

Awareness of Diagnostic Error among Japanese Residents: a Nationwide Study

Yuji Nishizaki, MD, MPH, PhD¹, Tomohiro Shinozaki, MPH, PhD², Kensuke Kinoshita, MD³, Taro Shimizu, MD, MPH, MBA⁴, and Yasuharu Tokuda, MD, MPH⁵

¹Medical Technology Innovation Center, Juntendo University, Tokyo, Japan; ²Department of Biostatistics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; ³Department of Medicine, Mito Kyodo General Hospital, University of Tsukuba, Tsukuba, Japan; ⁴Diagnostic and Generalist Medicine, Dokkyo Medical University, Mibu, Japan; ⁵Muribushi Okinawa for Teaching Hospitals, Urasoe City, Okinawa, Japan.

BACKGROUND: Residents' understanding of diagnostic error may differ between countries. We sought to explore the relationship between diagnostic error knowledge and self-study, clinical knowledge, and experience.

METHODS: Our nationwide study involved postgraduate year 1 and 2 (PGY-1 and -2) Japanese residents. The Diagnostic Error Knowledge Assessment Test (D-KAT) and General Medicine In-Training Examination (GM-ITE) were administered at the end of the 2014 academic year. D-KAT scores were compared with the benchmark scores of US residents. Associations between D-KAT score and gender, PGY, emergency department (ED) rotations per month, mean number of inpatients handled at any given time, and mean daily minutes of self-study were also analyzed, both with and without adjusting for GM-ITE scores. Student's *t* test was used for comparisons with linear mixed models and structural equation models (SEM) to explore associations with D-KAT or GM-ITE scores.

RESULTS: The mean D-KAT score among Japanese PGY-2 residents was significantly lower than that of their US PGY-2 counterparts (6.2 vs. 8.3, $p < 0.001$). GM-ITE scores correlated with ED rotations (≥ 6 rotations: 2.14; 0.16–4.13; $p = 0.03$), inpatient caseloads (5–9 patients: 1.79; 0.82–2.76; $p < 0.001$), and average daily minutes of self-study (≥ 91 min: 2.05; 0.56–3.53; $p = 0.01$). SEM revealed that D-KAT scores were directly associated with GM-ITE scores ($\beta = 0.37$, 95% CI: 0.34–0.41) and indirectly associated with ED rotations ($\beta = 0.06$, 95% CI: 0.02–0.10), inpatient caseload ($\beta = 0.04$, 95% CI: 0.003–0.08), and average daily minutes of study ($\beta = 0.13$, 95% CI: 0.09–0.17).

CONCLUSIONS: Knowledge regarding diagnostic error among Japanese residents was poor compared with that among US residents. D-KAT scores correlated strongly with GM-ITE scores, and the latter scores were positively associated with a greater number of ED rotations, larger caseload (though only up to 15 patients), and more time spent studying.

KEY WORDS: Diagnostic Error Knowledge Assessment Test (D-KAT); General Medicine In-Training Examination (GM-ITE); postgraduate education.

J Gen Intern Med 33(4):445–8
DOI: 10.1007/s11606-017-4248-y
© Society of General Internal Medicine 2017

INTRODUCTION

In 2004, the Japanese Ministry of Health, Labor and Welfare established a new postgraduate clinical training program, mandating that residents complete a 2-year internship after graduating from medical school and passing the National Examination for Medical Practitioners. In those 2 years, residents are expected to achieve a basic set of skills for primary care practice, obtained through rotations in internal medicine, surgery, and emergency medicine during the first year, and pediatrics, obstetrics and gynecology, psychiatry, and community medicine during the second year. In 2010, the required departments for rotation were reduced from seven to three: internal medicine, emergency department (ED), and community medicine. Residents are allowed to rotate on surgery, anesthesiology, pediatrics, OB-GYN, or psychiatry as electives. After this 2-year residency, they choose their subspecialty program.

Knowledge of diagnostic errors such as cognitive bias or heuristics is an important factor in preventing such errors. No studies to date have explored experiential variables, such as workload, ER rotations, or minutes of study per day, that might improve D-KAT scores. In a previous study, we found that GM-ITE scores were improved with optimal resident caseload, ED rotations, and access to online resources.¹ The purpose of the current study was to assess diagnostic error knowledge among residents throughout Japan and to compare the results with the benchmark scores in a previous US study.² We also evaluate the relationships between diagnostic error knowledge and self-study, clinical knowledge, and experience.

METHODS

We conducted a nationwide study of Japanese PGY-1 and -2 residents. The Diagnostic Error Knowledge Assessment Test (D-KAT) and General Medicine In-Training Examination (GM-ITE) were administered at the end of the 2014 academic year. Immediately after the D-KAT and GM-ITE were conducted, the participants completed a questionnaire regarding

Received July 21, 2017

Revised October 24, 2017

Accepted November 27, 2017

Published online December 18, 2017

their clinical experience and self-study habits, assessing ED rotations per month (the average frequency of night duty per month), the mean number of inpatients handled at one time, and mean daily minutes of self-study.

Diagnostic error knowledge was assessed using the D-KAT, a 13-item multiple-choice test developed and validated by Reilly et al. in 2010.² The GM-ITE is a clinical knowledge test that uses a methodology similar to that of the US Internal Medicine In-Training Examination (IM-ITE).³ The purpose of the IM-ITE is to provide residents and program directors with an objective, reliable, and valid assessment of clinical knowledge through a multiple-choice examination.^{4,5} Our GM-ITE, which was designed and written by a committee of experienced physicians assembled by the Japan Institute for Advancement of Medical Education Program (a non-profit organization in Tokyo, Japan), included 87 questions testing a wide range of clinical knowledge, from clinical skills and practical medical knowledge to psychosocial care of patients.⁶

The mean D-KAT scores were compared between Japanese and US residents using Student's *t* test.² In addition, the associations between D-KAT scores and gender, PGY, ED rotations per month, mean number of inpatients handled at any given time, and mean daily minutes of self-study were analyzed with or without adjusting for GM-ITE scores (calculated by the domains other than the D-KAT score). Linear mixed models were used to account for hospital variability as random intercepts, with Huber-White (sandwich) variance estimators. Answers of "unknown" (for ED duty and inpatient caseload) were retained in the analyses, whereas non-responders for either ED duty, inpatient caseload, or self-study ($N = 224$) were excluded from mixed

modeling. We also conducted the same analyses for the D-KAT score as a dependent variable, with and without adjusting for the GM-ITE score (except for the D-KAT score). A two-tailed *p* value of <0.05 was considered statistically significant. Analyses were performed using SAS version 9.4 software (SAS Institute Inc., Cary, NC). Structural equation models were built to test for the direct and indirect determinants of D-KAT performance using standardized estimates of effect and standard goodness-of-fit measures (STATA 14; StataCorp LP, College Station, TX).⁷ The study was approved by the Ethics Committee of the Tsukuba University Mito Medical Center.

RESULTS

A total of 2652 residents (1123 PGY-1 and 1529 PGY-2) from 258 teaching hospitals participated in the examination; 777 (29.3%) of participants were women. The voluntary survey response rate was 91.6% (2429/2652). Among respondents, the weighted average number of ED rotations was 3.94, with 3% having no rotations, 8% with 1–2, 70% with 3–5, and 17% with 6 or more rotations. By self-report, residents managed an average of 6.9 inpatients (weighted average value): 9% reported managing 4 or fewer, 64% reported 5–9, 19% noted 10–14, and 6% reported 15 or more. Residents reported spending an average of 42.2 min studying (weighted average value), with 39% spending less than 30 min, 38% between 31 and 60 min, 13% 61–90 min, and 4% 91 min or more of daily self-study.

The mean (\pm SD) D-KAT score among Japanese PGY-2 residents (6.2 ± 1.6) was significantly lower than that of their

Table 1 Linear Mixed Model Estimates for D-KAT Scores among Responding Doctors ($N = 2428$)

Category	Including GM-ITE score				Excluding GM-ITE score				
	Estimate	95% LL	95% UL	<i>P</i> value	Estimate	95% LL	95% UL	<i>P</i> value	
Fixed effects									
Female (vs. male)	−0.05	−0.20	0.09	0.48	−0.06	−0.21	0.09	0.43	
2nd PGY (vs. 1st)	0.12	0.00	0.24	0.05*	0.18	0.06	0.31	0.00*	
GM-ITE score (except D-KAT)	0.04	0.03	0.05	<.0001*					
ED rotations (per month)	0	Reference			Reference				
	1–2	−0.25	−0.60	0.11	0.17	−0.25	−0.62	0.13	0.19
	3–5	−0.18	−0.45	0.08	0.18	−0.10	−0.37	0.18	0.50
	≥6	−0.26	−0.58	0.06	0.12	−0.17	−0.51	0.17	0.32
	Unknown	−0.75	−1.32	−0.18	0.01*	−0.73	−1.34	−0.13	0.02*
Inpatient caseload (no. patients)	0–4	Reference			Reference				
	5–9	0.21	−0.03	0.44	0.09	0.29	0.05	0.52	0.02*
	10–14	0.04	−0.22	0.30	0.76	0.11	−0.15	0.37	0.39
	≥15	0.15	−0.19	0.49	0.39	0.21	−0.14	0.56	0.23
	Unknown	0.35	−0.27	0.97	0.27	0.36	−0.26	0.98	0.26
Self-study (min)	0–30	Reference			Reference				
	31–60	0.03	−0.11	0.17	0.71	0.06	−0.08	0.20	0.41
	61–90	−0.10	−0.30	0.10	0.33	−0.03	−0.23	0.18	0.80
	≥91	−0.15	−0.48	0.18	0.37	−0.07	−0.42	0.27	0.67
	None	−0.05	−0.31	0.22	0.72	−0.07	−0.35	0.21	0.63
Random components									
Var (random intercepts)	0.04	0.00	0.09	0.06	0.06	0.00	0.11	0.03*	
Var (residual)	2.52	2.37	2.66	<.0001*	2.57	2.42	2.72	<.0001*	

*Indicates statistical significance at $p < 0.05$. D-KAT = Diagnostic Error Knowledge Assessment Test, GM-ITE = General Medicine In-Training Examination, 95% LL = 95% lower limit, 95% UL = 95% upper limit, PGY = postgraduate year, ED = emergency department. Linear mixed models were used to account for hospital variability as random intercepts, with Huber-White (sandwich) variance estimators. Answers of "unknown" (for ED rotations and inpatients handled) were retained during analyses, while non-responders for either ED rotations, inpatients handled, or self-study ($N = 224$) were excluded from mixed modeling. All analyses were conducted using SAS version 9.4 (Cary, NC)

Table 2 Linear Mixed Model Estimates for GM-ITE Score (except for D-KAT) among Responding Doctors (N = 2428)

	Category	Including D-KAT score				Excluding D-KAT score			
		Estimate	95% LL	95% UL	P value	Estimate	95% LL	95% UL	P value
Fixed effects									
	Female (vs. male)	-0.10	-0.64	0.44	0.72	-0.13	-0.68	0.42	0.65
	2nd PGY (vs. 1st)	1.43	0.75	2.11	<.0001*	1.55	0.87	2.23	<.0001*
	D-KAT	0.64	0.45	0.83	<.0001*				
	ED rotations (per month)	0	Reference			Reference			
	1-2	0.01	-1.86	1.87	0.99	-0.16	-2.09	1.78	0.87
	3-5	1.94	-0.03	3.90	0.05	1.87	-0.17	3.92	0.07
	≥6	2.14	0.16	4.13	0.03*	2.02	-0.06	4.10	0.06
	Unknown	1.14	-0.89	3.18	0.27	0.72	-1.48	2.92	0.52
	Inpatient caseload (no. patients)	0-4	Reference			Reference			
	5-9	1.79	0.82	2.76	0.00*	2.00	1.07	2.94	<.0001*
	10-14	1.68	0.52	2.83	0.00*	1.77	0.65	2.89	0.00*
	≥15	1.12	-0.65	2.89	0.22	1.26	-0.56	3.07	0.17
	Unknown	0.36	-1.68	2.41	0.73	0.56	-1.43	2.55	0.58
	Self-study	0-30 min	Reference			Reference			
	31-60 min	0.80	0.17	1.42	0.01*	0.83	0.20	1.47	0.01*
	61-90 min	1.85	0.72	2.99	0.00*	1.82	0.68	2.96	0.00*
	≥91	2.05	0.56	3.53	0.01	1.97	0.45	3.50	0.01*
	None	-0.42	-1.61	0.77	0.49	-0.46	-1.68	0.77	0.46
Random components									
	Var (random intercepts)	7.07	4.73	9.41	<.0001*	7.48	5.02	9.94	<.0001*
	Var (residual)	42.96	40.44	45.48	<.0001*	43.88	41.31	46.45	<.0001*

*Same as shown in Table 1

US PGY-2 counterparts (8.3 ± 1.5 ; $p < 0.001$).² Similar to US resident scores, GM-ITE scores increased from PGY-1 to PGY-2 (6.1 to 6.2, respectively; $p = 0.018$).⁸ Using linear mixed models, the D-KAT scores were found to be associated with GM-ITE scores and inpatient caseload (Table 1). The GM-ITE scores were associated with D-KAT scores (estimate of change in score, 0.64; 95% confidence interval, 0.45–0.83; $p < 0.001$), ED rotations (≥ 6 rotations: 2.14; 0.16–4.13; $p = 0.03$), inpatient caseloads (5–9 patients: 1.79; 0.82–2.76; $p < 0.001$), and average daily minutes of self-study (≥ 91 min: 2.05; 0.56–3.53; $p = 0.01$; Table 2).

SEM models indicated a direct relationship between GM-ITE scores and D-KAT performance ($\beta = 0.37$, 95% CI: 0.34–0.41; Fig. 1), and suggested that D-KAT scores were indirectly

affected by the number of ED rotations ($\beta = 0.06$, 95% CI: 0.02–0.10), inpatient caseload ($\beta = 0.04$, 95% CI: 0.003–0.08), and average daily minutes of study ($\beta = 0.13$, 95% CI: 0.09–0.17). Our model was well fitted to our data (root mean square error of approximation [RMSEA]: 0.03; comparative fit index [CFI]: 0.98; Tucker-Lewis index [TLI]: 0.96).

DISCUSSION

It is unclear why Japanese residents scored more poorly on diagnostic error knowledge than US residents. We can infer, however, that graduate and postgraduate medical education on clinical reasoning has not reached a level similar to that in the

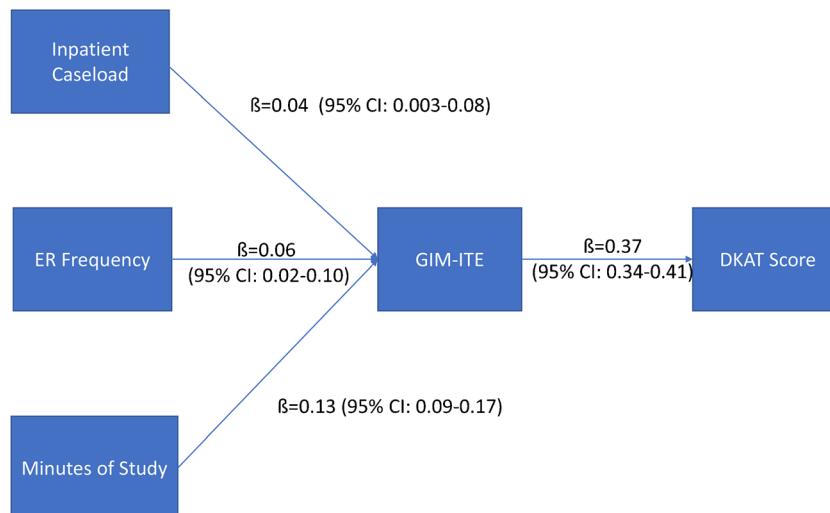


Figure 1 Structural equation modeling for D-KAT scores. D-KAT = Diagnostic Error Knowledge Assessment Test.

US. Medical education in clinical clerkships in Japan has long been provided during subspecialty rotations that occur in the PGY years following the first 2 years. This long-held educational tradition has brought fewer educational opportunities for clinical reasoning and even postgraduate clinical training because of the short training period in the department of general medicine (GM) or general internal medicine (GIM).⁶ Compared to subspecialty physicians, teachers of GIM or GM departments may be better able to provide their students and residents specific clinical problem-solving knowledge, including that for diagnostic error.

Our study suggests that self-study, time spent in ED rotations, and caseload are all directly associated with GM-ITE scores, and are indirectly associated with D-KAT scores through the relationship between GM-ITE and D-KAT scores. A greater number of ED rotations and minutes of self-study and a larger inpatient caseload appeared to contribute to general improvement in clinical knowledge and skill. An appropriate caseload in both emergency and inpatient settings may provide residents with the experience needed to reflect on various cognitive biases, leading to improved clinical and diagnostic reasoning. In addition, the development of independent curricula, such as simulation-based training, reflective practice, or active metacognitive review, may be useful.^{9–11} Our SEM models allowed us to explore the potential causal relationships between our variables, but given the nature of our cross-sectional data, our results can only be speculative. Our findings should be validated in a larger longitudinal study.

In conclusion, diagnostic error knowledge among Japanese residents was poor compared with that among US residents. GM-ITE performance was related to scores on D-KAT. Diagnostic error knowledge was indirectly related to increased clinical experience and self-study. Resident diagnostic error performance may improve with greater opportunities for reflection on inpatient and emergency patient conditions and additional independent curricula.^{10,11}

Acknowledgements: *The authors would like to thank Takeshi Asakura for help in translating this manuscript into English, and we*

thank all members of the GM-ITE production committee for their support. We would also like to thank Jeffrey Jackson, MD, MPH, for his editorial work and assistance with the SEM models.

Corresponding Author: Yasuharu Tokuda, MD, MPH; Muribushi Okinawa for Teaching Hospitals, Urasoe City, Okinawa, Japan (e-mail: yasuharu.tokuda@gmail.com).

Compliance with Ethical Standards:

Conflict of Interest: *The authors declare that they do not have a conflict of interest.*

REFERENCES

1. **Kinoshita K, Tsugawa Y, Shimizu T, et al.** Impact of inpatient caseload, emergency department duties, and online learning resource on General Medicine In-Training Examination scores in Japan. *Int J Gen Med.* 2015;8:355–60.
2. **Reilly JB, Ogdie AR, Von Feldt JM, et al.** Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ Qual Saf.* 2013;22:1044–50.
3. **McCoy CP, Stenerson MB, Halvorsen AJ, et al.** Association of volume of patient encounters with residents' in-training examination performance. *J Gen Intern Med.* 2013;28:1035–41.
4. **Kanna B, Gu Y, Akhuetie J, et al.** Predicting performance using background characteristics of international medical graduates in an inner-city university-affiliated Internal Medicine residency training program. *BMC Med Educ.* 2009;9:42.
5. **Perez JA, Greer S.** Correlation of United States Medical Licensing Examination and Internal Medicine In-Training Examination performance. *Adv Health Sci Educ Theory Pract.* 2009;14:753–8.
6. **Shimizu T, Tsugawa Y, Tanoue Y, et al.** The hospital educational environment and performance of residents in the General Medicine In-Training Examination: a multicenter study in Japan. *Int J Gen Med.* 2013;6:637–40.
7. **Kline RB.** Principles and practice of structural equation modeling. New York: Guilford Press; 1998.
8. **Kay C, Jackson JL, Frank M.** The relationship between Internal Medicine Residency Graduate performance on the ABIM certifying examination, yearly in-service examinations and attending evaluations. *Acad Med.* 2015;90(1):100–4.
9. **Schmidt HG, Mamede S.** How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Educ.* 2015;49:961–73.
10. **Graber ML, KISSAM S, Payne VL, et al.** Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf.* 2012;21:535–57.
11. **Singh H, Schiff GD, Graber ML, et al.** The global burden of diagnostic errors in primary care. *BMJ Qual Saf.* 2017;26:484–94.