

**Executive Summary  
Medical Council of Canada**

**The Quebec-Ontario Follow-up Study of the Association between Scores Achieved on the  
MCCQE Part II Examination and Performance in Clinical Practice**

**Investigators**

Robyn Tamblyn PhD  
McGill University

Michal Abrahamowicz PhD  
McGill University

Gillian Bartlett PhD  
McGill University

Nancy Winslade PharmD PhD candidate  
Maastricht University

Dr. André Jacques MD  
Quebec College of Physicians

Dr. Daniel Klass MD  
College of Physicians and Surgeons of Ontario

Elizabeth Wenghofer PhD  
College of Physicians and Surgeons of Ontario (now at Laurentian University)

Sydney Smee PhD  
Medical Council of Canada

Dr. Dale Dauphinee MD  
Medical Council of Canada (now at McGill University)

David Blackmore PhD  
Medical Council of Canada

Ilona Bartman MA  
Medical Council of Canada

Dr. David Buckeridge MD PhD  
McGill University

James Hanley PhD  
McGill University

## **Mandate**

The Medical Council of Canada funded this study to determine if the MCCQE Part 2 examination:

- 1) predicted the safety and quality of subsequent practice.
- 2) provided additional predictive information about the safety and quality of future practice to that provided by the MCCQE Part 1 examination.

The scientific team, comprised of investigators from the Medical Council of Canada, the College of Physicians and Surgeons of Ontario, the College of Physicians of Quebec, and McGill University, collaborated to conduct a multi-faceted assessment of different indicators of safety and quality of care in practice among physicians who took the Part 2 examination between 1993 as 1996 and entered practice in Ontario or Quebec.

## **Conceptual Framework for Assessment of Safety and Quality of Practice Indicators**

Three principles were used to select indicators of the safety and quality of practice: 1) the indicator measured an identified problem (e.g. complaints) or health problem where variability existed in the delivery of care that would have an impact on patient outcome (e.g. inappropriate antibiotic use), 2) physician decision-making and/or communication skills was expected to have a substantial influence on the problem/health care service (e.g. prescribing), and 3) there was a valid source of data to measure and attribute a problem/health service to a individual physician. On the basis of these principles, two measures were identified to assess overall quality/safety of care: complaints to the medical licensing authority, and in Ontario, the peer-review, on-site in practice assessment. Additional indicators were identified to assess safety and quality of care in three domains that may be sensitive to particular aspects of clinical competence that were assessed by the MCCQE2 examination. These included i) *acute care management* (appropriateness of antibiotic prescribing), ii) *chronic care management* (management of out-of-control asthmatics, choice of treatment, and adherence with anti-hypertensive treatment), and iii) *preventive care* (mammography screening for breast cancer). Due to data access constraints, these indicators could only be measured in the Quebec cohort of physicians, although theoretically they could be assessed in each province. While there was an attempt to assess similar domains of acute, chronic and preventive care using the Ontario peer-review assessment, limited numbers of physicians and high levels of performance on all peer-review assessments mitigated the utility of these data to address the primary research mandate.

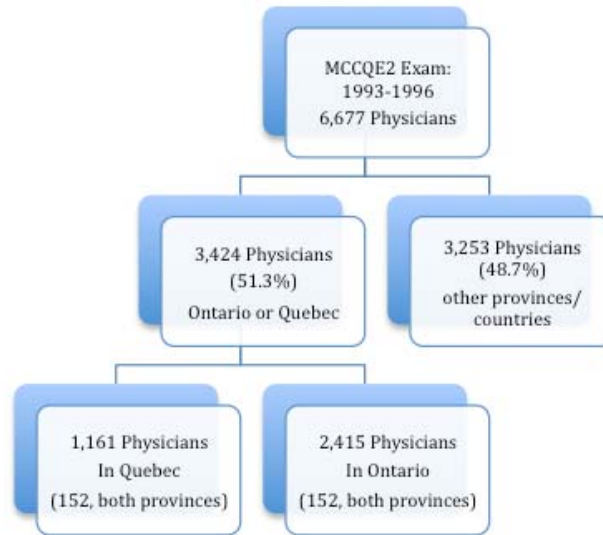
## **Methods**

A prospective cohort study was conducted. The safety and quality of clinical practice was assessed for the four examination cohorts who completed the MCCQE2 examination between 1993 and 1996 and entered practice in the province of Ontario or Quebec. Physicians were followed from the date they entered practice (earliest=1993) to 2007. The cohort was assembled in four steps. First the MCC identified all physicians who took the MCCQE2 examination between 1993-1996, and the MCCQE1 exam and MCCQE2 scores were retrieved for each physician. Second, postgraduate training information for each physician was determined by linking to the Canadian Postgraduate Training Registry. Third, location of practice was determined by linking to the licensure registry files of the Ontario and Quebec College of physicians. Fourth, for all physicians licensed in Ontario and Quebec, all outcome data was retrieved (complaints filed in both provinces, peer-review in Ontario, and billing information on practice activity in Quebec). Multivariate methods were used to provide an independent estimate

of the association between examination scores and practice outcomes, adjusted for physician demographic and training characteristics, as well as patient characteristics (e.g co-morbidity), when relevant.

## Results

Overall, 6 677 physicians took the QE2 examination between 1993 to 1996, and 3 424 (51.3%) entered practice in Ontario or Quebec.



Overall, 55.5% of the study population were men, most were between the ages of 25 – 30 years at the time of the clinical skills examination, 77.5% had graduated from Quebec or Ontario medical schools, 40.7% did post-graduate training in family medicine, and 16.1% trained in a surgical specialty, and two-thirds entered practice in Ontario. Overall, the population of physicians who enter practice in Ontario and Quebec had slightly higher scores on both the MCCQE2 and MCCQE1 than physicians entering practice in other provinces, however there was a range of approximately four standard deviations in the MCCQE1 scores and seven standard deviations in the MCCQE2 scores among physicians entering practice in Ontario and Quebec (Table E.1).

**Table E.1 The Characteristics of the 3424 physicians taking the National Postgraduate Clinical Skills Examination between 1993 and 1996 who were licensed to practice in Ontario and/or Quebec\***

Physician Characteristics	N (percent)
<b>Demographics</b>	
<b>Sex</b>	
Women	1525 (44.5)
Men	1899 (55.5)
<b>Age at the clinical skills examination<sup>†</sup></b>	
<25 years	159 (4.6)
25–30 years	2451 (71.6)
>30 years	814 (23.8)
<b>Undergraduate Medical Training</b>	
Quebec/Ontario Med. School	2655 (77.5)
Other Canadian Med. School	349 (10.2)
International Med. Graduate	420 (12.3)
<b>Postgraduate Specialty Program</b>	
Family/General Med	1393 (40.7)
Medical Specialty	1481 (43.3)
Surgical Specialty	550 (16.1)
<b>Practice Location</b>	
Ontario	2263 (66.1)
Quebec	1009 (29.5)
Both Provinces	152 (4.4)
<b>Clinical Skills Examination †</b>	
	<b>Mean (SD) [range]</b>
MCCQE 2 Score	525.1 (79.9) [50–749]
Communication subscore	510.9 (91.1) [31–723]
Data acquisition subscore	508.8 (90.7) [19–875]
Problem Solving subscore	541.6 (98.5) [170–864]
<b>Traditional Written Examination †</b>	
MCCQE 1 Score	526.5 (77.6) [338–787]
Multiple Choice Questions subscore	524.4 (82.7) [278–793]
Clinical Decision–Making subscore	525.0 (75.8) [221–739]

\* Data presented as No. (%) except as noted. Percentages may not total 100 due to rounding. † Scores standardized with a mean (SD) of 500 (100) for all first-time takers from Canadian medical schools in a given examination administration.

There was a moderate correlation between the MCCQE1 and MCCQE2 overall scores ( $r=0.40$ ; disattenuated  $r=0.47$ ) for study physicians. The lowest correlation was between the communication subscore of the MCCQE2 and the overall MCCQE1 examination score ( $r=0.14$ ; disattenuated  $r=0.23$ ), suggesting that two independent abilities were being assessed. Score reliability was highest for the MCCQE1 score (mean Cronbach alpha 1993-1996: 0.92), and lowest for the MCCQE2 communication subscore.

**Table E.2: The Correlation between the Medical Council of Canada Examination Scores and sub scores for the 3,424 Physicians in the Study Population**

MCCQE 2 (reliability)	Overall score (0.92)	MCCQE 1 (reliability)	
		Multiple choice questions (0.91)	Clinical Decision–Making Skills (0.64)
<b>Correlation between Scores*</b>			
Overall score (0.77)	0.40	0.36	0.33
Communication (0.41)	0.14	0.10	0.17
Data acquisition (0.66)	0.23	0.21	0.16
Problem–solving (0.54)	0.38	0.36	0.30

\* Pearson product-moment correlation coefficients. All correlations were statistically significant ( $p<0.001$ ).

1. For the traditional written exam scores, weighted Cronbach alpha in different examination administrations was overall score, 0.92; multiple choice questions, 0.91; clinical decision making, 0.64. For clinical skills examination scores, weighted Cronbach alpha was overall score, 0.77; communication, 0.41; data acquisition, 0.66; problem-solving, 0.54. Weights based on the number of candidates taking the examination in each administration. (36) Disattenuated correlations ( $r$ ) between the scores for the traditional written examination and clinical skills examination were

overall score, 0.47; communication, 0.23; between the traditional written examination clinical decision-making score and the clinical skills examination scores they were overall score, 0.47; communication score, 0.43.

The prevalence of safety and quality problems in the practices of the study physicians are summarized in Table E.3. Overall, 12.3% of the 3,424 in the study population had at least one complaint filed with the College of Physicians in Ontario or Quebec in the first three to ten years of practice. The overall rate of complaints was 4.9 complaints per 100 physicians per year for this newly licensed group of physicians. Within the Ontario cohort of practicing physicians, 208 physicians were randomly selected for the peer review quality of care assessment program in their first 3-10 years of practice. Among these 208 physicians, 15 (7.2%) were deemed to have unsatisfactory performance.

In the Quebec cohort of physicians, the prevalence of problems in the management of viral and bacterial infections was substantial; with 24.1% of patients receiving antibiotics for viral infections, and 35% of patients with first time bacterial infections receiving second or third line antibiotics. In the area of chronic disease management, we investigated the management of asthma and hypertension. Among patients presenting to study physicians with out of control asthma, 45.7% continued to overuse rescue medication in the six months following the first out-of-control visit, and 16.0% made at least one visit to the emergency room for their asthma in the following six months. Overall, only 69.4% of out-of-control patients used expected preventive therapy, inhaled steroids, in the six months following a visit to the study physician and of those who used inhaled steroids, only 23.9% had an inhaled steroid to fast acting beta agonist ratio of more than 75%.

In relationship to the management of hypertension, 44.8% of patients initially treated for uncomplicated hypertension and 90.5% of complicated patients received cost effective therapy according to Canadian guidelines. Among the 5,259 patients who were started on treatment for hypertension by study physicians, 29.2% discontinued all antihypertensive therapy in the first two months, and a further 19.4% discontinued all antihypertensive therapy by six months.

In relationship to breast cancer prevention, 31.8% of women 50-69 who visited study physicians received a screening mammogram, and 12.1% of the 127,812 women treated by study physicians received a diagnostic mammogram.

### E.3 The Prevalence of Safety and Quality Practice Outcomes in the Practices of Study Physicians

Overall Quality/ Safety of Practice				
Safety/Quality Practice Outcomes	Number of MDs	Number of Patients	Yes (N (%))	No (N (%))
Communication/Quality of Care Complaint	3424		422 (12,3%)	3002 (87.7%)
Negative Peer Review Assessment	208		15 (7,2%)	193 (92.8%)
Acute Care Management				
			Yes	No
Management of Infections				
Antibiotic for Viral Infection	440	137 833	33, 246 (24,1%)	104, 587 (75,9%)
Inappropriate Antibiotic-Bacterial Infections	440	213 385	74 ,752 (35,0%)	138, 633 (65,0%)
Chronic Care Management				
			Yes	No
Management of Out-of-Control Asthma				
Persistent Excess Rescue Med. Use	327	1 960	896 (45,7%)	1,064 (54,3%)
Number of ER Visits	327	1 960	313 (15,97%)	1,647 (84,0%)
Inhaled Steroid Use (any/none)	327	1 960	1,361 (69,4%)	599 (30,6%)
Ratio Inhaled Steroids-Rescue Med. (hi/lo)	327	1 960	325 (23,9%)	1,036 (76,1%)
Management of Hypertension				
Cost-Effective Therapy Choice			Yes	No
Uncomplicated	508	3315	1,484 (44,8%)	1,831 (55,2%)
Complicated	590	3887	3,519 (90,5%)	368 (9,5%)
Non-persistence with Treatment				
First 2 months	588	5259	1,533 (29,2%)	3,726 (70,8%)
First 6 months	588	3818	739 (19,4%)	3,079 (80,6%)
Prevention				
Breast Cancer Screening				
Screening Mammography	413	127 812	40, 646 (31,8%)	87,166 (68,2%)
Diagnostic Mammography	413	127 812	15, 423 (12,1%)	112,389 (87,9%)

## **Physician Training and Demographic Characteristics and Safety/ Quality of Care**

Physician demographic characteristics and medical training influenced the likelihood of having a safety quality problem in practice (Table E.4). Male physicians were 64% more likely to have a complaint filed against them with the College of Physicians and Surgeons for quality of care and communication problems (relative risk: 1.64; P-value < 0.001). Female physicians were more likely to prescribe cost-effective therapy for uncomplicated hypertension and screen women for breast cancer. There was a trend for international medical graduates to receive more complaints, and have an unsatisfactory peer review, as well have poorer outcomes in managing acute and chronic conditions, however only higher rates of non-persistence with hypertension treatment achieved statistical significance.

In relationship to post-graduate training, physicians who were trained in family medicine or surgery had a higher risk of having complaints filed with the CMQ and CPSO about their quality of care and communication. Medical and surgical specialists were less likely to make cost effective treatment selections for uncomplicated hypertension, whereas medical specialists were more likely, and surgical specialists were less likely to make cost effective treatment choices for complicated hypertension than family physicians. Patients started on antihypertensive treatment by surgeons were more likely to be non-persistent with treatment at six months relative to patients started on antihypertensive treatment by family physicians. Surgeons were more likely to screen women for breast cancer and order diagnostic mammograms. Compared to family physicians, pediatricians were 27% less likely to prescribe antibiotics for viral infections (relative risk: 0.73; p-value > 0.001).

#### E.4 The Relationship between Physician Demographic and Training Characteristics and Safety/Quality Practice Outcomes

Safety/Quality Practice Outcomes	Physician Sex		Medical School		Postgraduate Training		
	Women	Men	Canadian Medical Graduates	International Medical Graduate	Family Medicine	Medical Specialty	Surgery
<b>Overall Quality/ Safety of Practice</b>							
	Relative Risk ( <i>P-value</i> )		Relative Risk ( <i>P-value</i> )		Relative Risk ( <i>P-value</i> )		
Communication/Quality of Care Complaint	Reference Gp	<b>1.64 (&lt;0.001)</b>	Reference Gp.	1.61 (0.25)	<b>1.79 (&lt;0.001)</b>	Reference Gp.	<b>2.43 (&lt;0.001)</b>
Unsatisfactory Peer Review Assessment	0.84 (0.77)	Reference Gp.	Reference Gp.	3.05 (0.18)	2.67(0.11)	Reference Gp.	
<b>Acute Care Management</b>							
<b>Management of Infections</b>							
Antibiotic for a viral respiratory infection	1.04 (0.72)	Reference Gp.	Reference Gp.	1.33 (0.46)	Reference Gp.	<b>0.73 (&lt;0.001)</b>	
Inappropriate antibiotics for bacterial infection	<b>0.88 (0.02)</b>	Reference Gp.	Reference Gp.	1.21 (0.27)	Reference Gp.	1.02 (0.77)	
<b>Chronic Care Management</b>							
<b>Management of Asthma</b>							
Persistent excess FABA	0.92 (0.07)	Reference Gp.	Reference Gp.	1.02 (0.67)	Reference Gp.	<b>0.76 (0.02)</b>	-
ER visit Rate	1.28 (0.08)	Reference Gp.	Reference Gp.	1.45 (0.23)	Reference Gp.	0.94 (0.74)	-
ICS Use (Any vs None)	1.06 (0.11)	Reference Gp.	Reference Gp.	0.97 (0.71)	Reference Gp.	<b>1.10 (0.04)</b>	-
<b>Management of Hypertension</b>							
Cost-Effective Therapy Choice							
Uncomplicated	<b>1.28 (0.03)</b>	Reference Gp.	Reference Gp.	1.11 (0.06)	Reference Gp.	<b>0.48 (&lt;0.001)</b>	<b>0.38 (0.05)</b>
Complicated	1.09 (0.52)	Reference Gp.	Reference Gp.	0.096 (0.87)	Reference Gp.	<b>1.60 (0.02)</b>	<b>0.41 (0.01)</b>
Non-persistence with Treatment							
First 2 months	1.00 (0.94)	Reference Gp.	Reference Gp.	<b>1.24 (0.04)</b>	Reference Gp.	0.95 (0.56)	1.19 (0.46)
First 6 months	0.93 (0.40)	Reference Gp.	Reference Gp.	0.91 (0.53)	Reference Gp.	0.94 (0.57)	<b>2.01 (0.003)</b>
<b>Prevention</b>							
<b>Breast Cancer Screening</b>							
Screening Mammography	<b>1.23 (0.001)</b>	Reference Gp.	Reference Gp.	0.99 (0.93)	Reference Gp.	0.79 (0.25)	<b>1.60 (&lt;0.001)</b>
Diagnostic Mammography	1.09 (0.08)	Reference Gp.	Reference Gp.	0.98 (0.84)	Reference Gp.	0.73 (0.19)	<b>1.31 (0.01)</b>

1. The association between physician demographic and training characteristics is adjusted for the most significant MCCQE1 and MCCQE2 score, as well as patient characteristics, and clustering of patients within physicians when patients were used as the unit of analysis.



## **Examination Performance and Quality/Safety of Care**

### **Overall Quality of Care**

Licensing examination scores and their relationship to practice outcomes are summarized in Table E.5. The MCCQE1 total score and the MCCQE2 total score were associated with the likelihood of complaints. The risk of complaints increased by 39% for every 2 standard deviation reduction in the MCCQE1 overall score (relative risk: 1.39; p-value= 0.001). The MCCQE1 clinical decision-making sub-score was most strongly associated with the likelihood of receiving a complaint about a communication or quality of care problem. There was a 19% increase in the likelihood of receiving a complaint (relative risk 1.19; p-value=0.05) for every two standard deviation reduction in MCCQE2 score. The strength of this association was attenuated by the non-significant relationship between the data acquisition and problem-solving sub-scores and the likelihood of receiving a complaint. The only score that predicted the likelihood of complaint within the MCCQE2 examination was the communication sub-score (relative risk: 1.38; p-value < 0.01). When combined in the same predictive model, both the MCCQE1 clinical decision-making sub-score and the MCCQE2 communication sub-score provided a significant and independent contribution to predicting future complaints in practice.

Unsatisfactory peer review assessment was only explained by the MCCQE1 overall score, the multiple choice sub-score, and the MCCQE2 problem-solving sub-score. For every two standard deviation reduction in the overall QE1 score, the risk of an unsatisfactory peer review assessment increased by 566% (relative risk: 6.66; p-value= 0.02).

### **Management of Infections**

The association between MCCQE examination scores and the likelihood of receiving an antibiotic for a viral infection depended upon whether the treating physician was male or female. For male physicians, the risk of prescribing an antibiotic for a viral infection increased with higher scores on both the MCCQE1 and MCCQE2 examination. The opposite was true for female physicians where higher scores were associated with a reduction in the likelihood of receiving an antibiotic prescription for a viral infection. The only exception to this difference for male and female physicians was for the clinical decision-making sub-score of the QE1 examination, where higher scores were associated with the reduction of risk of receiving an antibiotic for viral infections for both male and female physicians. With respect to bacterial infections, the only statistically significant association was in relationship to the MCCQE 2 problem solving score, where every two standard deviation increase in score resulted in a 12% reduction in the likelihood of receiving second or third line treatment for a bacterial infection (relative risk: 0.88; p-value=0.02).

### **Management of Asthma**

Only the scores achieved on the clinical decision making sub-component of the MCCQE1 examination influenced the outcome of patients with out of control asthma. The risk of having an ER visit for asthma after a visit to the study physician was reduced by 36% (relative risk: 0.64; p-value 0.008) for every two standard deviation increase in the clinical decision making sub-score of the MCCQE1. All other examination scores had no significant relationship to outcomes for patients with out-of-control asthma.

In terms of the interim management steps that were relevant for the treatment of out of control asthma patients, there were significant associations between the MCCQE1 and MCCQE2 scores and the use of inhaled steroids. Physicians who received higher scores on the MCCQE1 and MCCQE2, specifically in the area of communication, were more likely to have patients use any

form of inhaled steroids in the six months following the first visit for out of control asthma. Specifically, out of control patients were 11% more likely to use inhaled steroids in the six months following a visit for out of control asthma (relative risk: 1.11; p-value=0.01) with every two standard deviation increases in the multiple-choice sub-score of the MCCQE1. Moreover, the likelihood of a patient using inhaled steroids increased by 12% (relative risk: 1.12; p-value=0.008) for every two standard deviation increases in communication sub-score on the MCCQE2 examination. However, the extent of use of inhaled steroids relative to rescue medications was not influenced by physician's examination performance.

### **Management of Hypertension**

The choice of cost effective treatment for both complicated and uncomplicated hypertension was not predicted by the MCCQE1 or MCCQE2 examination. However, the MCCQE1 And MCCQE2 examination scores predicted non-persistence with initial anti-hypertensive treatment, but different aspects of competence appear to influence early and later treatment dropouts. In the first two months, higher scores on the MCCQE1 examination reduced the likelihood of non-persistence. The risk of non-persistence with new antihypertensive treatment was reduced by 26% (relative risk: 0.74; p-value=0.002) for every two standard deviation increases in the MCCQE1 examination score. This association was partly explained by the tendency for high scoring physicians to make treatment changes (dose and drug changes) in the first month of treatment than lower scoring physicians. Among patients who stayed on treatment after the first two months, the risk of non-persistence in the next four months was related to the physician's communication ability. The risk of non-persistence in the first six months among patients who were not early dropouts from treatment was reduced by 20% for every two standard deviation increase in MCCQE2 communication sub-score; (relative risk: 0.80; p-value=0.01).

### **Prevention**

Achievement in two aspects of the examination influenced the likelihood a physician would order a screening mammogram for breast cancer for women between the ages of 50-69. First, higher scores on the MCCQE1 examination were associated with a 15% increase of the likelihood of receiving a screening mammogram for every two standard deviation increase in score (relative risk: 1.15; p-value=0.50). Second, higher scores in communication were associated with the largest increase in the likelihood of a screening mammogram, a 28% increase in mammography screening for every two standard deviation increases in score (relative risk: 1.28; p-value=0.000). In contrast, the likelihood of a diagnostic mammogram diminished among physicians who had higher scores on the MCCQE2 examination. There was no association between scores achieved on the MCCQE1 examination and the likelihood of receiving a diagnostic mammogram. The likelihood of receiving a diagnostic mammogram diminished by 13% (relative risk: 0.87; p-value=0.02) for every two standard deviation increases in the MCCQE2 overall score.

E.5 The Association between MCCQE1 and MCCQE2 examination scores and Performance on Indicators of Safety and Quality of Care in Practice in the First 3-10 Years

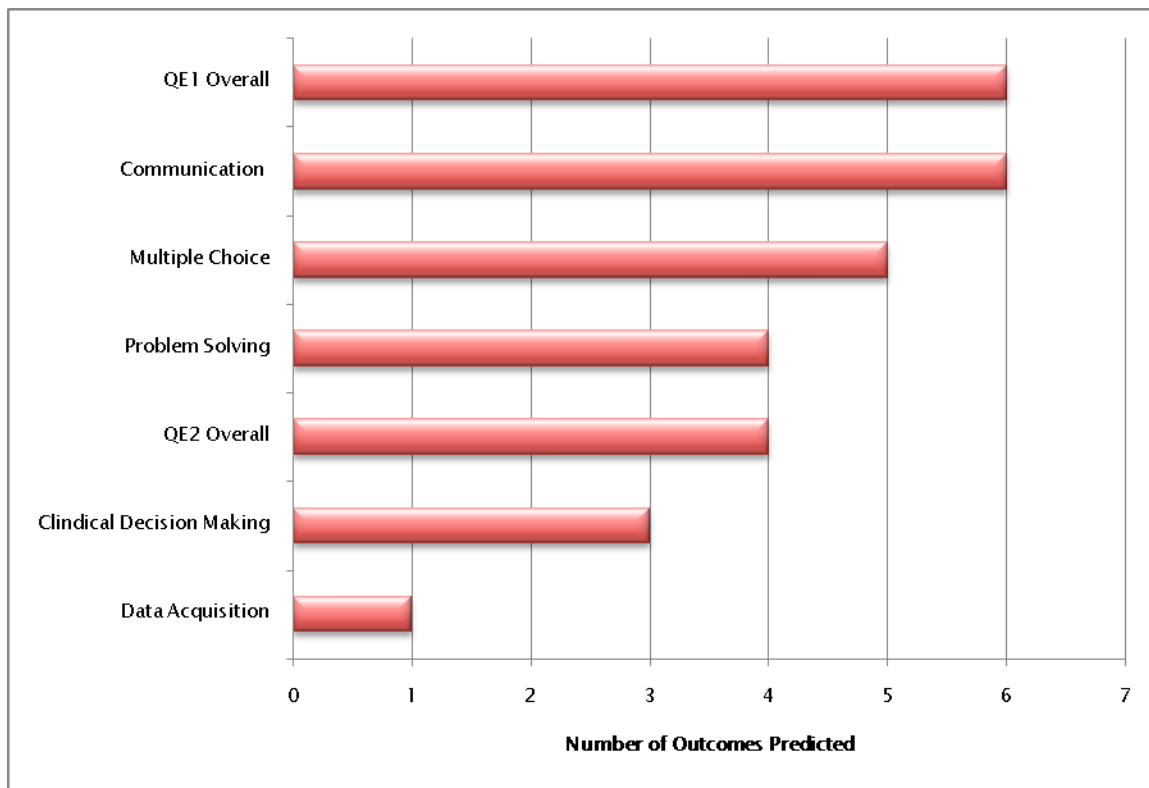
	Physicians (patients)	QE1 Overall	Multiple Choice	Clinical Decision- Making	QE2 Overall	Communi- cation	Data Acquisition	Problem- Solving
<b>Overall Performance</b>								
<b>Overall Quality/Safety of Care</b>	Number	<b>Relative Risk</b>	<b>Relative Risk</b>	<b>Relative Risk</b>	<b>Relative Risk</b>	<b>Relative Risk</b>	<b>Relative Risk</b>	<b>Relative Risk</b>
		<i>(P-value)</i>	<i>(P-value)</i>	<i>(P-value)</i>	<i>(P-value)</i>	<i>(P-value)</i>	<i>(P-value)</i>	<i>(P-value)</i>
Communication/Quality of Care Complaint <sup>1</sup>	3424	<b>1.39(0.001)</b>	<b>1.25(0.02)</b>	<b>1.50(&lt;0.001)</b>	<b>1.19(0.05)</b>	<b>1.38&lt;0.001)</b>	0.98(0.85)	1.02(0.76)
Unsatisfactory Peer Review Assessment <sup>1</sup>	208	<b>6.66(0.02)</b>	<b>8.33(0.01)</b>	0.93(0.91)	3.33(0.08)	0.68(0.55)	2.63(0.12)	<b>2.94(0.04)</b>
<b>Acute Care Management</b>								
<b>Management of Infection</b>								
Antibiotic for viral respiratory infection- Male MD	170 (73,519)	<b>1.32(0.03)</b>	<b>1.59 (0.006)</b>	0.64(0.08)	<b>1.44 (0.006)</b>	<b>1.30(0.03)</b>	1.32(0.55)	1.14 (0.16)
Antibiotic for viral respiratory infection- Female MD	266 (64,314)	0.74(0.06)	<b>0.71 (0.02)</b>	0.96(0.74)	0.77(0.06)	0.98(0.80)	0.89 (0.22)	<b>0.76 (0.01)</b>
Inappropriate antibiotics for bacterial infection	440 (151,109)	0.90(0.17)	0.86 (0.08)	1.06(0.51)	0.92(0.29)	1.12(0.07)	0.94(0.35)	<b>0.88 (0.02)</b>
<b>Chronic Care Management</b>								
<b>Management of Out-of-Control Asthma<sup>2</sup></b>								
Persistent Excess Rescue Med. Use	327 (1,960)	0.95(0.45)	0.95(0.36)	0.98(0.75)	0.99(0.85)	0.96 (0.46)	0.98(0.77)	1.00(0.94)
ER Visits for Asthma	327 (1,960)	1.07(0.70)	1.29(0.11)	<b>0.64(0.008)</b>	1.03(0.86)	1.23(0.25)	1.08(0.67)	0.92(0.44)
Inhaled Steroid Use (any/none)	327 (1,960)	<b>1.10(0.03)</b>	<b>1.11(0.01)</b>	1.06(0.27)	<b>1.12(0.009)</b>	<b>1.12(0.008)</b>	1.03(0.47)	1.03(0.38)
Ratio Inhaled Steroids-Rescue Med. (hi/lo)	327 (1,960)	1.05(0.27)	1.06(0.18)	1.02(0.74)	0.98(0.69)	0.98(0.60)	0.94(0.12)	1.03(0.36)
<b>Newly Treated Hypertension (HTN)<sup>2</sup></b>								
Cost-Effective Therapy Choice								
Uncomplicated HTN	508(3315)	1.09 (0.59)	1.01 (0.93)	1.27 (0.12)	1.01 (0.97)	0.86 (0.17)	1.00 (0.99)	0.98 (0.84)
Complicated HTN	590(3887)	1.17 (0.42)	1.10 (0.61)	1.28 (0.18)	1.03 (0.84)	1.12 (0.39)	0.91 (0.53)	1.08 (0.62)
Non-persistence with Treatment								
First 2 months	588(5259)	<b>0.74 (0.002)</b>	<b>0.78 (0.003)</b>	<b>0.81 (0.03)</b>	0.86 (0.06)	0.98 (0.76)	0.94 (0.50)	0.88 (0.07)
First 6 months	588(3818)	0.87 (0.20)	0.90 (0.29)	0.86 (0.19)	0.84 (0.11)	<b>0.80 (0.01)</b>	1.01 (0.91)	0.86 (0.09)
<b>Prevention</b>								
<b>Breast Cancer Screening<sup>2</sup></b>								
Screening Mammography	413 (127,812)	<b>1.15 (0.05)</b>	1.13 (0.07)	1.11 (0.16)	1.11 (0.13)	<b>1.28 (0.000)</b>	1.01 (0.85)	<b>1.11 (0.03)</b>
Diagnostic Mammography	413 (127,812)	0.93 (0.28)	0.93 (0.29)	0.95 (0.46)	0.87 (0.02)	0.86 (0.009)	0.87 (0.005)	0.94 (0.26)

Footnotes

1. Relative risk per 2 standard deviation **reduction** in score. Each score was modeled separately in a multivariate model that adjusted for other physician and patient characteristics, when patient was used as the unit of analysis, as well as clustering of patients within physician.
2. Relative risk per 2 standard deviation **increase** in score.

In evaluating multiple outcomes with multiple examination scores, there is a risk of finding significant results by chance. Therefore, we also examined overall trends in the associations between examination scores and practice outcomes, to provide a view that may be informative in considering future examination policy, Figure 1 presents an overall summary of the number of statistically significant associations between overall QE1 and QE2 examination scores and sub-scores and safety and quality outcomes in practice. Among the 12 practice outcomes examined, the overall MCCQE1 and the MCC communication sub-score were both independently associated with the greatest number of safety and quality practice outcomes; both being associated with six out of the twelve outcomes examined. The weakest relationships were for the data acquisition sub-score of the MCCQE2 examination in that it was associated with only one of the twelve practice outcomes examined. These frequency counts of statistically significant associations are not adjusted for differences in the reliability of the overall scores and sub-scores. With these adjustments, the communication sub-score, which was the least reliable score of the combined QE1 and QE2 examinations, would likely be more predictive of future practice outcomes.

**Figure 1 Summary of the Number of Significant Associations between MCCQE1 and MCCQE2 Examination Scores and Performance on Safety/Quality Indicators in the First 3 to 10 years of Practice**



## Summary and Recommendations

Both the MCCQE1 and MCCQE2 examinations provide important diagnostic information about the safety and quality of future practice, at least for the first 10 years. Lower scores on these examinations are associated with a greater likelihood of problems in practice, and higher scores are associated with better treatment outcomes in relationship to chronic disease management and prevention. The MCCQE2 examination does appear to provide additional information about future practice. The most predictive component of the MCCQE2 examination is the communication sub-score. It is important to note that the communication sub-score has the lowest reliability, a problem that would attenuate the magnitude of associations between this examination score and practice outcomes. If communication was measured more reliably, by increasing the sample of stations in which it was measured, the communication score should be even more predictive of future practice. The data acquisition sub-score of the MCCQE2 showed the least promise, as it was associated with only one of the twelve outcomes examined. Of interest, this sub-score was associated with the likelihood of potentially excessive diagnostic testing. As no other outcomes related to diagnostic testing were assessed in this study, the possibility that better scores in data acquisition are associated with less inappropriate and excessive diagnostic testing. Based on these results, the following recommendations are made:

1. As the communication sub-score of the MCCQE2 examination is not correlated with other aspects of either the MCCQE1 examination or the MCCQE2 examination, and is the most predictive score for future practice, consideration should be given to:
  - a. increasing the reliability of communication assessment, and
  - b. establishing a minimum passing standard for both communication ability and for medical competences assessed by other components of the examination.

A disjunctive scoring approach is currently used for the communication component of the USMLE clinical skills examination. The findings of this research fit with all prior evidence that shows that physician-patient communication is a separate ability from medical knowledge, diagnostic, and management abilities.

2. Consideration should be given to reviewing the approach that is taken to measurement of the MCCQE2 data acquisition subcomponent of the examination. The amount of test time devoted to data acquisition assessment relative to its predictive capacity should be improved. Indeed, it was difficult to identify outcomes in practice that may be related to deficiencies in data acquisition, and perhaps more strategic sampling of conditions where data acquisition by history and physical is an essential skill for optimal management should be considered.
3. The differences in the predictive capacity of QE1 and QE2 examination scores for antibiotic prescribing between male and female physicians is a very interesting finding which may reflect examination format effects or differences in the effects of clinical competence on approaches to practice. These findings require further investigation and research.

4. In collaboration with the Medical Licensing Authorities, consideration could be given to using scores achieved on the MCCQE1 and MCCQE2 examination as a risk stratification and management approach in practice. For example, practice review/ audit, or re-certification may be initially done within the first 4-5 years for physicians in the lowest score quartile, compared to every 7-10 years for others. Annual professional development could be targeted to areas where scores were the lowest that would have an impact in a physician's practice population.
5. The study examined one indicator of cost-effectiveness; choice of therapy for hypertension, and found no association with examination scores, possibly because we measured only one indicator, or because the examination was never intended to measure the cost-effectiveness of physician decision-making. The MCC needs to determine whether cost-effectiveness is a component of physician competence that needs to be assessed, and if so by what means.
6. The capacity for ongoing monitoring and the infrastructure needed to conduct follow-up evaluation in both residency and future practice is highly recommended as the foundations for future examination policy development. The conclusions and recommendations emanating from this study are based on the experience of four examination cohorts of physicians, practicing in two provinces. While this is the largest study done of practice outcomes, sound examination policy will be most effectively built by providing a mechanism for ongoing monitoring of the effects of changing examination policies on practice outcomes of Canadian physicians in multiple provinces. Establishing an examination blueprint that would enable test content to be adapted to safety and quality problems in current and future practice will be desirable. The Bordage study, recently funded by the MCC, will be a first step in that process.