Can Moral Reasoning Predict General Surgery Residents' Clinical Competence?

Jack Contessa, PhD, Tassos Kyriakides, PhD,† and Donald Kim, MD**

*Department of Surgery, Hospital of Saint Raphael, New Haven, Connecticut, and †Department of Internal Medicine, Yale School of Medicine, VA Cooperative Studies Program Coordinating Center (CSPCC), West Haven, Connecticut

BACKGROUND: When selecting residents for residency programs, there has been little success using standard academic criteria to predict their clinical performance. However, numerous studies in the past have found linkages between the nonstandard variable moral reasoning as measured by the instrument Defining Issues Test (Version 2; DIT-2) and clinical decision making with higher levels of moral reasoning linked to better clinical performance. This study sought to determine whether this important linkage exists in surgery residents.

METHODS: The DIT-2 was administered to 20 surgical residents to assess level of moral reasoning, and data were collected on all end-of-rotation evaluations of residents by faculty to measure resident clinical competence. Candidate rank list data were examined from the past 5 years to determine linkages between moral development as measured by the DIT-2 and clinical competence. Correlation analyses, univariate regression, and stepwise multivariate regression were carried out to assess the relationships of moral reasoning as measured by the DIT-2 with other study variables.

RESULTS: Analysis of data indicates a low correlation between DIT-2 scores and clinical competence. Rank list order showed a slight correlation with resident DIT-2 scores. Rank position was weakly correlated with individual competencies and aggregated scores for all competencies as measured by faculty evaluations. Rank position coupled with DIT-2 scores were predictive of 4 of the 6 ACGME competencies and predictive of clinical competence as measured by aggregated scores for all competencies.

CONCLUSIONS: No linkages emerged between DIT-2 scores and clinical competence. This reasons for this may include the size of the population studied and the assumption that end-of-rotation evaluations measure resident clinical competence. There were significant relationships demonstrated when DIT-2 scores were linked with resident ranking. Future research should be continued in this area but with more rigorous instrumentation to measure clinical performance and a larger sample size. (J Surg 69:17-22. © 2012 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: Defining Issues Test, (DIT-2), moral development, clinical competence, ACGME competencies, ranking candidates

COMPETENCIES: Medical Knowledge, Patient Care, Interpersonal and Communication Skills, Professionalism, Practice Based Learning and Improvement, Systems Based Practice

INTRODUCTION

General surgery residency programs are among the most highly competitive and sought after specialties with more individuals seeking positions than exist. Although there are numerous applicants, the process of selecting residents who will demonstrate a high degree of clinical competence is a challenge and one of the most critical tasks for the teaching faculty of a residency program. Most programs analyze applicant files on a subjective basis by members of a screening committee, who then reach consensus for ranking candidates.[1] This process frequently relies on criteria, such as medical school grades, election to AOA, test scores, and letters of recommendation, all of which represent certain biases and limitations.[2] The Medical College Admissions Test, the National Board of Medical Examiners, and the United States Medical Licensing Examination are used to select candidates that are expected to succeed with the academic challenges they face during a 5-year program and eventually in professional practice.

However, many studies using these criteria have demonstrated conflicting results or have shown little or no power to predict resident performance. For example, a study by Dirschl et al. of 58 orthopedic residents demonstrated that none of their predictor variables, which included scores on the United States Medical Licensing Examination (USMLE), Part I, had a significant correlation with Orthopaedic In-Training Examination

Correspondence: Inquiries to Jack Contessa, PhD, Department of Surgery, Hospital of Saint Raphael, 1450 Chapel Street New Haven, CT 06511; fax: (203) 867-5248; e-mail: jcontessa@srhs.org

or American Board of Orthopaedic Surgery Examination scores. Also, faculty rating of the 3 letters of recommendation had a poor correlation with all the outcomes variables, including faculty rating of overall resident performance.[1] Yet, in a study of 26 general surgery residents from 2003 to 2006, Brothers and Wetherholt[3] described a high correlation between USMLE and both American Board of Surgery In-Training Examination and American Board of Surgery Qualifying Examination scores. Additionally, the quality of reference letters assessed during faculty interviews correlated most strongly with subsequent clinical performance. A prospective study of 277 pediatric residents done between 1992 and 1999 demonstrated a significant correlation between a global assessment by an intern selection committee (ISC) and clinical performance at 24 and 36 months of training. ISC ratings also correlated significantly with in-service examination scores in the first 3 years of training.[4] Although these studies have produced conflicting evidence for predictors of resident performance using information found in the ERAS application and through faculty interviews, there is an important variable that has been studied extensively during the past several years that has been found consistently to be associated with exemplary clinical performance by residents.

The variable moral reasoning as measured using the Defining Issues Test (Version 2; DIT-2) has been cited in numerous studies in medicine, dentistry, nursing, and pharmacy, reporting a positive correlation between this variable and resident performance. One of the earliest studies to examine this variable was done by Sheehan and colleagues in 1980 with 244 pediatric residents. They found that residents as assessed by their supervisors for numerous criteria were found to be highly correlated with their levels of moral reasoning as measured by the DIT-2. A subsequent study by Sheehan et al. with 39 family practice residents confirmed the relationship between clinical performance and moral reasoning.[5]

There have been many other studies in the healthcare field that confirmed this relationship. In his study regarding renewing professionalism in dental education, Masella cited that in a review by Bebeau in 2002 of 33 studies from dentistry, medicine, veterinary medicine, nursing, and law, a significant relationship was demonstrated between moral reasoning and clinical performance in those studies involving dental and nursing students.[6] Self et al. confirmed that recent studies reported a significant relationship between moral reasoning and clinical performance in medical students, residents, as well as dental and nursing students. They recommended additional research be conducted to confirm this relationship before using it in the resident selection process.[5] In her study of physical therapy students using moral reasoning as a predictor of clinical practice, Sisola found that when moral reasoning was compared with 8 admission variables as predictors of clinical performance, that moral reasoning as measured by the DIT-2 was the only variable to demonstrate a significant relationship with clinical performance.[7]

Moral reasoning is based on Lawrence Kohlberg's Theory of Moral Development (see Table 1), which describes a model for making ethical judgments based on a series of stages that individuals advance along, each with its own characteristics. Individuals pass upward through these stages like steps on a staircase; theoretically stages cannot be skipped. Residents at different levels of moral development resolve social or moral dilemmas in different ways, which have implications for the patient care decisions they make. For example, at one of the earliest stages of moral development, individuals focus on self regardless of how that behavior affects others. Residents at this level may not provide high-quality care if the "costs" of doing so are seen as outweighing the benefits to them. At the highest stage of moral development, residents would most likely provide high-quality care despite conflicting situations. This means that residents reasoning at a higher moral level and using a more advanced conceptual framework are more likely to deliver better quality patient care demonstrating higher clinical competence than those reasoning at a lower level. Although higher staged individuals are not necessarily more intelligent than lower staged individuals, they have a better framework to guide their decision making.[3] In at least one study referenced by Sisola, Sheehan et al. concluded that a high level of moral rea-

TABLE 1. Six Stages in the Concept of Moral Development

Level I

Stage 1: Obedience and Punishment
- Do what you are told.
- Obeying the rules, which are viewed as absolute and unchangeable, is important to avoid punishment.

Stage 2: Self-Interest and Exchange
- Fair exchange or fair deals; "you scratch my back and I'll scratch yours."
- Follow rules when it serves one's needs.

Level II

Stage 3: Interpersonal Relationships and Conformity
- Seek approval, conformity, helping others; actions judged by intentions.
- Living up to social expectations and roles.

Stage 4: Law and Order
- Orientation transcends immediate group and considers society at large; no person is above the law.
- Follow rules, respect authority, obey/maintain social order.

Level III

Stage 5: Social Contract
- Individual action judged by standards (laws, rules) that have been agreed upon by the entire society.
- Laws, rules not rigid and can be changed for the greater good.

Stage 6: Universal Ethical Principles
- Action is directed not only by existing social standards, but also based on abstract reasoning using universal ethical principles.
- Principles of justice, reciprocity (the golden rule), equality, human dignity even if it means disobeying unjust laws and rules.

Source: Adapted from Crain WC. *Theories of Development*. Englewood Cliffs, NJ: Prentice-Hall; 1985, pp. 118-136.

soning "virtually excludes the possibility of poor performance" and "the highest level of clinical performance is rarely achieved by those at the lowest level of moral thought."[7] Surgical residency programs have been characterized by long hours, high stress, and a strong emphasis on individual responsibility and accountability. Recent changes have impacted the working conditions under which surgical residents train, including increasing bureaucracy, a trend toward ambulatory care, reduced in-hospital time for patients, and emphasis on inpatient care only for critically ill patients. It seems, therefore, that it is a vitally important and valid consideration to select applicants for surgical residency programs who have a high conceptual level of moral reasoning and are equipped to make high-quality patient care decisions.

As cited, numerous studies in the healthcare field have examined the linkages between moral reasoning and clinical competence. The empiric evidence reported thus far confirms this relationship. However, to date no single study has explored this relationship in a general surgery residency program. This study seeks to fill that gap and represents another step in this important area of research.

The purpose of this study was to investigate whether there was a correlation between moral reasoning and clinical performance in general surgery residents. Knowing whether this relationship exists for general surgery residents could help address the following questions:

1. Retrospectively, is there a relationship between moral development as measured by resident scores on the DIT-2 and their clinical competence as measured by faculty evaluations?
2. Retrospectively, is there a relationship between the past 5 years' ranking of residents who matched our program and their scores on the DIT-2?
3. Retrospectively, is there relationship between the past 5 years' ranking of residents who matched our program and their clinical competence?
4. What role could an objective instrument like the DIT-2 play in selection of general surgery residents?

Design, Methodology, and Instruments

Moral reasoning. The DIT-2 is a multiple-choice, paper-and-pencil instrument to measure moral reasoning that presents 5 hypothetical dilemmas, with each one followed by 12 issues for the subject to rate and rank in terms of their importance. The DIT-2 takes about 20–30 minutes to complete, can be group-administered, and can be computer scored. Using the DIT-2 is a practical way to assess moral reasoning in contrast to the labor-intensive process originally developed by Kohlberg and his associates, which include interviews where subjects are asked about several hypothetical dilemmas by trained judges who then assign a Stage score based on responses.[8]

Confirmatory factor analysis of more than 44,000 subjects indicated that the DIT-2 items cluster around 3 moral arguments appealing to:

- Personal interests (Personal Interest; corresponding to Kohlberg's Stages 2 and 3)
- Maintaining social laws and norms (Maintaining Norms; corresponding to Kohlberg's Stage 4)
- Moral ideals and/or theoretical frameworks for resolving complex moral issues (Postconventional—p-score corresponding to Kohlberg's Stages 5 and 6)

Researchers have typically reported scores using p-scores, which represent the percentage of items selected that appeal to post-conventional frameworks for making moral judgments.[9] The DIT-2 was administered as a one-time assessment, and the data reported in this study represent the p-scores.

Clinical Competence

Clinical competence was based on data from resident evaluations by faculty. End-of-rotation resident evaluations are completed online through the E*Value residency management software program. Faculty rate residents on 17 questions related to the 6 ACGME competencies (Medical Knowledge—MK, Patient Care—PC, Interpersonal and Communication Skills—IPC, Practice-based Learning and Improvement—PBLI, Professionalism—Prof, and Systems-Based Practice—SBP). Ratings range from 1 to 9, with 1 being the lowest score and 9 the highest. The total number of evaluations on file was used to aggregate data for each resident. As a result, chief residents had significantly more responses than interns. Evaluation reports were generated beginning with each resident's July 1 start date through July 1, 2010. The number of evaluations recorded for any competency for a single resident ranged from 8 (PGY1) to 330 (PGY 5). Individual resident scores were grouped by competency and averaged to yield a mean score for each competency and a grand mean for all competencies.

Statistical Procedures

Correlation analyses, univariate regression, and stepwise multivariate regression were carried out to assess the relationships of the DIT-2 variables with other study variables. As the USMLE score was recorded before the DIT-2 assessment of residents, it was used as an independent variable to predict outcomes like DIT-2 scores; the ABSITE score was used as an outcome variable whose prediction from DIT-2 scores was assessed. All analyses were performed using SAS version 9.1 (SAS Institute, Cary, NC).

RESULTS

Table 2 summarizes the univariate analysis for ACGME and DIT-2 score variables. The first question sought to determine whether there is a relationship between moral reasoning scores as measured by the Defining Issues Test (DIT-2) and clinical competence as measured by faculty evaluations of residents. Pearson product moment correlation analysis yielded relatively low relationships between the 2 variables. This was true for the

TABLE 2. Univariate Analysis of ACGME and DIT-2 Variables

Variable	n	Mean	Standard Deviation	Minimum	Maximum
MK mean	20	6.1282500	0.6422130	4.7100000	7.4800000
PC mean	20	6.3283750	0.6658178	5.2100000	7.9750000
IPC mean	20	6.3534500	0.7437725	5.0775000	8.0075000
PBLI mean	20	6.3722500	0.6938290	5.2700000	7.8400000
Prof mean	20	6.7574500	0.6536737	5.7366667	8.2420000
SBP mean	20	6.5212500	0.6225814	5.3600000	8.0000000
Grand mean	20	6.4101708	0.6577638	5.2429167	7.9240833
p-score	20	41.0225000	15.7665408	16.0000000	72.0000000

relationship between DIT-2 scores and clinical competence as measured by individual competencies (Medical Knowledge, Patient Care, etc.) and between DIT-2 scores and clinical competence as measured by aggregated scores for all competencies. Table 3 shows the correlation analyses results. Additional analysis was performed using univariate and multivariate stepwise regression. The results indicated that the DIT-2 score was not a predictor of resident clinical competence as measured by their evaluations from faculty. Table 4 provides the results of univariate and multivariate regression analyses. The second question queried whether there was a relationship between ranking of residents for the past 5 years and scores on the DIT-2. The rank list order showed a weak correlation with resident DIT-2 scores. The third question sought to discover whether a relationship existed between ranking of residents for the past 5 years and clinical competence as measured by their evaluations from faculty. Rank position was weakly correlated with individual competencies as measured by evaluations and between rank position and aggregated scores for all competencies. However, in a multivariate stepwise regression comprising rank position and DIT-2, DIT-2 scores were significant predictors of all individual measures of competency except Practice-Based Learning and Improvement and Systems Based Practice and predictive for clinical competence as measured by aggregated scores for all competencies. The final question queried the role of the DIT-2 in selection of general surgery residents. The current study found no linkages.

CONCLUSIONS

Numerous studies in the medical field have demonstrated relationships between DIT-2 scores and clinical competence. However, in this study, no linkages emerged solely between these 2 variables. There may be several reasons for this, including the size of the population being studied and the assumptions made about the clinical competence of residents based on end-of-rotation evaluation scores. The instrument used to measure resident clinical competence, end-of-rotation evaluations, may have low reliability and validity and may fall short of measuring performance. This assumption is supported in a study by Herbers et al. They concluded that when using the Clinical Evaluation Exercise (CEX) recommended by the American Board of Internal Medicine (ABIM) to assess resident clinical performance, evaluators vary substantially in their observations, the degree to which their findings are documented, and their standards for resident performance. They recommend that an accurate assessment of resident clinical competence requires multiple approaches. They state that their results strongly suggest that untrained evaluators are too variable and document too little to establish residents' clinical competence.[10]

Although a significant relationship was demonstrated when DIT-2 scores were coupled with resident ranking, the DIT-2 score alone was not a predictor of resident clinical competence. Residency programs struggle to find objective, predictive data to use for selecting candidates. Rank lists are the most commonly used tool to make this determination. Our rank list for the past 5 years showed a weak correlation with scores on the DIT-2, which measures moral reasoning and, based on previous research, is a proxy for measuring clinical competence. Additionally, this study found a similar weak correlation with resident evaluations by faculty. As a result, this study suggests that rank lists, although frequently used to make resident selections, may not be useful for predicting future residents' clinical competence. Despite the small sample size, some trends can be seen when assessing DIT-2 variables vis-à-vis other routinely collected variables in our residency program.

There are obviously limitations to this study. The DIT-2 questionnaire was administered only once to each resident, whereas the rest of the variables were collected at multiple times (and this number varied depending on the residency level). In addition, because each resident completed the DIT-2 questionnaire at different times in their residency program, inferences on transitions on DIT-2 scores as residents move from the first to latter years cannot be made. In addition, the number of assessments for other variables varied among residents and the mean values for each of these variables were used for analysis.

Recommendations for Further Study

After having completed the research and having correlated findings, conclusions, and implications, the following recommendations are made:

1. Continued research should be directed toward confirming a relationship between moral reasoning and clinical perfor-

TABLE 3. Correlation Analyses Between DTI Scores and ACGME Mean Scores

Pearson Correlation Coefficients Prob > r Under HO: Rho = 0 Number of Observations	p-Score	MK Mean	PC Mean	IPC Mean	PBLI Mean	Prof Mean	SBP Mean	Grand Mean	Rank Pos
p-Score	1.00000	−0.49338	−0.40833	−0.41793	−0.35455	−0.43955	−0.44858	−0.43384	0.03135
		0.0271	0.0739	0.0667	0.1251	0.0525	0.0473	0.0560	0.9049
	20	20	20	20	20	20	20	20	17
MK Mean	−0.49338	1.00000	0.93805	0.92879	0.90291	0.89068	0.93381	0.94959	0.35588
	0.0271		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.1609
	20	20	20	20	20	20	20	20	17
PC Mean	−0.40833	0.93805	1.00000	0.98840	0.97114	0.96450	0.97006	0.99114	0.33656
	0.0739	<0.0001		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.1865
	20	20	20	20	20	20	20	20	17
IPC Mean	−0.41793	0.92879	0.98840	1.00000	0.98119	0.96776	0.98023	0.99377	0.31384
	0.0667	<0.0001	<0.0001		<0.0001	<0.0001	<0.0001	<0.0001	0.2199
	20	20	20	20	20	20	20	20	17
PBLI Mean	−0.35455	0.90291	0.97114	0.98119	1.00000	0.97339	0.96780	0.98538	0.32586
	0.1251	<0.0001	<0.0001	<0.0001		<0.0001	<0.0001	<0.0001	0.2018
	20	20	20	20	20	20	20	20	17
Prof Mean	−0.43955	0.89068	0.96450	0.96776	0.97339	1.00000	0.95958	0.97817	0.25614
	0.0525	<0.0001	<0.0001	<0.0001	<0.0001		<0.0001	<0.0001	0.3210
	20	20	20	20	20	20	20	20	17
SBP Mean	−0.44858	0.93381	0.97006	0.98023	0.96780	0.95958	1.00000	0.98718	0.21255
	0.0473	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001		<0.0001	0.4128
	20	20	20	20	20	20	20	20	17
Grand Mean	−0.43384	0.94959	0.99114	0.99377	0.98538	0.97817	0.98718	1.00000	0.30645
	0.0560	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001		0.2316
	20	20	20	20	20	20	20	20	17
Rank Pos	0.03135	0.35588	0.33656	0.31384	0.32586	0.25614	0.21255	0.30645	1.00000
	0.9049	0.1609	0.1865	0.2199	0.2018	0.3210	0.4128	0.2316	
	17	17	17	17	17	17	17	17	17

TABLE 4. Univariate and Multivariate Regression

Regression	Outcome	Independent Variable	p-Value
Univariate	ACGME mean score	p-score	0.056
	ACGME mean score	rank position	0.2316
	MK mean	USMLE score	0.6796
	MK mean	ABSITE score	0.2714
	p-score	USMLE score	0.6041
	ABSITE score	p-score	0.1451
Multivariate Stepwise	ACGME mean score	p-score, rank	0.0334, 0.1497
	MK mean score	p-score, rank	0.0287, 0.0886
	PC mean score	p-score, rank	0.0296, 0.1175
	IPC mean score	p-score, rank	0.0311, 0.1378
	PBLI mean score	p-score, rank	0.0705, 0.1458
	Prof mean score	p-score, rank	0.0316
	SBP mean score	p-score	0.0287
	ABSITE score	USMLE score	0.0005

mance. However, resident clinical performance should be measured using multiple instruments and by trained observers.

2. Attempts should be made to replicate this study with a larger sample size, where residents entering their first year are prospectively followed until the end of their residency program; a repeated measures analyses would assist in definitely answering some of the relevant questions.
3. Studies should be conducted to determine whether objective instruments like the DIT-2 help select "better" residents, for example, comparing a year when these instruments are used to select residents with a year when they are not used to determine whether there is a difference in resident performance.

REFERENCES

1. Dirschl DR, Dahners LE, Adams GL, Crouch JH, Wilson FC. Correlating selection criteria with subsequent performance as residents. *Clin Orthop Relat Res*. 2002;399:265-271.
2. White AA. Resident selection: are we putting the cart before the horse? *Clin Orthop Relat Res*. 2002;399:255-259.
3. Brothers TE, Wetherholt S. Importance of faculty interview during resident application process. *J Surg Educ*. 2007;64:378-385.
4. Ozuah PO. Predicting residents' performance: a prospective study. *BMC Med Educ*. 2002;2:7.
5. Self DJ, Baldwin DC, Bunch W. Should moral reasoning serve as criterion for student and resident performance? *Clin Orthop Relat Res*. 2000;378:115-123.
6. Masella RS. Renewing professionalism in dental education: overcoming the market environment. *J Dent Educ*. 2007;71:205-216.
7. Sisola S. Moral reasoning as a predictor of clinical practice: the development of physical therapy students across the professional curriculum. *J Phys Therapy Educ*. 2000;14:26-34.
8. Latif D. Moral reasoning: should it serve as a criterion for student and resident selection in pharmacy? *Am J Pharm Educ*. 2001;65.
9. Bebeau M, Thoma S. *Guide for DIT-2*. Tuscaloosa, AL: Office for the Study of Ethical Development, University of Alabama; 2003.
10. Herbers J Jr., Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? *J Gen Intern Med*. 1989;4:202-208.