

Update in Internal Medicine Residency Education: A Review of the Literature in 2010 and 2011

JOHN E. EATON, MD
 DARCY A. REED, MD, MPH
 BRIAN M. ABOFF, MD
 STEPHANIE A. CALL, MD, MSPH
 PAUL R. CHELMINSKI, MD, MPH
 UMA THANARAJASINGAM, MD, PhD
 JASON A. POST, MD
 KRIS G. THOMAS, MD
 DENISE M. DUPRAS, MD, PhD
 THOMAS J. BECKMAN, MD
 COLIN P. WEST, MD, PhD
 CHRISTOPHER M. WITTICH, MD
 ANDREW J. HALVORSEN, MS
 FURMAN S. McDONALD, MD, MPH

Abstract

Background Evidence-based practice in education requires high-quality evidence, and many in the medical education community have called for an improvement in the methodological quality of education research.

Objective Our aim was to use a valid measure of medical education research quality to highlight the methodological quality of research publications and provide an overview of the recent internal medicine (IM) residency literature.

Methods We searched MEDLINE and PreMEDLINE to identify English-language articles published in the United States and Canada between January 1, 2010, and December 31, 2011, focusing on IM residency education. Study quality was assessed using the Medical Education Research Study Quality Instrument (MERSQI), which has demonstrated reliability and validity. Qualitative articles were excluded. Articles were ranked by quality score, and the top 25% were examined for common themes, and 2

articles within each theme were selected for in-depth presentation.

Results The search identified 731 abstracts of which 223 articles met our inclusion criteria. The mean (\pm SD) MERSQI score of the 223 studies included in the review was 11.07 (\pm 2.48). Quality scores were highest for data analysis (2.70) and lowest for study design (1.41) and validity (1.29). The themes identified included resident well-being, duty hours and resident workload, career decisions and gender, simulation medicine, and patient-centered outcomes.

Conclusions Our review provides an overview of the IM medical education literature for 2010–2011, highlighting 5 themes of interest to the medical education community. Study design and validity are 2 areas where improvements in methodological quality are needed, and authors should consider these when designing research protocols.

John E. Eaton, MD, is Instructor and Fellow in Gastroenterology in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Darcy A. Reed, MD, MPH**, is Associate Professor, Associate Program Director, and Consultant in the Division of Primary Care Internal Medicine in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Brian M. Aboff, MD**, is Clinical Associate Professor and Program Director in the Department of Medicine, Christiana Care Health System; **Stephanie A. Call, MD, MSPH**, is Associate Professor, Program Director, and Associate Chair for Education in the Department of Internal Medicine, Virginia Commonwealth University; **Paul R. Chelminski, MD, MPH**, is Clinical Associate Professor and Associate Program Director in the Department of Internal Medicine, University of North Carolina at Chapel Hill; **Uma Thanarajasingam, MD, PhD**, is Instructor and Fellow in Rheumatology in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Jason A. Post, MD**, is Instructor and Consultant in the Division of Primary Care Internal Medicine in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Kris G. Thomas, MD**, is Associate Professor, Associate Program Director, and Consultant in the Division of Primary Care Internal Medicine in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Denise M.**

Dupras, MD, PhD, is Assistant Professor, Associate Program Director, and Consultant in the Division of Primary Care Internal Medicine in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Thomas J. Beckman, MD**, is Professor of Medicine and Medical Education, Associate Program Director, and Consultant in the Division of General Internal Medicine in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Colin P. West, MD, PhD**, is Associate Professor, Associate Program Director, and Consultant in the Division of General Internal Medicine in the Department of Internal Medicine and Associate Professor of Biostatistics in the Department of Health Sciences Research, Mayo Clinic College of Medicine; **Christopher M. Wittich, MD**, is Assistant Professor and Consultant in the Division of General Internal Medicine in the Department of Internal Medicine, Mayo Clinic College of Medicine; **Andrew J. Halvorsen, MS**, is Biostatistician and Data and Project Manager for the Internal Medicine Residency Office of Educational Innovations in the Department of Internal Medicine, Mayo Clinic College of Medicine; and **Furman S. McDonald, MD, MPH**, is Associate Professor of Medicine and Medical Education, Program Director, Consultant in the Divisions of General and Hospital Internal Medicine, and Associate Chair in the Department of Internal Medicine, Mayo Clinic College of Medicine.

Introduction

To promote best practices in education, it is important to be familiar with the highest level of evidence in the contemporary literature. However, the quality of medical education literature is variable and there is a need to improve quality and identify where further research is needed.¹⁻⁷ Leaders in medical education have suggested that authors should use validated instruments for assessing study quality when planning and assessing their work.⁸ The Medical Education Research Study Quality Instrument (MERSQI) is 1 method for assessing the quality of quantitative medical education research. Validity of MERSQI scores includes high interrater, intrarater, and internal consistency reliability, as well as content, criterion, and predictive validity.^{9,10} MERSQI scores correlate with article citation rates and journal editors' ratings of article quality.¹⁰

Sorting through the landscape of medical education literature can be a daunting task for the busy clinical educator. Residency educators need to remain current on relevant and timely topics in medical education; therefore, we conducted a focused review and quality assessment of the recent literature examining internal medicine (IM) residency education.

Methods

Data Sources and Searches

In consultation with a reference librarian, we conducted an electronic search of MEDLINE and PreMEDLINE (which includes articles not yet indexed) using the following terms: "Internship and Residency" or "Education, Medical, Graduate" and "Internal Medicine." Limits included studies published between January 1, 2010, and December 31, 2011; studies published in the English language; and studies involving humans. Abstracts from the *Journal of Graduate Medical Education* and *Canadian Medical Education Journal* were read manually.

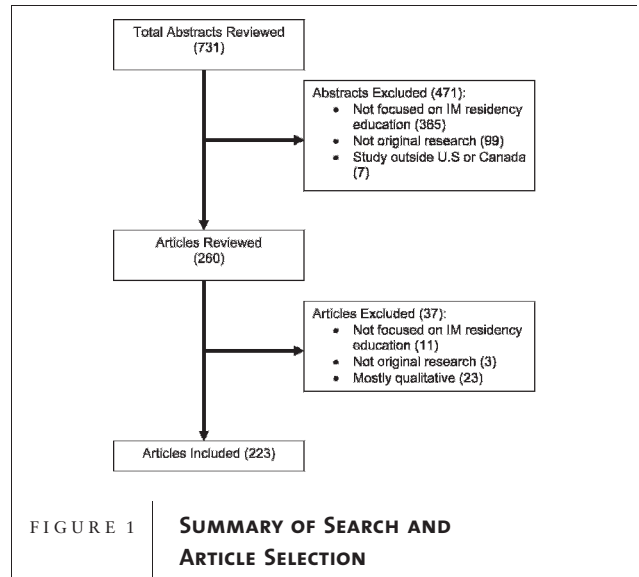
Funding: This study was supported in part by the Mayo Clinic Internal Medicine Residency Office of Educational Innovations as part of the Accreditation Council for Graduate Medical Education Educational Innovations Project.

While we are reporting results presented at an Association of Program Directors in Internal Medicine (APDIM) workshop, we are not presuming to speak for the organization, and our paper does not constitute an official policy statement of APDIM or any other organization with which any of the authors may be affiliated.

Corresponding author: Furman S. McDonald, MD, MPH, Department of Internal Medicine, Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, mcdonald.furman@mayo.edu

Received August 14, 2012; revisions received November 11, 2012, and December 19, 2012; accepted January 25, 2013.

DOI: <http://dx.doi.org/10.4300/JGME-D-12-00238.1>



Study Selection and Data Abstraction

We included studies published in the United States and Canada that were focused on IM residency education. This review did not include articles about subspecialty fellowship training. We included only original research studies (experimental, quasi-experimental, and observational research). We excluded reviews (meta-analysis, systematic, and narrative reviews), perspectives, qualitative studies, commentaries, editorials, and letters.

FIGURE 1 shows the results of the search strategy and number of articles excluded based on each criterion. Of a total of 731 abstracts, 417 were excluded based on title and abstract. The remaining 260 articles were reviewed in full, and 223 articles that met our inclusion criteria were included.

Articles were assigned a numerical code and randomly assigned to the reviewers. Reviewers who were authors of included papers did not score their own articles. All data were abstracted using a standard template that included the title, first author, and the 10 MERSQI items. Articles were ranked by MERSQI score. Next, we examined those articles with scores in the top 25th percentile, based on MERSQI score. This included 59 articles (because some articles had the same MERSQI score). These 59 articles were examined for content themes of importance to the IM education community. The themes that were selected for this review were represented by at least 2 articles in the top 25th percentile of MERSQI scores. Within themes, articles selected for in-depth presentation in this review were determined by reviewer consensus. The process of reviewer consensus consisted of discussions during teleconferences and voting among 5 coauthors (J.E.E., B.M.A., S.A.C., P.R.C., and F.S.M.) who represented 4 different institutions.

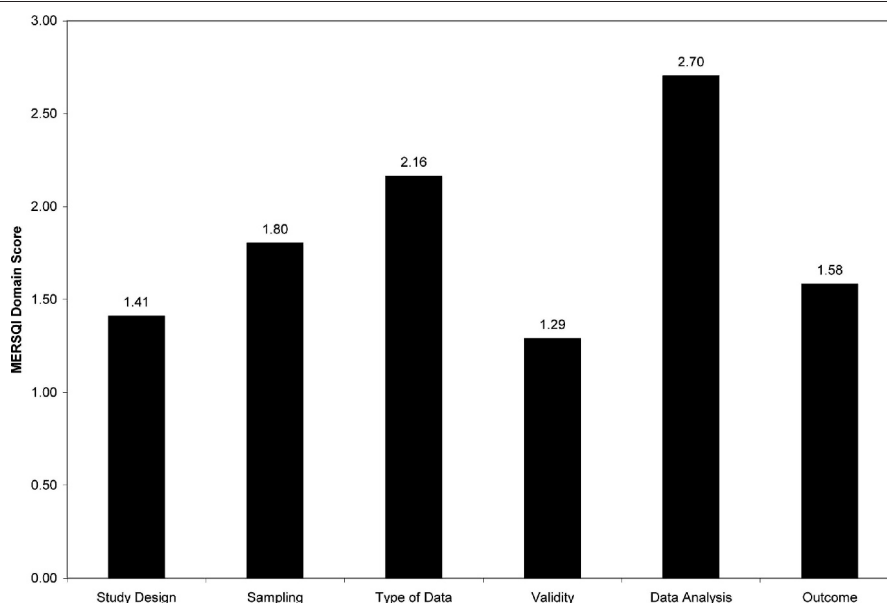


FIGURE 2 | MEAN SCORES FOR THE 6 METHODOLOGICAL DOMAINS AMONG 223 ARTICLES PUBLISHED 2010–2011

After achieving consensus among 5 coauthors, the remaining coauthors approved the article selection without additional changes. This process allowed us to select high-quality studies that reflect important topics for medical educators. Articles selected in the 5 important topics were presented at the Association of Program Directors in Internal Medicine annual spring meetings in 2011 and 2012.

Quality Assessment

The MERSQI is an instrument designed to assess the methodological quality of quantitative medical education research studies. It consists of 10 items that reflect 6 domains of study quality: sampling data, data type, study design, data analysis, outcomes, and validity of assessments.⁹ Each of the 6 domains within the MERSQI has a maximum score of 3. Total MERSQI scores range from 5 to 18 with higher scores indicating higher overall methodological quality. MERSQI scores were summarized using means (\pm SD).

All reviewers participating in this study were trained to use the MERSQI. This training consisted of a calibration exercise in which several articles were reviewed by all coauthors and results were discussed to further minimize interrater variability.

Results

The mean (\pm SD) MERSQI score for the 223 quantitative studies was 11.07 (\pm 2.48). FIGURE 2 shows the mean scores for the 6 methodological domains. The domains in which studies received the highest scores were data analysis (2.70) and reporting of objective data (2.16). Articles scored lowest in

outcomes (1.58), study design (1.48), and validity of assessments (1.29). This finding reflects previous MERSQI studies, which also demonstrated low scores for these domains.^{9,10}

Major themes in the 2010–2011 IM residency education literature included resident well-being, duty hours and resident workload, career decisions and gender, simulation medicine, and patient-centered outcomes. Examples of other specific topics that were identified but not discussed in this review included medical knowledge, professionalism, scholarship, residency match, quality improvement, cultural and racial diversity, signout, night float, communication, supervision, evaluation, resident attitudes and perceptions, systems-based practice, and health care costs. Several articles could not be categorized further and were labeled as miscellaneous.

Resident Well-Being

West CP, et al. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. *JAMA*. 2011;306(9):952–960.

The authors performed a cross-sectional study of IM residents who took the IM training examination (IM-ITE) in the 2008–2009 academic year in order to measure resident well-being and how it relates to other covariates (MERSQI score = 14.5).¹¹ Seventy-four percent of all eligible US IM residents (more than 16,000) were represented in the sample. Nearly 15% of residents reported that their quality of life (QOL) was “as bad as it can be,” and more than half of the respondents reported burnout, which tended to be lower in international medical

graduates than in US medical graduates (odds ratio [OR], 0.70; 99% confidence interval [CI], 0.63–0.77); $P < .001$) but higher in those with a greater educational debt (OR, 1.72; 99% CI, 1.49–1.99; $P < .001$; for debt greater than \$200,000). Those with greater debt (more than \$200,000) had a mean IM-ITE score that was 5 points lower—a difference that remained similar across all years of training. Similarly, IM-ITE scores were lower by 2.7 and 4.2 points if residents reported their QOL “as bad as it can be” and had daily symptoms of emotional exhaustion, respectively. Moonlighting was not associated with burnout.

The sample in this study is strongly representative of the national IM resident population. The observed associations between educational debt, resident distress, and a core Accreditation Council for Graduate Medical Education competency (medical knowledge) are compelling, but the study design does not demonstrate causality (study design domains). Limitations of this study include the possibility of recall bias and the inability to correlate QOL with outcomes of patient care or resident performance (validity and outcome domains).

Billings M, et al. The effect of the hidden curriculum on resident burnout and cynicism. *J Grad Med Educ.* 2011;3(4):503–510.

This study assessed the effect of the hidden curriculum on resident burnout and cynicism (MERSQI score = 12.5).¹² Between 2008 and 2010, 337 of 708 (48%) residents at 2 academic medical centers responded to a survey that examined the effects of the “hidden curriculum” (unprofessional conduct, which can lead to conveying an unprofessional culture to learners) on burnout and professional behavior (cynicism).

Similar to the study by West et al,¹¹ 45% of residents met the criteria for burnout. Those with burnout had a higher hidden curriculum score (median, 26 [interquartile range, 19.0–34.0]) versus median of 19.0 (interquartile range, 10.0–28.0; $P < .001$) and cynicism score (median, 12.7 [interquartile range, 9.7–16.2]) versus median of 8.7 (interquartile range, 6.7–10.8; $P < .001$). The hidden curriculum score varied significantly by study site, and men had higher cynicism scores than women.

Burnout is particularly relevant to the medical education community as it may impact other behaviors and performance measures. Although this study does not demonstrate causality (study design domain), it reinforces the importance of identifying and improving the hidden curriculum, which may be associated with resident burnout and markers of unprofessionalism such as cynicism. The potential for recall or response bias, low response rate, lack of correlation to other outcomes, and decreased external

validity are limitations of this study (sampling and validity domains).

Duty Hours and Resident Workload

Coit MH, et al. The effect of workload reduction on the quality of residents' discharge summaries. *J Gen Intern Med.* 2010;26(1):28–32.

The authors performed a nonrandomized study that compared 2 team designs (intervention versus control) in an IM inpatient setting to examine the impact of workload reduction on discharge summaries (MERSQI score = 14.7).¹³ Compared to the control group, the intervention teams had 2 staff and 2 residents versus 1 staff and 1 resident, less frequent resident admission periods (coverage 1 night in 4 [1:4] until 10 PM and night coverage versus 1:2 until 7 PM and night coverage), and more interns with decreased admission periods (3 interns with 1:6 overnight versus 2 interns with 1:4 until 10 PM).

A total of 142 summaries produced by 61 residents were assessed for quality metrics by 2 reviewers (kappa = 0.90). The instrument was based on Joint Commission standards. The intervention team summaries had more of the required discharge elements than the controls (74% versus 65%, $P < .001$).

Discharge summaries are critical safe transfers of care and important markers of communication. This study illustrated that resident work quality can improve with workload reduction. Limitations include the fact that this study was conducted at a single community hospital-affiliated academic center, potentially limiting the generalizability of its results; it was not randomized; and the possible confounding effect of the additional staffing on the outcomes (sampling and study design domains).

Mourad M, et al. Shifting indirect patient care duties to after hours in the era of work hours restrictions. *Acad Med.* 2011;86(5):586–590.

A study to assess the indirect effects of the duty hour limits measured times of initiation and termination of dictations and compared them to 39 residents' schedules to determine if indirect patient care tasks such as dictating were a source of duty hour violations (MERSQI score = 13.2).¹⁴ This study was conducted at a single academic center.

Most residents (82% with definite violations) dictated beyond their required duty period; however, only 5% of residents reported duty hour violations on a Web-based self-reporting system. The proportion who dictated on days off was similar to that of those who dictated after a 30-hour duty period. Longer dictations, a greater time spent on dictations, and team census were associated with increased work hour violations.

There is concern that residents will spend a significant proportion of their time on indirect patient care and that reductions in duty hours may shift this work to time out of the counted duty hours. This study illustrates the fact that indirect patient care tasks are an important source of duty hour violations. Programs may need to make systematic changes to prevent the need for completion of indirect patient care tasks outside the required duty period. It also provides evidence in support of the long-standing suspicion that self-reported duty hour violations may differ from actual frequency of duty hour violations. This reinforces the need for better systems to monitor duty hours. Primary limitations of this study are generalizability (sampling domain), outcome measurement, and nonrandomized design.

Career Decisions and Gender

Willett LL, et al. Do women residents delay childbearing due to perceived career threats? *Acad Med.* 2010;85(4):640–646.

The researchers surveyed 424 residents across 3 institutions to assess gender differences and influencing factors related to having children during residency (MERSQI = 12.0).¹⁵ The response rate was 77%. In the multivariate model, women were less likely than men to plan to have children during residency (OR = 0.46; [95% CI: 0.25–0.84]), and women were more likely to report a belief that pregnancy would be complicated or pose a threat to their career by extended training or the loss of fellowship positions.

This multicenter survey found that women chose to postpone pregnancy due to perceived career threats. Awareness of these gender differences is important when counseling residents. Limitations are that some factors likely to influence childbearing decisions were not measured (validity), the outcome focused primarily on behavior (outcome), and it was a cross-sectional survey (study design domain).

Halvorsen AJ, et al. Gender and future salary: disparate trends in internal medicine residents. *Am J Med.* 2010;123(5):470–475.

This study investigated associations between gender and medical school and subspecialty choice and salary potential by examining fellowship match data and national data from IM In-Training Examination results (MERSQI = 10.0).¹⁶ From the programs that participate in the In-Training Examination, 17 015 residents responded to questions regarding their gender, medical school, and career plans. There was a negative correlation between the proportion of women fellows and mean salary potential ($R = -0.83$, $P = .005$). Career choice trends and future mean salary were similar between female US medical graduates and international medical graduates. Endocrinology had

the second lowest mean salary and highest proportion of women (67%); cardiology had the highest mean salary and lowest proportion of women (19%). The effect of medical school origin did not contribute significantly to future salary and specialty choice when gender was taken into account.

The findings suggest that women tend to pursue subspecialties that have a lower mean salary more often than men do. This study examined a highly representative sample of residents; however, as this was an observational study (study design domain), causality cannot be inferred, and it primarily focused on behaviors rather than health and patient outcomes (outcome domain).

Simulation and Medical Education

Khouli H, et al. Performance of medical residents in sterile techniques during central vein catheterization. *Chest.* 2011;139(1):80–87.

A study of the impact of simulation training on effective central vein catheterization randomized IM residents to receive video training alone or simulation-based plus video training, and their sterile techniques were measured in a postintervention test (MERSQI = 15.0).¹⁷ The study also compared the incidence of catheter-related bloodstream infections (CRBSI) before and after simulation-based training within the medical intensive care unit and compared them to those in the surgical intensive care unit control group.

Baseline scores in sterile technique between groups were equally poor; however, after simulation-based plus video training, the scores were significantly higher than those in the video group (92% versus 75%, respectively, $P < .001$). There was a 70% reduction in the incidence of CRBSI following implementation of simulation-based plus video training.

Simulation-based medicine is gaining increasing acceptance. In contrast to the traditional “see one, do one, teach one” model, it enables physicians in training to develop their skills while exposing patients to minimal risk. The findings of the article illustrate the potential of simulation-based medicine to improve resident compliance with sterile technique and improve patient outcomes by reducing the incidence of CRBSI. The primary limitation to this study was that it was restricted to a single institution (sampling domain).

Sekiguchi H, et al. A prerotational, simulation-based workshop improves the safety of central venous catheter insertion. *Chest.* 2011;140(3):652–658.

The researchers conducted a prospective cohort study at a single urban teaching hospital over a 6-month period to examine the effectiveness of a prerotational, simulation-based workshop with ultrasound-guided central venous catheter placement in reducing mechanical complications

TABLE	SUMMARY OF THE KEY FINDINGS
Resident Well-Being	
<ul style="list-style-type: none"> West et al.¹¹ MERSQI = 14.5. There is a high prevalence of burnout among residents. Burnout was higher in those with greater educational debt, and debt correlated with lower standardized test scores. Billings et al.¹² MERSQI = 12.5. Residents with burnout had higher hidden curriculum scores and cynicism scores. Men (compared to women) had higher cynicism scores. 	
Duty Hours and Resident Workload	
<ul style="list-style-type: none"> Coit et al.¹³ MERSQI = 14.7. Resident workload reduction can improve the quality of residents' discharge summaries. Mourad et al.¹⁴ MERSQI = 13.2. Indirect patient care tasks such as documentation are a source of delay beyond the required duty period. Self-reported duty hour violations differed markedly from the actual frequency of duty hour violations. 	
Career Decisions and Gender	
<ul style="list-style-type: none"> Willett et al.¹⁵ MERSQI = 12.0. Women were less likely than men to plan to have children during residency and more likely to report a belief that pregnancy would adversely affect their career. Halvorsen et al.¹⁶ MERSQI = 10.0. Based on a large survey, women were more likely than men to pursue subspecialties with a lower mean salary. 	
Simulation Medicine	
<ul style="list-style-type: none"> Khouli et al.¹⁷ MERSQI = 15.0. Sterile technique in placing central lines improved and catheter-related bloodstream infections decreased after simulation plus video training compared to video training alone. Sekiguchi et al.¹⁸ MERSQI = 14.4. Implementation of a simulation-based, ultrasound-guided central line workshop improved the safety of central venous catheter placement. 	
Patient-Centered Outcomes	
<ul style="list-style-type: none"> Record et al.¹⁹ MERSQI = 14.0. The number of admissions was reduced in the intervention team to allow residents more time to follow-up with patients after discharge from the hospital. This was associated with a decrease in the rate of deaths or 30-day readmissions for heart failure. Hess et al.²⁰ MERSQI = 14.0. A large multicenter survey concluded that elderly patients were less likely to receive key aspects of geriatric care if seen by residents than if seen by practicing physicians. Resident patients rated their care lower than patients seen by practicing physicians. 	

Abbreviation: MERSQI, Medical Education Research Study Quality Instrument.

associated with the procedure (MERSQI = 14.4).¹⁸ Data for complication rates postintervention were compared to those for preintervention rates.

In the preintervention time period, 334 procedures were performed compared to 402 procedures in the postintervention phase. Prior to the workshop, postgraduate year-1 (PGY-1) residents were more likely to cause mechanical complications than fellows and staff. However, this trend did not occur in the postintervention period. Following the workshop, ultrasound usage increased (61% versus 3%, respectively, $P < .01$), and the overall complication rate decreased (23% versus 33%, respectively, $P < .01$).

Implementation of a prerotational workshop significantly improved the safety of central venous catheter placement for inexperienced operators. It is likely that an increase in ultrasound use following the workshop contributed to this outcome. A single center, nonrandomized design (sampling and study design domains) is the primary limitation to this study.

Patient-Centered Outcomes

Record JD, et al. Reducing heart failure readmissions by teaching patient-centered care to internal medicine residents. *Arch Intern Med.* 2011;171(9):858–859.

This study compared the effect of a patient-centered curriculum on 1 inpatient service versus 3 control services to assess the effect of the intervention on 30-day rehospitalizations for heart failure patients, using a non-randomized study design (MERSQI = 14.0).¹⁹

The number of patients admitted to the intervention team was reduced by half to allow residents more time to participate in the new curricular activities, which consisted of (1) medication adherence assessment, (2) telephone calls to outpatient providers, and (3) telephone calls to each patient after discharge. The intervention team visited selected patients at home following discharge to assess patients' perspectives regarding their transition of care. The study was conducted at a single academic center. The teams

had similar baseline characteristics, and 52 patients with heart failure were admitted to the intervention and 323 patients with heart failure made up the control group. The rate of death or 30-day readmission for heart failure was lower in the intervention group (14% versus 4%, respectively, $P = .046$).

This type of educational intervention could be applied to other medical conditions and may affect patient care and quality metrics such as hospital readmissions. The effect that a reduced census had on improved outcomes for the intervention team (versus change in curriculum) is unclear. This was a nonrandomized design (study design domain), and the ability to generalize these findings elsewhere might be limited because of the sample size and because it was conducted at a single center (sampling domain).

Hess BJ, et al. Listening to older adults: elderly patients' experience of care in residency and practicing physician outpatient clinics. *J Am Geriatr Soc.* 2011;59(5):909–915.

The authors administered a survey to resident patients across 42 training programs (34 IM and 12 family medicine) and to the patients of 144 practicing physicians in order to compare patient care performance measures (MERSQI = 14.0).²⁰ The authors used a survey that is part of the American Board of Internal Medicine Practice Improvement Module, which focused on the care of vulnerable elderly. A panel of geriatricians who represented various societies helped select various geriatric specific process measures for inclusion in the patient survey.

The characteristics of the training programs (academic versus nonacademic) and practicing physicians (private practice versus academic or hospital-owned office-based practice) were diverse. Data from 4204 patient surveys from 144 practicing physicians were compared to those from 2213 patient surveys collected from 42 residency programs. The response rate for an individual clinic or physician was not assessed (sampling domain). Patients seen in residents' clinics were less likely to receive care and guidance for key aspects of geriatric care. For example, fewer patients reported that residents provided assistance to help prevent falls or imbalance (42% versus 62%, respectively, $P < .001$), and resident patients were less likely to rate their overall care as high (78% versus 89%, respectively, $P < .001$).

This large, multicenter survey detected disparities within care processes and patient perceptions among older adults cared for by resident physicians. This population will continue to increase, and programmatic changes will be required to ensure the needs of older adults are being met by the physician workforce of tomorrow. A limitation of this study is that the primary outcomes were measured by patient survey (study design) domain. In addition, it could not identify resident or programmatic factors that could

explain if and why some programs and individuals performed better than others (validity domain).

Discussion

To our knowledge, this is the first study to perform a systematic review of the IM graduate medical education quantitative literature and apply a quality assessment tool (MERSQI) with strong validity evidence to identify studies of high methodological quality for the busy clinical educator. Study design, validity assessment, and outcomes are 3 domains that need improvement in the IM graduate medical education literature. Most of these studies scored low in study design and outcomes because the design was nonrandomized and did not measure patient- or health-related outcomes. Validity was a limitation often because authors did not comment on interrater variability of their instrument, describe how their instrument was developed and tested, or comment on relationship to other variables. We would encourage authors to consider these measures of quality when designing future research. The use of the MERSQI aids in quantifying the quality of research and highlights articles that reflect important topics for the medical education community.

Limitations

Our study has several limitations. First, MERSQI scores were not adjudicated by a second person. To minimize interrater variability, we held a calibration exercise where a series of articles were reviewed, scored, and discussed. Second, we were able to highlight only several of the many important papers published in 2010–2011, and qualitative studies were excluded as the MERSQI could not be applied to them. However, article selection was based on the MERSQI score and after discussion with members of the IM medical education community who represented 4 different institutions. Other articles discussed at our literature review workshops at the Association of Program Directors in Internal Medicine annual spring meetings in 2011 and 2012 that the reader may find useful are cited below.^{21–28} We limited our search to articles focusing on IM residents for several reasons. First, IM residents are the largest specialty trainee group. Second, the themes highlighted in this review are not unique to IM and are applicable to many medical and surgical specialties. Therefore, this body of work has broad applicability to other areas of medicine.

Conclusion

Important themes in the medical education literature published in 2010–2011 highlighted in this review included resident well-being, duty hours and resident workload, career decisions and gender, simulation and medical

education, and patient-centered outcomes (TABLE). Although the studies were primarily focused on IM residents, the themes and data summarized in this review are important for training physicians across multiple specialties.

References

- 1 Dauphinee WD, Wood-Dauphinee S. The need for evidence in medical education: the development of best evidence medical education as an opportunity to inform, guide, and sustain medical education research. *Acad Med.* 2004;79(10):925–930.
- 2 Wartman SA. Research in medical education: the challenge for the next decade. *Acad Med.* 1994;69(8):608–614.
- 3 Lurie SJ. Raising the passing grade for studies of medical education. *JAMA.* 2003;290(9):1210–1212.
- 4 Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med.* 2004;79(10):955–960.
- 5 Whitcomb ME. Using clinical outcomes data to reform medical education. *Acad Med.* 2005;80(2):117.
- 6 Murray E, Gruppen L, Catton P, Hays R, Woolliscroft JO. The accountability of clinical education: its definition and assessment. *Med Educ.* 2000;34(10):871–879.
- 7 Golub RM. Medical education theme issue 2012 call for papers. *JAMA.* 2012;307(6):616.
- 8 Sullivan GM. A primer on the validity of assessment instruments. *J Grad Med Educ.* 2011;3(2):119–120.
- 9 Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. *J Gen Intern Med.* 2008;23(7):903–907.
- 10 Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA.* 2007;298(9):1002–1009.
- 11 West CP, Shanafelt TD, Kolars JC. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. *JAMA.* 2011;306(9):952–960.
- 12 Billings M, Lazarus M, Wenrich M, Curtis JR, Engelberg RA. The effect of the hidden curriculum on resident burnout and cynicism. *J Grad Med Educ.* 2011;3(4):503–510.
- 13 Coit MH, Katz JT, McMahon GT. The effect of workload reduction on the quality of residents' discharge summaries. *J Gen Intern Med.* 2010;26(1):28–32.
- 14 Mourad M, Vidyarthi AR, Hollander H, Ranji SR. Shifting indirect patient care duties to after hours in the era of work hours restrictions. *Acad Med.* 2011;86(5):586–590.
- 15 Willett LL, Wellons MF, Hartig JR, Roenigk L, Panda M, Dearing AT, et al. Do women residents delay childbearing due to perceived career threats? *Acad Med.* 2010;85(4):640–646.
- 16 Halvorsen AJ, Kolars JC, McDonald FS. Gender and future salary: disparate trends in internal medicine residents. *Am J Med.* 2010;123(5):470–475.
- 17 Khouli H, Jahnnes K, Shapiro J, Rose K, Matthew J, Gohil A, et al. Performance of medical residents in sterile techniques during central vein catheterization. *Chest.* 2011;139(1):80–87.
- 18 Sekiguchi H, Tokita JE, Minami T, Eisen LA, Mayo PH, Narasimhan M. A prerotational, simulation-based workshop improves the safety of central venous catheter insertion. *Chest.* 2011;140(3):652–658.
- 19 Record JD, Rand C, Christmas C, Hanyok L, Federowicz M, Bilderback A, et al. Reducing heart failure readmissions by teaching patient-centered care to internal medicine residents. *Arch Intern Med.* 2011;171(9):858–859.
- 20 Hess BJ, Lynn LA, Conforti LN, Holmboe ES. Listening to older adults: elderly patients' experience of care in residency and practicing physician outpatient clinics. *J Am Geriatr Soc.* 2011;59(5):909–915.
- 21 Babbott SF, Beasley BW, Reddy S, Duffy FD, Nadkarni M, Holmboe ES. Ambulatory office organization for internal medicine resident medical education. *Acad Med.* 2010;85(12):1880–1887.
- 22 Byrne JM, Kashner M, Gilman SC, Aron DC, Cannon GW, Chang BK, et al. Measuring the intensity of resident supervision in the department of veterans affairs: the resident supervision index. *Acad Med.* 2010;85(7):1171–1181.
- 23 Wetz RV, Seelig CB, Khoueiry G, Weiserbs KF. Out-of-match residency offers: the possible extent and implications of prematching in graduate medical education. *J Grad Med Educ.* 2010;2(3):327–333.
- 24 Patel MS, Katz JT, Volpp KG. Match rates into higher-income, controllable lifestyle specialties for students from highly ranked, research-based medical schools compared with other applicants. *J Grad Med Educ.* 2010;2(3):360–365.
- 25 Fancher TL, Keenan C, Meltvedt C, Stocker T, Harris T, Morfin J, et al. An academic-community partnership to improve care for the underserved. *Acad Med.* 2011;86(2):252–258.
- 26 West CP, Halvorsen AJ, McDonald FS. Scholarship during residency training: a controlled comparison study. *Am J Med.* 2011;124(10):983–987.
- 27 Cullen MW, Reed DA, Halvorsen AJ, Wittich CM, Kreuziger LM, Keddiss MT, et al. Selection criteria for internal medicine residency applicants and professionalism ratings during internship. *Mayo Clin Proc.* 2011;86(3):197–202.
- 28 Didwania A, McGaghie WC, Cohen ER, Butter J, Barsuk JH, Wade LD, et al. Progress toward improving the quality of cardiac arrest medical team responses at an academic teaching hospital. *J Grad Med Educ.* 2011;3(2):211–216.