ADVANCING ASSESSMENT AND SELF-ASSESSMENT

The State of Evaluation in Internal Medicine Residency

Saima I. Chaudhry, MD, MSHS^{1,4}, Eric Holmboe, MD², and Brent W. Beasley, MD³

¹Department of Medicine and Office of Graduate Medical Education, North Shore University Hospital, Manhasset, NY, USA; ²American Board of Internal Medicine, Philadelphia, PA, USA; ³University of Missouri-Kansas City, Kansas City, MO, USA; ⁴Office of Graduate Medical Education, North Shore University Hospital, Manhasset, NY, USA.

BACKGROUND: There are no nationwide data on the methods residency programs are using to assess trainee competence. The Accreditation Council for Graduate Medical Education (ACGME) has recommended tools that programs can use to evaluate their trainees. It is unknown if programs are adhering to these recommendations.

OBJECTIVE: To describe evaluation methods used by our nation's internal medicine residency programs and assess adherence to ACGME methodological recommendations for evaluation.

DESIGN: Nationwide survey.

PARTICIPANTS: All internal medicine programs registered with the Association of Program Directors of Internal Medicine (APDIM).

MEASUREMENTS: Descriptive statistics of programs and tools used to evaluate competence; compliance with ACGME recommended evaluative methods.

RESULTS: The response rate was 70%. Programs were using an average of 4.2–6.0 tools to evaluate their trainees with heavy reliance on rating forms. Direct observation and practice and data-based tools were used much less frequently. Most programs were using at least 1 of the Accreditation Council for Graduate Medical Education (ACGME)'s "most desirable" methods of evaluation for all 6 measures of trainee competence. These programs had higher support staff to resident ratios than programs using less desirable evaluative methods.

CONCLUSIONS: Residency programs are using a large number and variety of tools for evaluating the competence of their trainees. Most are complying with ACGME recommended methods of evaluation especially if the support staff to resident ratio is high.

KEY WORDS: graduate medical education; residency; ACGME; competency.

J Gen Intern Med 23(7):1010–5

DOI: 10.1007/s11606-008-0578-0

© Society of General Internal Medicine 2008

INTRODUCTION

Effective trainee evaluation is a professional responsibility of medical educators. In 1999, the Accreditation Council for Graduate Medical Education (ACGME) endorsed the concept

of the core competencies and mandated programs to evaluate their trainees in 6 broad areas: Patient Care, Medical Knowledge, Professionalism, Communication, Practice-Based Learning, and Systems-Based Practice. In 2002, programs were expected to start implementing competency evaluation. By 2006, the competencies were to be fully integrated into the curriculum and evaluation of all trainees in the United States.¹

Although the Council has largely let programs decide how to best assess each competency, they have provided recommendations to help programs tackle the important task of evaluation.

Such recommendations are included in the ACGME Outcomes Project, and more specifically, in its Toolbox of Assessment Methods, which provides information about the psychometric qualities and feasibility of a variety of evaluative tools.²

Several investigators have studied the use of these tools, usually providing a description of a particular evaluation at a single institution, and other times providing outcomes assessment.^{3–7} These studies provide programs with practical insight regarding different evaluative methods. However, there has been no nationwide description of the state of evaluation in graduate medical education. Without such an analysis, we cannot fully understand how programs determine the competence of their graduates or whether they are adhering to ACGME methodological recommendations for evaluation.

The purpose of this study is to describe how competence is evaluated in Internal Medicine graduate medical education in the United States during the early stages of the evolution toward outcomes-based education. We describe the frequency with which various tools are used to assess trainees, along with the number of tools used to measure each competency. We also describe how well residency programs are complying with the ACGME's recommended evaluation methodologies. Finally, we compare the characteristics of programs that use the ACGME's recommended evaluation methods for assessing all 6 competencies versus fewer than 6 competencies.

METHODS

The Survey Task Force of the Association of Program Directors in Internal Medicine (APDIM) developed a 74-item questionnaire to obtain information about our nation's residency programs. Our main outcome included tools used to assess each of the 6 competencies. Our survey was comprised of 5 sections assessing: 1) characteristics of the program including program resources, 2) characteristics of faculty and staff, 3) characteristics of residents, 4) characteristics of the program director, and 5) tools used to assess each competency.

We e-mailed the questionnaire in March of 2005 to each member of APDIM (total=391 programs), representing nearly 100% of the training programs in Internal Medicine nationwide. We sent second and third request emails in May and July of the same year. The cover letter contained definitions used within the questionnaire. The letter did not mention or refer to the ACGME's recommendations for competency assessment. A program administrator or an associate program director could complete baseline characteristics. We asked the program director to review and sign-off on this section, and then personally complete the remaining questions. The survey was confidential with respondents tracked by numerical codes.

In September 2000, the ACGME published their Toolbox of Assessment Methods, which lists each of the 6 competencies (e.g., Patient Care) and their subsequent domains.² Each competency is comprised of a varying number of domains (e.g., "caring and respectful behavior" is a domain for Patient Care). The Toolbox contains a list of evaluation methods (e.g., Portfolio) used to evaluate each competency's domain. These methods are ranked as "most desirable", the "next best", or "potentially applicable" for each domain. For example, in the competency of Patient Care, the ACGME lists a 360-degree evaluation as the "most desirable" method to evaluate "working within a team", but lists Chart-stimulated Recall as the "most desirable" method to evaluate "informed decision-making," both important domains of Patient Care. The "most desirable" tools recommended to evaluate all domains of each of the 6 competencies are listed in Table 1.

In our survey, we asked Program Directors to indicate which of the following 12 tools they used to evaluate each of the 6 competencies: ABIM evaluation form, Local forms, In-training Exam, Mini-CEX, Standardized Patients/Objective Structured Clinical Exams (OSCEs), Video with Patient encounter, Computer Simulation, Peer Evaluation, Nurse Evaluation, Patient Satisfaction Survey, Portfolios, Chart-stimulated Recall. These are not an exact match of those on the ACGME website, as our Survey Task Force generated a list that was reflective of Internal Medicine residencies' practice in 2005. However, 9 of the 12 tools from our survey are listed in the Toolbox of Assessment Methods. Program Directors were instructed to choose all tools they used to evaluate each competency. We did not ask them to specify what domains, if any, they were evaluating with each tool.

We also asked Program Directors to indicate whether they used "other" methods of evaluation. For those programs that chose "other" as a tool to evaluate any of the competencies, we

analyzed written responses describing the alternative tool and incorporated them into existing categories as appropriate. If no category could be assigned, a new one was created. For example, "in-house x-ray and ECG exam", "monthly quizzes", and "tests on rotations" were all collapsed into the category of "Written Exams." Two members of the Survey Task Force independently assigned categories to these hand written answers and adjudicated discrepancies as needed until consensus was attained.

DATA ANALYSIS

We entered each survey into a Microsoft Access database and double checked answers for errors. We used STATA for Windows 8.2 (Copyright © STATA Inc. 1984–2003) for all statistical analyses. We combined response categories for variables when we identified sparsely selected responses. For example, when few respondents chose a particular response category for an ordinal variable (e.g., salary <90K, 90–130K, 131–150K, etc), we lumped categories to facilitate presentation of the data (e.g., 90–150K). We examined continuous variables for evidence of skewness, outliers, and nonnormality, and described them using distributions, means, medians, standard deviations, and ranges.

We listed tools in four broad categories using a framework presented in a previous study.⁸ Foundational Tools included end-of-block evaluation forms, locally generated forms, and the In-Training Exam. Direct observation included the Mini-CEX, standard patients/OSCEs, videos with patient encounters, and computer simulations. Non-Faculty Perspectives included peer and nurse evaluations and patient satisfaction surveys. Practice and Data-based Tools included the portfolio and chart-stimulated recall.

We calculated the frequency with which a specific tool was used to measure each competency and the average number of tools used to assess each competency. We calculated the proportion of programs using a "most desirable" method to evaluate each competency as defined in the ACGME Toolbox of Assessment Methods. We also calculated the proportion of programs using all of the recommended "most desirable" methods for each competency, thereby "comprehensively" evaluating this competency.

Programs were dichotomized into those using a "most desirable" method of evaluation for all 6 competencies and those using this method for fewer than 6 competencies. We

Table 1. "Most Desirable Tools" Recommended by ACGME to Evaluate Each Competency*

Tool	Patient Care	Medical Knowledge	Professionalism	Communication	Practice-based Learning	Systems-based Practice
Record Review	X				X	
Chart-stimulated Recall	X	X			X	
Standard Patient	X			X		
Objective-structured Clinical Exam	X		X	X		
Simulations	X					
360 Evaluation	X		X		X	X
Patient Survey	X		X	X		X
Oral Exam		X			X	
Multiple-Choice Exam		X			X	X
Portfolio					X	
Checklist						X

*There are a variable number of domains in each competency. For a complete list of all domains of each competency, see ACGME website, <http://www.acgme.org/outcome/assess/toolbox.asp>

dichotomized our data into 6 versus fewer than 6 to ensure an equal number of programs in each group (53% versus 47%) and also because we thought programs that evaluated all 6 competencies as per the ACGME's recommended "most desirable" method might be different than programs evaluating fewer than 6 competencies this way. We compared our dichotomized programs on their 1) program characteristics including type (e.g., community hospital, university hospital, or military hospital), size (e.g., number of approved ACGME residency positions, hospital beds), and accreditation cycle length, 2) program resources (number of full-time equivalent faculty and full-time equivalent residency support staff), and 3) program director characteristics (experience, time spent on administrative activities for program). These demographic characteristics were obtained from sections 1, 2, and 4 of our survey as described in the methods.

For all analyses we used the Mann-Whitney *U* test. To avoid multiple comparison difficulties, we report only bivariate associations that are significant at the Bonferroni corrected $p < 0.001$ level.⁸

RESULTS

Out of the 391 programs surveyed nationwide, 5 were excluded because they were in the process of closing. Of the remaining 386 programs that were included in the analysis, 272 responded to our questionnaire, giving us a response rate of 70%.

The majority of programs were sponsored by a hospital (57%) in contrast to a medical school (30%). The primary teaching hospital for each residency was privately owned for 67%, university owned for 14%, and government owned for 16%. These numbers are representative of programs nationally.⁹ The average number of ACGME approved residency spots was 41 ± 36 (range 0–187) and the average number of teaching beds per hospital was 302 ± 247 (range 30–2100). The average ACGME accreditation cycle length was $3.8 \text{ years} \pm 1.2$ (range 0–5).

Program Directors had been at their post for an average of $6.8 \text{ years} \pm 5.9$ (range 0–32).

Table 2 reports the frequency of tools used to assess each of the 6 competencies. The average number of tools programs reported using to evaluate each competency was 6.0 ± 2.0 (range 0–11) for Patient Care, 5.1 ± 2.0 (range 0–11) for Medical Knowledge, 5.2 ± 1.9 (range 0–10) for Professionalism, 5.2 ± 1.9 (range 0–10) for Communication, 4.4 ± 2.1 (range 0–10) for Practice-based Learning, and 4.2 ± 2.1 (range 0–10) for Systems-based Practice. Over half are using "home grown" instruments such as local forms and nurse evaluations and almost one quarter are using "other" methods of competency assessment.

Table 3 reports the proportion of programs using the ACGME's "most desirable" method of competency evaluation. Most programs were using the "most desirable" tools for evaluating some domain of all the competencies. For example, 64% of programs were using a "most desirable" tool to measure some domain in Communication and 98% were using a "most desirable" tool to measure some domain of Patient Care. The proportion of programs using all of the recommended "most desirable" tools possible for each competency ranged from 1.5% for Patient Care to 14% for Communication.

Programs using 1 of the "most desirable" methods for all 6 competencies were similar to programs using the "most desirable" method for fewer than 6 competencies in most program and program director characteristics. However, the former programs had more full-time equivalent (FTE) support staff per resident. (See Table 4).

Only 8% of all responses concerning competency tools involved the use of "other" methods of assessment thereby obviating the need for re-analysis using data listing "other". Of these responses, 30% could not be collapsed into an existing tool with Systems-based Practice and Practice-based Learning comprising the majority (88%) of other non-collapsible methods of assessment. Some examples of tools that we could not collapse included "remediation committee", "mini M + M", "QI project",

Table 2. Proportion of Programs (N=272) Using Various Tools to Assess Trainee Competence in Internal Medicine Residency Programs*

Category	Tool	Patient Care (%)	Medical Knowledge (%)	Professionalism (%)	Communication (%)	Practice-based Learning (%)	Systems-based Practice (%)
Foundational	ABIM evaluation form	81	81	80	78	77	77
	Local Forms	52	52	49	50	55	51
	In-Training Exam	70	91	28	24	33	27
Direct Observation	Mini-CEX	90	76	73	82	53	48
	Standard Pt (OSCE)	26	19	23	28	13	13
	Video with Patient encounter	17	9	12	16	5	4
Non-faculty Perspective	Computer Simulation	6	6	2	2	3	2
	Peer Evaluation	82	65	81	79	60	58
	Nurse Evaluation	71	38	76	73	40	53
Practice and Data-based	Patient satisfaction survey	46	19	49	49	20	24
	Portfolio	35	25	24	21	34	27
	Chart-stimulated recall	16	12	5	6	14	13
Other	Other	10	18	16	16	34	33

Bold items are considered the "most desirable" method of evaluation by ACGME for each particular competency.

*ABIM American Board of Internal Medicine, Mini-CEX mini clinical evaluation exercise, OSCE objective-structured clinical exam

Table 3. Programs Use of ACGME “Most Desirable” Method of Competency Evaluation for Each of the 6 Competencies (N=272)*

	Patient Care (%)	Medical Knowledge (%)	Professionalism (%)	Communication (%)	Practice-based Learning (%)	Systems-based Practice (%)
Number of ACGME recommended “most desirable” methods	7	3	3	3	6	4
Programs using a “most desirable” method for measuring this competency (%)	98	94	92	64	89	84
Programs using all recommended “most desirable” methods for measuring this competency (%)	1.5	4.4	10	14	1.8	2.9

“communications course”, “morning report”, “evidence-based medicine rounds”, and “scholarship”.

CONCLUSIONS

This study describes the tools residency programs are using to evaluate trainee competence. We find that programs are using a large number and variety of tools to evaluate the competence of their trainees, averaging 4.2–6.0 tools per competency. This finding is consistent with ACGME guidelines, which encourage programs to use a comprehensive system of evaluation employing more than 1 tool, as no individual competency can be thoroughly judged utilizing just 1 instrument.¹

Foundational Tools are the most popular instruments for resident competency assessment and include rating-based forms such as the ABIM end of rotation form (used by a low of 77% of programs for Systems-based Practice and a high of 81% of programs for Patient Care) as well as “home grown” local forms (range of use 49–55%), and the In-Training Exam (range of use 24–91%).

It is worth noting that the ABIM form was the single most popular tool for all 6 competencies. However, global rating forms are only recommended by the ACGME to evaluate the competencies of Practice-based Learning and Patient Care. Furthermore, for both of these competencies, rating forms are only considered a “potentially applicable method.”¹⁰ Global evaluation forms provide a retrospective subjective assessment

Table 4. Bivariate Analysis of Programs Using ACGME “Most Desirable” Method of Competency Evaluation (N=272)*

	Programs using a “Most Desirable” Method to Evaluate All 6 Competencies	Programs using a “Most Desirable” Methods to Evaluate Fewer than 6 Competencies	P value
N (% programs)	145 (53%)	127 (47%)	
Program Characteristics			
% Type			
University	50	50	NS
Community	58	42	
Federal/military	40	60	
	Mean (95% confidence interval)		
Number ACGME approved spots	44.5 (38.0, 51.5)	37.9 (32.1, 43.7)	.10
Number Teaching Beds	302 (255, 348)	303 (260, 346)	.37
Cycle Length in years	3.8 (3.6, 4.0)	3.8 (3.6, 4.0)	.74
Program Resources	Mean (95% confidence interval)		
Number teaching faculty	88.2 (71.8, 104.0)	79.5 (65.0, 94.0)	.61
Number FTE support staff for program [†]	2.4 (2.2, 2.6)	2.5 (2.0, 2.9)	.21
Teaching faculty/ACGME approved spots	2.8 (2.3, 3.3)	2.0 (1.6, 2.5)	.02
FTE/ACGME spots	0.10 (0.09, 0.1)	0.07 (0.06, 0.09)	.001
Teaching Beds/ACGME approved spot	12.5 (9.6, 15.5)	9.35 (7.72, 11.0)	.22
Program Director Characteristics	Mean (95% confidence interval)		
Years as PD [‡]	6.7 (5.8, 7.7)	6.9 (5.9, 8.0)	.80
% time PD spends on program related activity	79.6 (74.7, 84.5)	73.7 (67.6, 79.8)	.26

*P values calculated by rank sum test

[†]FTE full-time equivalent support staff

[‡]PD program director

usually at the end of a clinical experience rather than an objective measure of specific skills and tasks. Ease of use may explain their popularity despite documented problems with discriminatory ability, reliability, and validity.^{11–14}

Given the current disconnect between actual practice and recommended practice in using global rating forms, we suggest either a “culture” change by programs in their use of rating forms or added help by the ACGME to improve the psychometric quality of these forms and train raters on their proper use.

Non-faculty perspectives such as peer and nurse evaluations were also commonly used. Although we did not specifically ask, we assume local forms, peer, and nurse evaluations are all scale-based rating forms that have not been psychometrically validated. In addition, there appears to be heavy reliance on “other” methods of competency assessment (used on average by 21% of programs).

More than 10 years ago, it was predicted that direct observation of trainees would prevail over rating scales as a means of evaluation in graduate medical education.¹¹ Our study does not support this notion. Direct observation of trainees occurred via the mini-CEX quite often (used by 90% of programs to assess Patient Care), but very infrequently for the Standard Patient/OSCE (range of use 13–28%), video of patient encounters (range of use 4–17%), or simulations (2–6%). In a similar way, Practice- and Data-based tools such as chart-stimulated recall (range of use 5%–16%) and portfolios (range of use 21%–34%) were infrequently used. It should be noted, however, that direct observation may be the basis for which rating-based forms are completed. We did not ask respondents to comment on the data they used to fill out their rating-based forms.

Certain tools may be used very infrequently because they are labor or resource intensive. We confirmed this suspicion by asking a subset of the APDIM Survey Task Force membership (program directors and associate program directors) to list tools in increasing order of difficulty. Although not a true validation, the task force rated videos, computer simulations, and standardized patients/OSCEs to be the most burdensome tools for evaluating residents. Not surprisingly, the results from our survey showed these tools to be very infrequently used by programs nationwide.

In terms of compliance with ACGME recommendations, the data reveal that half of all programs (53%) are using at least 1 of the “most desirable” methods to measure all 6 competencies. These data are encouraging as the ACGME only began assessing competency evaluation in 2002.

Not surprisingly, very few programs were able to employ all of the “most desirable” tools to evaluate each competency comprehensively (1.5 to 14%). Note that competencies with the fewest number of “most desirable” tools (e.g., 3 tools) were easier to comprehensively assess. Future studies should ascertain if using 1 of the “most desirable” tools encourages using other “most desirable” tools, perhaps creating a “change threshold” for a particular competency.

The ACGME does not expect programs to evaluate every domain of every competency. Instead, it allows programs to decide which domains are most important to their locality and which tools are most feasible to use.¹ Although programs are still a long way from comprehensively evaluating each competency, our results do show many programs to be using multiple tools for competency assessment.

Our bivariate analysis finds that programs using the “most desirable” methods of evaluation for all 6 competencies have

more full-time equivalent (FTE) support staff per resident than programs using the “most desirable” method for evaluating fewer than 6 competencies. Having more teaching faculty per resident approached significance.

It is widely recognized that programs and program staff have struggled to devote the added time and effort needed to effectively teach and evaluate the competencies since their required implementation began in 2002.¹⁵ Our data demonstrates that programs evaluating all 6 competencies with a “most desirable” method are utilizing 1 FTE per 10 residents and 2.8 teaching faculty per resident. However, programs that were not able to evaluate all 6 competencies using a “most desirable” method employed 1 FTE per 14.3 residents and 2.0 teaching faculty per resident. It is interesting to note that the experience of the program director or the time he/she spends on administrative activities was not related to the number of competencies being evaluated with a “most desirable” method.

Several strengths and limitations in this study should be acknowledged. This is the first nationwide analysis describing the current state of resident evaluation processes for graduate medical education in Internal Medicine in the United States. We report a reasonably high response rate and provide prevalence data on the use of tools to measure trainee competence. We caution the reader that our data represent process measures, not outcomes. Our study shows how trainee competence is assessed, not whether trainee competence has occurred.

Other limitations include the exclusion of 4 tools in the ACGME Toolbox of Assessment Methods from our survey (record review, checklist, oral exams, and procedures/case logs). Such omissions limit our ability to comment on compliance with ACGME “most desirable” methods. However, we did allow programs to indicate “other” methods of competency assessment and found that these tools were in fact, infrequently used (only mentioned in 11% of the few comments made), thereby maintaining the validity of our results.

We cannot infer why certain programs are or are not compliant with ACGME recommendations. We cannot comment on social desirability bias and cannot verify the accuracy of responses provided to us by survey participants. Finally, and importantly, we cannot comment on the competence with which programs are using each tool. This analysis will be important for future studies on evaluation methods.

Competency evaluation in graduate medical education is still in its infancy. With the help of the ACGME, programs have started the complex task of assessment with a strong foothold. However, we are far away from a comprehensive evaluation of trainee competence and are not yet adept at using all the tools the ACGME suggests we use for evaluating our trainees. Such evaluation is imperative if medical education is to join the quality movement that aims to provide high-quality care through high quality physicians.¹⁶

Acknowledgment: The authors acknowledge the thoughtful review of this manuscript provided by Drs. Lawrence Smith, MD and Stephen Kamholz, MD.

Conflict of Interest: There are no conflicts of interest to report for Drs. Saima Chaudhry and Brent Beasley. Dr. Eric Holmboe is employed by the American Board of Internal Medicine, a non-profit organization that performs assessment of physicians.

Corresponding Author: Saima I. Chaudhry, MD, MSHS; Office of Graduate Medical Education, North Shore University Hospital, 300 Community Drive, 4DSU, Manhasset, NY 11030, USA (e-mail: schaudhr@nshs.edu).

REFERENCES

1. Accreditation Council for Graduate Medical Education (ACGME). Outcome Project. <<http://www.acgme.org/outcome/project/proHome.asp>>. Accessed January 20, 2008.
2. Accreditation Council for Graduate Medical Education (ACGME). Outcomes Project Toolbox of Assessment Methods <<http://www.acgme.org/outcome/assess/toolbox.asp>>. Accessed January 20, 2008.
3. Hassett JM, Zinnerstrom K, Nawotniak RH, Schimpfhauser F, Dayton MT. Utilization of standardized patients to evaluate clinical and interpersonal skills of surgical residents. *Surgery*. 2006;140(4):633-9.
4. Kligler B, Koithan M, Maizes V, et al. Competency-based evaluation tools for integrative medicine training in family medicine residency: a pilot study. *BMC Med Educ*. 2007;7(1):7.
5. Lynch DC, Swing SR, Horowitz SD, Holt K, Messer JV. Assessing practice-based learning and improvement. *Teach Learn Med*. 2004; 16(1):85-92.
6. Ottestad E, Boulet JR, Lighthall GK. Evaluating the management of septic shock using patient simulation. *Clinical investigations. Crit Care Med*. 2007;35(3):769-75.
7. Wallenstein JN, Ander DS. Use of an Advanced Skills OSCE to evaluate core competencies in an emergency medicine clerkship. *Acad Emerg Med*. 2007;14(5S):215.
8. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995;310(6973):170.
9. Blumenthal D, Gokhale M, Campbell EG, et al. Preparedness for clinical practice: reports of graduating residents at academic health centers. *JAMA*. 2001;286:1027-34.
10. Accreditation Council for Graduate Medical Education/American Board of Medical Specialties Joint Initiative. Attachment/Toolbox of Assessment Methods Version 1.1, September 2000 <<http://www.acgme.org/Outcome/assess/ToolTable.pdf>>. Accessed January 20, 2008.
11. Gray JD. Global rating scales in residency education. *Acad Med*. 1996 Jan;71(1 Suppl):S55-63.
12. Ringsted C, Østergaard D, Ravn L, Pedersen JA, Berlac PA, Van Der Vleuten CPM. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. *Med Teach*. 2003;25(6):654-8.
13. Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski JJ. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med*. 2004;79(6):549-56.
14. Swing SR. Assessing the ACGME general competencies: general considerations and assessment methods. *Acad Emerg Med*. 2002;9(11):1278.
15. Gordon P, Tomasa L, Kerwin J. ACGME Outcomes Project: selling our expertise. *Fam Med*. 2004;36(3):164-7.
16. Goroll AH, Sirio C, Duffy FD, et al. A new model for accreditation of residency programs in internal medicine. *Ann Intern Med*. 2004;140(11):902-9.