

The Impact of the Hospital Volume on the Performance of Residents on the General Medicine In-Training Examination: A Multicenter Study in Japan

Atsushi Mizuno¹, Yusuke Tsugawa², Taro Shimizu³, Yuji Nishizaki⁴, Tomoya Okubo⁵, Yusuke Tanoue⁶, Ryota Konishi⁷, Toshiaki Shiojiri⁸ and Yasuharu Tokuda⁹

Abstract

Objective Although several studies have been conducted worldwide on factors that might improve residents' knowledge, the relationship between the hospital volume and the internal medicine residents' knowledge has not been fully understood. We conducted a cross-sectional study to compare the relationships of the hospital volume and hospital resources with the residents' knowledge assessed by the In-training Examination.

Methods We conducted a retrospective survey and a clinical knowledge evaluation of postgraduate year 1 and 2 (PGY-1 and -2) resident physicians in Japan by using the General Medicine In-training Examination (GM-ITE) in 2014. We compared the ITE score and the hospital volume.

Results A total of 2,015 participants (70.6% men; age, 27.3±2.9 years old) from 208 hospitals were retrospectively analyzed. Generalized estimating equations were used, and the results revealed that an increasing number of hospitalizations, decreasing staff number, decreasing age and PGY-2 were significantly associated with higher GM-ITE scores.

Conclusion The hospital volume, such as the number of hospitalizations, is thus considered to have a positive impact on the GM-ITE scores.

Key words: in-training examination, general medicine, hospital volume, education

(Intern Med 55: 1553-1558, 2016)

(DOI: 10.2169/internalmedicine.55.6293)

Introduction

Standardizing education for residents remains challenging. Theoretically, the more cases residents experience, the better their knowledge. Although the impact of patient numbers on the learning curve was shown to have a ceiling effect for several reasons such as the workload, a previous study also showed that the number of patient encounters was related to residents' knowledge, which was assessed using the In-training Examination (ITE) scores (1, 2). In addition, not

only the number of patient encounters but also the quality of clinical practice must be considered.

The hospital volume is a valid proxy for the quality of not only surgery and interventional procedures but also several types of care (3-5). Therefore, we hypothesized that the hospital volume could also be a valid proxy for the quality and quantity of residents' education. To the best of our knowledge, no studies on the relationship between the hospital volume and the internal medicine residents' knowledge have been conducted.

In Japan, previous reports showed that the residents in

¹Department of Cardiology, St. Luke's International Hospital, Japan, ²Department of Health Policy and Management, Harvard T. H. Chan School of Public Health, USA, ³Department of General Internal Medicine, Tokyo Joto Hospital, Japan, ⁴Department of Cardiology, Juntendo University Graduate School of Medicine, Japan, ⁵Research Division, The National Center for University Entrance Examinations, Japan, ⁶Center for Clinical Epidemiology, St. Luke's Life Science Institute, Japan, ⁷Department of General Internal Medicine, Kanto Rosai Hospital, Japan, ⁸Department of General Internal Medicine, Asahi General Hospital, Japan and ⁹Japan Community Healthcare Organization, Japan

Received for publication August 4, 2015; Accepted for publication September 13, 2015

Correspondence to Dr. Atsushi Mizuno, atmizu@luke.ac.jp

university hospitals have less experience than those in non-university ones (6). Therefore, residents have preferred non-university-affiliated hospitals since new postgraduate medical education policies were implemented, despite the smaller hospital size (6). Under these circumstances, it is important to know what really affects residents' knowledge, such as the ITE score. We conducted a cross-sectional study to compare the relationships between the hospital volume and hospital resources and the residents' knowledge assessed by the ITE.

Materials and Methods

Participants and data collection

We conducted a retrospective analysis and a clinical knowledge evaluation involving postgraduate year 1 and 2 (PGY-1 and -2) resident physicians at 208 teaching hospitals in Japan by using the General Medicine In-training Examination (GM-ITE) administered in March 2014 (this exam is administered at the end of each academic year). Individual data collected for each resident included age, gender, PGY and test score, after informed consent had been obtained. The institutional review board of Tsukuba University approved the study.

Hospital volume and measures

All hospital data were collected via the REIS website (Residency Electronic Information System, <http://reis.mhlw.go.jp/common/ad0.php>) and the PMET website (Foundation for the Promotion of Medical Training, <http://guide.pmet.jp/web2014/index.html>). These websites include data on the hospital education and the hospital volume, which are updated regularly for matching applicants. These available data include the numbers of staff doctors, beds, emergency department patients, total hospitalizations and new outpatients during the year.

General Medicine In-training Examination (GM-ITE)

A methodology similar to that used for the GM-ITE was applied to develop the U.S. GM-ITE. The purpose of the U.S. ITE is to provide medical residents and program directors with an objective, reliable and valid assessment of each resident's personal performance on a written, multiple-choice examination, and the performance of each residency program is compared with that of its peers (7). The GM-ITE includes 100 questions that cover a wide range of issues, from clinical skills and pragmatic medical knowledge to psychosocial care of the patient, which are categorized into four major topics, namely, semiology/clinical diagnoses, physical examinations/procedures, medical interviews/professionalism and subspecialties. A three-hour fixed time limit was set for the exam. Every year, the examination questions are reviewed and the content validity is confirmed by a committee of experienced attending physicians organized by the Japan Organization of Advancing Medical Education

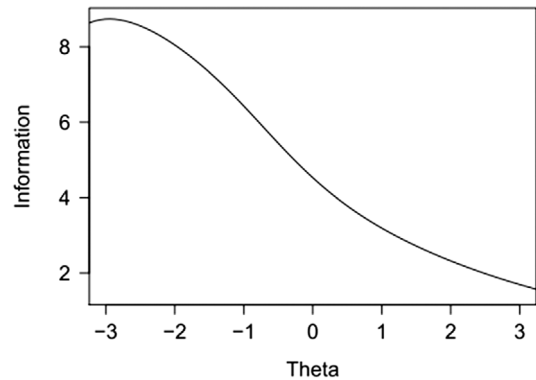


Figure 1. GM-ITE test information function. This graph shows that a lower theta (lower level) was associated with more information on the GM-ITE.

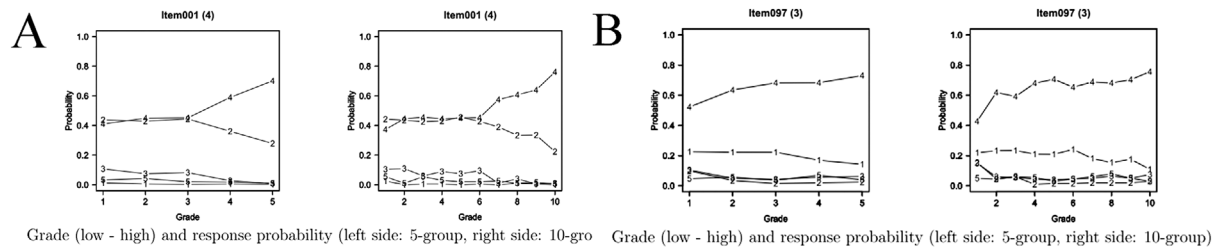
Program (JAMEP, a non-profit organization).

Score calculation

The examination gives a maximum score of 100 and a minimum score of zero; higher scores indicate an increased knowledge base in internal medicine. To check the reliability of the GM-ITE, we calculated the pass rate, the choice probability for each choice and the polyserial correlation coefficients of each question (8). The pass rate was calculated by the number of correct answers divided by the total number of participants. A trace line was drawn based on the choice probability for each choice of each quintile (and decile) of the total score. The polyserial correlation coefficient is based on the choice probability for each choice of each quintile (and decile) of the total score. The test information function also showed that the GM-ITE administered in 2014 was more effective at the lower level than at the middle and upper levels (Fig. 1). Samples of appropriate questions and inappropriate questions are shown in Fig. 2. We concluded that Question No. 97 was inappropriate since it had a negative correlation coefficient, and the pass rate was very low (6.3%).

Statistical analyses

We tested the hypothesis of a significant association between several hospital volume measures and the GM-ITE score by using generalized estimating equations (GEE) in order to account for the clustering of residents within each hospital. Continuous variables such as the number of total hospitalizations, staff doctors, emergency department patients, beds and new outpatients were divided into quartiles (Q1-Q4), avoiding linearity assumptions. Finally, we performed an additional analysis assessing whether there are differences in the hospital characteristics between university hospitals and community hospitals (Supplementary material). All analyses were conducted in the statistical environment R 3.1.2 (R Development Core Team, 2014; <http://www.r-project.org>).



Question number No.1
Pass rate was 52.0%

Polyserial correlations coefficient; 0.199

Question number No.97
Pass rate was 6.3%

Polyserial correlations coefficient; -0.067

Figure 2. Trace lines of appropriate and inappropriate tests. A shows that the selection of a correct answer (the correct answer was 4) was associated with higher-scoring groups, which showed good discrimination. B shows that Question No. 97 did not discriminate well among the respondents. One plausible reason for this is that choice 4 was too effective as a distracting option.

Table 1. Baseline Characters of Test Examinee and Working Hospitals.

Participants	n=2,015
Male, n(%)	1,422(70.6)
Age, y	27.3±2.9
Second year residency, n(%)	1,207(60.0)
<i>Hospital data (208 hospitals)</i>	
University, n(%)	14(6.7)
Salary per month, 10,000yen	34±7.5
Staff numbers, n	120.6±96.6
Number of ER patients, n	14,888.2±9,905.9
Bed numbers, beds	456.3±182.2
Length of stay, days	13.6±3.1
Number of hospitalizations, n	3,835.6±2,012.4
Number of new outpatients, n	7,682.1±15,032.4
<i>Test scores</i>	
Total score (maximum of 100)	65.7±8.2
Total score excluding inappropriate question (maximum of 99)	65.6±8.2
Test score about Diagnosis	17.0±2.5
Test score about Physical examination	15.3±2.9
Test score about Professionalism	16.9±2.5
Test score about Subspecialty	16.4±2.8

ER: emergency room

Results

A summary of the baseline characteristics and the hospital data is shown in Table 1. A total of 2,015 participants (70.6% men; age, 27.3±2.9 years old) from 208 hospitals were retrospectively analyzed. Sixty percent of the participants were second-year residents, while the rest were in the first year. Of the 208 hospitals, 14 were university hospitals and the others were community hospitals. The volume data for these hospitals were as follows: about 450 beds with 120 staff doctors and about 7,500 outpatients, and 15,000 emergency patients annually. The average GM-ITE score was 65.6±8.2 after the inappropriate question was excluded. The scores did not differ much between each category, and the diagnosis test score was the highest.

The results of a multivariate analysis (Table 2) showed that an increasing number of hospitalizations, decreasing age and second-year residency were significantly associated with higher GM-ITE scores.

A further categorical analysis revealed that age and PGY-1 were associated with all subcategories (Table 3). The number of hospitalizations was associated with the physical examination test score only. Staff numbers were negatively associated with the subspecialty score. Men scored lower than women on professionalism. We also found that university hospitals had lower test scores in spite of a larger number of staff compared with community hospitals (Supplementary material).

Discussion

We found a positive relationship between the hospital volume, such as the number of hospitalizations, and the GM-ITE score for physical examinations. Of the residents' characteristics, a decreasing age and PGY-2 had a positive impact on the GM-ITE score.

Several studies on the that improve residents' knowledge have been conducted (1, 2). Patient encounters are an important factor in improving the knowledge of internal medicine. However, the quality of patient care should also be considered. A teaching hospital can offer high-quality care, which might result in high educational quality (9, 10). The hospital volume, which is usually assessed via patient numbers or the number of procedures, is thought to be related to the quality of care (5, 11). Therefore, we hypothesized that the hospital volume might be an important predictor of a high ITE score. To our knowledge, no studies have examined the relationship between the test score and the hospital volume. Our study revealed that a larger number of hospitalizations was associated with a higher GM-ITE score, which was reasonable according to our hypothesis. We could not conclude that the number of hospitalizations was directly associated with the GM-ITE score only from the findings of this study,

Table 2. The Results of the Generalized Estimating Equations Model for the GM-ITE Score.

		Estimate (95%CI)		
Intercepts		62.04		
Hospital variables				
Presence of GIM department		0.7(-0.71, 2.1)		
		Q2 vs Q1	Q3 vs Q1	Q4 vs Q1
Hospitalizations	ref	1.19(-0.21, 2.59)	-0.17(-1.78, 1.44)	1.98(0.02, 3.94)*
Staff numbers	ref	-0.83(-2.16, 0.51)	-1.02(-3, 0.96)	-1.69(-4.27, 0.89)
Emergency patients	ref	-0.44(-1.89, 1.01)	-0.48(-1.9, 0.95)	0.3(-1.57, 2.17)
Bed numbers	ref	0.61(-0.78, 2)	0.65(-1.15, 2.44)	1.35(-0.88, 3.58)
Newly outpatients	ref	-0.71(-2.19, 0.77)	-0.32(-1.72, 1.08)	-0.19(-1.92, 1.55)
Resident profile				
		Age (26)	Age (27)	Age (≥27)
Age [comparing with Age(<26)]		-1.05(-2.15, 0.05)	-2.63(-3.8, -1.47)*	-3.99(-5.3, -2.68)*
Male gender (vs female)		0.13(-0.49, 0.75)		
PGY-2 (vs PGY-1)		2.62(1.8, 3.44)		

GM-ITE: general medicine in-training examinations, SE: standard error, GIM: general internal medicine, PGY: post-graduate year, Q1-Q4: quantiles of each variable, * means p<0.05

Table 3. The Results of the Mixed Effects Linear Regression Model for the GM-ITE Score.

	Test score about Diagnosis	Test score about Physical examination	Test score about Professionalism	Test score about Subspeciality
	Estimate (95%CI)	Estimate (95%CI)	Estimate (95%CI)	Estimate (95%CI)
Intercepts				
Hospital variables				
Presence of GIM department				
	0.11(-0.24, 0.47)	0.35(-0.08, 0.78)	0.19(-0.23, 0.61)	0.13(-0.26, 0.52)
Hospitalizations				
Q2 vs Q1	0.32(-0.04, 0.69)	0.56(0.09, 1.03)*	0.1(-0.39, 0.58)	0.1(-0.35, 0.54)
Q3 vs Q1	-0.02(-0.42, 0.38)	0.04(-0.44, 0.53)	-0.13(-0.7, 0.43)	-0.01(-0.51, 0.49)
Q4 vs Q1	0.17(-0.39, 0.74)	1.02(0.4, 1.64)*	0.39(-0.24, 1.03)	0.43(-0.15, 1.01)
Staff numbers				
Q2 vs Q1	-0.06(-0.45, 0.34)	-0.06(-0.46, 0.34)	-0.31(-0.71, 0.08)	-0.45(-0.92, 0.01)
Q3 vs Q1	-0.03(-0.54, 0.48)	-0.28(-0.89, 0.34)	-0.14(-0.75, 0.47)	-0.69(-1.25, -0.12)*
Q4 vs Q1	-0.45(-1.06, 0.17)	-0.65(-1.39, 0.1)	-0.09(-0.92, 0.73)	-0.72(-1.43, -0.01)*
Emergency patients				
Q2 vs Q1	-0.16(-0.55, 0.24)	-0.11(-0.56, 0.35)	-0.05(-0.46, 0.35)	-0.04(-0.5, 0.42)
Q3 vs Q1	-0.1(-0.49, 0.28)	-0.26(-0.74, 0.21)	-0.04(-0.48, 0.4)	-0.08(-0.54, 0.38)
Q4 vs Q1	0.12(-0.33, 0.56)	0.35(-0.28, 0.98)	-0.37(-0.97, 0.22)	0.29(-0.21, 0.79)
Bed numbers				
Q2 vs Q1	0.1(-0.26, 0.45)	0.14(-0.3, 0.57)	0.22(-0.27, 0.7)	0.28(-0.13, 0.69)
Q3 vs Q1	0.29(-0.16, 0.75)	-0.15(-0.67, 0.37)	-0.003(-0.6, 0.59)	0.52(-0.01, 1.05)
Q4 vs Q1	0.37(-0.18, 0.92)	0.3(-0.4, 1)	0.15(-0.57, 0.88)	0.48(-0.22, 1.18)
Newly outpatients				
Q2 vs Q1	-0.33(-0.74, 0.08)	-0.29(-0.72, 0.14)	-0.12(-0.59, 0.35)	-0.16(-0.57, 0.25)
Q3 vs Q1	0.07(-0.28, 0.42)	-0.47(-0.96, 0.02)	-0.09(-0.5, 0.32)	0.24(-0.17, 0.65)
Q4 vs Q1	0.14(-0.28, 0.57)	-0.44(-1.01, 0.12)	0.06(-0.46, 0.58)	0.03(-0.53, 0.58)
Resident profile				
Age [comparing with Age(<26)]				
Age (26)	-0.41(-0.72, -0.11)*	-0.54(-0.94, -0.15)*	0.06(-0.27, 0.39)	-0.16(-0.51, 0.19)
Age (27)	-0.85(-1.21, -0.49)*	-0.96(-1.38, -0.54)*	-0.27(-0.62, 0.08)	-0.56(-0.9, -0.22)*
Age (≥27)	-0.93(-1.28, -0.59)*	-1.55(-2.02, -1.08)*	-0.63(-1, -0.25)*	-0.96(-1.36, -0.56)*
Male gender (vs female)	0.15(-0.06, 0.35)	0.23(-0.01, 0.46)	-0.22(-0.43, 0)*	-0.004(-0.25, 0.24)
PGY-2 (vs PGY-1)	0.35(0.1, 0.6)*	0.99(0.71, 1.28)*	0.31(0.05, 0.57)*	1.06(0.76, 1.35)*

GM-ITE: general medicine in-training examinations, SE: standard error, GIM: general internal medicine, PGY: post-graduate year, Q1-Q4: quantiles of each variable, * means p<0.05

but residents who worked at hospitals with more hospitalizations tended to have higher scores.

Learning how to perform physical examinations remains challenging. Recent education systems ignore the importance of education about physical examinations. In a U.S. survey, 4% of students said that they had never taken a history or conducted a physical examination under the observation of a faculty member, and 20% said that they had been directly

observed only two or fewer times (12). The same can be said of Japan (13). Although many simulation and web-based models have been used to improve physical examination skills and knowledge, the results were inconsistent (14, 15). We still have much clinical knowledge to learn from real patients through hands-on patient care (16). Several reports have also shown that patient encounters were associated with ITE scores, especially in PGY-3 (1). Our study

results showed that increased patient encounters might help improve the physical examination scores in particular. These results might help administrators to understand how to teach residents how to undertake physical examinations more systematically.

PGY-2 and age showed relationships with each category of test score, which is comparable with previous reports (17). First, the year-to-year increase in the ITE score could be explained by the increasing number of patient encounters. A previous study showed that an average of 263 patient encounters was associated with a 5% higher internal medicine ITE score. Although we did not have data on the number of direct patient encounters in this study, the same could be said of Japan. Second, the fact that a younger age was associated with higher test results has been discussed previously (18). However, Brateanu et al. reported that not age but each postgraduate year affected the board test results (19). Thus, age might affect the test results as a sole factor, but years of training might overcome this disadvantage.

Our study also revealed that an increased number of staff had a negative impact on the GM-ITE score. We speculated on the reasons for this result as follows. First, Japan has been very successful in training subspecialists but not generalists and primary care doctors (20). Under these circumstances, staff doctors could not have a positive impact on the ITE score. Second, even if the number of staff was high, the ratio of staff to patients treated might not be appropriate, which might result in a negative impact on the GM-ITE score. In Japan, there are many hospitals, but the number of doctors is insufficient. Third, the impact of staff numbers might also have a ceiling effect on the quality of education. We suspected that these factors might have affected the result of a lack of a positive relationship between staff numbers and ITE scores. The Japan Primary Care Association (JPCA), the largest academic association of generalists in Japan, was established on April 1, 2010. We hope that more highly trained primary care doctors will emerge in the future. A more positive impact of true primary care physicians on future scores should thus be anticipated.

Finally, we analyzed the hospital characteristics and compared them between university hospitals and community hospitals. We found that the university hospitals had a higher number of staff, a smaller number of emergency patients, more beds and a higher number of new outpatients. We also found that residents working at university hospitals had a smaller number of patient encounters than those working at community hospitals, which is consistent with previous reports (6). However, the relatively small number of university hospitals included in our data (only 14 university hospitals were analyzed) precluded us from generalizing our findings to other contexts.

Our study is associated with several limitations. First, there may have been a population bias. The program directors at each hospital were responsible for the decision to participate in this study, which resulted in a bias towards

higher proportions of participation among teaching hospitals with program directors with more motivation and subsequently better programs. Only 10% of the residents in Japan participated in this study, and almost all (86%) worked at community hospitals. Among this population, the test information function analysis revealed the effectiveness of the GM-ITE in the lower-score group. Therefore, our findings might not be generalizable to all residents throughout Japan. However, the chief aim of our study was to examine the relationship between the hospital volume and the score, which was successfully achieved. Previous reports have shown that residents who work at non-university hospitals have more experience and are more confident about their clinical skills than those who work at university hospitals. Therefore, more motivated young physicians could have better effects after being exposed to a larger number of patients. Second, the hospital volume might be a potential cause of the selection of residents with increased knowledge. Emergency medicine is the most appealing department for students; thus, the high number of patient encounters might lead to more highly educated and motivated residents. High-scoring residents were concentrated at higher-volume centers, which might result in a positive cycle of higher volume and more talented young physicians. Third, our study did not include data on several variables that might affect the results, such as the following: patient encounter data evaluated by each participant, quality of life data during residency and staff doctors' workload. Future studies are thus needed to confirm the direct relationship between the hospital volume and the test scores after adjusting for several variables.

In summary, we found that the hospital volume, especially the number of emergency patients, affected the level of general medical knowledge.

The authors state that they have no Conflict of Interest (COI).

Acknowledgement

We thank Nobuo Kikuhara (JAMEP) for his excellent assistance.

References

1. McCoy CP, Stenerson MB, Halvorsen AJ, Homme JH, McDonald FS. Association of volume of patient encounters with residents' in-training examination performance. *J Gen Intern Med* **28**: 1035-1041, 2013.
2. Haney EM, Nicolaidis C, Hunter A, Chan BK, Cooney TG, Bowen JL. Relationship between resident workload and self-perceived learning on inpatient medicine wards: a longitudinal study. *BMC Med Educ* **6**: 35, 2006.
3. Kuwabara K, Matsuda S, Fushimi K, et al. Quantitative assessment of the advantages of laparoscopic gastrectomy and the impact of volume-related hospital characteristics on resource use and outcomes of gastrectomy patients in Japan. *Ann Surg* **253**: 64-70, 2011.
4. O'Brien SM, Delong ER, Peterson ED. Impact of case volume on hospital performance assessment. *Arch Intern Med* **168**: 1277-1284, 2008.
5. Ross JS, Normand SL, Wang Y, et al. Hospital volume and 30-day

- mortality for three common medical conditions. *N Engl J Med* **362**: 1110-1118, 2010.
6. Nomura K, Yano E, Mizushima S, et al. The shift of residents from university to non-university hospitals in Japan: a survey study. *J Gen Intern Med* **23**: 1105-1109, 2008.
 7. Perez JA Jr, Greer S. Correlation of United States Medical Licensing Examination and Internal Medicine In-Training Examination performance. *Adv Health Sci Educ Theory Pract* **14**: 753-758, 2009.
 8. Thissen D, Steinberg L, Mooney JA. Trace lines for Testlets: a use of multiple-categorical-response models. *J Educ Meas* **26**: 247-260, 1989.
 9. Allison JJ, Kiefe CI, Weissman NW, et al. Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI. *JAMA* **284**: 1256-1262, 2000.
 10. Opila DA. The impact of feedback to medical housestaff on chart documentation and quality of care in the outpatient setting. *J Gen Intern Med* **12**: 352-356, 1997.
 11. McGrath PD, Wennberg DE, Dickens JD Jr, et al. Relation between operator and hospital volume and outcomes following percutaneous coronary interventions in the era of the coronary stent. *JAMA* **284**: 3139-3144, 2000.
 12. Goldstein EA, Maclaren CF, Smith S, et al. Promoting fundamental clinical skills: a competency-based college approach at the University of Washington. *Acad Med* **80**: 423-433, 2005.
 13. Murai M, Kitamura K, Fetters MD. Lessons learned in developing family medicine residency training programs in Japan. *BMC Med Educ* **5**: 33, 2005.
 14. Grundman JA, Wigton RS, Nickol D. A controlled trial of an interactive, web-based virtual reality program for teaching physical diagnosis skills to medical students. *Acad Med* **75**: S47-S49, 2000.
 15. Lee CA, Chang A, Chou CL, Boscardin C, Hauer KE. Standardized patient-narrated web-based learning modules improve students' communication skills on a high-stakes clinical skills examination. *J Gen Intern Med* **26**: 1374-1377, 2011.
 16. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. *Med Educ* **43**: 303-311, 2009.
 17. McDonald FS, Zeger SL, Kolars JC. Factors associated with medical knowledge acquisition during internal medicine residency. *J Gen Intern Med* **22**: 962-968, 2007.
 18. Meskauskas JA, Webster GD. The American Board of Internal Medicine recertification examination: process and results. *Ann Intern Med* **82**: 577-581, 1975.
 19. Brateanu A, Yu C, Kattan MW, Olender J, Nielsen C. A nomogram to predict the probability of passing the American Board of Internal Medicine examination. *Med Educ Online* **17**: 18810, 2012.
 20. Teo A. The current state of medical education in Japan: a system under reform. *Med Educ* **41**: 302-308, 2007.