerkship Performance Scores:							
Psychiatry			1	1			
OBGYN		1		1		1	
Surgery		1			1	1	
Internal Medicine		1	1	1	1	1	
Pediatrics				1	1		
Anesthesiology (Unique)						1	
Emergency Medicine (Unique)						1	
Total	**5**	**5**	**6**	**6**	**7**	**8**	**4**

Northern ontario school of medicine university

The final model using NOSM University data had a sample size of 178. The first step explained 15.3% of the variance and included one common variable (Age) and two variables unique to NOSM U: Minority Status (coded as non-Indigenous or non-Francophone versus other) and Rural background (Yes versus No). It was significantly predictive of MCCQE1 scores ($F(6,156)=5.87$, $p<0.001$). In the second step, Age and Minority Group were retained as significant predictors and GPA was added. This model explained 15.8% of the variance ($F(8,169)=5.15$, $p<0.001$). The change from the first step was significant ($\Delta F(5,169)=2.30$, $p=0.047$). In step 3 ($F(5,173)=21.05$, $p<0.001$) only GPA was retained, with one common variable (Biomedical Knowledge) and one unique variable (year 1 performance in Clinical Skills in Health Care) added. Step 3 showed significant improvement from the step 2 ($\Delta F(2,173)=28.95$, $p<0.001$), and explained 36.0% of the variance. The fourth step ($F(5,170)=21.93$, $p<0.001$) retained all the variables in step 3 and added one common variable (OSCE-2 z-scores). Step 4 was a significant improvement from step 3 ($\Delta F(2,170)=4.83$, $p=0.009$). The explained variance increased slightly to 37.4%. Finally, the fifth step ($F(10,167)=17.64$, $p<0.001$) retained all of the variables in step 4 and added two common clerkship variables (Psychiatry and Internal Medicine). Step 5 showed significant improvement from the step 4 ($\Delta F(6,167)=6.80$, $p<0.001$). This final model explained 49% of the variance. The final list of variables associated with the performance on the MCCQE1 at NOSM includes: GPA, Biomedical Knowledge, Clinical Skills in Health Care year 1, OSCE2 and two clerkship scores, Psychiatry, and Internal Medicine.

Queen's university

The final model developed with Queen's University data involved a sample size of 283. In the first step there were no variables retained at within the final model. In the second step, a unique variable (MCAT Total Score) was significantly associated with the MCCQE1 scores ($F(1,281)=21.17$, $p<0.001$) and explained 7% of the variance. The third step ($F(2,280)=46.12$, $p<0.001$), which included a common variable (Biomedical Knowledge), showed a small but significant improvement from the second step ($\Delta F(1,280)=66.16$, $p<0.001$) and explained 25% of the variance. The fourth step retained no significant variables. The fifth step ($F(6,276)=49.60$, $p<0.001$) included clerkship variables (Psychiatry, OBGYN, Medicine, Pediatrics). It showed significant improvement from the third step ($\Delta F(4,276)=38.87$, $p<0.001$) and explained 27% more variance. Overall, this final model explained 52% of the variance. The final list of variables associated with the performance on the MCCQE1 at Queen's University includes: MCAT Total Score, Biomedical Knowledge, and four clerkship scores, Psychiatry, OBGYN, Internal Medicine, Pediatrics.

University of Ottawa

The final model developed with University of Ottawa data involved a sample size of 458. In the first step, 3% of the variance was explained by a common variable (Sex), which was significantly associated with the MCCQE1 scores ($F(1,456)=13.82$, $p<0.001$). In the second step no admission variables were retained. The third step ($F(2,455)=235.10$, $p<0.001$) included a common variable (Biomedical Knowledge) and showed significant

improvement from the first/second step (Δ F(1,455)=442.99, p=<0.001) and an overall explained variance of 51%. The fourth step (F(4,453)=145.499, $p<0.001$) included two common variables (OSCE1; OSCE2). It showed significant improvement from the third step (Δ F(2,453)=28.00, $p<0.001$) and explained 5.4% more variance. The fifth step, (F(10,447)=68.81, $p<0.001$), included clerkship variables (Surgery, Medicine, Pediatrics). This step showed significant improvement from the fourth step (Δ F(6,447)=8.30, $p<0.001$) and explained 4.4% more of the variance. Overall, this final model explained 60% of the variance. The final list of variables associated with the performance on the MCCQE1 at University of Ottawa includes: Sex, Biomedical Knowledge, OSCE1, OSCE2 and three clerkship scores, Surgery, Internal Medicine, Pediatrics.

University of Toronto

The final model developed involved a sample size of 699. In the first step, Gender; Age, along with the unique variable of Geographic Status were significant against MCCQE1 (F(1,693)=11.638, $p<0.001$). These explained 3% of variance. In the second step, Age and Geographic status were no longer significant but Gender remained along with MCAT VR, MCAT BS and showed a significant improvement from step 1 (Δ F(8,662)=9.9, $p=0.001$. In total these explained 13.8% of variance (F (3, 692)=9.5, $p<0.001$). In the third step, caused Gender to drop out of the model though MCAT VR and BS were significant. Additionally the common variable of Biomedical grade was significant and the unique variables of Determinants of Community Health and the Art and Science of Medicine were significant. This was a significant change from step 2 (Δ F(6,685)=40.10, $p=001$). In total the variables in step 3 explained 34.3% of variance (F(4,691)=34.091, $p<0.001$). The fourth step included addition of OSCEs but caused resulted in no significant variables. The fifth step of clerkship variables resulted in Art and Science and MCAT BS to drop out of the model. The ObsGyn, Medicine, Surgery, Anasthesiology, and Emerg clerkship all significant and resulted in a significant change (Δ F(6,683)=21.56, $p=0.001$)with a final variance explained of 46% (F(9,685)=41.60, $p<0.001$) The final list of variables associated with the performance on the MCCQE1 at University of Toronto includes: MCAT VR, Biomedical Knowledge, Doc1, and the following clerkships: Surgery, OBGYN, Internal Medicine, Anesthesiology, and Emergency Medicine.

Western university

The final model developed with Western University data involved a sample size of 466. The first step no demographic variables were retained. The second step explained 0.8% of the variance. It included a unique variable reflecting an MCAT examination sub-score (MCAT Verbal Reasoning), which was significantly associated with the MCCQE1 scores (F(1,464)=3.80, $p=0.05$). The third step (F(2,463)=214.04, $p<0.001$) included a common variable (Biomedical Knowledge) and showed significant improvement from the second step (Δ F(1,463)=420.83, $p<0.001$). This step generated an estimate of 47.8% for overall explained variance. The fourth step (F(4,461)=115.87, $p<0.001$) included common variables (OSCE 1; OSCE 2). It showed significant improvement from the second step (Δ F(2,462)=9.68, $p<0.001$) and explained 2.1% more variance. As no variable were retained in the fifth step, this final model represented variables retained at the end of step four and

explained 50% of the variance. The final list of variables associated with the performance on the MCCQE1 at Western University includes: MCAT Verbal Reasoning, Biomedical Knowledge, OSCE1, and OSCE2.

Discussion

The research objectives encompassed two main aspects: firstly, identifying the admissions and assessment variables commonly employed by the six Ontario medical schools that exhibit a significant association with stronger performance on the MCCQE1, and secondly, identifying the admissions and assessment variables unique to each of the six medical schools that also demonstrate a significant association with success on the MCCQE1. While the results section primarily focuses on the final models, it is crucial to acknowledge that certain variables, which initially held significance, were overshadowed by more proximal variables closely aligned with the MCCQE1 as they were introduced.

Retained variables through each step of all models developed included Age and Sex. These variables registered significant association with the MCCQE1 outcomes - with younger persons and those who identify as female within medical school applications being more likely to perform better on the exam. Similar age effects have been noted in the context of other physician credentialing examinations (e.g., Gauer & Jackson, 2018; Grierson et al., 2017). Notably, the drivers of the effect have not been determined. It has been speculated that older students are likely to have greater family, financial, and social responsibilities that decrease their available time and energy for examination preparation. While previous research has also shown that females often perform better on credentialling examinations than their male counterparts (Cuddy et al., 2007, 2011; Grierson et al., 2017; Swygert et al., 2012), individual schools (McMaster, Queen's, and Western) reveal no significant relationship (see also: Gauer et al., 2018; Rubright et al., 2019). As such considering both what matters in the overall collection of data is variable at the school level. Whether these variables are significant or not is important to consider at the individual school level with the caveat that the relative effect of these variables is minimal when compared to other variables (e.g., Clerkship).

In the model building process, admission variables were initially found to be significant, however their relative effect to the overall model diminished during the final steps. However, this does not imply that admission variables are irrelevant. In fact, admission variables exhibited considerable strength, with notable associations. Nevertheless, more closely aligned variables were included in later steps of the model showed stronger associations. When admission variables were retained, they tended to be unique variables reflecting admissions testing (e.g. MCAT). It is well-documented in the literature that the MCAT, specifically the verbal reasoning section, has been linked to performance on licensing exams (Barber et al., 2018; Donnon et al., 2007; Gauer et al., 2018; Gullo et al., 2015; Hanson et al., 2022; Raman et al., 2019). We suspect that the lack of retained admission variables may be partly attributed to range restriction, as the study included only those who were already admitted to medical school, thus limiting the full scale of variation to admitted individuals. This issue has been acknowledged in other health professions research as a limitation when estimating post-admissions performance effects.

In discussion of school level assessments that typically take place in the first 2 years of medical school are also associated to MCCQE1. This was consistent across both inter-institutional analyses of common variables as well as within the context of school-specific analyses. Research (Barber et al., 2018; Kleshinski et al., 2009; White et al., 2009) has found this to be important in considering performance and risk. We may speculate at this stage there may be a strong association between what is being assessed on the MCCQE1. Further we also believe that the process of accreditation may play a role in this strong association to performance on licensing exams and the consistency of this across all schools. Notably, the complementary unique professional competence variable was not associated with MCCQE1. However, this is not to say that professional competence is not important rather that the focus of the MCCQE1 is likely more aligned with content of knowledge-based assessments.

At this stage our results are in line and predominantly support the well cited study by Hecker, & Violato (2009). In their research they found an association between curriculum variables and United States Medical Licensing Examination (USMLE) scores. Similar to our results these associations tended to be smaller in nature when compared to performance that was more proximal to the outcomes. However, there are differences between the Hecker & Violato (2009) and the study presented in this paper for example. They examined the three steps of the USMLE as a composite, while we only examined association to MCCQE1. A closer comparison would be to examine Step1 scores as an outcome and with undergraduate performance as an independent variable. What is important to highlight is the proximity of the assessments and strength of association. Hecker & Violato (2009) found very high correlations between steps 1,2 and 3, suggesting that an outcome on an initial step is a strong predictor of an outcome on the next step.

Similarly in this study, the performance, and workplace-based assessments, which were included in steps 4 and 5 (OSCE1[in only two schools], OSCE2 and clerkship), and tend to take place closer to the MCCQE1 revealed a very strong association to MCCQE1. In some cases, these variables reflected of more than 20% of variance explained. These assessments are a strong indicator of future performance on the MCCQE1. These findings are consistent with research that has shown an association between these variables and the United States Medical Licensing Examination (USMLE) and MCCQE1 (Escudero, 2011; Monteiro et al., 2017). It is notable that the association of performance, workplace-based, and biomedical knowledge to MCCQE1, contribute the bulk of our effects. This pattern of results suggests that the school assessments are well aligned to MCCQE1, which has a predominant focus on biomedical and clinical practice content. While we are making this substantial claim, there exists differences in these associated effects at the school level. Some have more strongly associated OSCEs performances while others have more strongly associated clerkship assessments. Thus, these results and processes ought to be considered in light of improving the quality of local assessments. This may be a particularly meaningful finding for schools that might have put less effort into their school-level OSCEs because of cancelation of Part 2 of the MCCQE. Furthermore, the results ought to be considered in relation to the unique pedagogy that might be offered at each school.

Overall, the objective of this work has been to gain a better understanding of the connections that contribute to high-quality patient care, operating under the assumption that performance on licensing exams is associated with quality care. When considering our individual school analyses, which encompassed a comprehensive range of clerkship variables

and several unique variables, we observed that they collectively accounted for an average of ~51% of the variance in MCCQE1 outcomes. This is comparable to the ~32% variance explained by our analysis of collated interinstitutional data. The difference in the proportions of variance explained may be attributed to the inclusion of unique variables, but it could also be influenced by sample sizes. The larger collated data set, with its greater amount of data, tends to dampen the impact of extreme values, while smaller (school-level) data may exhibit more pronounced extreme values. This phenomenon, often referred to as the "law of small numbers" in the literature, (Tversky, & Kahneman, 1971,) which suggests that effects tend to be overestimated in smaller data sets, whereas an underestimation may occur when examining the full disaggregated data. (McNeish & Stapleton, 2016; Osborne, 2000) Unfortunately, due to ethics and data sharing agreements, schools could not be identified in the collated data set, which would have allowed for a multilevel model approach to account for school-level clustering effects. As it stands, we can minimally account for approximately 32% of the variance with the variables identified in this study. However, individual schools have the potential to explain more of the variance, and with a greater number of higher quality assessments, a larger portion of this variance can be accounted for at both the school level and overall.

Given the significance of these results and recognizing that this Canadian based study is one of the very few multi-institutional studies, it is crucial to consider some of the limitations. Our primary objective in this research study was to initiate progress towards a common goal of enhancing our understanding of school-level data and its association with performance on licensing exams. Despite our initial enthusiasm, we swiftly encountered a lack of established procedural elements that could facilitate multi-institutional educational research in medical schools. Consequently, a significant portion of our collaborative effort involved establishing processes, procedures, policies, and data sharing agreements to enable the group to work collectively towards this shared objective. However, this paper demonstrates the feasibility and documents that schools can indeed collaborate and share educational data to support educational endeavors across regions.

It's important to note that, although the results are mostly confirmatory, the process of collaborating and organizing across seven organizations was not straightforward. Our group initiated these conversations at the 2016 Association for Medical Education in Europe(AMEE) conference in Barcelona, Spain. While we successfully pulled together our initial ideas (see Chahine et al., 2018), secured grants, and engaged in key knowledge mobilization activities (dataconnection.ca), the process of sharing data was not simple. The intensive process required legal advice and clearance from each institution, along with the creation of a comprehensive data governance document and body. However, despite being time-consuming, it's essential to highlight that everyone involved valued the importance of this work, and the issues mainly centered around confirming data protection, maintaining institutional independence, and anticipating possible consequences. It's worth noting that now there is a useful consensus statement on data sharing and big data in health professionals education that may serve as a guide (Kulasegaram et al., 2024).

After we established our collaborative procedure, we then encountered technical limitations related to the lack of commonality across the schools. Demographic and admissions variables displayed consistency, substantial variability was observed in school assessments and educational data at the individual school level. This variation was particularly noticeable in courses typically offered during the first two years of medical school. After

extensive discussion and collaboration among the schools, we decided to create a variable representing biomedical knowledge. However, this level of reduction may have minimized the overall impact of individual courses. This was a substantial limitation as we reduced the number of pre-clerkship variables to only one. While this served the purposes of this study, having more data points between admissions and clerkship could have increased the percentage of variance explained. Yet, given the variability between schools, this was our best, albeit limited, approach to having some data to work with. Furthermore, we were unable to account for courses focused on professional and clinical skills due to how assessments were captured in individual school data systems. Workplace-based assessments showed slight improvement, and we identified some common ground in this area. Nevertheless, incorporating these assessments into a common regression model required significant data manipulation. While presenting the results of this study, we acknowledge the potential compromise to the validity of interpretations that may be made. Therefore, we strongly urge schools to collaborate in developing common frameworks for reporting data related to course-based, performance-based, and workplace undergraduate medical education data.

Having a common framework would not only facilitate multi-institutional educational research studies but also assist in studying the effects of interventions or new programming across schools. The collaboration was framed around the concepts of Big Data (Chahine et al., 2018), aiming to coordinate individual data sets into a much larger integrated data set. While our long-term goals involve integrating education data with healthcare system data, this study only represents a small fraction of the potential possibilities. In addition to the direct implications of these findings, our research process (Grierson et al., 2022) provided us with valuable insights into the challenges and advantages of interinstitutional data sharing in support of education scholarship. However, we encountered significant obstacles at a fundamental level due to the lack of alignment among our individual school systems. Despite the existence of conversations about Big Data in medical education for over a decade (Arora, 2018; Ellaway et al., 2014, 2019; Murdoch & Detsky, 2013; Pusic et al., 2015) our attempt to create a data set encompassing three years of medical school performance and medical exam scores was not flawless. While theoretically we should be capable of creating large-scale educational data sets, we still have a long way to go as a community to achieve these endeavors. Unless substantial efforts are dedicated to facilitating the technical aspects of establishing compatible data sets and implementing common reporting frameworks, while simultaneously respecting individual school autonomy and individuality, medical schools risk falling behind other industries that utilize data and information to enhance quality.

Conclusion

There is immense potential for data to revolutionize educational practices. Our study provides a glimpse into the possibilities of generalizing the factors that influence licensure exams. Through collaboration and data collation, we were able to explore the cumulative effects of education data in predicting licensing performance. However, our results also reveal what is missing. Specifically, we lacked assessment data with significant pre-clerkship effects. This deficiency does not stem from a lack of assessments, but rather from the challenge of scaling high quality assessment data for aggregate purposes. With the advancement of technology in medical school programs, scaling assessments becomes a viable

option. This paper can serve as a guide, not only in terms of its results, but also in outlining the process. It highlights the potential for schools to work together and create large-scale, multi-institutional data sets that can be utilized to transform education.

Declarations

References

Adams, J., & Neville, S. (2020). Program evaluation for health professionals: What it is, what it isn't and how to do it. *International Journal of Qualitative Methods*, *19*, 1609406920964345.

Arora, V. M. (2018). Harnessing the power of big data to improve graduate medical education: Big idea or bust? *Academic Medicine*, *93*(6), 833–834.

Asch, D. A., Nicholson, S., Sriniva, S., Herrin, J., & Epstein, A. J. (2009). *Evaluating Obstetrical Residency Programs Using Patient Outcomes JAMA*.;302(12):1277–1283.

Barber, C., Hammond, R., Gula, L., Tithecott, G., & Chahine, S. (2018). In search of black swans: Identifying students at risk of failing licensing examinations. *Academic Medicine*, *93*(3), 478–485.

Barber, C., Burgess, R., Mountjoy, M., Whyte, R., Vanstone, M., & Grierson, L. (2022). Associations between admissions factors and the need for remediation. *Advances in Health Sciences Education*, *27*(2), 475–489.

Casey, P. M., Palmer, B. A., Thompson, G. B., Laack, T. A., Thomas, M. R., Hartz, M. F., & Grande, J. P. (2016). Predictors of medical school clerkship performance: A multispecialty longitudinal analysis of standardized examination scores and clinical assessments. *BMC Medical Education*, *16*(1), 1–8.

Chahine, S., Kulasegaram, K. M., Wright, S., Monteiro, S., Grierson, L. E., Barber, C., & Touchie, C. (2018). A call to investigate the relationship between education and health outcomes using big data. *Academic Medicine*, *93*(6), 829–832.

College of Family Physicians of Canada, Working Group on Curriculum Review. CanMEDS-Family Medicine: a framework of competencies in family medicine. Mississauga, ON: College of Family Physicians of Canada (2009). Available from: www.cfpc.ca/uploadedFiles/Education/CanMeds%20FM%20Eng.pdf. Accessed 2023 Jun 21.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Coumarbatch, J., Robinson, L., Thomas, R., & Bridge, P. D. (2010). Strategies for identifying students at risk for USMLE step 1 failure. *Family Medicine*, *42*(2), 105–110.

Cuddy, M. M., Swanson, D. B., & Clauser, B. E. (2007). A multilevel analysis of the relationships between examinee gender and United States Medical Licensing exam (USMLE) step 2 CK content area performance. *Academic Medicine*, *82*(10), S89–S93.

Cuddy, M. M., Swygert, K. A., Swanson, D. B., & Jobe, A. C. (2011). A multilevel analysis of examinee gender, standardized patient gender, and United States medical licensing examination step 2 clinical skills communication and interpersonal skills scores. *Academic Medicine*, *86*(10), S17–S20.

De Champlain, A. F., Ashworth, N., Kain, N., Qin, S., Wiebe, D., & Tian, F. (2020). Does pass/fail on medical licensing exams predict future physician performance in practice? A longitudinal cohort study of Alberta physicians. *Journal of Medical Regulation*, *106*(4), 17–26.

Deng, F., Gluckstein, J. A., & Larsen, D. P. (2015). Student-directed retrieval practice is a predictor of medical licensing examination performance. *Perspectives on Medical Education*, *4*, 308–313.

Donnon, T., Paolucci, E. O., & Violato, C. (2007). The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine*, *82*(1), 100–106.

Dore, K. L., Reiter, H. I., Kreuger, S., & Norman, G. R. (2017). CASPer, an online pre-interview screen for personal/professional characteristics: Prediction of national licensure scores. *Advances in Health Sciences Education*, *22*(2), 327–336.

Ellaway, R. H., Pusic, M. V., Galbraith, R. M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical Teacher*, *36*(3), 216–222.

Ellaway, R. H., Topps, D., & Pusic, M. (2019). Data, big and small: Emerging challenges to medical education scholarship. *Academic Medicine*, *94*(1), 31–36.

Escudero, C. (2011, September). Do the National Board of Medical Examiners (NBME) subject examinations predict success on the MCCQE part I. In *The International Conference on Residency Education*.

Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education*, *43*, 767–775.

Eva, K. W., Reiter, H. I., Rosenfeld, J., Trinh, K., Wood, T. J., & Norman, G. R. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *Jama*, *308*(21), 2233–2240.

Fowler, N., Oandasan, I., & Wyman, R. (Eds.). (2022). *Preparing our Future Family Physicians. An educational prescription for strengthening health care in changing times*. College of Family Physicians of Canada.

Frank, J. R., Snell, L. S., Cate, O. T., Holmboe, E. S., Carraccio, C., Swing, S. R., & Harris, K. A. (2010). Competency-based medical education: Theory to practice. *Medical Teacher*, *32*(8), 638–645.

Gauer, J. L., & Jackson, J. B. (2018). Relationships of demographic variables to USMLE physician licensing exam scores: A statistical analysis on five years of medical student data. *Advances in Medical Education and Practice*, 39–44.

Gohara, S., Shapiro, J. I., Jacob, A. N., Khuder, S. A., Gandy, R. A., Metting, P. J., & Kleshinski, J. (2011). Joining the conversation: Predictors of success on the United States Medical Licensing examinations (USMLE). *Learning Assistance Review*, *16*(1), 11–20.

Grierson, L. E., Mercuri, M., Brailovsky, C., Cole, G., Abrahams, C., Archibald, D., & Schabort, I. (2017). Admission factors associated with international medical graduate certification success: A collaborative retrospective review of postgraduate medical education programs in Ontario. *Canadian Medical Association Open Access Journal*, *5*(4), E785–E790.

Grierson, L., Kulasegaram, K., Button, B., Lee-Krueger, R., McNeill, K., Youssef, A., & Cavanagh, A. Data-Sharing in Canadian Medical Education Research: Recommendations and Principles for Governance. Hamilton, Ontario: Jan. 6, 2022. Available from: www.dataconnection.ca.

Gullo, C. A., McCarthy, M. J., Shapiro, J. I., & Miller, B. L. (2015). Predicting medical student success on licensure exams. *Medical Science Educator*, *25*, 447–453.

Hecker, K., & Violato, C. (2009). Medical school curricula: Do curricular approaches affect competence in medicine. *Family Medicine*, *41*(6), 420–426.

Holmboe, E. S., & Batalden, P. (2015). Achieving the desired transformation: Thoughts on next steps for outcomes-based medical education. *Academic Medicine*, *90*(9), 1215–1223.

Janssen, A., Kay, J., Talic, S., Pusic, M., Birnbaum, R. J., Cavalcanti, R., & Shaw, T. (2022). Electronic Health Records that Support Health Professional reflective practice: A missed opportunity in Digital Health. *Journal of Healthcare Informatics Research*, *6*(4), 375–384.

Kawasumi, Y., Ernst, P., Abrahamowicz, M., & Tamblyn, R. (2011). Association between physician competence at licensure and the quality of asthma management among patients with out-of-control asthma. *Archives of Internal Medicine*, *171*(14), 1292–1294.

Kimbrough, M. K., Thrush, C. R., Barrett, E., Bentley, F. R., & Sexton, K. W. (2018). Are surgical milestone assessments predictive of in-training examination scores? *Journal of Surgical Education*, *75*(1), 29–32.

Klamen, D. L., Williams, R. G., Roberts, N., & Cianciolo, A. T. (2016). Competencies, milestones, and EPAs–Are those who ignore the past condemned to repeat it? *Medical Teacher*, *38*(9), 904–910.

Kleshinski, J., Khuder, S. A., Shapiro, J. I., & Gold, J. P. (2009). Impact of preadmission variables on USMLE step 1 and step 2 performance. *Advances in Health Sciences Education*, *14*, 69–78.

Kulasegaram, K., Grierson, L., Barber, C., Chahine, S., Chou, F. C., Cleland, J., Ellis, R., Holmboe, E. S., Pusic, M., Schumacher, D., Tolsgaard, M. G., Tsai, C., Wenghofer, E., & Touchie, C. (2024). Data sharing and big data in health professions education: Ottawa consensus statement and recommendations for scholarship. *Medical Teacher*, 1–15.

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, *20*, 1–20.

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, *28*, 295–314.

Medical Council of Canada (2023). Medical Council of Canada Qualifying Examination Part 1. Available from https://mcc.ca/examinations/mccqe-part-i/. Accessed 2023 Jun 21.

Monteiro, K. A., George, P., Dollase, R., & Dumenco, L. (2017). Predicting United States Medical Licensure Examination Step 2 clinical knowledge scores from previous academic indicators. *Advances in Medical Education and Practice*, 385–391.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama*, *309*(13), 1351–1352.

Norcini, J., Grabovsky, I., Barone, M. A., Anderson, M. B., Pandian, R. S., & Mechaber, A. J. (2023). The associations between United States Medical Licensing Examination Performance and outcomes of Patient Care. *Academic Medicine*, 10–1097.

Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment Research and Evaluation*, *7*(1), 1.

PhillipsJr, R. L., George, B. C., Holmboe, E. S., Bazemore, A. W., Westfall, J. M., & Bitton, A. (2022). Measuring graduate medical education outcomes to honor the social contract. *Academic Medicine*, *97*(5), 643–648.

Pusic, M. V., Boutis, K., Hatala, R., & Cook, D. A. (2015). Learning curves in health professions education. *Academic Medicine*, *90*(8), 1034–1042.

Reiter, H. I., Eva, K. W., Rosenfeld, J., & Norman, G. R. (2007). Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education*, *41*(4), 378–384.

Rubright, J. D., Jodoin, M., & Barone, M. A. (2019). Examining demographics, prior academic performance, and United States Medical Licensing Examination scores. *Academic Medicine*, *94*(3), 364–370.

Salim, S. Y., & White, J. (2018). Swimming in a tsunami of change. *Advances in Health Sciences Education*, *23*, 407–411.

Santen, S. A., Hamstra, S. J., Yamazaki, K., Gonzalo, J., Lomis, K., Allen, B., Lawson, L., Holmboe, E. S., Triola, M., George, P., Gorman, P. N., & Skochelak, S. (2021). Assessing the transition of training in health systems science from undergraduate to graduate medical education. *Journal of Graduate Medical Education*, *13*(3), 404–410.

Statistics Canada Postal Code Conversion File (PCCF) https://www150.statcan.gc.ca/n1/en/catalogue/92-154-X.

Swygert, K. A., Cuddy, M. M., van Zanten, M., Haist, S. A., & Jobe, A. C. (2012). Gender differences in examinee performance on the step 2 clinical skills data gathering (DG) and patient note (PN) components. *Adv Health Sci Educ Theory Pract*, *17*, 557–571.

Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D., et al. (2007). Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *Journal of the American Medical Association*, *298*(9), 993–1001.

Tannenbaum, D., Konkin, J., Parsons, E., Saucier, D., Shaw, L., Walsh, A. (2011). Triple C competency-based curriculum. Report of the Working Group on Postgraduate Curriculum Review—Part 1. Mississauga, ON: College of Family Physicians of Canada; Available from:www.cfpc.ca/uploadedFiles/Education/_PDFs/WGCR_TripleC_Report_English_Final_18Mar11.pdf. Accessed 2023 Jun 21.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105.

Violato, C., & Donnon, T. (2005). Does the Medical College Admission Test predict clinical reasoning skills? A longitudinal study employing the Medical Council of Canada clinical reasoning examination. *Academic Medicine*, *80*(10), S14–S16.

Wenghofer, E. F., Williams, A. P., & Klass, D. J. (2009). Factors affecting physician performance: Implications for performance improvement and governance. *Healthcare Policy*, 5(2), e141.

White, C. B., Dey, E. L., & Fantone, J. C. (2009). Analysis of factors that predict clinical performance in medical school. *Advances in Health Sciences Education*, *14*, 455–464.

Whitehead, C., & Kuper, A. (2017). Faith–based medical education. *Adv Health Sci Educ Theory Pract*, *22*(1), 1–3.

Authors and Affiliations

S. Chahine[1] · I. Bartman[2] · K. Kulasegaram[3,4] · D Archibald[5,6] · P. Wang[7] · C. Wilson[7] · B. Ross[8] · E. Cameron[8,9] · J. Hogenbirk[10,11] · C. Barber[12,13] · R. Burgess[12,14] · E. Katsoulas[1] · C. Touchie[15] · L Grierson[12,16]

✉   S. Chahine
    saad.chahine@queensu.ca

    I. Bartman
    ilona@mcc.ca

    K. Kulasegaram
    mahan.kulasegaram@utoronto.ca

    D Archibald
    darchibald@bruyere.org

    P. Wang
    Peter.Wang@lhsc.on.ca

    C. Wilson
    Claire.Parent@lhsc.on.ca

    B. Ross
    bross@nosm.ca

    E. Cameron
    ercameron@nosm.ca

    J. Hogenbirk
    jhogenbirk@laurentian.ca

    C. Barber
    c.barber@maastrichtuniversity.nl

    R. Burgess
    raquel.burgess@yale.edu

    E. Katsoulas
    eleni.katsoulas@queensu.ca

    C. Touchie
    CTOUCHIE@toh.ca

    L Grierson
    griersle@mcmaster.ca

[1]   Queen's University, Kingston, Canada

[2]   Medical Council of Canada, Ottawa, Canada

[3]   Wilson Centre, University of Toronto, Toronto, Canada

[4]   Department of Family & Community Medicine Temerty Faculty of Medicine, University of Toronto, Toronto, Canada

5   Department of Family Medicine, University of Ottawa, Ottawa, Canada

6   Bruyère Research Institute, Ottawa, Canada

7   Western University, London, Canada

8   North Ontario School of Medicine University, Thunder Bay, Canada

9   Dr. Gilles Arcand Centre for Health Equity, NOSM University, Thunder Bay, Canada

10   Centre for Rural and Northern Health Research, Laurentian University, Greater Sudbury, Canada

11   Postgraduate Education and Health Sciences, NOSM University, Thunder Bay, Canada

12   Department of Family Medicine, McMaster University, Hamilton, Canada

13   School of Health Professions Education (SHE), Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

14   Department of Social and Behavioral Sciences, Yale School of Public Health, Yale University, New Haven, USA

15   Departments of Medicine and of Innovation in Medical Education, University of Ottawa, Ottawa, Canada

16   McMaster Education Research, Innovation, and Theory Program, McMaster University, Hamilton, Canada