Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: http://www.tandfonline.com/loi/imte20

Clinical teaching as part of continuing professional development: Does teaching enhance clinical performance?

Jocelyn M. Lockyer, Carol S. Hodgson, Tzu Lee, Sonia Faremo, Bruce Fisher, William Dafoe, Verna Yiu & Claudio Violato

To cite this article: Jocelyn M. Lockyer, Carol S. Hodgson, Tzu Lee, Sonia Faremo, Bruce Fisher, William Dafoe, Verna Yiu & Claudio Violato (2016) Clinical teaching as part of continuing professional development: Does teaching enhance clinical performance?, Medical Teacher, 38:8, 815-822, DOI: 10.3109/0142159X.2015.1112895

To link to this article: http://dx.doi.org/10.3109/0142159X.2015.1112895

Published online: 30 Nov 2015.

2016, 38: 815–822 MEDICAL TEACHER

Clinical teaching as part of continuing professional development: Does teaching enhance clinical performance?

JOCELYN M. LOCKYER[1], CAROL S. HODGSON[2], TZU LEE[2], SONIA FAREMO[3], BRUCE FISHER[1], WILLIAM DAFOE[1], VERNA YIU[1] & CLAUDIO VIOLATO[4]

[1]University of Calgary, Canada, [2]University of Alberta, Canada, [3]Mount Royal University, Canada, [4]Wake Forest University, USA

Abstract

Introduction: Physicians identify teaching as a factor that enhances performance, although existing data to support this relationship is limited.

Purpose: To determine whether there were differences in clinical performance scores as assessed through multisource feedback (MSF) data based on clinical teaching.

Methods: MSF data for 1831 family physicians, 1510 medical specialists, and 542 surgeons were collected from physicians' medical colleagues, co-workers (e.g., nurses and pharmacists), and patients and examined in relation to information about physician teaching activities including percentage of time spent teaching during patient care and academic appointment. Multivariate analysis of variance, partial eta squared effect sizes, and Tukey's HSD post hoc comparisons were used to determine between group differences in total MSF mean and subscale mean performance scores by teaching and academic appointment data.

Results: Higher clinical performance scores were associated with holding any academic appointment and generally with any time teaching versus no teaching during patient care. This was most evident for data from medical colleagues, where these differences existed across all specialty groups.

Conclusion: More involvement in teaching was associated with higher clinical performance ratings from medical colleagues and co-workers. These results may support promoting teaching as a method to enhance and maintain high-quality clinical performance.

Introduction

Ensuring that medical school teachers are clinically current and demonstrate competence in patient care is critical to meeting patient and societal needs and expectations. As medical schools and residency programs have expanded and evolved, increasing numbers of physicians are assuming roles as preceptors. The association between teaching and the quality of patient care is difficult to ascertain. While there is some evidence that patient care is better when delivered by physician teachers, the narrow clinical focus of these studies limit the generalizability of the findings (Eid et al. 2010; Greer et al. 2011).

Why might involvement in teaching improve clinical performance? Physicians who teach medical students and residents believe teaching is a powerful stimulus for their own learning and staying current (Lockyer et al. 2011; Wenrich et al. 2011), and describe how teaching helps them by expanding their knowledge and skills, deconstructing their clinical

Practice points

- Physicians with higher scores on a multisource feedback assessment report spending more time teaching in patient care settings.
- Engaging in teaching may involve more explicit and reflective thinking about the content to be taught and practices to be followed and may explain the association between higher scores on multisource feedback assessment and involvement in teaching.
- Recognizing physician's contributions to teaching within recertification/revalidation and maintenance of competence programs can be justified on the basis of the association between multisource feedback scores and University/medical school appointments.
- Potential teachers may be motivated by the association with clinical performance.

Correspondence: Jocelyn Lockyer PhD, Senior Associate Dean, Faculty of Medicine, University of Calgary, 3280 Hospital Drive NW, Calgary, AB T2N 4Z6, Canada. Tel: 403 220 4248; Fax: 403 220 1828; E-mail: lockyer@ucalgary.ca

ISSN 0142-159X print/ISSN 1466-187X online/15/080815–8
DOI: 10.3109/0142159X.2015.1112895

experience (e.g., deepening, broadening, slowing one's practice), and fostering greater self-reflection (e.g., awareness of being a role model) (Wenrich et al. 2011). The results of a recent survey of nearly 3000 fellows of the Royal College of Physicians and Surgeons of Canada (RCPSC) show that teaching others is one of the primary ways physicians report enhanced behavior and performance (Campbell et al. 2009). A Dutch study also identified teaching as one of the most important methods for professional development (van de Wiel et al. 2011). Additional support for this relationship is provided by the results of another study which showed that residents' ratings of teachers' clinical excellence were correlated with teaching effectiveness (McOwen et al. 2007). That study also showed that clinical effectiveness was rated higher for clinical and tenure-track faculty members than for staff physicians, but teaching effectiveness was not significantly different. This may imply that physicians with more of a teaching role may be better clinicians, even though physicians from all groups were seen as good teachers (McOwen et al. 2007).

The study of the relationship between teaching and overall clinical competence is important for medical schools. Role modeling and the hidden curriculum affect what students learn. Having faculty whose clinical work, professionalism, and communication skills are of high quality is integral to developing the next group of learners. Multisource feedback (MSF) offers an approach to assessing broadly based physician skills. In MSF, data from medical colleagues, non-physician co-workers (e.g., nurses, pharmacists, technical staff), and patients are collected. Depending on the instrument and its domains, a broad range of competencies can be assessed.

In order to examine the association between physician clinical performance and teaching, we utilized MSF data collected by the College of Physicians and Surgeons of Alberta (CPSA), the Physician Achievement Review (PAR) program and two datasets pertaining to teaching, namely, academic appointment information from the Universities of Alberta and Calgary and self-reported data about time spent teaching during patient care held by the CPSA. Our research question was: do family physicians, medical specialists and surgeons, who vary by level of teaching (as measured by appointment type and percent teaching) differ in their clinical performance as measured by PAR scores from three sources (i.e., medical colleagues; patients, and co-workers)?

Methods

Databases

The CPSA, PAR program (College of Physicians and Surgeons of Alberta 2015) requires all physicians in Alberta to participate in MSF as part of continuous improvement every five years. The program provides questionnaire-based feedback about a broad range of competencies including medical expert, communicator, collaborator, professional, and manager. Different but comparable suites of instruments have been created for family medicine (Hall et al. 1999; Lockyer et al. 2009; College of Physicians and Surgeons of Alberta 2015), surgical practice (Violato et al. 2003; Lockyer et al. 2012; College of Physicians and Surgeons of Alberta 2015), and medical practice including pediatrics (Violato & Lockyer 2006; Violato et al. 2006; College of Physicians and Surgeons of Alberta 2015), internal medicine (Violato & Lockyer 2006; College of Physicians and Surgeons of Alberta 2015), and psychiatry (Violato et al. 2008a). Between 1999 and 2012, almost 11,000 MSF assessments of physicians had been completed in a population of physicians that continuously grew over that period to 8000 physicians. Of these, many physicians had 2 and some had 3 assessments and the data demonstrated physician performance improved between two iterations in line with earlier studies showing that physicians used the data to inform practice changes (Violato et al. 2008b). The largest datasets are for family medicine/general practice ($n = 4623$), the medical specialties ($n = 2911$), and surgery ($n = 1275$).

Pivotal Research Inc., the administrator of the PAR program, created three separate anonymous databases, for each of three instruments (medical colleague, co-worker, and patient). The data for individual physicians were drawn from three separate sources: Alberta PAR data, CPSA annual registration data, and academic appointment data from the Universities of Calgary and Alberta.

PAR data

MSF data for the family physician, medical specialties, and surgical specialties versions of the PAR surveys provided information about the physicians' clinical performance related to CanMEDS competencies (e.g., professionalism and communication). Data for each physician included the most recent assessments from 8 medical colleagues (other physicians), 8 co-workers (e.g., nurses, pharmacists), and 25 patients, collected between 2007 and 2011. Physicians select their own medical colleagues and co-workers and have their office staff invite consecutive patients to complete the questionnaire following their visit. Each specialty group has a different set of items and each survey contains between 19 and 40 items. Items are scored on 5-point scales as follows: (1) for the medical colleague and co-worker instruments, 1 = among the worst and 5 = among the best, with an "unable to assess" category and (2) for the patient instruments 1 = strongly disagree and 5 = strongly agree (for copies of the instruments, see: http://par-program.org/par/for-physicians/how-par-works/how-par-is-scored/par-questionnaires/).

CPSA annual registration data

The CPSA data collected during the Fall of 2011 for registration in 2012, provided information about physicians' practice location and percentage of time spent in different activities such as teaching, administration, and research. For the purposes of this study, only data related to percentages of time physicians spent teaching within patient care activities were included.

Academic appointment data

The Universities of Calgary and Alberta provided information about those physicians who held clinical (part-time) or full-time appointments within their Faculties of Medicine.

Clinical faculty are physician teachers in private practice who are recognized for their work by a university title (e.g., Clinical Assistant Professor) but receive little or no recompense. Academic appointment data were collected in July 2012 by going through the University records and coding all faculty members as part-time or full-time. By elimination, other MDs in the Pivotal dataset who had PAR data were identifiable as not having an appointment.

All physicians who participated in the PAR program during the last five years were included if PAR data and their 2011 CPSA annual registration data were available.

Data analysis

Pivotal Research held the PAR data with names and a unique ID provided by the CPSA. The CPSA registration data used the same ID number along with the physician's name. The University datasets provided the names of physicians and their status. Pivotal created one database and then removed the names and changed the unique ID number to ensure the data were anonymous.

Separate analyses were done for each of the three groups of physicians and three sets of instruments. The data from multiple surveys per physician were aggregated. Using subscales based on recent studies employing factor analyses conducted previously (Violato & Lockyer 2006; Lockyer et al. 2009, 2012) both subscale scores and total scale scores were calculated. The CPSA Annual Registration data on percent time teaching in patient care settings were collected in the following categories: 0%, 1–5%, 6–15%, 16–25%, 26–40%, 41–60%, 61–75%, 86–95%, and 96–100%. While participants did respond in all categories, the large majority of responses fell below 15%. For each specialty group, approximately 18% of participants indicated that they spent more than 15% of their time teaching during patient care, and most of those indicated that they spent less than 40% of their time teaching during patient care. For this reason, the data were recoded into the following categories: 0%, 1–5%, 6–15%, and >15%.

We employed multivariate analysis of variance (MANOVA) to explore between group differences. Independent variables were physician specialty, academic appointment, and percent time teaching during patient care. The dependent variables were the PAR total scores and PAR subscale scores (i.e., professionalism, patient care, clinical competence, and communication). Analyses were conducted using SPSS version 19. Due to recognized limitations in significance testing, especially with large sample sizes, we also calculated partial eta squared effect sizes to indicate the relative size of the effect [0.01 is a small effect size, 0.06 is medium effect and 0.14 is a large effect size (Nandy 2012)]. Tukey's HSD post hoc comparisons were conducted to test for significant differences by academic appointment (none, part-time, vs. full-time) and by percent teaching during patient care (0%, 1–5%, 6–15%, and >15%). A priori a significance level of 0.05 was considered for significance.

The study was approved by the University of Calgary Conjoint Health Ethics Research Board and by the University of Alberta Human Ethics Research Office.

Results

Data were available for 1831 family physicians, 1510 medical specialists, and 542 surgeons. The results of the one- and two-way MANOVA analyses are presented in Tables 1–3 summarized by instrument source (medical colleague, co-worker, or patient) for the three groups of physicians (family medicine, medical specialists, and surgeons). The original factor analyses (Violato & Lockyer 2006; Lockyer et al. 2009, 2012) identified different factors for each of the specialties and each of the instruments. In each table, the subscales for each instrument (based on the previously identified factors) are presented in the left column. The overall mean scores on each PAR subscale and the total aggregate mean (on the original 5-point scale) are presented for each group and each level of appointment (no appointment, part-time or full-time) and percent time teaching during patient care (0%, 1–5%, 6–15%, and >15%). Significance levels and effect sizes (partial eta squared, η^2) are also included.

The analysis of the medical colleague data shows that for appointment status, results were statistically significant for all three specialty types. Effect sizes for the total aggregate mean were $\eta^2 = 0.036$ for family physicians, $\eta^2 = 0.013$ for medical specialties, and $\eta^2 = 0.02$ for surgeons, all of which are considered small effect sizes (Table 1). Post hoc results indicate that for family medicine and medical specialists, physicians who had either a part-time or full-time appointment had significantly higher PAR scores on all subscales than those with no faculty appointment. For surgeons, the results were somewhat more mixed, but generally having a part-time or full-time appointment also resulted in higher PAR scores than those with no faculty appointment. Those with a self-reported higher percent time teaching had higher PAR scores on all the subscales and on the instruments overall for family physicians and medical specialists, but none were significant for surgeons (Table 1). Effect sizes for the total aggregate mean were $\eta^2 = 0.013$ for family physicians and $\eta^2 = 0.010$ for medical specialists, both of which are small effect sizes. Post hoc results indicate that family medicine physicians who spent more than 5% time teaching during patient care had higher PAR scores than those who did not teach at all. For medical specialists, those who did any teaching had significantly higher PAR scores than those who did not teach (Table 1).

For the co-worker data, the total aggregate mean scores were not statistically significant although several subscales were significant with small effect sizes (Table 2). For appointment status, family physicians and medical specialists with any faculty appointment (i.e., part-time or full-time) had higher scores than someone with no appointment (Table 2). Generally, for all specialist types, those with greater than 5% time teaching had higher PAR scores on some of the subscales (Table 2).

For the patient data, there were no significant differences by appointment status. For percent time teaching, there were a few subscales with overall significance, but effect sizes were extremely small, and there were no significant post hoc comparisons (Table 3).

Table 1. Mean PAR scale scores, overall significance level, effect size, and post hoc comparisons are shown by specialty and appointment and by specialty and percent teaching both during patient care for the medical colleague instruments.

Specialty/scales (total number of surveys administered)	Appointment, $n = 3875$					Percent teaching during patient care, $n = 3437$					
	None	Part-time	Full-time	*p* Level*	Effect size	0%	1–5%	6–15%	>15%	*p* Level*	Effect size
	($n = 1870$) Mean (SD)	($n = 1543$) Mean (SD)	($n = 462$) Mean (SD)		(η^2)**	($n = 492$) Mean (SD)	($n = 1228$) Mean (SD)	($n = 1075$) Mean (SD)	($n = 642$) Mean (SD)		(η^2)**
Family physicians (13,526)	($n = 1337$)	($n = 461$)	($n = 29$)			($n = 393$)	($n = 594$)	($n = 407$)	($n = 175$)		
Professionalism	66.90 (4.13)[a]	68.33 (3.98)[b]	69.71 (3.33)[b]	0.001	0.028	67.03 (3.82)[a]	67.22 (4.10)[a,b]	67.79 (4.29)[b,c]	68.45 (3.94)[c]	0.001	0.010
Clinical competence	52.89 (3.31)[a]	54.27 (3.10)[b]	55.44 (2.35)[b]	0.001	0.039	52.95 (3.03)[a]	53.25 (3.26)[a,b]	53.71 (3.44)[b,c]	54.12 (3.17)[c]	0.001	0.012
Psychosocial	17.46 (1.37)[a]	17.92 (1.23)[b]	18.67 (1.07)[c]	0.001	0.032	17.45 (1.33)[a]	17.57 (1.34)[a,b]	17.79 (1.33)[b,c]	17.98 (1.24)[c]	0.001	0.015
Total (aggregate mean)	4.43 (0.27)[a]	4.53 (0.26)[b]	4.64 (0.21)[b]	0.001	0.036	4.43 (0.25)[a]	4.45 (0.27)[a,b]	4.49 (0.28)[b,c]	4.53 (0.26)[c]	0.001	0.013
Medical specialists (11,253)	($n = 368$)	($n = 781$)	($n = 357$)			($n = 76$)	($n = 455$)	($n = 489$)	($n = 334$)		
Patient management	75.93 (4.28)[a]	76.96 (4.15)[b]	76.52 (4.46)	0.001	0.010	75.31 (4.20)[a]	76.78 (4.13)[b]	76.76 (4.31)[b]	76.99 (4.03)[b]	0.05	0.007
Clinical assessment	27.02 (1.67)[a]	27.39 (1.54)[b]	27.42 (1.74)[b]	0.001	0.010	26.90 (1.72)[a]	27.43 (1.54)[b]	27.38 (1.60)	27.39 (1.53)		0.005
Professional development	39.80 (2.74)[a]	40.57 (2.51)[b]	41.04 (2.34)[c]	0.001	0.029	39.23 (2.78)[a]	40.59 (2.42)[b]	40.69 (2.47)[b]	40.78 (2.37)[b]	0.001	0.019
Communication	26.71 (1.78)[a]	27.14 (1.68)[b]	26.94 (1.78)	0.001	0.010	26.60 (1.70)[a]	26.97 (1.65)	27.06 (1.79)	27.25 (1.60)[b]	0.05	0.006
Total (aggregate mean)	4.46 (0.26)[a]	4.53 (0.25)[b]	4.53 (0.26)[b]	0.001	0.013	4.42 (0.25)[a]	4.52 (0.24)[b]	4.53 (0.25)[b]	4.54 (0.24)[b]	0.01	0.010
Surgeons (4073)	($n = 165$)	($n = 301$)	($n = 76$)			($n = 23$)	($n = 179$)	($n = 179$)	($n = 133$)		
Patient care	81.80 (4.21)	82.42 (4.10)	82.97 (3.79)		0.009	82.95 (3.97)	82.53 (4.07)	82.14 (4.14)	82.46 (3.69)	–	0.002
Clinical competence	22.98 (1.34)[a]	23.36 (1.10)[b]	23.48 (1.11)[b]	0.001	0.026	23.28 (1.14)	23.35 (1.09)	23.27 (1.13)	23.23 (1.17)	–	0.002
Communication and humanistic	31.36 (2.00)[a]	31.62 (1.89)	32.13 (1.73)[b]	0.05	0.016	31.43 (2.01)	31.58 (1.92)	31.60 (1.87)	31.91 (1.69)	–	0.002
Professional development	17.67 (1.33)[a]	18.08 (1.18)[b]	18.69 (0.96)[c]	0.001	0.065	17.80 (1.60)	17.96 (1.26)	18.17 (1.10)	18.16 (1.15)	–	0.009
Total (aggregate mean)	4.52 (0.24)[a]	4.57 (0.22)	4.63 (0.21)[b]	0.01	0.020	4.57 (0.23)	4.57 (0.22)	4.56 (0.22)	4.58 (0.20)	–	0.000

*p Levels are only provided for those with significance levels at $p < 0.05$ or less.

**Partial eta squared.

[a,b,c]Different letters indicate Tukey's HSD post hoc comparisons that are different at $p < 0.05$. If the same letter or no letter is given (e.g., ''a'' vs. ''a'' or ''a'' vs. ''a,b''), there is no significant difference between levels. If one or more letters are different then there is a significant difference at $p < 0.05$.

Table 2. Mean PAR scale scores, overall significance level, effect size, and post hoc comparisons are shown by specialty and appointment and by specialty and percent teaching both during patient care for the co-worker instruments.

Specialty/scales (total number of surveys administered)	Appointment, $n = 3815$					Percent teaching during patient care, $n = 3382$					
	None	Part-time	Full-time	*p* Value*	Effect size	0%	1–5%	6–15%	>15%	*p* Value*	Effect size
	($n = 1822$) Mean (SD)	($n = 1535$) Mean (SD)	($n = 458$) Mean (SD)		(η^2)**	($n = 468$) Mean (SD)	($n = 1216$) Mean (SD)	($n = 1063$) Mean (SD)	($n = 635$) Mean (SD)		(η^2)**
Family physicians (12,740)	($n = 1312$)	($n = 459$)	($n = 29$)			($n = 380$)	($n = 590$)	($n = 403$)	($n = 173$)		
Professionalism	63.65 (4.43)[a]	64.32 (4.22)[b]	64.37 (2.94)	0.05	0.005	63.57 (4.81)[a]	63.80 (4.19)	64.40 (4.08)[b]	64.41 (3.67)	0.05	0.007
Communication	13.15 (1.15)[a]	13.42 (1.03)[b]	13.46 (0.91)	0.001	0.011	13.17 (1.23)	13.21 (1.11)	13.36 (1.07)	13.42 (1.01)	0.05	0.007
Total (aggregate) mean	4.52 (0.31)	4.57 (0.29)	4.58 (0.22)	–	0.006	4.51 (0.34)	4.53 (0.30)	4.57 (0.29)	4.58 (0.26)	–	0.007
Medical specialists (10,895)	($n = 347$)	($n = 776$)	($n = 354$)	–		($n = 66$)	($n = 447$)	($n = 482$)	($n = 331$)	–	
Humanistic and psychosocial	63.93 (4.48)[a]	64.56 (3.79)[b]	64.66 (3.70)[b]	0.05	0.005	63.00 (4.55)[a]	64.09 (4.10)[a]	64.50 (3.89)[b]	65.27 (3.52)[c]	0.001	0.015
Co-worker collegiality	27.66 (1.98)	27.77 (1.90)	27.76 (1.77)	–	0.001	27.32 (1.93)[a]	27.59 (2.08)[a]	27.74 (1.82)	28.03 (1.70)[b]	0.01	0.007
Written communication	8.65 (0.86)	8.61 (0.90)	8.65 (0.82)	–	0.001	8.56 (0.92)[ab]	8.51 (0.91)[a]	8.65 (0.84)[b]	8.87 (0.75)[c]	0.001	0.017
Total (aggregate) mean	4.56 (0.31)	4.59 (0.27)	4.59 (0.26)	–	0.003	4.49 (0.31)	4.55 (0.30)	4.59 (0.27)	4.64 (0.25)	–	0.014
Surgeons (4002)	($n = 163$)	($n = 300$)	($n = 75$)	–		($n = 22$)	($n = 179$)	($n = 178$)	($n = 131$)	–	
Humanistic and psychosocial	31.42 (2.52)[a]	31.87 (2.03)	32.22 (1.33)[b]	0.05	0.016	31.16 (2.40)	31.65 (2.32)	31.71 (2.18)	32.12 (1.77)	–	0.006
Co-worker collegiality	31.48 (2.70)[a]	31.99 (2.11)	32.52 (1.52)[b]	0.01	0.022	30.95 (2.75)[a]	31.68 (2.38)[a]	31.89 (2.33)	32.41 (1.88)[b]	0.01	0.016
Written communication	22.20 (1.79)	22.43 (1.54)	22.69 (1.13)	–	0.010	21.93 (1.75)	22.29 (1.66)	22.43 (1.62)	22.63 (1.35)	–	0.009
Total (aggregate) mean	4.48 (0.36)	4.54 (0.28)	4.60 (0.18)	–	0.018	4.42 (0.35)	4.51 (0.32)	4.53 (0.31)	4.59 (0.24)	–	0.011

**p* Levels are only provided for those with significance levels at $p < 0.05$ or less.

**Partial eta squared.

[a,b,c]Different letters indicate Tukey's HSD post hoc comparisons that are different at $p < 0.05$. If the same letter or no letter is given (e.g., ''a'' vs. ''a'' or ''a'' vs. ''ab''), there is no significant difference between levels. If one or more letters are different then there is a significant difference at $p < 0.05$.

Table 3. Mean PAR scale scores, overall significance level, effect size, and post hoc comparisons are shown by specialty and appointment and by specialty and percent teaching both during patient care for the patient instruments.

Specialty/scales (total number of surveys administered)	Appointment, $n = 3722$					Percent teaching during patient care, $n = 3297$					
	None	Part-time	Full-time	*p* Level*	Effect size	0%	1–5%	6–15%	>15%	*p* Level*	Effect size
	($n = 1805$) Mean (SD)	($n = 1474$) Mean (SD)	($n = 443$) Mean (SD)		(η^2)**	($n = 475$) Mean (SD)	($n = 1187$) Mean(SD)	($n = 1026$) Mean(SD)	($n = 609$) Mean (SD)		(η^2)**
Family physicians (42,680)	($n = 1303$)	($n = 451$)	($n = 28$)			($n = 387$)	($n = 577$)	($n = 391$)	($n = 170$)		
Professionalism and communication	110.84 (4.51)	111.00 (4.05)	110.98 (5.32)	–	0.000	111.56 (4.35)	110.81 (4.38)	111.12 (4.11)	111.25 (3.89)	–	0.005
Office personnel	27.38 (1.28)	27.47 (1.12)	27.89 (0.90)	–	0.003	27.50 (1.26)	27.39 (1.21)	27.52 (1.17)	27.39 (1.23)	–	0.002
Access to physician	20.17 (1.91)	20.02 (1.74)	20.54 (1.38)	–	0.002	20.51 (1.91)	20.13 (1.78)	20.27 (1.80)	19.84 (1.90)	0.001	0.008
Physical office	22.66 (1.12)	22.52 (1.15)	22.49 (0.83)	–	0.003	22.77 (1.13)	22.62 (1.11)	22.67 (1.02)	22.46 (1.09)	0.05	0.004
Total (aggregate mean)	4.53 (0.19)	4.53 (0.17)	4.55 (0.18)	–	0.000	4.56 (0.18)	4.52 (0.18)	4.54 (0.17)	4.52 (0.17)		0.006
Medical specialists (32,561)	($n = 342$)	($n = 724$)	($n = 339$)			($n = 66$)	($n = 432$)	($n = 459$)	($n = 307$)	–	
Patient care	124.46 (5.04)	124.28 (4.93)	124.66 (6.80)	–	0.001	124.00 (5.60)	124.09 (6.50)	124.50 (5.05)	125.29 (4.43)	0.05	0.004
Technical communication	9.22 (0.47)	9.24 (0.45)	9.25 (0.52)	–	0.000	9.21 (0.50)	9.22 (0.52)	9.25 (0.44)	9.29 (0.42)	–	0.003
Staff and office function	50.79 (1.94)	50.57 (2.03)	50.52 (2.21)	–	0.003	50.60 (2.19)	50.59 (2.37)	50.55 (1.90)	50.78 (1.76)	–	0.000
Total (aggregate mean)	4.61 (0.18)	4.60 (0.18)	4.61 (0.23)	–	0.001	4.60 (0.20)	4.60 (0.22)	4.61 (0.18)	4.63 (0.15)	–	0.003
Surgeons (12,775)	($n = 160$)	($n = 299$)	($n = 76$)	–		($n = 22$)	($n = 178$)	($n = 176$)	($n = 132$)	–	
Information giving	37.29 (1.50)	37.31 (1.47)	37.53 (1.30)	–	0.003	37.15 (1.60)	37.19 (1.49)	37.24 (1.47)	37.64 (1.37)	0.05	0.006
Office staff	53.82 (2.73)	53.87 (2.49)	53.58 (2.35)	–	0.001	54.14 (2.36)	53.90 (2.45)	53.65 (2.57)	53.82 (2.68)	–	0.002
Humanistic	37.30 (1.32)	37.37 (1.31)	37.23 (1.28)	–	0.002	37.36 (1.31)	37.28 (1.31)	37.33 (1.26)	37.37 (1.40)	–	0.001
Physical office	22.62 (1.01)	22.59 (1.12)	22.17 (0.93)	0.01	0.020	22.63 (0.89)	22.68 (0.99)	22.38 (1.17)	22.50 (1.04)	–	0.012
Personal communication	27.69 (1.28)	27.69 (1.15)	27.82 (1.05)	–	0.001	27.76 (1.29)	27.58 (1.21)	27.62 (1.16)	27.96 (1.10)	0.05	0.006
Total (aggregate mean)	4.58 (0.18)	4.59 (0.17)	4.57 (0.16)	–	0.001	4.59 (0.18)	4.58 (0.17)	4.57 (0.17)	4.60 (0.17)	–	0.000

*p Levels are only provided for those with significance levels at $p < 0.05$ or less.
**Partial eta squared.

Discussion

This study utilized a large dataset of nearly 4000 physicians from family medicine and the medical and surgical specialties with data provided by their medical colleagues, co-workers (e.g., nurses, pharmacists), and patients. The dataset was unique, combining data from three sources, namely MSF scores, physician registration information about work, and academic appointment information.

Our results suggest a relationship may exist between involvement in teaching and clinical performance, particularly for the data from medical colleagues, even though the effect sizes are generally small. The results add further evidence to support the thesis that teaching activities may enhance or maintain clinical performance, as provided in physician self-report data (Campbell et al. 2009; Lockyer et al. 2011; Wenrich et al. 2011). We recognize that the effect sizes are small.

It may be that engaging in teaching involves more explicit and reflective thinking about the content to be taught (Ericsson 2004, 2008; Wenrich et al. 2011; Kulasegaram et al. 2013). Clinician educators who provide scaffolding (or support) and coaching in authentic clinic contexts to promote learner development, and role modelling and comparisons to their own practice, may be stimulated to self-reflect on their own clinical work. Physicians who do not have students and residents may be less likely to have access to such stimuli for feedback and reflection on their daily medical practice. However, the relationship between teaching and clinical performance may not be causal and other explanations are warranted. The relationship may reflect a sampling effect whereby physicians who are the most energetic and committed to their clinical practices may also be the ones that self-select for teaching. As well, though some of the data are derived from full-time academic career physicians who chose their career, the majority of the physicians are likely those who are recruited as clinical faculty by both universities and are obligated to teach as they work within or in conjunction with a teaching hospital or outpatient facility that has medical students and residents. These physicians may have access to seminars on effective teaching, but they would not have had any exposure to interventions to enhance their clinical abilities. The differences between teaching involvement and clinical performance is most evident from the data provided by medical colleagues related to academic appointment status for all three groups and for percent time teaching in conjunction with patient care for family physicians and medical specialists. No significant results were found for surgeons for teaching during patient care, a result which may be due to the small number of surgeons who are not involved with teaching: $n = 23$ (4.3% of surgeons; Table 1). This small sample may result in a lack of statistical power to detect any significant differences. The co-worker data also show some differences between clinical performance and academic appointments and teaching. As before, the general lack of differences for surgeons for teaching during patient care is likely due to the small sample size ($n = 23$) and those who do not teach during patient care (Table 2).

The data provided by patients show very few differences between teaching involvement and performance. The items on the patient questionnaire are different than those on the colleague or co-worker ones and focus on the unique experiences of patients. The rating scales are also different; patients are asked how much they agreed with the statements whereas colleagues and co-workers are asked to compare the physician to other physicians. Moreover, patient data provide the highest ratings (mean = 4.60) compared to medical colleagues (mean = 4.45) and co-workers (mean = 4.50). The high patient scores likely reflect a ceiling effect which is a measurement artifact that precludes detecting statistical differences between groups. The item content differences, the rating scale difference, and the ceiling effects may all combine to result in the very few significant differences between appointment and teaching involvement for all three specialty groups that are evident in Table 3.

In addition to demonstrating the differences between clinical performance and teaching for colleague and co-worker data, the present findings are important for a number of reasons. The PAR program was originally developed in Alberta for Alberta physicians. It has since been adopted in other Canadian provinces. The results to date continue to support the robustness of the data and its use in informing physicians. In Canada, both the Royal College of Physicians and Surgeons of Canada and the College of Family Physicians of Canada enable physicians to document teaching, curriculum development, and assessment of students and trainees as part of the documentation for Maintenance of Competence/ Proficiency study credits. This study provides some evidence that these decisions were warranted and supports the continued allocation of study credits for teaching. The present findings also provide additional evidence of construct validity for the MSF PAR instruments since the expected between group differences were found: more involvement in teaching generally results in higher clinical performance ratings from medical colleagues and co-workers. At the Universities of Calgary and Alberta, like in many medical schools, clinical teachers provide a significant amount of the clinical teaching to medical students and residents. Whilst, the present results do not prove causality, they do affirm an encouraging association between individual teaching contributions and clinical performance.

There are several limitations to the study. The study is a naturalistic correlation study and as such it is not possible to control for the many variables that may influence the results. The study involved data from only one province in Canada and two medical schools. In the future, it may be possible to replicate the study with other provinces that have implemented the PAR program. It was done with three large specialty groups. Due to issues of confidentiality and group size, it was not possible to examine specific specialty or subspecialty groups within the medical and surgical specialties, nor to access socio-demographic or other data (e.g., year or school of graduation, practice location, or gender) that might have been illuminating. Further, we looked at two measures of teaching data and these are gross measures of teaching. Future work may include such variables as well as more granular aspects of teacher characteristics (e.g., teacher credentials, age, years of teaching experience, continuing medical education activity) as well as the quality of teaching (e.g., student ratings, peer ratings, direct observation).

Conclusion

This study provides objective evidence of a relationship between clinical performance and teaching. The findings are consistent with the ways that physicians describe the important role that teaching plays in the maintenance of competence (Campbell et al. 2009; Lockyer et al. 2011; Wenrich et al. 2011; van de Wiel et al. 2011), although other explanations of the findings, such as a sampling effect, should also be acknowledged. This is one of the few studies that have utilized data provided by medical colleagues, co-workers, and patients to examine the relationship between involvement in teaching and clinical performance.

Notes on contributors

JOCELYN M. LOCKYER, PhD, is Professor, Department of Community Health Sciences and Senior Associate Dean, Education, Cumming School of Medicine, University of Calgary, Calgary, Alberta.

CAROL S. HODGSON, PhD, is J Alan Gilbert Chair in Medical Education Research and Associate Professor, Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta.

TZU K. LEE, MB, BS, MSc (Exp Med) FRCPC, FRCP (UK), MACP, is Clinical Professor Emeritus, Department of Medicine (Division of General Internal Medicine), Faculty of Medicine of Dentistry, University of Alberta, Edmonton, Alberta.

SONIA FAREMO, PhD, is Research Proposal Developer, Health and Community Studies, Mount Royal University, Calgary, Alberta.

BRUCE FISHER, MD, MSc, FRCPC, is Professor Department of Medicine (Division of General Internal Medicine) and Associate Dean, Faculty Development, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta.

WILLIAM DAFOE, MD, FRCPC, is Professor, Department of Medicine (Division of Cardiology), Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta.

VERNA YIU, MD, FRCPC, is Professor, Department of Pediatrics (Division of Pediatric Nephrology), Faculty of Medicine and Dentistry, University of Alberta and Vice President Quality and Chief Medical Officer, Alberta Health Services, Alberta.

CLAUDIO VIOLATO, PhD, Professor, Department of Medicine, School of Medicine, Wake Forest University, Winston-Salem, North Carolina, USA.

Acknowledgements

The authors thank John Swiniarski, former Assistant Registrar, College of Physicians and Surgeons of Alberta and Steven Dennis and Deanne Hennig, Pivotal Research Inc. for the work required to create the databases. Other members of the Alberta Medical Education Study Team include: Herta Fidler, MSc, Research Associate, Continuing Medical Education and Professional Development, Faculty of Medicine, University of Calgary (Retired), Calgary, Alberta; Trevor Theman, MD FRCSC, Registrar, College of Physicians and Surgeons of Alberta, Edmonton, Alberta; and Gregg Trueman, BScN, MN, PhD Associate Professor, Department of Nursing, Mount Royal University, Calgary, Alberta.

Declaration of interest: The authors report no declarations of interest. The Standing Committee on Continuing Professional Development of the Association of Faculties of Medicine of Canada provided a grant for the study.

Ethical Approval: University of Calgary Conjoint Health Ethics Research Board and the University of Alberta Human Ethics Research Office.

References

Campbell CM, Lockyer JM, Fidler H. 2009. Maintenance of Certification Program Evaluation, Phase 2, Descriptive Analysis, Electronic Survey Responses, Internal Report for Royal College of Physicians and Surgeons of Canada, 20 May 2009.

College of Physicians and Surgeons of Alberta. 2015. Physician Achievement Review Program. [Accessed 16 June 2015] Available from www.par-program.org.

Eid SM, Boueiz A, Paranji S, Motivo C, Landis R, Abougergi MS. 2010. Patterns and predictors of proton pump inhibitor overuse among academic and non-academic hospitalists. Internal Med 49(23): 2561–2568.

Ericsson KA. 2004. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 79: S70–S81.

Ericsson KA. 2008. Deliberate practice and acquisition of expert performance: A general overview. Acad Emerg Med 15:988–994.

Greer RC, Powe NR, Jaar BG, Troll MU, Boulware LE. 2011. Effect of primary care physicians' use of estimated glomerular filtration rate on the timing of their subspecialty referral decisions. BMC Nephrol 14(12):1.

Hall W, Violato C, Lewkonia R, Lockyer J, Fidler H, Toews J, Jennett P, Donoff M, Moores D. 1999. Assessment of physician performance in Alberta: The Physician Achievement Review Project. CMAJ 161:52–57.

Kulasegaram KM, Grierson LE, Norman GR. 2013. The roles of deliberate practice and innate ability in developing expertise: Evidence and implications. Med Educ 47(10):979–989.

Lockyer J, Armson H, Chesluk B, Dornan T, Holmboe E, Loney E, Mann K, Sargeant J. 2011. Feedback data sources that inform physician self-assessment. Med Teach 33(2):e113–e120.

Lockyer J, Violato C, Wright B, Fidler H, Chan R. 2012. An analysis of long term outcomes for surgeons from three and four year curriculum. Can J Surgery 55(4):S163–S170.

Lockyer JM, Violato C, Wright B, Fidler HM. 2009. An analysis of long term outcomes of the impact of curriculum: A comparison of the 3 and 4 year medical school curricula. Acad Med 84(10):1342–1347.

McOwen KS, Bellini LM, Shea JA. 2007. Residents' ratings of clinical excellence and teaching effectiveness: Is there a relationship? Teach Learn Med 19(4):372–377.

Nandy K. 2012. Understanding and quantifying effect sizes. Available from http://nursing.ucla.edu/workfiles/research/Effect%20Size%204-9-2012.pdf.

van de Wiel MW, Van den Bossche P, Janssen S, Jossberger H. 2011. Exploring deliberate practice in medicine: How do physicians learn in the workplace? Adv Health Sci Educ Theory Pract 16(1):81–95.

Violato C, Lockyer JM, Fidler H. 2003. Multi source feedback: A method of assessing surgical practice. BMJ 326:546–548.

Violato C, Lockyer J, Fidler H. 2006. The assessment of pediatricians by a regulatory authority. Pediatrics 117:796–802.

Violato C, Lockyer J. 2006. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: Implications for self-directed learning. Adv Health Sci Educ 11:235–244.

Violato C, Lockyer J, Fidler H. 2008a. The assessment of psychiatrists in practice through multi source feedback. Can J Psychiatry 53(8):525–533.

Violato C, Lockyer JM, Fidler H. 2008b. Changes in performance: A 5-year longitudinal study of participants in a multi-source feedback programme. Med Educ 42(10):1007–1013.

Wenrich MD, Jackson MB, Ajam KS, Wolfhagen IH, Ramsey PG, Scherpbier AJ. 2011. Teachers as learners: The effect of bedside teaching on the clinical skills of clinician-teachers. Acad Med 86(7):846–852.