

Association Between Funding and Quality of Published Medical Education Research

Darcy A. Reed, MD, MPH

David A. Cook, MD, MHPE

Thomas J. Beckman, MD

Rachel B. Levine, MD, MPH

David E. Kern, MD, MPH

Scott M. Wright, MD

STAKEHOLDERS IN MEDICAL EDUCATION have maintained that the quality of medical education research is inadequate. Professional organizations, journal editors, and education researchers have appealed for greater methodological rigor,¹⁻⁵ larger multi-institutional studies,⁶⁻⁹ and more meaningful and clinically relevant outcomes.¹⁰⁻¹⁴

Despite the need to improve the quality of medical education research, funding opportunities are limited.^{6,15} Approximately two-thirds of published medical education studies are not funded, and those with support are substantially underfunded.¹⁵ Investigators believe that increased funding will enhance the quality of medical education research by facilitating the use of stronger study designs and multi-institutional collaborations,¹⁵ but an association between education research funding and study quality has not been shown. An evidentiary link between funding and study quality is needed to justify greater resource allocation for medical education research.

The lack of valid measures for evaluating the quality of medical education research has precluded attempts to demonstrate associations between study quality and funding. Although guidelines exist for conducting and report-

Context Methodological shortcomings in medical education research are often attributed to insufficient funding, yet an association between funding and study quality has not been established.

Objectives To develop and evaluate an instrument for measuring the quality of education research studies and to assess the relationship between funding and study quality.

Design, Setting, and Participants Internal consistency, interrater and intrarater reliability, and criterion validity were determined for a 10-item medical education research study quality instrument (MERSQI). This was applied to 210 medical education research studies published in 13 peer-reviewed journals between September 1, 2002, and December 31, 2003. The amount of funding obtained per study and the publication record of the first author were determined by survey.

Main Outcome Measures Study quality as measured by the MERSQI (potential maximum total score, 18; maximum domain score, 3), amount of funding per study, and previous publications by the first author.

Results The mean MERSQI score was 9.95 (SD, 2.34; range, 5-16). Mean domain scores were highest for data analysis (2.58) and lowest for validity (0.69). Intraclass correlation coefficient ranges for interrater and intrarater reliability were 0.72 to 0.98 and 0.78 to 0.998, respectively. Total MERSQI scores were associated with expert quality ratings (Spearman ρ , 0.73; 95% confidence interval [CI], 0.56-0.84; $P < .001$), 3-year citation rate (0.8 increase in score per 10 citations; 95% CI, 0.03-1.30; $P = .003$), and journal impact factor (1.0 increase in score per 6-unit increase in impact factor; 95% CI, 0.34-1.56; $P = .003$). In multivariate analysis, MERSQI scores were independently associated with study funding of \$20 000 or more (0.95 increase in score; 95% CI, 0.22-1.86; $P = .045$) and previous medical education publications by the first author (1.07 increase in score per 20 publications; 95% CI, 0.15-2.23; $P = .047$).

Conclusion The quality of published medical education research is associated with study funding.

JAMA. 2007;298(9):1002-1009

www.jama.com

ing research,¹⁶⁻²⁴ these are limited because they apply only to specific study types, such as randomized trials. Thus, these guidelines cannot be used to compare research quality across studies of various designs. We are unaware of published instruments for assessing the quality of both experimental and observational medical education studies. Such instruments would enable educators, reviewers, and journal editors to make comparisons across the spectrum of evidence in medical education.

We hypothesized that funding would be associated with higher-quality medi-

cal education research studies. Therefore, the objectives of this study were to (1) develop an instrument to measure the methodological quality of education research studies and determine its reliability and validity, and (2) identify relationships between funding and study quality.

Author Affiliations: Mayo Clinic College of Medicine, Rochester, Minnesota (Drs Reed, Cook, and Beckman); and Johns Hopkins University School of Medicine, Baltimore, Maryland (Drs Levine, Kern, and Wright). **Corresponding Author:** Darcy A. Reed, MD, MPH, Division of Primary Care Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN 55901 (reed.darcy@mayo.edu).

See also p 1038.

METHODS

We conducted a validity study of an instrument to measure the quality of medical education research and a cross-sectional study using that instrument to identify associations between funding and study quality. The Mayo Foundation institutional review board deemed this study exempt from review. Authors' written response to the survey indicated consent to participate.

Development and Evaluation of the Medical Education Research Study Quality Instrument

A medical education research study quality instrument (MERSQI) was designed to measure the quality of experimental, quasi-experimental, and observational studies. MERSQI content was determined by a comprehensive literature review of reports on research quality and critical discussion and instrument revision among the study authors. Items were selected that reflected research quality rather than reporting quality (eg, clarity of writing); elements such as "importance of research questions" and "quality of conceptual frameworks" were excluded because they are subject to individual interpretation and vary with time. MERSQI items were operationally defined and modified according to repeated pilot testing using medical education studies not included in the study sample.

The final MERSQI included 10 items, reflecting 6 domains of study quality: study design, sampling, type of data (subjective or objective), validity, data analysis, and outcomes (TABLE 1). MERSQI items were scored on ordinal scales and summed to determine a total MERSQI score. The maximum score for each domain was 3, producing a maximum possible MERSQI score of 18 and potential range of 5 to 18. The total MERSQI score was calculated as the percentage of total achievable points (accounting for "not applicable" responses) and then adjusted to a standard denominator of 18 to allow for comparison of scores across studies.

Item definitions and response categories were based on available evidence, with an attempt to avoid arbitrary cut-offs. For example, study design (item 1) was scored according to established hierarchies of research designs.^{23,25} There is general agreement that multi-institutional studies are preferable to single-institution studies (item 2)^{6,7,9} and that objective measurements are preferable to subjective measures in quantitative research (item 4).^{13,26} Although there is no consensus on what percentage constitutes an adequate response rate (defined as the percentage of participants who completed the evaluation component of the study), response rate categories (item 3) were chosen a priori, anticipating roughly equal differentiation among published studies. Items 5, 6, and 7 reflect an established validity framework²⁷⁻²⁹ and include the 3 most commonly reported categories of validity evidence: internal structure, content, and relationships to other variables ("criterion validity").³⁰ Accuracy and integrity of data analysis (items 8 and 9) is recognized as an important aspect of study quality; some journals require a written statement of responsibility for data analysis as a condition of publication.³¹ The rating of outcomes (item 10) is based on Kirkpatrick's³² widely accepted hierarchy.¹⁶ The highest score was assigned to health care- or patient-related outcomes in response to requests for clinically relevant outcomes-based research in medical education.^{10-14,26}

MERSQI Validity and Reliability

Principal Components Analysis and Reliability. MERSQI score dimensionality was examined by using principal components analysis with orthogonal rotation. Components with eigenvalues greater than 1 were retained and confirmed by inspecting the corresponding scree plot.³³ Items with factor loadings greater than 0.4 were retained.³⁴

Cronbach α was used to determine internal consistency of individual components and the overall MERSQI (all items combined). Intraclass correlation coefficients (ICCs) were used to as-

sess interrater and intrarater reliability for all items. ICCs were interpreted according to Landis and Koch³⁵: less than 0.4, poor; 0.4 to 0.75, fair to good; and greater than 0.75, excellent.³⁶

Criterion Validity. Criterion validity evidence was demonstrated by the association of MERSQI scores with 3 criterion variables. First, MERSQI scores were correlated with global quality ratings from 2 independent experts who are nationally recognized authorities in medical education research and current or former editors of leading medical education journals. The experts had no knowledge of the MERSQI. Experts were asked to make a global assessment of methodological quality for 50 studies randomly selected from the overall sample, using a scale of 1 to 5 (1=very poor; 5=excellent). To blind experts to the authors and affiliations, acknowledgments, and journal, the text of each article was electronically copied into a uniform format before evaluation. Expert interrater agreement was determined with ICC. Spearman ρ was used to measure correlation between median expert quality ratings and total MERSQI scores.

Second, the association between MERSQI scores and the 3-year citation rate (number of times the study was cited in the first 3 years after publication) was measured. Citations were obtained from the Scopus database.³⁷ To adjust for potential citation rate inflation,³⁸ self-citations were manually excluded.

Third, the association between MERSQI scores and the impact factor³⁹ of the publishing journal in the year the study was published was measured. Simple linear regression was used to measure associations between total MERSQI scores and citation rate and impact factor.

Study Sample and Data Extraction

The MERSQI was applied to a sample of 210 medical education research studies published between September 1, 2002, and December 31, 2003, in 13 peer-reviewed journals. Studies were selected from 2 journals encompassing all

medical specialties (*JAMA, New England Journal of Medicine*), 4 medical education journals (*Academic Medicine, Medical Education, Teaching and Learning in Medicine, Medical Teacher*), and 7 journals representing the core specialty areas of internal medicine, general surgery, pediatrics, family medicine, obstetrics and gynecology, and emergency medicine (*Annals of Internal Medicine, Journal of General Internal Medicine, American Journal of Sur-*

Table 1. MERSQI Domain and Item Scores for 210 Medical Education Research Studies

Domain	MERSQI Item	Studies, No. (%) ^a	Score		Mean (SD)	
			Item	Maximum Domain	Item	Domain
Study design	1. Study design			3	1.28 (0.47)	1.28 (0.47)
	Single group cross-sectional or single group posttest only	140 (66.7)	1			
	Single group pretest and posttest	33 (15.7)	1.5			
	Nonrandomized, 2 group	31 (14.8)	2			
	Randomized controlled trial	6 (2.9)	3			
Sampling	2. No. of institutions studied			3	0.84 (0.46)	1.90 (0.65)
	1	135 (64.3)	0.5			
	2	8 (3.8)	1			
	>2	67 (31.9)	1.5			
	3. Response rate, %				1.06 (0.44)	
	Not applicable	30 (14.3)				
	<50 or not reported	60 (33.3) ^b	0.5			
	50-74	39 (21.7) ^b	1			
	≥75	81 (45.0) ^b	1.5			
Type of data	4. Type of data			3	1.91 (0.99)	1.91 (0.99)
	Assessment by study participant	114 (54.3)	1			
	Objective measurement	96 (45.7)	3			
Validity of evaluation instrument ^c	5. Internal structure			3	0.25 (0.44)	0.69 (0.93)
	Not applicable	25 (11.9)				
	Not reported	138 (74.6) ^d	0			
	Reported	47 (25.4) ^d	1			
	6. Content				0.29 (0.45)	
	Not applicable	25 (11.9)				
	Not reported	132 (71.4) ^d	0			
	Reported	53 (28.6) ^d	1			
	7. Relationships to other variables				0.15 (0.36)	
	Not applicable	25 (11.9)				
Not reported	157 (84.9) ^d	0				
	Reported	28 (15.1) ^d	1			
Data analysis	8. Appropriateness of analysis	29 (13.8)		3	0.86 (0.35)	2.58 (0.65)
	Data analysis inappropriate for study design or type of data	181 (86.2)	0			
	Data analysis appropriate for study design and type of data		1			
	9. Complexity of analysis	58 (27.6)			1.72 (0.45)	
	Descriptive analysis only	152 (72.4)	1			
	Beyond descriptive analysis		2			
Outcomes	10. Outcomes	102 (48.6)		3	1.44 (0.50)	1.44 (0.50)
	Satisfaction, attitudes, perceptions, opinions, general facts	41 (19.5)	1			
	Knowledge, skills	62 (29.5)	1.5			
	Behaviors	5 (2.4)	2			
	Patient/health care outcome		3			
Total Score				18	9.95 (2.34)	

Abbreviation: MERSQI, medical education research study quality instrument.

^aPercentages may not total 100 due to rounding.

^bPercentage based on the 180 studies without a "not applicable" rating for the "response rate" item.

^cApplies to a new or referenced instrument.

^dPercentage based on the 185 studies without a "not applicable" rating for the "validity of evaluation instrument" items.

ger, *Pediatrics, Family Medicine, American Journal of Obstetrics and Gynecology, Academic Emergency Medicine*). Medical education research was defined as any original research study pertaining to medical students, residents, fellows, faculty development, or continuing medical education for physicians. Original research was defined as an educational intervention or trial, curriculum evaluation with subjective or objective outcomes, evaluation of an educational instrument or tool, and surveys. We included experimental, quasi-experimental, and observational studies (including case-control, cohort, and cross-sectional design) and validation studies. Qualitative studies were not included, because fundamental differences in study design, sampling, evaluation instruments, and analysis preclude summative comparison to other study types.^{24,40} Meta-analyses and systematic reviews were also excluded.

The 210 studies were randomly assigned to pairs of researchers who used the MERSQI to independently rate the studies. Individual raters' scores were used to determine interrater reliability, and then disagreements between raters were resolved by consensus to determine final scores. Raters were blinded to funding data and expert global quality ratings; however, they were not blinded to study authors and journals. Each study was scored at the highest possible level for each of the 10 MERSQI items. For example, if a study reported more than 1 outcome, the rating for the outcome that yielded the highest score was recorded, regardless of whether this outcome was a primary or secondary outcome. The criteria for scoring validity of evaluation instruments were the same whether a new or established instrument was used; however, to receive full credit, studies using established instruments were required to reference the instrument and indicate the type of validity evidence established for the instrument. To examine intrarater reliability, each study was rescored a second time by the same rater between 3 and 5 months after the first rating.

Study Funding and Author Experience

In January 2004, we conducted a cross-sectional survey of the first authors of the medical education research studies in the current study sample.¹⁵ The survey assessed author experience (including self-report of number of previous publications, academic rank, advanced degrees, and fellowship training of the first author), amount of funding obtained for the study, and an estimate of resources used for the study. Details of the study cost estimation are provided elsewhere¹⁵; in brief, it was calculated by multiplying the authors' percentage of effort dedicated to the study by the national median salary for each author, according to specialty and academic rank, and then adding the costs of resources used.¹⁵ Responses were received from authors of 243 of 290 studies (84%) in the initial survey. We used the survey data from 210 studies in this sample (after excluding 33 qualitative studies) to identify associations between MERSQI scores and study funding, study costs, and author experience.

Data Analysis

Bivariate and multivariate linear regression were performed to identify factors associated with study quality. The primary outcome was the total MERSQI score, calculated as the percentage of total achievable points, standardized to a denominator of 18 to account for "not applicable" responses. Variables included amount of funding obtained for the study (US dollars, grouped as <\$20 000 vs ≥\$20 000 [the median amount for funded studies]); study cost (US dollars); and first-author academic rank (student, resident, or fellow; instructor; assistant professor; associate professor; or professor), fellowship training (yes or no), advanced degrees (1 advanced degree vs more than 1 advanced degree, eg, MD and PhD), overall number of previous publications, and number of previous medical education publications. Stepwise forward selection was applied to model building. Variables were added to the multivariate model according to level of significance ($P < .10$) in bivari-

ate analysis. Model variables were examined for evidence of collinearity and interactions. With a sample size of 210 studies, a multivariate linear regression model with $\alpha = .05$ and 5 covariates was estimated to have 90% power to detect an R^2 of 0.08, or greater than 99% power to detect an R^2 of 0.15.

Preplanned subgroup analyses were conducted with χ^2 tests to examine associations between amount of funding (grouped a priori as <\$20 000 vs ≥\$20 000 [median funding] and <\$100 000 vs ≥\$100 000 [top quartile of funding]) and individual MERSQI items. For χ^2 analyses, MERSQI items were dichotomized a priori as study design (single-group cross-sectional and pretest and posttest vs 2-group with or without randomization); institutions (single vs 2 or more); response rate (<75% vs ≥75%); type of data (assessment by study participant vs objective measurement); internal structure, content, and relationships to other variables (not reported vs reported); appropriateness of data analysis (inappropriate vs appropriate); complexity of analysis (descriptive vs beyond descriptive); and outcome (satisfaction and knowledge or skills vs behaviors and health care outcome). Similar analyses using the full scales for MERSQI items showed the same results and are not reported here.

For all analyses, a 2-tailed $P < .05$ was considered statistically significant. Data were analyzed with SAS version 9.1 (SAS Institute Inc, Cary, North Carolina).

RESULTS

Study Quality

Total MERSQI scores among the 210 studies ranged from 5 to 16, with a mean (SD) of 9.95 (2.34). Mean domain scores were highest for data analysis (2.58), type of data (1.91), and sampling (1.90) domains; they were lowest for validity evidence (0.69) and study design (1.28) (Table 1).

Two-thirds of studies used single-group cross-sectional or single-group posttest-only designs (Table 1). Less than one fifth (17.7%) of studies included a

Table 2. Interrater and Intrarater Reliability of MERSQI Scores

MERSQI Item	Intraclass Correlation Coefficient (95% CI)	
	Interrater Reliability ^a	Intrarater Reliability ^b
Study design	0.86 (0.83-0.89)	0.998 (0.996-0.999)
Institutions	0.98 (0.97-0.99)	1.00
Response rate	0.89 (0.87-0.91)	0.996 (0.994-0.997)
Type of data	0.90 (0.88-0.92)	0.995 (0.992-0.997)
Validity		
Internal structure	0.91 (0.89-0.93)	0.97 (0.96-0.98)
Content	0.89 (0.87-0.92)	0.95 (0.94-0.96)
Relationships to variables	0.72 (0.65-0.78)	0.96 (0.94-0.97)
Appropriateness of analysis	0.94 (0.93-0.95)	0.92 (0.89-0.95)
Sophistication of analysis	0.86 (0.83-0.89)	0.97 (0.96-0.98)
Outcome	0.83 (0.79-0.86)	0.78 (0.70-0.85)

Abbreviations: CI, confidence interval; MERSQI, medical education research study quality instrument.

^aCalculated from 170 studies after 40 studies with "not applicable" ratings for 1 or more items were removed.

^bCalculated from 210 studies.

comparison group and 2.9% were randomized. Approximately one-third of studies were multi-institutional and 45.7% included objective data. Few studies measured a behavioral (29.5%) or health care (2.4%) outcome.

An example of a study that received a very high total MERSQI score (15.5) is a randomized controlled trial of a "residents-as-teachers curriculum" that assessed residents' teaching skills by using an objective structured teaching examination with high reliability, content validity, and predictive validity.⁴¹ This study received the highest possible scores for study design (randomized controlled experiment), response rate (all participants completed the evaluation component of the study), type of data (objective observation-based assessment by trained and blinded raters), validity of evaluation instruments (internal structure, content, and relationships to other variables all established), and data analysis (comparative analyses appropriately conducted). However, this study was conducted at a single university and the outcome was skills (residents' teaching skills); therefore, it did not receive maximum possible scores for "institutions" and "outcome" items. In contrast, many of the studies that received the lowest MERSQI scores were cross-sectional studies with low response rates, measuring opinions or perceptions and lacking validity evidence for evaluation instruments.

MERSQI Validity and Reliability

Principal Components Analysis and Reliability. Principal components analysis revealed 5 factors with eigenvalues greater than 1. This 5-factor model accounted for 71% of the total variance among the variables. The first factor, describing "validity evidence," was composed of internal structure, content, and relations to other variables items (Cronbach $\alpha=0.92$). The second factor, representing "method and data characteristics," included 4 items: study design, type of data, complexity of data analysis, and outcomes (Cronbach $\alpha=0.57$). Factor 3 included "institutions" and "appropriateness of data analysis." Factor 4 contained a single item, "response rate." Factor 5 was considered insignificant because it contained only 2 items (institutions and outcomes) that loaded more heavily on other factors. Cronbach α for the overall MERSQI (all 10 items) was 0.6.

Interrater reliability (ICC range, 0.72-0.98) and intrarater reliability (ICC range, 0.78-0.998) for all MERSQI items was excellent (TABLE 2).

Criterion Validity. Total MERSQI scores were highly correlated with the median quality rating of the 2 independent experts ($\rho=0.73$; 95% confidence interval [CI], 0.56-0.84; $P<.001$). Agreement between the 2 experts was excellent (ICC, 0.80; 95% CI, 0.49-0.85). MERSQI scores were also significantly associated with

3-year citation rate (0.8 increase in score per 10 citations; 95% CI, 0.03-1.30; $P=.003$) and journal impact factor (1.0 increase in score per 6-unit increase in impact factor; 95% CI, 0.34-1.56; $P=.003$).

Associations Among Study Quality, Funding, and Author Experience

Of the 210 studies, 149 (71%) did not have any funding, 30 (14%) had between \$1 and \$19 999, and 31 (15%) had \$20 000 or more. Among funded studies, the median amount of funding obtained was \$20 000 (interquartile range [IQR], \$5000-\$100 000). The median cost of studies was \$23 179 (IQR, \$9892-\$50 308). First authors had a mean (SD) of 23.5 (29.0) previous publications overall and 8.2 (13.9) previous medical education publications. Thirty-three (15.7%) first authors were students, residents, or fellows, 14 (6.7%) were instructors, 67 (31.9%) were assistant professors, 56 (26.7%) were associate professors, 28 (13.3%) were professors, and 12 (5.7%) were not in academics or reported that academic rank was "not applicable." Ninety-two (43.8%) authors had completed fellowship training, and 61 (29%) authors had more than 1 advanced degree.

In bivariate analysis, attainment of \$20 000 or more in funding was significantly associated with an increase in total MERSQI score of 1.29 (95% CI, 0.40-2.17; $P=.005$). Studies with higher costs also received higher quality scores (0.36 increase in score per \$100 000; 95% CI, 0.10-0.62; $P=.007$) (TABLE 3).

The level of experience of the first author was also associated with study quality. Higher MERSQI scores were found in studies conducted by first authors with higher numbers of overall previous peer-reviewed publications (0.38 increase in score per 20 publications; 95% CI, 0.01-0.75; $P=.048$) and higher numbers of previous peer-reviewed medical education publications (1.46 increase in score per 20 publications; 95% CI, 0.43-2.50; $P=.006$). There was no association between MERSQI score and the first author's fellowship training, possession of more

Table 3. Association of Funding and Author Publication Record With Total MERSQI Scores (n = 210)

Variable	Bivariate Analysis		Multivariate Analysis	
	Coefficient (95% CI) ^a	P Value	Coefficient (95% CI) ^a	P Value
Study funding ≥\$20 000	1.29 (0.40 to 2.17)	.005	0.95 (0.22 to 1.86)	.045
Study cost, per \$100 000 ^b	0.36 (0.10 to 0.62)	.007	0.20 (−0.08 to 0.47)	.15
Total previous peer-reviewed publications by first author, per 20	0.38 (0.01 to 0.75)	.048	0.19 (−0.21 to 0.58)	.34
Medical education previous peer-reviewed publications by first author, per 20	1.46 (0.43 to 2.50)	.006	1.07 (0.15 to 2.23)	.047
Academic rank of first author ^c	0.05 (−0.22 to 0.31)	.74		
Fellowship training of first author	−0.27 (−0.92 to 0.37)	.39		
More than 1 advanced degree for first author	0.31 (−0.40 to 1.02)	.38		

Abbreviations: CI, confidence interval; MERSQI, medical education research study quality instrument.

^aEstimated effect of predictor variable on total MERSQI score.

^bStudy cost was calculated by multiplying the authors' percentage of effort dedicated to the study by the national median salary for each author, according to specialty and academic rank, and then adding the costs of resources used.

^cStudent, resident, or fellow; instructor; assistant professor; associate professor; or professor. Each academic rank category has a value of 1. Calculated for 198 first authors after excluding 12 "not applicable" responses.

than 1 advanced academic degree (eg, MD and PhD), or academic rank (Table 3).

Variables significantly associated with MERSQI scores in bivariate analysis at $P < .10$ were included in the multivariate model (Table 3). After multivariate adjustment, attainment of \$20 000 or more in funding was independently associated with an increase of 0.95 in MERSQI score (95% CI, 0.22-1.86; $P = .045$). In a preplanned subgroup analysis of the 31 studies that received \$20 000 or more, every additional \$20 000 in obtained funding beyond the initial \$20 000 was associated with an increase in the MERSQI score of 0.12 (95% CI, 0.06-0.18; $P < .001$) independent of all covariates in the multivariate model. The number of previous peer-reviewed medical education publications by the first author was also independently associated with total MERSQI scores (1.07 increase per 20 publications; 95% CI, 0.15-2.23; $P = .047$).

Eighteen of 31 studies (58.1%) with at least \$20 000 in funding were multi-institutional compared with 57 of 179 studies (31.8%) with less than \$20 000 in funding (difference, 26.3%; 95% CI, 7.6%-45.0%; $P = .005$, χ^2 test; odds ratio, 2.96; 95% CI, 1.36-6.46). Seven of 18 studies (38.9%) with at least \$100 000 in funding used a 2-group controlled or a randomized controlled design compared with 30 (15.6%) of 192 studies with less than \$100 000 in

funding (difference, 23.3%; 95% CI, 0.2%-46.4%; $P = .01$, χ^2 test; odds ratio, 3.43; 95% CI, 1.23-9.57). No associations were found between amount of funding and the other MERSQI items.

COMMENT

Inadequate funding for medical education research is often cited as a reason for methodological shortcomings in published studies, yet to our knowledge the association between funding and education research quality has not been previously established. Our results show a significant association between funding and study quality (as measured by the MERSQI), providing evidence to support the call to increase funding for medical education research.

In a recent report on the advancement of scientific research in education, the National Academy of Sciences recommended that to promote research quality, education journals and federal funding agencies should identify reliable and valid metrics for scoring the quality of medical education research.⁴² The MERSQI is an example of such a metric. Strong evidence for MERSQI score validity includes content evidence based on expert consensus and authoritative literature supporting instrument items; internal structure evidence based on multiple, interpretable factors and excellent interrater and intrarater reliability; and

criterion validity evidence demonstrated by significant associations between MERSQI scores and quality ratings by independent experts, citation rate, and journal impact factor. The MERSQI may be a useful tool for assessing the quality of the medical education literature and may be a guide for investigators who wish to conduct high-quality medical education research.

The overall quality of studies in this sample was modest. Studies scored highest in the data analysis domain, likely because major analytic flaws are corrected by reviewers before publication or studies with fatal flaws are rejected.⁴³ MERSQI scores were lowest in the validity domain, which is consistent with previous observations that studies infrequently explore validity²⁸ and many categories of validity evidence are underreported in medical education studies.³⁰ Few studies in this sample measured patient or health care outcomes, highlighting the need to advance outcomes research in medical education.^{10,13}

Our results suggest that attainment of \$20 000 or more in funding is associated with higher-quality medical education research. In the absence of previous studies examining associations between funding and study quality in medical education, it is difficult to interpret this finding in the context of previous work. However, there are comparators in clinical research. Several

studies have shown an association between attainment of funding and the methodological quality of randomized clinical trials,⁴⁴⁻⁴⁷ whereas other studies evaluating clinical trials⁴⁸ and other study designs^{49,50} have found no association. Lee et al⁵⁰ found that clinical research studies disclosing a funding source achieved higher quality scores, but this association was not statistically significant, possibly because published disclosures of funding are incomplete.^{48,51} Surveying authors, as was done in this study, may be an alternative to obtain more accurate funding data. One systematic review concluded that the methodological quality of industry-funded studies was superior to that of studies funded by other sources,⁵² but this review did not examine comparisons with unfunded studies.

Studies with greater amounts of funding support were more likely to be multi-institutional and use controlled designs. These findings are important, given that medical education research is frequently criticized for lack of generalizability and rigor.^{1,3,5} Experts have responded by encouraging the development of multi-institutional research networks^{3,42,53} and greater use of stronger study designs.⁴² Although educators continue to debate the role of controlled trials in medical education research, particularly randomized controlled trials,^{54,55} use of an appropriate control group and random assignment may allow for causal inferences when trials are rigorously conducted in the proper context.⁵⁶

Although we found a significant association between study funding and quality, the causal direction of the findings cannot be determined from these data. Although we theorize that funding allows researchers to conduct more rigorous studies, it is conceivable that the process of applying for funding, rather than the funding itself, is responsible for the observed associations. Studies examining the effect of grant application processes on research quality in medical education may further our understanding of this relationship.

This analysis was limited to quantitative original medical education research. Qualitative studies, which comprised 13% of the initial sample, were excluded because fundamental differences in study design, sampling, and analysis precluded comparison using the MERSQI. However, this exclusion does not imply a devaluation of qualitative methods. On the contrary, qualitative methods are vital to the advancement of medical education research and for certain research endeavors may be more appropriate than quantitative methods.⁵⁷ Future studies using qualitative, quantitative, and mixed approaches are needed to enhance our understanding of relationships between funding and quality in medical education research.

This study has several limitations. First, although MERSQI items are based on published evidence, the definitions and cut points for some items, such as response rate, were based on expert consensus only. Second, although the MERSQI is intended to reflect methodological quality rather than quality of reporting, the evaluation of quality was performed with published reports. Limitations in reporting of medical education research^{43,58} and length restrictions imposed by journals may have affected the scores. Third, although the study raters were blinded to funding data and expert assessments, they were not blinded to study authors and journals. Fourth, the MERSQI does not encompass all aspects of study quality. In particular, the quality of the conceptual framework⁵⁹ and the importance of the research question²⁰ were not included, because the quality of these elements is subjective (ie, influenced by individual perceptions and preferences), variable over time, and dependent on authors' abilities to describe them in articles, potentially reflecting writing ability rather than research quality. Similarly, we did not evaluate the extent to which the studies influenced medical education theory or practice. Given recent discussions surrounding what constitutes meaningful research,^{60,61} this may merit further study. Fifth, journal impact factors and study citation rates

are imperfect proxies for study quality.⁶²⁻⁶⁵ We attempted to decrease potential citation rate inflation by excluding self-citations. Finally, data on funding, study costs, and author experience relied on self-report by first authors.

These limitations notwithstanding, we have described a reliable and valid means for assessing the quality of evidence in medical education. The MERSQI may be a useful tool for educators, reviewers, and journal editors to assess the quality of medical education research. In addition, we have demonstrated an association between the amount of funding obtained for medical education research and the methodological quality of the corresponding studies.

This study supports the need for greater funding for medical education research. The Institute of Medicine has recommended that Congress create a fund to provide grants for educational innovations.⁵³ Others have suggested developing a national center to support education research⁶⁶ and collaborative networks for sharing of resources and expertise.³ These visions have not been realized. Policy reform that increases funding support may promote high-quality medical education research.

Author Contributions: Dr Reed had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Reed, Cook, Beckman, Levine, Kern, Wright.

Acquisition of data: Reed, Cook, Beckman.

Analysis and interpretation of data: Reed, Cook, Beckman.

Drafting of the manuscript: Reed, Beckman, Levine, Kern, Wright.

Critical revision of the manuscript for important intellectual content: Cook, Beckman, Levine, Kern, Wright.

Statistical analysis: Reed, Beckman.

Obtained funding: Reed, Cook.

Administrative, technical, or material support: Cook, Beckman.

Study supervision: Cook, Levine, Kern, Wright.

Financial Disclosures: None reported.

Funding/Support: Dr Wright received support as an Arnold P. Gold Foundation Associate Professor of Medicine. Dr Wright is a Miller-Coulson Scholar; the support is associated with the Johns Hopkins Center for Innovative Medicine. Drs Reed and Wright received support from the Society of General Internal Medicine Research and Education Mentorship Award.

Role of the Sponsor: No funding organization or sponsor had any role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; nor in the preparation, review, or approval of the manuscript.

Additional Contributions: We would like to thank Larry Gruppen, PhD, University of Michigan Medical School, and Addeane Caellegh, MA, University of Virginia School of Medicine, for providing expert quality ratings, and Stephen Cha, MS, Mayo Clinic College of Medicine for assistance with data analysis. These colleagues received compensation for their contributions.

REFERENCES

- 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.
- Wartman SA. Research in medical education: the challenge for the next decade. *Acad Med.* 1994;69(8):608-614.
- Carney PA, Nierenberg DW, Pipas CF, Brooks WB, Stukel TA, Keller AM. Educational epidemiology: applying population-based design and analytic approaches to study medical education. *JAMA.* 2004;292(9):1044-1050.
- Shea JA, Arnold L, Mann KVA. RIME perspective on the quality and relevance of current and future medical education research. *Acad Med.* 2004;79(10):931-938.
- Lurie SJ. Raising the passing grade for studies of medical education. *JAMA.* 2003;290(9):1210-1212.
- Carline JD. Funding medical education research: opportunities and issues. *Acad Med.* 2004;79(10):918-924.
- Regehr G. Trends in medical education research. *Acad Med.* 2004;79(10):939-947.
- Lim JK, Golub RM. Graduate medical education research in the 21st century and JAMA on call. *JAMA.* 2004;292(23):2913-2915.
- Golub RM. Medical education theme issue 2007: call for papers. *JAMA.* 2006;296:2857.
- Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med.* 2004;79(10):955-960.
- Whitcomb ME. Research in medical education: what do we know about the link between what doctors are taught and what they do? *Acad Med.* 2002;77(11):1067-1068.
- Whitcomb ME. Using clinical outcomes data to reform medical education. *Acad Med.* 2005;80(2):117.
- Prystowsky JB, Bordage G. An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. *Med Educ.* 2001;35(4):331-336.
- Murray E, Gruppen L, Catton P, Hays R, Woollicroft JO. The accountability of clinical education: its definition and assessment. *Med Educ.* 2000;34(10):871-879.
- Reed DA, Kern DE, Levine RB, Wright SM. Costs and funding for published medical education research. *JAMA.* 2005;294(9):1052-1057.
- Harden RM, Grant J, Buckley G, Hart IR. BEME guide No. 1: best evidence medical education. *Med Teach.* 1999;21(6):553-562.
- Morrison JM, Sullivan F, Murray E, Jolly B. Evidence-based education: development of an instrument to critically appraise reports of educational interventions. *Med Educ.* 1999;33(12):890-893.
- Education Group for Guidelines on Evaluation. Guidelines for evaluating papers on educational interventions. *BMJ.* 1999;318(7193):1265-1267.
- Green ML. Identifying, appraising, and implementing medical education curricula: a guide for medical educators. *Ann Intern Med.* 2001;135(10):889-896.
- Bordage G, Dawson B. Experimental study design and grant writing in eight steps and 28 questions. *Med Educ.* 2003;37(4):376-385.
- Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA.* 2001;285(15):1987-1991.
- Des Jarlais DC, Lyles C, Crepaz N; TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health.* 2004;94(3):361-366.
- West S, King V, Carey TS, et al. *Systems to Rate the Strength of Scientific Evidence.* Rockville, MD: Agency for Healthcare Research and Quality; 2002. Evidence Report/Technology Assessment No. 47. AHRQ publication 02-E016.
- Stacy R, Spencer J. Assessing the evidence in qualitative medical education research. *Med Educ.* 2000;34(7):498-500.
- Harris RP, Helfand M, Woolf SH, et al; Methods Word Group, Third U.S. Preventive Services Task Force. Current methods of the U.S. Preventive Services Task Force: a review of the process. *Am J Prev Med.* 2001;20(3)(suppl):21-35.
- Belfield C, Thomas H, Bullock A, Eynon R, Wall D. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach.* 2001;23(2):164-170.
- Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830-837.
- Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 2006;119(2):166.e7-16.
- American Educational Research Association; American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association; 1999:9-24.
- Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med.* 2005;20(12):1159-1164.
- JAMA instructions for authors. <http://jama.ama-assn.org/misc/ifora.dtl#DataAccessandResponsibility>. Accessed February 5, 2007.
- Kirkpatrick DI. Evaluation of training. In: Craig R, Bittel I, eds. *Training and Development Handbook.* New York, NY: McGraw-Hill; 1976:18-2.
- DeVellis RF. *Scale Development: Theory and Applications.* 2nd ed. London, England: Sage Publications; 2003:113-115.
- Gorsuch RL. *Factor Analysis.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1983:208-210.
- Landis JR, Koch GG. The measure of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
- Fleiss JL, Leven B, Paik MC. *Statistical Methods for Rates and Proportions.* 3rd ed. Hoboken, NJ: John Wiley & Sons; 2003:604.
- SCOPUS. <http://www.info.scopus.com>. Published 2007. Accessed January 8, 2007.
- McVeigh ME. Journal self-citation in the journal citation reports. Institute for Scientific Information Web site. <http://scientific.thomson.com/free/essays/journalcitationreports/selfcitation2002>. Published 2002. Accessed January 8, 2007.
- Institute for Scientific Reports. The ISI impact factor. <http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor>. Accessed January 8, 2007.
- Côté L, Turgeon J. Appraising qualitative research articles in medicine and medical education. *Med Teach.* 2005;27(1):71-75.
- Morrison EH, Rucker L, Boker JR, et al. A pilot randomized, controlled trial of a longitudinal residents-as-teachers curriculum. *Acad Med.* 2003;78(7):722-729.
- Towne L, Wise LL, Winters TM, eds; Committee on Research in Education, National Research Council. *Advancing Scientific Research in Education.* Washington, DC: National Academies Press; 2004:3.
- Bordage G. Reasons reviewers reject and accept manuscripts: the strengths and weaknesses in medical education reports. *Acad Med.* 2001;76(9):889-896.
- Als-Nielsen B, Chen W, Glud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA.* 2003;290(7):921-928.
- Djulbegovic B, Lacey M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet.* 2000;356(9230):635-638.
- Kjaergard LL, Glud C. Funding, disease area, and internal validity of hepatobiliary randomized clinical trials. *Am J Gastroenterol.* 2002;97(11):2708-2713.
- Kjaergard LL, Nikolova D, Glud C. Randomized clinical trials in Hepatology: predictors of quality. *Hepatology.* 1999;30(5):1134-1138.
- Clifford TJ, Barrowman NJ, Moher D. Funding source, trial outcome and reporting quality: are they related? results of a pilot study. *BMC Health Serv Res.* 2002;2(1):18.
- Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med.* 1996;124(5):485-489.
- Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. *JAMA.* 2002;287(21):2805-2808.
- McCrary SV, Anderson CB, Jakovljevic J, et al. A national survey of policies on disclosure of conflicts of interest in biomedical research. *N Engl J Med.* 2000;343(22):1621-1626.
- Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ.* 2003;326(7400):1167-1170.
- Institute of Medicine. *Academic Health Centers: Leading Change in the 21st Century.* Washington, DC: National Academies Press; 2003:3.
- Norman G. RCT=results confounded and trivial: the perils of grand educational experiments. *Med Educ.* 2003;37(7):582-584.
- Prideaux D. Researching the outcomes of educational interventions: a matter of design. *BMJ.* 2002;324(7330):126-127.
- Torgerson CJ. Researching outcomes of educational interventions. *BMJ.* 2002;324(7346):1155-1156.
- Irby DM. Shifting paradigms of research in medical education. *Acad Med.* 1990;65(10):622-623.
- Cook DA, Beckman TJ, Bordage G. Quality of reporting experimental studies in medical education: a systematic review [published online July 4, 2007]. *Med Educ.* doi:10.1111/j.1365-2923.2007.02777.x.
- Prideaux D, Blich J. Research in medical education: asking the right questions. *Med Educ.* 2002;36(12):1114-1115.
- Chen FM, Burstin H, Huntington J. The importance of clinical outcomes in medical education research. *Med Educ.* 2005;39(4):350-351.
- Albert M, Hodges B, Regehr G. Research in medical education: balancing service and science. *Adv Health Sci Educ Theory Pract.* 2007;12(1):103-115.
- Semenzato G, Rizzato G, Agostini C. Impact factor as measure of scientific quality. *Am J Respir Crit Care Med.* 2004;169(9):1070-1071.
- Linardi PM, Coelho PM, Costa HM. The "impact factor" as a criterion for the quality of scientific production is a relative, not absolute, measure. *Braz J Med Biol Res.* 1996;29(5):555-561.
- Walter G, Bloch S, Hunt G, Fisher K. Counting on citations: a flawed way to measure quality. *Med J Aust.* 2003;178(6):280-281.
- Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA.* 2005;293(19):2362-2366.
- Wartman SA. Revisiting the ideas of a national center for health professions education research. *Acad Med.* 2004;79(10):910-917.