
Does Pass/Fail on Medical Licensing Exams Predict Future Physician Performance in Practice?

A Longitudinal Cohort Study of Alberta Physicians

André F. De Champlain, PhD; Nigel Ashworth, MB.ChB. MSc; Nicole Kain, MPA, PhD; Sirius Qin, MSc; Delaney Wiebe, MPH; Fang Tian, PhD

ABSTRACT: The purpose of this longitudinal study was to gather extrapolation evidence of validity by assessing whether performance on a national medical licensing exam, in addition to practice and socio-demographic variables, is predictive of future physician performance in practice. The study focused on a cohort of 3,404 physicians who were registered with the College of Physicians and Surgeons of Alberta (CPSA) and who completed the Medical Council of Canada Qualifying Examination (MCCQE) Parts I and II between 1992–2017. Separate multivariate quasi-Poisson regression models were run to assess the degree of relationship between first-time pass/fail status on the MCCQE I and II, and several CPSA socio-demographic variables and several CPSA socio-demographic variables, in addition to complaints/physician and various prescribing flags. Candidates who failed the MCCQE I on their first attempt had 27% more complaints lodged against them, compared to those who passed. Physicians who failed the MCCQE II on their first attempt prescribed 2+ benzodiazepines and 2+ opioids to 30% more patients than those who passed. Conclusions: Performance on the MCCQE Part I and II is an important predictor of physician performance. Combined with other critical variables, these measures provide important evidence to aid in risk modeling efforts and to guide educational interventions for physicians at an early stage of their careers.

Introduction

The primary intent of any medical licensing examination is to assure the public that the licensee has demonstrated the core knowledge and skills (competency) necessary for safe and effective performance in practice. In Canada, one requirement for entry into independent practice is the successful completion of the *Medical Council of Canada Qualifying Examination (MCCQE)*. This bipartite program includes the MCCQE Part I, usually completed prior to entering post-graduate training, and the MCCQE Part II, a requirement for entering independent practice in Canada.

Using Kane's argument-based approach to validation, several sources of evidence to support the plausibility of key claims within the framework of all licensing examination programs can be collected.¹⁻³

Recent interest in exploring whether medical licensing exams can be used to inform secondary interpretations, akin to "off-label" uses of medications in clinical medicine, has emerged; and the use of exams to predict future practice performance has, in particular, been expressed.⁴ Using Kane's validation framework, the latter falls under the *extrapolation* argument. To what extent does performance on an examination, in a circumscribed

number of domains, extrapolate to the full complement of physician performance? We might posit that an examinee's communication score generalizes to communication skills in a breadth of settings as a practicing physician, including end-of-life, patient education and other situations. Such an argument's complexity underscores the importance of clearly delineating the score- or decision-based inference with provision of pertinent supporting (empirical) evidence.

Most stakeholders assume that competence, as reflected by a medical licensing exam score, predicts quality of care and performance in actual practice.⁵ However, in the absence of clear arguments and supporting evidence, the latter reduces to a highly fallible claim or in the worst case, mere conjecture.

Previous research explored whether performance on medical licensing exams extrapolated to key educational and clinical outcomes. One investigation followed more than 200 practicing physicians and reported that candidates who performed poorly on the MCCQE had a greater than threefold increase in the risk of an unacceptable quality-of-care peer assessment outcome.⁶ Another study tracked the performance of more than 2000 United States Medical Licensing Examination (USMLE) candidates

and found that performance on the Step 2 Clinical Skills data interpretation sub-score predicted post-graduate ratings in history-taking and physical examination.⁷ Finally, a Canadian investigation reported that MCCQE Part I scores, for a sample of international medical graduates, were a significant predictor of subsequent performance on a family medicine specialty certification examination.⁸

Another study, focusing on more than 3,400 physicians who completed the MCCQE between 1993 and 1996, reported that candidates who performed poorly on the communication sub-score

TO WHAT EXTENT DOES PERFORMANCE ON AN EXAMINATION, IN A CIRCUMSCRIBED NUMBER OF DOMAINS, EXTRAPOLATE TO THE FULL COMPLEMENT OF PHYSICIAN PERFORMANCE?

of the MCCQE Part II had a significantly higher number of patient complaints.⁹ Similarly, one study reported that a one-standard-deviation increase in Step 2 Clinical Knowledge scores was associated with a 25% decrease in the odds of receiving a patient complaint.¹⁰

Another investigation explored the relationship between scores on the College of Family Physicians of Canada certification examination and health-care resource usage by primary care physicians in Québec,¹¹ noting that exam scores were significantly associated with consultations and mammography screening rates, as well as prescribing behaviors. Similarly, an additional study found that higher scores on the MCCQE Part I were linked to better primary care clinical outcomes.¹²

While helpful in supporting the *extrapolation argument* to validation, some of these past investigations suffer from a number of methodological limitations. Chief among these are the use of: 1) un-equated scores, namely, scores that were unadjusted to account for varying levels of difficulty of test forms analyzed; and (2) unreliable sub-scores in the predictive models.^{6,9,11-12} Furthermore, most of the latter investigations were carried out on licensing exams that are more than 25 years old and no longer reflect the current state of medicine and medical education. Clearly, there is a need to replicate and extend past studies.

Our longitudinal study expands upon earlier work by investigating the extent to which physician performance

on the MCCQE predicts patient complaints, as well as prescribing behavior for a sample of physicians registered with the College of Physicians and Surgeons of Alberta (CPSA). Specifically, our study focused on answering the following questions:

- (1) Does first-time pass/fail standing on the MCCQE Part I and/or Part II predict overall patient complaint rates for a sample of CPSA-registered physicians?
- (2) Does first-time pass/fail standing on the MCCQE Part I and/or Part II predict opioid and benzodiazepine prescribing behavior for this sample?

Since earlier forms of the MCCQE were not fully equated (prior to 2013), it was not possible to use actual exam scores as predictors in our models. We therefore restricted our analyses to Pass/Fail (P/F) standing on the first attempt for each exam in our various statistical models. This measure does provide some differentiation, as typically 10% to 20% of candidates might not succeed on either exam on their first attempt.

Rates of patient complaints and potentially harmful patterns in opioid and benzodiazepine prescribing were used as proxy measures of physician performance in practice. Alberta has some of the highest rates of prescribed opioids in the world, and, unsurprisingly, high rates of opioid overdoses and deaths.¹³⁻¹⁴ Alberta also has high rates of benzodi-

OUR LONGITUDINAL STUDY EXPANDS UPON EARLIER WORK BY INVESTIGATING THE EXTENT TO WHICH PHYSICIAN PERFORMANCE ON THE MCCQE PREDICTS PATIENT COMPLAINTS, AS WELL AS PRESCRIBING BEHAVIOR FOR A SAMPLE OF PHYSICIANS...

azepine prescriptions, including to individuals over 65 years of age who are vulnerable to potentially harmful consequences of these medications.^{14,15} The CPSA houses a database of complaint information for all registered physicians in the province. Opioids and benzodiazepines are medications with high potential for misuse and/or diversion; prescriptions have been monitored by the CPSA since 1986 and 2015, respectively.^{14,16}

We also included CPSA practice and socio-demographic variables in our models, such as years since medical school graduation; country of medical

training; specialty or type of practice; and gender; as such outcomes have been detailed as potential factors relating to physician performance.¹⁷⁻¹⁸ Ethics approval was obtained from the University of Alberta's Health Research Ethics Board — Health Panel (Study ID: Pro00079279).

Method

Study sample

Our study sample included a matched group of 3,404 physicians who were registered to practice in Alberta within the past 15 years and who completed the MCCQE Parts I and II between

ALBERTA HAS SOME OF THE HIGHEST RATES OF PRESCRIBED OPIOIDS IN THE WORLD, AND, UNSURPRISINGLY, HIGH RATES OF OPIOID OVERDOSES AND DEATHS.

1992–2017. We excluded data that could not be matched from the study (for any reason). For the purposes of our study, family medicine and general practitioners were grouped together as “family medicine (FM).” “Primary specialty (PS)” included internal medicine, dermatology, pediatrics and obstetrics/gynecology. “Other specialty (OS)” included remaining physicians. Table 1 provides a breakdown of our sample.

As outlined in Table 1, our sample primarily included males (57.1%), Canadian medical graduates (60.8%) and FM physicians (48.8%). The mean age of our sample was 43.3 (SD=7.6) with a mean number of years in practice of 11.7 years (SD=4.0). The mean age of physicians in Alberta is slightly lower than the Canadian average (in 2019, the Canadian average was 49.4 years)¹⁹ likely owing to a younger population in that province compared to other Canadian jurisdictions.

The MCCQE Part I and Part II

The MCCQE Part I measures the medical knowledge and clinical decision-making skills at a level expected of a physician entering supervised practice (residency) in Canada. The exam, for the years under study, was comprised of 196 multiple-choice items and up to 55 clinical decision-making cases. These cases were further subdivided into short-menu (pick-*N*) items, as well as short-answer constructed-response items. The MCCQE Part I was a paper-and-pencil exam until the year 2000, at which point it began to be computer-delivered.

The MCCQE Part II assesses clinical skills at a level expected of a physician entering independent practice in Canada. Exam forms included in the present study assessed data gathering, physical exam skills, communication skills, and considerations of cultural communication, legal, ethical and organizational aspects of medical practice. Though

Table 1
Breakdown of Study Sample: Physicians Registered to Practice in Alberta, 1992–2017
 (N = 3,404)
 by Gender, CMG/IMG Status, Specialty Certification, Age and Years Since Registration

	Number	Group	Percent		
Gender*	3,402	Female	1,458 (42.9%)		
		Male	1,944 (57.1%)		
CMG/IMG	3,404	CMG ^a	2,069 (60.8%)		
		IMG ^b	1,335 (39.2%)		
Specialty certification	3,404	Family medicine	1,662 (48.8%)		
		Primary specialty	747 (21.9%)		
		Other specialty	995 (29.2%)		
	Number	Mean	SD	Min	Max
Male	71	29	18%	100	100
Female	39	21	13%	60	60

*=Two physicians in the study group did not specify a gender.

a=Canadian medical graduate

b=International medical graduate

the number of stations varied in the early offerings of the MCCQE Part II, they remained fixed at 12 stations as of 2003.

Decision accuracy and consistency estimates, appropriate to report with criterion-referenced exams where the reliability of a pass/fail decision is of paramount importance, varied between the high 0.80s and low 0.90s for the test forms under study.

CPSA registration variables

In addition to exam performance, the following 13 CPSA registration variables were also included in our regression models, based on previous related research and evaluation:

- (1) Years since registration
- (2) Specialty (FM, PS, OS)
- (3) Days per week providing medical services
- (4) Canadian/International Medical Graduate (CMG/IMG)
- (5) Accepting new patients
- (6) Gender
- (7) Age
- (8) Performing procedures requiring sedation/ anesthesia
- (9) Use of an electronic medical record (EMR)
- (10) Hospital privileges
- (11) Practicing exclusively as a locum
- (12) Teaching in a non-clinical environment
- (13) Group practice.⁵⁻⁶

Outcomes

The following four outcome variables were selected for our models:

- (1) Total number of patient complaints
- (2) Total number of patients for whom 90+ Oral Morphine Equivalents (90 OME) were prescribed
- (3) Total number of patients for whom three times the defined daily dose of benzodiazepines (3xDDD) was prescribed
- (4) Total number of patients for whom 2+ opioids and 2+ benzodiazepines (2+2) were prescribed

Analyses

Separate multivariate quasi-Poisson regression models were run for each outcome variable. We were unable to run common Poisson regression models due to over-dispersion of the count data for one of the models (complaints) and under-dispersion for another model (2+2). The Poisson model only

FINDINGS SHOW THAT THE FIRST-ATTEMPT P/F STANDING ON THE MCCQE PART I WAS A SIGNIFICANT PREDICTOR OF THE TOTAL NUMBER OF PATIENT COMPLAINTS.

has one free parameter and therefore does not allow variance to be adjusted independently of the mean. In quasi-Poisson modeling, the variance is a linear function of the mean.²⁰

Results

Table 2 provides a breakdown of first-attempt P/F rates on each exam, available for 3,283/3,404 (96.5%) candidates on the MCCQE Part I and 2,909/3,404 (85.5%) on the MCCQE Part II. Respectively, 87.8% and 83% of our sample passed the MCCQE Part I and II on their first attempt.

For two out of the four multivariate quasi-Poisson regression models, MCCQE first-time P/F standing was a significant predictor. Results for these two

Table 2
First-attempt Pass/Fail Rates on the MCCQE Part I and MCCQE Part II for Study Sample: Physicians Registered to Practice in Alberta, 1992–2017 (N=3,404)

Exam	Number	Group	Percent
MCCQE Part I (first attempt)	3,283	Fail	401 (12.2%)
		Pass	2,882 (87.8%)
MCCQE Part II (first attempt)	2,909	Fail	495 (17.0%)
		Pass	2,414 (83.0%)

models are provided in the next section of the paper.

Table 3 provides the results of the multivariate quasi-Poisson regression model, predicting the total number of patient complaints. Regression parameter estimates, incidence rate ratios (IRR), and Holm-Bonferroni corrected Type I error rates (nominal Type I error rate = 0.05) to control for multiplicity effects, are provided. For a binary variable, the IRR is the ratio of the number of events in one category to the number of events in the other category. For a categorical variable (with more than two categories), the IRR is the ratio of the expressed category to the base category.

Findings show that the first-attempt P/F standing on the MCCQE Part I was a significant predictor of the total number of patient complaints. Candidates who failed the MCCQE Part I on their first attempt, on average, received 27% more complaints than passing candidates (mean number of complaints [SD] per physician was 1.34 [2.07] and 0.62 [1.37], respectively, for those who failed and passed the MCCQE Part I on their first attempt), controlling for all other variables in the model.

However, passing or failing the MCCQE Part II on the first attempt was not significantly related to patient complaint rates.

All 13 CPSA registration variables were significantly associated with a higher frequency of patient complaints. Specifically, a larger number of complaints was associated with:

- (1) A higher number of years since initial registration (7% more complaints per year)
- (2) FM (over double the number of complaints compared to OS physicians)
- (3) More days of providing medical services per week (26% more complaints per day)
- (4) IMGs (44% more complaints than CMGs)
- (5) Accepting new patients (33% more complaints than physicians not accepting new patients)
- (6) Male physicians (33% more complaints than female physicians)
- (7) Older physicians (2% more complaints per year of age)

Table 3
Predicting the Total Number of Patient Complaints for Study Sample of Physicians Registered to Practice in Alberta, 1992–2017 (N=3,404)
 Results of a Multivariate Quasi-Poisson Regression Model

Predictor	β	IRR ^a	Type I Error ^b
Years since registration	0.07	1.07	<0.001
Specialty—family medicine	0.78	2.17	<0.001
Days/week providing medical services	0.23	1.26	<0.001
Canadian medical graduate	-0.59	0.56	<0.001
Not accepting new patients for office practice in Alberta	-0.40	0.67	<0.001
Female	-0.40	0.67	<0.001
Age	0.02	1.02	<0.001
Does not perform procedures that require sedation/Local Anesthetic/General Anesthetic	-0.32	0.73	<0.001
MCCQE Part I first attempt Fail standing	0.24	1.27	<0.001
Does not use an EMR system	-0.27	0.77	0.001
Does not have privileges in any capacity	0.20	1.22	0.001
Does not practice exclusively as a locum	0.51	1.66	0.003
Teaching with no provision of medical services*	0.01	1.01	0.009
Specialty—Primary specialty	-0.23	0.79	0.019
Does not share patient information with other physicians	0.14	1.15	0.033

a=Incidence rate ratio

b=Holmes-Bonferroni adjusted Type I error rates

*=e.g., non-clinical teaching environment

- (8) Performing procedures requiring sedation/ anaesthesia (27% more complaints for those physicians performing procedures requiring sedation/anaesthesia)
- (9) Using an EMR (23% more complaints for those physicians using an EMR)
- (10) Not having hospital privileges (22% more complaints for those physicians not having privileges)
- (11) Not practicing exclusively as a locum (66% more complaints for those physicians not practicing exclusively as a locum)
- (12) Non-clinical teaching (1% more complaints for those physicians that teach with no provision of medical services)
- (13) PS physicians (21% fewer complaints than OS physicians)
- (14) Not practicing in a group (15% more complaints for those physicians who did not practice in a group, e.g., solo practitioners)

Table 4 provides the results of the multivariate quasi-Poisson regression model predicting the total number of patients for whom 2+2 were prescribed.

Results show that the first-attempt P/F standing on the MCCQE Part II was a significant predictor

of this outcome variable. Candidates who failed the MCCQE Part II on their first attempt, on average, prescribed 2+2 to 30% more patients than passing candidates (mean number of patients for whom 2+2 were prescribed [SD] was 0.27 [1.11] and 0.11 [0.54], respectively, for those who failed and passed the MCCQE Part II on their first attempt.) Note that passing or failing the

OF INTEREST IS THAT THOSE PHYSICIANS USING EMRs (ELECTRONIC MEDICAL RECORDS) HAD A HIGHER RATE OF PATIENT COMPLAINTS.

MCCQE Part I on the first attempt was not related to this prescribing outcome.

Ten out of the 14 CPSA registration variables were also significantly associated with this outcome. Prescribing 2+2 to a higher number of patients was related to:

- (1) More days of providing medical services per week (48% more patients were prescribed 2+2 for each day of providing services)
- (2) FM (over 79 times the number of patients were prescribed 2+2 compared to OS physicians)

Table 4
Predicting Total Number of Patients Prescribed 2+ Opioids and 2+ Benzodiazepines for Study Sample of Physicians Registered to Practice in Alberta, 1992–2017 (N=3,404)
 Results from a Multivariate Quasi-Poisson Regression Model

Predictor	β	IRR ^a	Type I Error ^b
Days/week providing medical services	0.39	1.48	<0.001
Specialty—family medicine	4.37	79.32	<0.001
Female	-0.40	0.67	<0.001
Does not perform procedures that require sedation/Local Anesthetic/General Anesthetic	0.85	2.34	<0.001
Canadian medical graduate	-0.37	0.69	<0.001
Not accepting new patients for office practice in Alberta	-0.30	0.74	<0.001
Age	-0.02	0.98	<0.001
Does not practice exclusively as a locum	2.63	13.94	<0.001
Teaching with no provision of medical services*	-0.06	0.95	0.001
Does not use an EMR system	-0.47	0.63	0.001
MCCQE Part II first attempt fail standing	0.26	1.30	0.002

a=Incidence rate ratio

b=Holmes-Bonferroni adjusted Type I error rates

*=e.g., non-clinical teaching environment

- (3) Male physicians (33% more patients were prescribed 2+2)
- (4) Performing procedures requiring sedation/ anaesthesia (over twice the number of patients were prescribed 2+2)
- (5) Being an IMG (31% more patients were prescribed 2+2 than CMG)
- (6) Accepting new patients (26% more patients were prescribed 2+2)
- (7) Younger physicians (for each decrease in one year of age, 2% more patients were prescribed 2+2)
- (8) Not practicing exclusively as a locum (14 times more patients were prescribed 2+2)
- (9) Teaching with provision of medical services (5% more patients were prescribed 2+2)
- (10) Using an EMR (37% more patients were prescribed 2+2)

Discussion

The severity of the opioid crisis that has prevailed over the past two decades cannot be understated. In 2017, 47,600 U.S. residents, or more than 130 per day, died of overdoses related to both illicit and prescription opioids.²¹ Though perhaps less visible in the public press, benzodiazepine prescriptions have also skyrocketed. In a study using National Survey on Drug Use and Health data from 2015-16, Maust, Lin and Blow reported that 30.6 million U.S. adults disclosed using benzodiazepines during the study period.²² More alarming was misuse, which accounted for 20% of overall use.

In light of this public health emergency, our investigation was aimed at gathering evidence to assess whether performance on licensing exams, in addition to a number of in-practice registration variables, could be informative in predicting future professional outcomes, especially as they pertain to patient complaints as well as opioid and benzodiazepine prescribing behaviors. Our study supports validity arguments that fall beyond the primary aim of any medical licensing examination program but that nonetheless jibe with public expectations. That is, physicians who are licensed to practice medicine presumably can demonstrate the knowledge and skills necessary for safe and effective care. Furthermore, there is an implicit assumption (albeit often unverified) that there is some

relationship between performance on licensing exams and actual practice.⁵

Consequently, gathering evidence to bolster any such extrapolation argument becomes particularly critical; i.e., does performance in core competencies expected of all physicians, irrespective of specialty, as measured by licensing examinations, extrapolate to in-practice behaviors? Though stakeholders readily recognize that licensing exams are designed and used for a primary purpose (i.e., initial entry into independent practice), the extent to which the latter might inform performance in residency, at different stages of independent practice and throughout a physician's career, assumes added importance.

Our study showed that performance on the MCCQE Part I was related to the number(s) of patient complaints. Of interest, complaints seem to be more related to cognitive domains (knowledge, application of knowledge, clinical decision-making, etc.) which form the centerpiece of the MCCQE Part I, compared to the MCCQE Part II, which focuses on communication, professionalism and affective skills. A systematic review of complaints carried out on 59 studies suggests that the latter are equally divided between clinical (quality of care, safety incidents) and management (bureaucratic issues, access and admissions), as well as relationship (communication, caring, patient rights) issues.²³ Most sustained patient complaints in the jurisdiction under study stemmed from patient dissatisfaction with treatment.²⁴ Our findings seem to substantiate this statement. Results noted with CPSA registration variables were largely expected (e.g., higher complaints are associated with older

FROM A REGULATORY PERSPECTIVE, PHYSICIANS WHO ARE UNSUCCESSFUL ON THEIR FIRST ATTEMPT AT MCCQE PART II COULD PERHAPS BE DIRECTED TO CONTINUOUS PROFESSIONAL DEVELOPMENT OPPORTUNITIES RELATED TO PRESCRIBING...

male physicians, trained abroad, etc.).²⁵ Of interest is that those physicians using EMRs had a higher rate of patient complaints. Previous studies have also highlighted this unintended consequence; that is, the doctor-patient relationship can be negatively impacted by the use of an EMR, presumably due to the attention placed on completing

the record at the detriment of communicating with the patient.²⁶⁻²⁷

It is also interesting to note that performance on the MCCQE Part II was significantly associated with potentially dangerous opioid and benzodiazepine prescribing behavior. Given the greater emphasis placed on higher-order cognitive skills, as well as core non-medical competencies (communication, professionalism, etc.), it is likely that this risky prescribing behavior is due to a combination of

FAMILY MEDICINE PHYSICIANS WERE PARTICULARLY AT GREATER RISK OF RECEIVING A PATIENT COMPLAINT AND PRESCRIBING DANGEROUS COMBINATIONS OF OPIOIDS AND BENZODIAZEPINES.

poor clinical reasoning skills as well as deficits in affective domains.²⁸ Many of the same CPSA registration variables previously noted with the patient complaints model were also significantly associated with this risky prescribing behavior. From a regulatory perspective, physicians who are unsuccessful on their first attempt at MCCQE Part II could perhaps be directed to continuous professional development opportunities related to prescribing of these and other potentially harmful medications — as a way to mitigate risk of future poor prescribing habits and negative patient outcomes.

FM physicians were particularly at greater risk of receiving a patient complaint and prescribing dangerous combinations of opioids and benzodiazepines.²⁹⁻³² Logically, this makes sense because in Alberta, family medicine physicians account for almost half of all physicians in that province. They also tend to have more points of contact with patients over longer periods of time (compared to, for example, an anesthesiologist or emergency room physician), perhaps allowing for increased noted risks.

Our study also identified an interesting albeit unanticipated result: Younger physicians were more likely to have at least one patient to whom they prescribed two or more opioids and two or more benzodiazepines. One study in the literature found a positive relationship between physician age and prescribing of non-psychotropic medications, contrary to our results.³³ This appears to be a

complex issue, warranting further investigation as to why younger physicians in Alberta are prescribing more of these potentially dangerous combinations (2+2) to patients.

Finally, from a “practical significance” point of view, it is important to underscore that the value added by including first-attempt P/F standing on the MCCQE Part I and II is largely comparable to that associated with most CPSA registration variables. Though some of the latter measures were more strongly associated with patient complaints and 2 + 2 benzo/opioid prescribing behavior (e.g.: Specialty — Family Medicine for both outcomes), the incidence rate ratio values for first-time P/F standing were higher than 12 out of the remaining 13 CPSA registration variables for the complaints outcome and six out the remaining 10 CPSA predictors for the 2 + 2 benzo/opioid prescribing outcome, for the MCCQE Parts I and II, respectively.

Our study is not without limitations. First, only those physicians who eventually passed the MCCQE Part I and II were included in our analyses. Restriction of range and other effects impacted the findings reported in this study. Second, with regard to data from the CPSA, the sample retained for our investigation may not be fully representative of the population of practicing physicians in Alberta. Though we did not denote any major deviations,

OUR STUDY ALSO IDENTIFIED AN INTERESTING ALBEIT UNANTICIPATED RESULT: YOUNGER PHYSICIANS WERE MORE LIKELY TO HAVE AT LEAST ONE PATIENT TO WHOM THEY PRESCRIBED TWO OR MORE OPIOIDS AND TWO OR MORE BENZODIAZEPINES.

it would nonetheless be important to formally undertake this comparison as it impacts the generalizability of our findings. The CPSA complaints data includes *all* complaints — including those that may have been dismissed soon after receipt. Normally these dismissed grievances account for approximately 25% to 50% of all complaints received by CPSA, so our results may indicate a modest inflation of actual complaints that are processed. Third, we were unable to model actual MCCQE Part I and II scores as predictors, given the absence of formal score equating prior to 2013, leading to a loss of information associated with

using first-time P/F status. It was not possible to identify a large enough sample (post-2013) to complete score-based analyses in the present investigation. However, present research is replicating this work using actual equated scores. Fourth, though not possible to address in our current study, future research should aim to assess whether any licensing exam-patient complaint/prescribing behavior relationships have changed with the evolution of exam-content to better reflect current health concerns, as well as a greater recognition of these areas by medical practitioners. Fifth, our study was exploratory and to a certain extent, “opportunistic.” Though our findings confirmed expected patterns, future research should be stated in an *ex-ante* manner, as opposed to *ex-post facto*. Confirmatory models postulating effects based on theoretical considerations would greatly strengthen this type of research. Finally, the list of predictors examined

...WE BELIEVE THAT OUR WORK MEANINGFULLY CONTRIBUTES TO PREVIOUS STUDIES BY DEMONSTRATING THAT STANDING ON MEDICAL LICENSING EXAMS CAN PREDICT ASPECTS OF PHYSICIAN PERFORMANCE IN INDEPENDENT PRACTICE; A FORM OF ‘OFF-LABEL’ USE FOR SUCH ASSESSMENTS.

in our initial investigation is by no means comprehensive and should be supplemented with additional measures in any future research.

Despite these caveats, we believe that our work meaningfully contributes to previous studies by demonstrating that standing on medical licensing exams can predict aspects of physician performance in independent practice; a form of “off-label” use for such assessments. However, supplementary replication and extension of this research should be undertaken to further develop and support additional validation arguments. ■

About the Authors

André F. De Champlain, PhD, is Director, Department of Psychometrics and Assessment Services, Medical Council of Canada.

Nigel Ashworth, MB.ChB. MSc, is Senior Medical Advisor, Research and Evaluation Unit, College of Physicians and Surgeons of Alberta.

Nicole Kain, MPA, PhD, is Program Manager, Research and Evaluation Unit, College of Physicians and Surgeons of Alberta.

Sirius Qin, MSc, is Senior Statistical Analyst, Department of Psychometrics and Assessment Services, Medical Council of Canada.

Delaney Wiebe, MPH, is Research Associate, Research and Evaluation Unit, College of Physicians and Surgeons of Alberta.

Fang Tian, PhD, is Senior Psychometrician, Department of Psychometrics and Assessment Services, Medical Council of Canada.

References

1. Kane MT. Validation. In: Brennan R, ed. *Educational Measurement*. 4th ed. Westport, CT: American Council on Education and Praeger; 2006:17–64.
2. Kane M. Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*. 2013a;50:1-73.
3. Kane, M. The Argument-Based Approach to Validation. *School Psychology Review*. 2013b;42:448-457.
4. Norcini J, Boulet J, Opalek A, Dauphinee W. The Relationship Between Licensing Examination Performance and the Outcomes of Care by International Medical School Graduates. *Academic Medicine*. 2014; 89:1157–1162.
5. Rethans J, Sturmans F, Drop R, van der Vleuten C, Hobus P. Does Competence of General Practitioners Predict Their Performance? Comparison Between Examination Setting and Actual Practice. *BMJW*. 1991; 303:1377-1380.
6. Wenghofer E, Klass D, Abrahamowicz M, et al. Doctor Scores on National Qualifying Examinations Predict Quality of Care in Future Practice. *Medical Education*. 2009; 43:1166–1173.
7. Cuddy M, Winward ML, Johnston MM, Lipner RS, Clauser BE. Evaluating Validity Evidence for USMLE Step 2 Clinical Skills Data Gathering and Data Interpretation Scores: Does Performance Predict History-Taking and Physical Examination Ratings for First-Year Internal Medicine Residents? *Academic Medicine*. 2016; 91:133-139.
8. De Champlain A, Streefkerk C, Roy M, Tian F, Qin S, Brailovsky C. Predicting Family Medicine Specialty Certification Status Using Standardized Measures for a Sample of International Medical Graduates Engaged in a Practice-Ready Assessment Pathway to Provisional Licensure. *Journal of Medical Regulation*. 2014; 100:8-16.
9. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities. *JAMA*. 2017; 298:993–1001.
10. Cuddy M, Young A, Gelman A, et al. Exploring the Relationships Between USMLE Performance and Disciplinary Action in Practice: A Validity Study of Score Inferences from a Licensure Examination. *Academic Medicine*. 2017; 92: 1780-1785.

11. Tamblyn R, Abrahamowicz M, Brailovsky C, et al. Association Between Licensing Examination Scores and Resource Use and Quality of Care in Primary Care Practice. *JAMA*. 1998; 280:989–996.
12. Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association Between Licensure Examination Scores and Practice In Primary Care. *JAMA*. 2002; 288:3019–3026.
13. Dong KA, Villa-Roel C, Rowe BH, et al. Community-Based Naloxone: A Canadian Pilot Program. *CMAJ Canadian Journal of Addiction Medicine*. 2012;3(2):4-9.
14. Kain NA, Ashworth NL, Jess EH, Mazurek KA. Opioids, Benzodiazepines and Z-Drugs: Alberta Physicians' Attitudes and Opinions upon Receipt of their Personalized Prescribing Profile. *Journal of Medical Regulation*. 2018;104(4):8-13.
15. Morgan SG, Hunt J, Rioux J, Proulx J, Weymann D, Tannenbaum C. Frequency and Cost of Potentially Inappropriate Prescribing for Older Adults: A Cross-Sectional Study. *CMAJ Open*. 2016;4(2):E346-E51.
16. Dormuth CR, Miller TA, Huang A, Mamdani MM, Juurlink DN. Effect of a Centralized Prescription Network on Inappropriate Prescriptions for Opioid Analgesics and Benzodiazepines. *CMAJ: Canadian Medical Association Journal = Journal De L'association Medicale Canadienne*. 2012;184(16):E852-E6.
17. Yen W, Thakkar N. State of the Science on Risk and Support Factors to Physician Performance: A Report from the Pan-Canadian Physician Factors Collaboration. *Journal of Medical Regulation*. 2019;105(1):6-21.
18. Kain NA, Hodwitz K, Yen W, Ashworth N. Experiential Knowledge of Risk and Support Factors for Physician Performance in Canada: A Qualitative Study. *BMJ open*. 2019;9(2):e023511.
19. Canadian Institute for Health Information. Physicians in Canada, 2019. Ottawa, ON: CIHI; 2020.
20. Gardner W, Mulvey EP, Shaw EC. Regression Analyses of Counts and Rates: Poisson, Over-Dispersed Poisson, and Negative Binomial Models. *Psychological Bulletin*. 1995; 118:392-404.
21. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and Opioid-Involved Overdose Deaths — United States — 2013–2017. *WR Morb Mortal Wkly Rep*. ePub: 21 December 2018.
22. Maust DT, Lin LA, Blow FC. Benzodiazepine Use and Misuse Among Adults in the United States. *Psychiatr Serv*. 2019; 70:97-106.
23. Reader, TW., Gillespie, A., & Roberts, J. Patient Complaints in Healthcare Systems: A Systematic Review and Coding Taxonomy. *BMJ Quality & Safety*. 2014; 23: 678-689.
24. College of Physicians and Surgeons of Alberta. 2014 Annual Report.; 2014:26.
25. Kohatsu ND, Gould D, Ross LK, Ross, PJ. Characteristics Associated with Physician Discipline. *Archives of Internal Medicine*. 2004; 164:653-658.
26. Bowman, S. Impact of Electronic Health Record Systems on Information Integrity: Quality and Safety Implications. *Perspect Health Inf Manag*. 2013; 10:1C.
27. Evans, RS. Electronic Health Records: Then, Now, and in the Future. *Yearb Med Inform*. 2016; 25:s48-s61.
28. Dineen KK, DuBois JM. Between a Rock and a Hard Place: Can Physicians Prescribe Opioids to Treat Pain Adequately While Avoiding Legal Sanction? *Am J Law Med*. 2016; 42:7-52.
29. Wilson M, Randhawa H, Lavis J. Rapid Synthesis: Identifying Risk and Protective Factors for Quality Clinical Practice. *McMaster Health Forum*. 2015.
30. Khaliq A DH, Huang C, Narine L, Smego R. Disciplinary Action Against Physicians: Who is Likely to Get Disciplined? *The American Journal of Medicine*. 2005;118:773-7.
31. Neal D K, Dawn G, Leslie K R, Patrick J F. Characteristics Associated With Physician Discipline: A Case-Control Study. *Archives of Internal Medicine*. 2004(6):653.
32. Alam A, Klemensber J, Griesman J, Bell C. The Characteristics of Physicians Disciplined by Professional Colleges in Canada. *Open MEd*. 2011;5(4):e166-e72.]
33. Hadsall RS, Freeman RA, Norwood GJ. Factors Related to the Prescribing of Selected Psychotropic Drugs by Primary Care Physicians. *Social Science and Medicine*. 1982; 16(20):1747-56.