

Performance of emergency medicine residents on a novel practice examination using visual stimuli

Danielle Blouin, MD, MHPE;* Luigi Eugene Dagnone, MD;† Robert McGraw, MD, MEd‡

ABSTRACT

Objectives: In 2004, the format of the Royal College of Physicians and Surgeons of Canada certification exam in Emergency Medicine was modified to include an audiovisual session. Residents' performance on such stimuli is unknown. This study aims to determine the performance of Emergency Medicine postgraduate year 5 (PGY-5) on a novel practice examination using visual stimuli, in an effort to guide residents' preparatory study, and to help educators focus the teaching related to visual stimuli.

Methods: In this prospective observational study, 30 PGY-5 residents from 12 Emergency Medicine programs across Canada participated in a practice examination in which they had to answer 133 questions related to 80 visual stimuli. The stimuli included electrocardiograms, medical imaging and clinical photographs. Three experienced faculty provided "reference standard" answers for the visual stimuli. For analysis purpose, stimuli were classified into 9 clinical domains, outlined in *The Model of the Clinical Practice of Emergency Medicine*.

Results: Insufficient question numbers in 3 domains prevented reliable inferences. Correct answers were given for 65% of remaining questions, with the following domain distribution: 69% respiratory, 64% trauma, 73% cardiovascular, 62% head-eyes-ears-nose-throat (HEENT), 63% musculoskeletal and 69% cutaneous. Seventy-four percent of ECGs, 64% of imaging, and 62% of photograph-related questions were answered correctly. Incorrect answers were related to a lack of knowledge as well as imprecise or incomplete responses.

Conclusions: PGY-5 subjects scored lowest in the HEENT and musculoskeletal domains. Medical educators should emphasize these domains in their teaching, and insist on precision and completeness of answers.

Key words: postgraduate education; visual stimulus; teaching

RÉSUMÉ

Objectifs : En 2004, la structure de l'examen de certification en médecine d'urgence du Collège Royal des médecins et chirurgiens du Canada a été modifiée pour inclure une partie audiovisuelle. La performance des résidents devant de tels stimuli est inconnue. La présente étude a pour but de déterminer la performance de résidents en médecine d'urgence dans leur 5^e année postdoctorale lors d'un nouvel examen de pratique utilisant des stimuli visuels, dans un effort de guider l'étude préparatoire des résidents et d'aider les éducateurs à préciser l'enseignement lié aux stimuli visuels.

Méthodes : Lors de cette étude d'observation prospective, 30 résidents dans leur 5^e année postdoctorale dans 12 programmes de médecine d'urgence partout au Canada ont participé à l'examen de pratique dans le cadre duquel ils ont répondu à 133 questions liées à 80 stimuli visuels.

*Assistant Professor, †Professor, and ‡Associate Professor, Department of Emergency Medicine, Queen's University, Kingston, Ont.

Received: Feb. 23, 2005; final submission: Nov. 22, 2005; accepted: Nov. 23, 2005

This article has been peer reviewed.

Can J Emerg Med 2006;8(1):21-6

Ces stimuli comprenaient des électrocardiogrammes, des clichés d'imagerie médicale et des photographies cliniques. Trois enseignants chevronnés ont fourni les réponses servant de «norme de référence» pour les stimuli visuels. Pour les fins d'analyse, les stimuli furent classés dans neuf domaines cliniques, décrits dans *le Modèle de pratique clinique de la médecine d'urgence*.

Résultats : Un nombre insuffisant de questions dans trois domaines a empêché d'établir des inférences fiables. Des réponses exactes furent données à 65 % des autres questions, les résultats dans les différents domaines étant les suivants : 69 % respiratoire, 64 % trauma, 74 % cardiovasculaire, 62 % tête-yeux-oreilles-nez-gorge, 63 % musculo-squelettique et 69 % cutané. Quant aux questions liées aux ECG, à l'imagerie et aux photographies, le pourcentage de réponses exactes était de 74 %, 64 % et 62 % respectivement. Les réponses inexactes étaient attribuables à un manque de connaissances ainsi qu'à la nature imprécise ou incomplète de ces réponses.

Conclusions : Les sujets en 5^e année postdoctorale ont obtenu les plus mauvais résultats dans le domaine tête-yeux-oreilles-nez-gorge et dans le domaine musculo-squelettique. Les enseignants en médecine devraient mettre l'accent sur ces domaines dans le cadre de leur enseignement et insister sur l'importance de donner des réponses précises et complètes.

Introduction

Photographs, radiographs and electrocardiograms have always been a component of the Royal College of Physicians and Surgeons of Canada (RCPSC) certification examinations in Emergency Medicine (EM). Until 2004, these visual stimuli were presented within the context of clinical cases that candidates had to manage on oral examinations. In 2004, the format of the exam was modified to include an audiovisual "circuit" with 12–14 stations, each providing an audio or visual stimulus linked to 1 or more short-answer questions. Visual stimuli include photographs, electrocardiograms, laboratory data, radiographs and computed tomography (CT) images. Candidates rotate through this circuit at timed intervals.

In 2003, a review course including visual stimulus sessions was designed and implemented to help final-year (i.e., PGY-5) EM residents from all Canadian residency programs prepare for their upcoming RCPSC certification exams. Because visual stimulus questions were a new component of the RCPSC examinations, no prior information was available describing the performance of EM residents on visual stimuli. The course was repeated in October 2004.

This study sought to determine the performance of final-year EM residents on visual stimulus stations during the 2004 course. Identification of weaker performance in certain clinical domains or on specific types of stimulus will guide residents in their preparatory study and will allow faculty to better focus their teaching around visual stimuli.

Methods

Design and subjects

This prospective observational study was conducted at

Queen's University, Kingston, Ont., Canada, in October 2004. All PGY-5 residents from all 12 RCPSC Emergency Medicine residency programs were invited to the course. Thirty of 40 eligible residents chose to register and they constitute the study population.

Study protocol

Registration cost each resident \$250. Transportation to Kingston was provided free of charge (thanks to an unrestricted grant from Roche Canada), but residents were responsible for their own accommodation. The course was held over a 3½-day period, and interactive didactic sessions were presented during the first 2 days. The last day-and-a-half focused on oral examination practices, followed by feedback.

Visual stimuli

Stimuli included medical imaging (radiographs, ultrasound, CT scans), electrocardiograms and clinical pictures that were representative of the practice of Emergency Medicine, with a level of difficulty targeted at graduating residents. Each stimulus was linked to 1 or more questions, and clinical scenarios were provided if necessary. The questions ranged from simple identification of an abnormality to management of the condition. Stimuli were based on actual cases encountered by Queen's University EM faculty, and were typical of diagnostic and therapeutic issues confronting emergency physicians. To ensure face validity and develop reference standard responses, the stimuli and related questions were piloted on 3 experienced EM faculty volunteers who were not involved with the design of the course. The reference standard answer to each stimulus question was that given by at least 2 of these physicians.

Study procedures

Five visual stimuli sessions were interspersed throughout the course, with a total of 80 stimuli. These stimuli were introduced in random order and were independent of the material discussed during didactic sessions. Individual participants rotated from one station to the next on a 3-minute bell ring, anonymously answering questions on index cards. At the ring of the bell, participants placed their answer card in an opaque envelope before moving to the next station. Individual results were not provided to residents, but a debriefing session was held by a faculty member after each session, where the expected answers to the session stimuli were presented and residents' comments and concerns were addressed. All residents provided written consent to have their responses used for study and publication purposes, and this study was approved by the Queen's University Institutional Review Board.

Data analysis

Stimuli were classified into the clinical domains described in *The Model of Clinical Practice of Emergency Medicine (EM Model)* by the American Board of Emergency Medicine (ABEM) Core Content Task Force II.¹ Answers to visual stimulus questions were compared to the reference standard responses provided by the EM faculty members. Our primary outcome measures were the overall percentage of correct answers and the percentage of correct answers in each clinical domain. Secondary outcome measures were the percentages of correct answers by type of visual stimulus. The percent agreement between the faculty members was also reported.

Results

The 30 participants in the course came from 12 of the 13 RCPSC residency programs. The 13th program did not have 5th-year graduates in 2004. No information is avail-

able on the 10 residents who elected not to participate except that some were French-speaking, and might not have felt confident attending an English-language course.

Two stimuli were excluded from the analysis because of poor technical quality, leaving 78 stimuli for analysis. One of the excluded stimuli accounted for 1 question in the respiratory domain; the second excluded stimulus accounted for 3 questions in the cutaneous domain. The remaining stimuli included 37 medical images, 15 electrocardiograms and 26 clinical pictures. The exclusions described above left 129 questions within 9 clinical domains.

Table 1 summarizes the number of questions in each domain and presents measures of internal reliability based on agreement between the 3 faculty experts. Table 2 shows that, overall, 65% of questions were answered correctly (standard deviation [SD] 6.9), ranging from 61.4% (SD 11.8) in the head-eyes-ears-nose-throat (HEENT) domain to 72.5% (SD 10.6) in cardiovascular.

Table 3 summarizes the distribution of correct answers by type of visual stimulus, showing that residents scored worst on questions related to clinical pictures and best on questions related to electrocardiograms.

Three domains (infectious diseases, neurology and renal-urology) lacked a sufficient number of questions to allow reliable inferences regarding residents' performance. Within these domains, the overall correct response rate was 54% (SD 11.7). Table 4 summarizes resident responses on questions related to the excluded items and domains.

Discussion

This is the first study describing the performance of PGY-5 Emergency Medicine residents on a visual stimulus examination. Our results indicate weaker performances in the clinical areas of musculoskeletal disorders, HEENT and trauma, and on medical imaging and clinical picture types of stimulus. These results should prompt greater teaching

Table 1. Metrics of the visual stimulus examination, by domain

Clinical domain	No. of questions	Cronbach alpha	Agreement of experts, no. (and %) of questions	
			2 experts	All 3 experts
Respiratory	5	0.08	3 (60)	2 (40)
Trauma	43	0.36	8 (18.6)	35 (81.4)
Cardiovascular	30	0.11	6 (20)	24 (80)
HEENT	22	0.06	1 (5)	21 (95)
Musculoskeletal	8	-0.18	8 (100)	0 (0)
Cutaneous disorders	10	0.25	5 (50)	5 (50)

HEENT = head-eyes-ears-nose-throat

emphasis in these clinical domains and the use of these types of stimulus.

In order to help residents prepare for their certification exam and to guide program directors and faculty in their teaching efforts, we will discuss performance results from 3 perspectives based on clinical domain, type of stimulus and analysis of incorrect answers.

Clinical domain

Of the tested domains, musculoskeletal, HEENT and trauma ranked lowest, and cardiovascular ranked highest. Pathologies represented in the HEENT domain were typical emergency medicine cases easily identified by our expert faculty (e.g., neck soft tissue abscesses, epiglottitis and various fundoscopic abnormalities). This highlights the need for additional emphasis on emergency HEENT conditions during residency training.

Performance on the trauma-related stimuli was relatively low given that trauma is one of the defining fields of emergency medicine. These stimuli consisted largely of ortho-

pedic trauma cases, requiring residents to identify radiographs of various bony fractures and describe appropriate management. The relatively poor performance seen could be, in part, explained by the fact that our review course was offered early in the 5th year of residency, at a time where residents might not yet have sufficient factual knowledge in these areas.

Cardiovascular emergencies yielded the highest correct score. A more structured approach to ECG interpretation and greater clinical exposure during training may account for these results.

Type of stimulus

In this study, residents were most successful on ECG-related questions and least successful on medical imaging and clinical picture related questions. Our residents scored only 64% on questions related to medical imaging stimuli — substantially lower than the 86% to 99% rates of “correct” image interpretation reported for emergency physicians (relative to radiologists) in the published literature.²⁻⁹ Sev-

Table 2. Correct answers, by clinical domain

Clinical domain	No. of questions	% of correct answers, mean (and SD)	Range, %
Respiratory	5	69.0 (21.4)	25–100
Trauma	43	63.5 (10.0)	39–88
Cardiovascular	30	72.5 (10.6)	48–93
HEENT	22	61.4 (11.8)	38–83
Musculoskeletal	8	64.3 (17.4)	25–100
Cutaneous disorders	10	69.2 (20.2)	29–100

SD = standard deviation; HEENT = head–eyes–ears–nose–throat

Table 3. Correct answers, by type of visual stimulus

Stimulus type	No. of stimuli	No. of questions	% correct, mean (and SD)	Range, %
Medical imaging	37	55	63.6 (9.6)	45–83
Clinical pictures	26	27	62.4 (8.1)	43–76
Electrocardiograms	15	47	73.7 (10.8)	48–92

HEENT = head–eyes–ears–nose–throat

Table 4. Correct answers to visual stimuli in domains excluded from analysis

Clinical domain	No. of questions in the domain	% correct, mean (and SD)	Range, %
Infectious disorders	5	56.5 (28.5)	0–100
Neurology	3	34.3 (37.2)	0–100
Renal-Urology	3	51.7 (35.3)	0–100

eral factors may explain this apparent discrepancy. First, the radiographs used in our study were not representative of everyday practice but were specifically selected because of their level of reading difficulty. Unlike previous studies (and real life), in which normal or near normal radiographs make up the bulk of readings, our participants were dealing specifically with abnormal images. Our data are compatible with previous studies showing that level of training is directly related to accuracy of interpretation, with junior residents being less accurate than senior residents, and senior residents less accurate than faculty.^{2,4-6,9} Finally, the stress associated with the timed station session format might have negatively influenced resident performance. Previous work suggests that educational interventions directed at image interpretation are useful.⁹ In view of our findings, these teaching sessions should be developed for PGY-5 residents.

In this study, residents were often unable to identify pathologies depicted in clinical pictures, at least in the 4 clinical domains studied (trauma, HEENT, musculoskeletal and cutaneous disorders). Unfortunately, we could not locate any prior study addressing this.

Study subjects were most successful (74%) on questions related to ECG stimuli, which included evidence of cardiac ischemia, arrhythmias and toxic ingestions. This compares favorably with rates reported in the literature,¹⁰⁻¹³ although it is difficult, for several reasons, to compare ECG interpretation rates across studies. Some investigators have exposed study subjects to consecutive ECGs obtained in the emergency department,¹⁰⁻¹⁴ and others have used ECGs from patients with chest pain¹⁵⁻¹⁸ or specific dysrhythmias.^{19,20} Differences between studies in what constitutes ECG misinterpretation further complicate attempts at comparison. Most studies define misinterpretation as a disagreement between emergency physician and cardiologist readings, and consider minor findings such as sinus arrhythmias, atrial hypertrophy, indeterminate axis, occasional premature atrial complexes or non-specific ST-T changes as discrepancies. The rate of potentially clinically important misinterpretation is much lower, from 0.8% to 8.6%,¹⁰⁻¹⁴ and misreads leading to a change in patient management are particularly uncommon; however, in 1 study focusing on 19 cases of "missed" acute cardiac ischemia, cardiologists disagreed with the emergency physician reading in 2 cases (11%).¹⁵ The 24% ECG error rate seen in our study is relatively high but may reflect the fact that the study ECGs were all challenging.

Analysis of incorrect answers

We classified incorrect answers into 3 categories: errors resulting from a lack of knowledge, a lack of precision and a

lack of completeness. Examples in the "lack of knowledge" category included misidentification of the carpal bones, confusion in the Salter–Harris classification, misnaming of rashes (e.g., erythema marginatum instead of migrans), wrong association of diagnoses (e.g., writing both Wenckebach and Mobitz II for the same stimulus) and wrong treatments (e.g., lidocaine for supraventricular arrhythmias). To detect important gaps, teaching faculty should not simply assume a solid knowledge base; rather, they should continuously challenge residents on their basic knowledge.

A "lack of precision" error might be stating only "pneumonia" as a diagnosis. Specifying upper- versus lower-lobe pneumonia demonstrates to the examiners an awareness of the potential different etiologic considerations. Similarly, residents were expected to specify the type of shoulder dislocation as posterior or anterior, and in the latter case, which subtype — not merely diagnose "shoulder dislocation." Their answers should convey a clear appreciation of the prognostic and therapeutic factors associated with each diagnosis.

Another example of a "specificity error" related to management of a traumatic injury was "referral to ortho." Although referral might be appropriate, examiners need a sense of how urgently referral will be made (e.g., while the patient is still in the emergency department or in 2 weeks), and to what purpose (reduction and fixation, or merely follow-up). When evaluating resident performance, teaching faculty should insist on clear and precise diagnoses and management plans, and should have learners explicitly state their concerns for each case.

"Lack of completeness" refers to failing to identify additional injuries once 1 has been found, leading to non-recognition of a Galeazzi fracture or missing a pneumothorax in the presence of a hemothorax. Certification examinations are unlikely to present an uncomplicated isolated radial fracture. Residents need to increase their exposure to more complex, but still realistic and typical, cases. Developing a bank of these cases and using them frequently in practice exams would accustom residents to always look for additional diagnoses.

Limitations

Our results were based on a small number of participants; however, they do represent 75% of all PGY-5 residents enrolled in RCPSC EM residency programs. Participants came from 12 of the 13 Canadian programs, allowing safe generalization of our results to all PGY-5 Emergency Medicine residents.

The low Cronbach alpha coefficients indicate that questions within a particular domain are successful at assessing

different facets of that domain. Given the wide range of clinical conditions that compose each domain, an impractical number of questions and visual stimuli would be necessary to saturate each domain.

Residents' performance on our examination might not predict their performance on the RCPSC examination. The format and content of the RCPSC examination are confidential and thus could not inform our design. Our goal was to expose residents to visual stimuli felt to be representative of the clinical conditions seen in Emergency Medicine and targeted at graduating residents, with the hope that the RCPSC examination shares this purpose.

The transferability of residents' performance on our examination to the clinical setting is unknown. For example, residents scored lower on clinical pictures and medical imaging. It is possible that the residents had difficulty with the technical quality of clinical pictures and photographs used, and that they might have performed better with real-life images and actual radiographs or CT scans. However, our expert faculty had no difficulty in answering the related questions and eliminated only 2 stimuli because of concerns about poor quality. This is not unique to our examination, and the same comment could be made of the RCPSC certification exam. The determination of an exam's predictive validity for clinical competence implies the existence of agreed upon measures of clinical competence. These measures are unfortunately missing.

Several domains did not have a sufficient number of questions to allow inferences on residents' performance in these domains. A further edition of the visual stimulus sessions should take this into consideration in its design.

Conclusions

The performance of Emergency Medicine residents in their last year of training on visual stimuli varies by clinical domains and by type of visual stimuli presented. In our examination, performance was weaker in HEENT and in trauma, and on pictures and medical imaging interpretation. Residents' mistakes result from a lack of knowledge and from a lack of precision and completeness in their answers.

Competing interests: None declared.

References

- Hockberger RS, Binder LS, Graber MA, et al. The model of the clinical practice of emergency medicine. *Ann Emerg Med* 2001; 37(6):745-70.
- Brunswick JE, Ilkhanipour K, Seaberg DC, et al. Radiographic interpretation in the emergency department. *Am J Emerg Med* 1996;14(4):346-8.
- Espinosa JA, Nolan TW. Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study. *BMJ* 2000;320:737-40.
- Warren JS, Lara K, Connor PD, et al. Correlation of emergency department radiographs: results of a quality assurance review in an urban community hospital setting. *J Am Board Fam Pract* 1993;6(3):255-9.
- Mayhue FE, Rust DD, Aldag JC, et al. Accuracy of interpretations of emergency department radiographs: effect of confidence levels [review]. *Ann Emerg Med* 1989;18(8):826-30.
- Eng J, Mysko WK, Weller GE, et al. Interpretation of emergency department radiographs: a comparison of emergency medicine physicians with radiologists, residents with faculty, and film with digital display. *AJR Am J Roentgenol* 2000; 175(5):1233-8.
- Walsh-Kelly CM, Melzer-Lange MD, Hennes HM, et al. Clinical impact of radiograph misinterpretation in a pediatric ED and the effect of physician training level. *Am J Emerg Med* 1995;13 (3):262-4.
- Lufkin KC, Smith SW, Matticks CA, et al. Radiologists' review of radiographs interpreted confidently by emergency physicians infrequently leads to changes in patient management [published erratum appears in *Ann Emerg Med* 1998;32(3 pt 1):390; PMID: 9758569]. *Ann Emerg Med* 1998; 31(2):202-7.
- Ilkhanipour K, Seaberg DC. Radiographic interpretation by emergency medicine residents. *Acad Emerg Med* 1995;2(12): 1113-4.
- Zappa M, Smith M, Li S. How well do emergency physicians interpret ECGs? [abstract]. *Ann Emerg Med* 1991;20(4):463.
- Kuhn M, Morgan MT, Hoffman JR. Quality assurance in the emergency department: evaluation of the ECG review process. *Ann Emerg Med* 1992;21(1):10-5.
- Westdrop EJ, Gratton MC, Watson WA. Emergency department interpretation of electrocardiograms. *Ann Emerg Med* 1992;21 (5):541-4.
- Todd KH, Hoffman JR, Morgan MT. Effect of cardiologist ECG review on emergency department practice. *Ann Emerg Med* 1996;27(1):16-21.
- Snoey ER, Housset B, Guyon P, et al. Analysis of emergency department interpretation of electrocardiograms. *J Accid Emerg Med* 1994;11(3):149-53.
- Pope JH, Aufderheide TP, Ruthazer R, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. *N Engl J Med* 2000;342(16):1163-70.
- Brady WJ, Perron A, Ullman E. Errors in emergency physician interpretation of ST-segment elevation in emergency department chest pain patients. *Acad Emerg Med* 2000; 7(11):1256-1260.
- McCallion WA, Templeton PA, McKinney LA, et al. Missed myocardial ischaemia in the accident & emergency department: E.C.G. a need for audit? *Arch Emerg Med* 1991;8(2):102-7.
- Brady WJ, Perron AD, Chan T. Electrocardiographic ST-segment elevation: correct identification of acute myocardial infarction (AMI) and non-AMI syndromes by emergency physicians. *Acad Emerg Med* 2001;8(4):349-60.
- Herbert ME, Votey SR, Morgan MT, et al. Failure to agree on the electrocardiographic diagnosis of ventricular tachycardia. *Ann Emerg Med* 1996;27(1):35-8.
- Isenhour JL, Craig S, Gibbs M, et al. Wide-complex tachycardia: continued evaluation of diagnostic criteria. *Acad Emerg Med* 2000;7(7):769-73.

Correspondence to: Dr. Danielle Blouin, Emergency Department, Queen's University, 76 Stuart St., Kingston ON K7L 2V7