Using In-Training Examination data in residency program quality assurance

Oksana Babenko[2], Shelley Ross[2], Shirley Schipper[2], John Chmelicek[2], Kimberley Duerksen[2], Denise Campbell-Scherer[2]

Abstract

In-Training Examinations (ITEs) are an established part of many residency programs; however, their use varies across programs. In our Family Medicine residency program, we use the ITE in two ways: as a formative assessment as part of residency training, and as a way to monitor and inform curriculum. When examining residents’ overall and domain performance on the ITE, several factors are at play in the multi-year ITE data: the year of residency (PGY1 or PGY2), residents’ undergraduate medical training background, urban/rural residency placement, as well as the fact that residents take the ITE twice during their training. Such a plethora of factors presents difficulties in monitoring the effectiveness of curricular changes over the years. To address this, we demonstrate the use of the Generalized Estimating Equations (GEE) method with the ITE data for the purpose of program quality assurance.

Keywords: In-Training Examination (ITE), Family Medicine, Residency, Generalized Estimating Equations (GEE)

Introduction

In the United States (US) and in specialty programs in Canada, In-Training Examinations (ITE) have long been used as standardized assessments of residents’ medical knowledge and clinical reasoning. Recently, ITEs have become more common in Canadian family medicine residency programs; however, their implementation and use vary across the country. In our residency program, we use the US Family Medicine ITE for two purposes. The first is as a formative assessment tool to provide residents with external evidence on their strengths and weaknesses in their medical knowledge and clinical reasoning as they progress through their program and prepare for their board certification examinations. The other is to inform the residency program about curricular opportunities and monitor the effectiveness of the implemented changes over time.
Our 2-year residency program offers two streams of residency training, urban and rural. Each year we have on average 72 residents admitted into the program. Although the majority of the admitted residents are Canadian medical graduates (CMG), approximately 15% of our residents are international medical graduates (IMG) who obtained their undergraduate medical degrees from non-Canadian medical schools. Our residency program was the first in Canada to use the US Family Medicine ITE, with the first ITE administration in 2009. The ITE offers our program the opportunity to gauge how well our curriculum is delivering content (as measured by how well our residents perform), as well as a way to determine if there are any areas of concern or opportunity among certain populations of residents (by cohort, residency placement, or CMG/IMG).

In November each year, our first and second year (PGY1/PGY2) residents take the ITE. Within two months of writing the examination, residents receive their individual results on the full examination and high-level domains. The residency program director receives only program averages (for PGY1 and PGY2 cohorts), which provide information on how residents’ performance in our program compares to residents’ performance in other programs in the same testing year. When examining the variation in residents’ performance across the years, the program director is interested to know how much of the variation in the ITE scores is attributed to program factors to be able to make changes in curriculum based on the identified concerns in residents’ performance and examine the effect of such changes overtime. To address this, we demonstrate the use of the Generalized Estimating Equations (GEE) method with the ITE data for the purpose of program quality assurance. Institution Review Board ethics was obtained for this study.

Methods

Measure

American Board of Family Medicine In-Training Examination (ABFM® ITE; www.theabfm.org) is a timed computer-based assessment that consists of 240 multiple-choice questions. The ITE assesses residents’ knowledge on eight high-level domains: Adult Medicine, Care of Surgical Patients, Maternity Care, Care of Children and Adolescents, Mental Health Care, Care of the Elderly, Care of Female Patients, and Emergent and Urgent Care. The ITE score range is 200-800, with the first year (PGY1) residents’ scores expected to be in the 300s; the second year (PGY2) residents are expected to score 40-60 points higher.

Data

Data from the residents who wrote the ITE in their first and second years of residency in our program in 2011-2015 were used in the analyses (n=480 out of 505 residents registered to write the ITE (95%) but did not write it due to personal circumstances (e.g., illness, maternity leave)).

Analysis

Given the rather uncommon dataset used in the current study, a design that combines cross-sectional and longitudinal data, we employed the Generalized Estimating Equations (GEE) analyses (Hardin & Hilbe, 2003). The GEE takes into account correlations present in the data due to the longitudinal design (PGY1 residents in one testing year become PGY2 residents in the next testing year). In addition, the GEE approach accounts for differences in number of times each resident wrote the examination over the course of his/her residency training. This is important as a very small number of residents wrote the exam only once during their residency due to personal circumstances. Additionally, the PGY2 residents in 2011 and the PGY1 residents in 2015 have only one test result in the 2011-2015
More importantly, the GEE allows modeling and examining residents' performance on the ITE, using residents' characteristics and factors/variables as they pertain to the residency program. For example, over the years the percent of IMG residents admitted into the program varies. Similarly, the percent of residents in the rural placement of the program also differs from year to year. These (CMG/IMG status, rural/urban placement) and other factors can be entered into the GEE analysis and accounted for when examining residents' performance longitudinally.

As a semi-parametric technique, the GEE does not require the assumptions of normality, homogeneity of variances and homoscedasticity to be satisfied as it is the case in regression analysis and the analysis of variance. In the GEE analysis, estimated means and mean differences ($md$), which have been adjusted for factors of interest, are reported together with their standard errors (for constructing confidence intervals) and are used to examine residents’ performance across academic years and between residency years (PGY1/PGY2). We used IBM SPSS Statistics 22.0 (Armonk, NY: IBM Corp) to perform the GEE analyses. Significance level was set to .05, with Bonferroni correction used for multiple comparisons.

Results

As shown in Figure 1, there were significant differences in residents' performance across the academic years (2011-2015) within each residency year ($p<.01$). Of note, some domains such as Care of Children and Adolescents and Care of the Elderly had more variable performances across the academic years than other domains. Irrespective of the testing (academic) year, residents' overall performance steadily improved as residents transitioned from PGY1 to PGY2 of their residency ($md=28\pm3.8; p<.01$), with significant differences in performance in 5 out of 8 domains: Adult Medicine ($md=29\pm4.6; p<.01$), Care of Surgical Patients ($md=25\pm11.8; p<.05$), Maternity Care ($md=41\pm12.2; p<.01$), Care of Children and Adolescents ($md=34\pm8.1; p<.01$), and Care of the Elderly ($md=29\pm8.2; p<.01$).
Urban residents’ overall performance was on average comparable to rural residents’ performance (\(md=3.5\pm8.1;\) \(p>.05\)) in 2011-2015. However, urban residents performed on average higher than rural residents on Mental Health Care (\(md=30\pm14.9;\) \(p<.05\)), whereas rural residents outperformed urban residents in Maternity Care (\(md=44\pm17.3;\) \(p<.01\)) and Emergent and Urgent Care (\(md=25\pm11.7;\) \(p<.05\)).

Across the testing years, CMGs’ overall performance was on average consistently higher than IMGs’ performance (PGY1 \(md=33\pm11.8\) & PGY2 \(md=23\pm9.9;\) \(p<.01\)), with significant differences in performance in 6 out of 8 domains: Care of Surgical Patients (PGY1 \(md=61\pm25.9\) & PGY2 \(md=55\pm23.4;\) \(p<.01\)), Care of Children and Adolescents (PGY1 \(md=56\pm18.7\) & PGY2 \(md=42\pm22.7;\) \(p<.01\)), Mental Health (PGY1 \(md=77\pm28.4\) & PGY2 \(md=42\pm22.7;\) \(p<.01\)), Care of the Elderly (PGY1 \(md=47\pm17.1\) & PGY2 \(md=27\pm18.9;\) \(p<.01\)), Care of Female Patients (PGY1 \(md=59\pm26.6\) & PGY2 \(md=29\pm24.5;\) \(p<.05\)), and Emergent and Urgent Care (PGY1 \(md=64\pm23.7\) & PGY2 \(md=60\pm18.1;\) \(p<.01\)). Of note, the differences in performance between CMGs and IMGs tended to decrease as residents transitioned from PGY1 to PGY2 of their residency.

**Discussion**

Regular assessment of learners from year to year is important for identifying opportunities in curriculum and informing educational policies. Residency programs are no exceptions to this. In-Training Examinations are one of the sources of external evidence for both residents and the residency program director. With many program and
resident factors (i.e., year of residency, CMG/IMG status, urban/rural training) at play in the multi-year ITE data, GEE allows for analysis of data collected over the span of several years and makes it possible to examine the contribution of various factors to residents' performance on the full examination and high-level domains.

In our program in 2011-2015, residents' overall performance steadily improved between the first and second years of residency, with significant differences in performance in 5 out of 8 domains. Canadian graduates performed consistently higher than international graduates, irrespective of the residency year, with significant differences in performance in 6 out of 8 domains. The demonstrated approach has informed our program on potential curricular changes and learning opportunities based on what concerns in performance have been identified. For example, lower performance on certain domains (e.g., Care of Children and Adolescents, Care of Female Patients) has prompted efforts to increase exposure and training that target those areas, resulting in higher performance in these areas in subsequent years. Similarly, given lower performance in Care of the Elderly and Maternity Care in the last two testing years, efforts to increase exposure in these areas will be implemented and monitored to see if there is a change.

In this way, the GEE allows us to use the ITE data for a "bigger picture" view of performance of our residents both within cohorts and over time. With this data, we can identify those areas of the program where changes are warranted, as well as confirm where the curriculum is working well. Further, the ITE data used in this way can allow us to get a sense of the effects of changes to the curriculum over time, in terms of effects on resident examination performance for specific domains.

**Conclusion**

The demonstrated approach can be used as a model for other programs in how to use their ITE data for the purpose of program quality assurance, and for identifying curricular opportunities and monitoring the effectiveness of implemented curricular changes over time.

**Take Home Messages**

**Notes On Contributors**

Dr. Oksana Babenko, PhD, is an Assistant Professor and Medical Education Researcher with the Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada.

Dr. Shelley Ross, PhD, is an Associate Professor and Medical Education Researcher with the Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada.

Dr. Shirley Schipper, MD, is an Associate Professor and the Assessment Director with the Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada.

Dr. John Chmelicek, MD, CCFP, FAAFP, FAWM, is an Assistant Professor and the Postgraduate Program Director with the Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada.
Ms. Kimberley Duerksen, MSc, is a Research Coordinator with the Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada.

Dr. Denise Campbell-Scherer, MD, PhD, CCFP, FCFP, is an Associate Professor with the Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada.

Acknowledgements

We wish to thank the Northern Alberta Academic Family Medicine Fund (NAAFMF) for funding this research. We also wish to thank Ms. Tanya Barber, MA, for assisting with data collection on this study.

Bibliography/References


Appendices

Declarations

The author has declared that there are no conflicts of interest.

This has been published under Creative Commons “CC BY 4.0” (https://creativecommons.org/licenses/by-sa/4.0/)