

Article details: 2017-0073	
Title	Admission factors associated with international medical graduate certification success: a collaborative retrospective review of post-graduate medical education programs in Ontario
Authors	Lawrence E. M. Grierson PhD, Mathew Mercuri PhD, Carlos Brailovsky MD MA(Ed), Gary Cole PhD, Caroline Abrahams MPA, Douglas Archibald PhD, Glen Bandiera MD MEd, Susan P. Phillips MD, Glenna Stirrett MD, J. Mark Walton MD, Eric Wong MD, Inge Schabort MB ChB
Reviewer 1	Dr. Bart J. Harvey
Institution	Dalla Lana School of Public Health, University of Toronto, Toronto, Ont.
General comments (author response in bold)	<p>Grierson and colleagues have conducted an interesting, Ontario-based analysis to attempt to gain further insights regarding possible factors associated with the outcomes of International Medical Graduates (IMG) on the CFPC and RCPSC certification examinations. While I believe the results of this study provide valuable insights, I do have a few comments and suggestions for the authors and CMAJ Open editors to consider prior to the manuscript being accepted for publication.</p> <p>1. To begin, it appears that the outcome data (i.e., pass/fail on applicable certification exam) involves multiple years (~2007 to 2014?). If this is correct, I wonder what, if any, adjustment or exploratory approaches the authors used to determine if year of challenging the exam might have affected the overall results. Of note, the authors mention that the IMG pass rate on the CCFP exam was only 66.7% in 2015 but rose to 88.8% a year later. This appears to be a sizeable change from one year to another, which raises the question about what effects there might be from combining multiple years of results together?</p> <p>This review comment is very intriguing. Indeed there is year-over-year variation in the pass rate for the exams. However, for our purposes, and more broadly for the Colleges' purposes, these exams are conceptualized as standardized so as to dissociate competent residents from incompetent residents reliably and validly. Thus, we have to assume that all residents, regardless of exam year, have effectively written the same test and that any differences in the year-over-year pass rate is reflective of natural variations that occur in cohorts and not in the exam. This position is supported by the high degree of consistency in the pass rates of candidates with Canadian medical degrees taking the exam for the first time with training in Canadian postgraduate institutions. The purpose of our model is then to illuminate the factors that contribute to IMG resident variation.</p> <p>2. Second, this analysis appears to be limited to only those who ultimately challenged the applicable certification examination. I would think that never challenging the exam would be as important an outcome as failing to pass it, no? As such, I would suggest the authors more clearly identify the full extent of the inception cohort (i.e., all those who entered an Ontario residency program between 2005 and 2012) and account for all of these residents not just the 912 who ultimately challenged the exam. I would expect that some number might be "still in training" and others "never attempted exam" and perhaps some withdrew from training?</p> <p>We agree that there is certainly more to know and learn about this cohort of IMG residents, and that there are undoubtedly numerous individuals that studied in Ontario residency programs within our timeline that did not or have not yet written the examinations. However, this particular request seems to be unfeasible given our current constraints on data access. That is, while we are able to include data pertaining to residents that have participated in residency programs via our institutional partners and match that data to certification results from the Colleges, the inclusion of residents that did not write the exams extends beyond the data sharing agreement that underpins this analysis. In this regard, we consider it a major victory to have been able to generate a provincial dataset that is this comprehensive. Yet, as we move further with this work, we will consider the reviewer's comment. In particular, we would like to extend the parameters of our data sharing agreements so as to include even more information. Part of this may be to look at those individuals that completed residency programs but did not write the exam (and perhaps identify factors associated with that). Although, in considering this we recognize the dangers of making a value judgment that we are not necessarily in position to levy. In particular, we cannot and should not assume that those that did not write the exam would have failed, and that there may be many legitimate and appropriate reasons why one may postpone the test.</p> <p>3. Third, perhaps I missed it in my review of this manuscript, but it is not clear to me how those who failed at least once but who were ultimately successful were handled in this analysis. Or, perhaps, it is only the result after their first attempt at the applicable exam that was analyzed? If so, it would also be important to know how many tried the exam subsequent times, and what number (and proportion) of them ultimately passed. In addition, how are residents who may have challenged multiple exams handled? For example, many Public Health & Preventive Medicine (PH&PM) residents will meet the requirements of the CCFP during their residency training so will be eligible to challenge both the CCFP and RCPSC PH&PM exams.</p> <p>The analyses included only the first attempt of each examination. This is stated in the revised manuscript and is highlighted in the marked copy. This is again a function of the data sharing agreement that was formed between the schools and Colleges, and which we believe represents a tremendous first step in large scale inter-institutional educational data sharing despite its limits (or potential for growth, depending on one's perspective). It is</p>

worth noting, however, that first time pass rate is a benchmark of success and candidates that fail the examination on first attempt have a reduced chance of passing on future attempts.

4. Fourth, I wonder if this study design isn't a "retrospective cohort study"? I'm also not sure that these are "predictive" models but perhaps "explanatory" ones - to better inform the limited interventional options available? I also wondered if the authors had considered doing a split-half analysis of this data to better enable the creation and validation of a truly "predictive" model? (1st paragraph of Methods)

We agree with the reviewer that our work is better conceptualized as associative and not predictive. We have worked to amend the relevant language throughout the manuscript.

5. Fifth, I wondered why 'institution' was not included as a potentially-explanatory variable in the modelling? Also, was Canadian citizenship included as a potentially-explanatory variable in the model? Both of these aspects should be made explicit and described more fully under Study Population.

Central to our exclusion of the Program of Post-Graduate Training factor (i.e., institution) in our analyses was a collaborative reticence to positioning the results of this work as a comparison between programs. Our team all agreed that it was of little value, and potentially damaging, to show that IMG's from one program performed significantly better or worse than those at other programs. Moreover, this type of comparison would not contribute to our main purpose, which was to identify pre-admissions features associated with examination success among IMG residents in Ontario. This view was shared by our College partners.

6. Sixth, I would suggest revising the phrase to "6 Ontario post-graduate residency programs" at the beginning of the Methods 4th paragraph to make it more explicit that this is an analysis limited to Ontario. I would also suggest that "previous professional experience" be better described. For example, would this include the completion of a graduate degree in Canada?

We have made the suggested edit and added language that further describes "previous professional experience."

7. Seventh, in addition to the assessment of the p-value associated with each variable assessed in the modelling, it is not clear if the authors also assessed overall model fit when determining which variables should be included and excluded while building the various multivariable models. Given the relatively modest sample size (by my calculation, only 67 who failed CCFP and 103 who failed RCPSC), I wonder if the ultimate models might have differed if a more liberal exclusion p-value (e.g., 0.15) was used? I would also suggest the authors calculate and include 95% confidence intervals with all ORs presented so readers can explicitly appreciate the precision of each of these measures. Why was there missing age data for 12 residents? How would this be possible? Perhaps this should also be explicitly explained. I would suggest that "country of medical school" not be used alone but always with HDI clearly indicated (or instead of it). Also, it appears that HDI is measured as a continuous variable in the model. This should also be clarified and more fully described. If I have understood the results correctly, I'm not sure why "previous internship" should be a factor suggesting failure on the exam. Perhaps a clearer description of what OR values greater and lesser than null (i.e., 1) would further assist readers' understanding.

The model produced Odds Ratios and 95% Confidence Intervals. These are included for all factors in the provided tables. At the Editor's request, we have also included the outcomes of the univariate analyses, which will allow the reader to see that the inclusion cutoff of $p < .1$ from the univariate analyses, while arbitrary, is appropriate. Furthermore, we are unable to explain why age data is missing for 12 individuals and attempts to acquire that data were unsuccessful. Lastly, we have worked to ensure that the HDI is appropriately identified as a continuous variable and have provided a citation to the reader that is interested in the value.

8. Eighth, in the Discussion I would suggest "identify the predictors of" be changed to "identify the factors associated with". Also, I'm not sure what a "robust predictor of performance" is. Perhaps it could be revised to "strongly associated with"? I would also suggest changing "the current work refutes" to perhaps "the current work does not support"? I would also note that Schabort and colleagues study (13th reference) was a very modest, dare I say pilot, study that only included 69 CCFP and 85 RCPSC IMG trainees in its analysis. I would also suggest more moderate wording be used in the final paragraph of the Discussion. For example, perhaps "critical" could be replaced by "important" and "to identify factors associated with" instead of "for determining the factors that have the most substance in explaining" and "relatively large data set" as it would be interesting if such an analysis could be repeated using a longer time period and pan-Canadian data. I would suggest alternate wording to "sets the table" ("enables"?) and also for "depressed exam performance" (perhaps just delete "depressed"?) be revised.

We thank the Reviewer for the suggestions. We have incorporated many.

	<p>9. Ninth, on more minor notes, I believe that "who" should be used rather than "that" in the last line of the 3rd paragraph of Results as well as in the 2nd line of "Study Population" and the title for reference number 13 does not appear to be listed correctly.</p> <p>We have made these edits. Thank-you.</p>
Reviewer 2	Dr. Roy Thomas Dobson
Institution	College of Pharmacy and Nutrition, University of Saskatchewan Saskatoon, Sask.
General comments (author response in bold)	<p>Thank you for the opportunity to review "Predictors of international medical graduate certification success". Overall this is a well written manuscript. The authors do a good job presenting the issue of higher failure rates among IMG compared to CMS graduates. The selection of variables used in the modelling appears appropriate and justified by the literature presented. The results are presented in a concise and organized manner and some good reflection evident in the discussion.</p> <p>1. That being said, the results were somewhat underwhelming in that most variables were not found to be significantly associated with the dependent variables. Given the strong theoretical basis for their inclusion in the study, and the large number of subjects included in the models, this was somewhat surprising. However, it might represent an opportunity for the authors to reflect on these results to a great extent than they did in the discussion.</p> <p>The authors and our affiliated institutions have all reflected on the results of this study - significant or otherwise - with great interest, and we are happy to include further discussion on our non-significant associations. However, given the Journal's word limits for this type of article and the current length of our present submission we are concerned about "over-editorializing." As such, we have opted to add a single statement into the discussion highlighting that many of the factors analyzed were not associated with examination success. We will be glad to revisit this and add additional commentary at the Journal's request.</p> <p>2. Given the number of models and the inconsistent strength of the associations (excepting age), there is a concern for Type 1 error and I would suggest some consideration of this in limitations.</p> <p>We have now stated this as a potential limit to interpretation in the revised manuscript, and take the opportunity to remind the reader that this work should be considered hypotheses-generating rather than hypothesis-confirming.</p> <p>3. Finally, I would like to know how some of these results compare to CMS graduates. Are age, gender and other comparable factors similar or different compared to IMGs? A discussion of the pertinent literature might enhance the discussion section, and by extension, the quality of the manuscript.</p> <p>We appreciate the Reviewer's interest in these ideas. We also had interest in this comparison but were not able to include the necessary information in our data sharing agreement. We do hope that this will be a topic for subsequent research, which now has potential due to the success of the partnership that has yielded the data used in the presented analysis and this manuscript. Indeed we could go into great editorial depth in our discussion regarding a number of topics and comparisons related to the current study. Again, however, we recognize the structure limits to papers of this type published in CMAJ Open. In this regard, we have opted not to considerably expand the discussion in this regard but are open to doing so at the Journal's request.</p>