

Physician Ratings of Appropriate Indications for Three Procedures: Theoretical Indications vs Indications Used in Practice

ROLLA EDWARD PARK, PhD, ARLENE FINK, PhD, ROBERT H. BROOK, MD, ScD, MARK R. CHASSIN, MD, KATHERINE L. KAHN, MD, NANCY J. MERRICK, MD, JACQUELINE KOSECOFF, PhD, AND DAVID H. SOLOMON, MD

Abstract: We previously reported substantial disagreement among expert physician panelists about the appropriateness of performing six medical and surgical procedures for a large number of theoretical indications. A recently completed community-based medical records study of about 4,500 patients who had one of three procedures—coronary angiography, upper gastrointestinal endoscopy, and carotid endarterectomy—shows that many of the theoretical indications are seldom or never used in practice. However, we find that there is also substantial disagreement (5, 25, or 32 per cent

for angiography, endoscopy, or endarterectomy, respectively) about the appropriateness of indications used in actual cases if disagreement is defined by first discarding the two extreme of nine ratings, then looking for at least one rating near the bottom (1 to 3) and one near the top (7 to 9) of the 9-point scale. Patients should know that a substantial percentage of procedures are performed for indications about which expert physicians disagree. (*Am J Public Health* 1989; 79:445-447.)

Introduction

We previously reported on ratings by panels of physicians of the appropriateness of a large number of detailed indications for the use of six medical and surgical procedures.^{1,2} The panelists assigned appropriateness ratings to indications in a process that was designed to encourage but not to require agreement. We found that, even in their final ratings, the panelists disagreed about the appropriateness of substantial fractions of the indications. One possible reason for the disagreement is that disagreed-upon indications may not occur in practice; our comprehensive lists included many indications that might be used only in theory.

Our purpose in this paper is to compare the panelists' agreement and disagreement about the theoretical indications with their agreement and disagreement about the indications actually assigned to cases in a recently completed medical records study.³⁻⁵ That study determined the actual reasons for doing three of the procedures (coronary angiography, diagnostic upper gastrointestinal endoscopy, and carotid endarterectomy) in several thousand actual cases in five large areas of the United States. In this paper, we also compare the distribution by appropriateness or inappropriateness of the theoretical indications with that of the actual cases.

Methods

We have previously described in detail our two-stage modified Delphi process for obtaining appropriateness ratings.^{1,2} The process used nationally representative panels consisting of nine expert physicians, including both generalists and specialists. The ratings on a nine-point scale ranged from extremely inappropriate (1), through equivocal (5), to extremely appropriate (9). Here we adopt the same four definitions of agreement that we previously used:

A9S: All nine of the ratings fell within a single three-point region—1 to 3, 4 to 6, or 7 to 9.

A9R: All nine of the ratings fell within *any* three-point range.

A7S: After discarding one extreme high and one extreme low rating, the remaining seven ratings all fell within a single three-point region—1 to 3, 4 to 6, or 7 to 9.

A7R: After discarding one extreme high and one extreme low rating, the remaining seven ratings all fell within *any* three-point range.

We also adopt the four previously used definitions of disagreement, and add a fifth. We used the fifth definition in our initial report of the medical records study:⁴

D9S: Considering all nine ratings, at least one was a 1 and at least one was a 9.

D9R: Considering all nine ratings, at least one fell in the lowest three point region (1 to 3) and at least one fell in the highest (7 to 9).

D7S: After discarding one extreme high and one extreme low rating, at least one of the remaining seven ratings was a 1 and at least one was a 9.

D7R: After discarding one extreme high and one extreme low rating, at least one of the remaining seven ratings fell in the lowest three point region (1 to 3) and at least one fell in the highest (7 to 9).

D5R: After discarding two extreme high and two extreme low ratings, at least one of the remaining five ratings fell in the lowest three point region (1 to 3) and at least one fell in the highest (7 to 9).

Deciding what to count as disagreement is necessarily subjective. We added definition D5R and used it in reporting the results of the medical records study³ after examining patterns of individual panelist's ratings for representative indications. Many patterns that D7R would classify as disagreement, but D5R would not, seemed to us to show substantial clustering around the median rating—enough clustering so that we did not at that time want to count them as disagreement. We include all five definitions here so that readers can make their own choices.

In addition, we use our previous definitions of three appropriateness categories:²

Appropriate: a median rating of 7 to 9 without disagreement (using D7R to define disagreement for this purpose in this paper).

Address reprint requests to Rolla Edward Park, PhD, Rand Corporation, 1700 Main Street, Santa Monica, CA 90406-2138. Drs. Fink and Kosecoff are with Fink and Kosecoff, Inc; Dr. Brook is with the Systems Science Department, Rand Corp, and the Departments of Medicine and Public Health, University of California at Los Angeles; Dr. Chassin is with the Systems Science Department, Rand Corp; Dr. Kahn is with the Systems Science Department, Rand Corp and the Department of Medicine, UCLA; Dr. Merrick is with the Prospective Payment Assessment Commission; Dr. Solomon is with the Department of Medicine, UCLA. This paper, submitted to the *Journal* December 9, 1987, was revised and accepted for publication June 27, 1988.

Inappropriate: a median rating of 1 to 3 without disagreement.
Equivocal: a median rating of 4 to 6 without disagreement, or any median rating with disagreement.

During the medical records study, we abstracted hospital and office records for random samples averaging about 1,500 cases for each of the three procedures. The patient population consisted of people over age 65 covered by Medicare Part B. From the detailed abstracts, we determined which indication or indications applied to each case; if more than one indication applied, we chose the one with the highest median rating as the primary indication.

We compared the panelists' ratings of the full lists of indications that theoretically could occur, with their ratings of indications assigned to cases that actually did occur in practice. For each case, we noted the primary indication, and what category that indication fell into: agreement or disagreement; appropriate, equivocal, or inappropriate. We then calculated the per cent of cases whose primary indication fell into each category.

Note that the same ratings were used for the theoretical indications and for the actual cases; the panels did not separately rate the actual cases. The comparison we make in this paper is between an unweighted average for all of the indications on the theoretical list, and a weighted average in which each indication is weighted by the number of actual cases for that indication.

Results

Only a small number of indications for each procedure accounted for most of the cases found in the medical records study (see Appendix Table A). For example, the single most frequently occurring indication for angiography (preoperative evaluation of patients undergoing cardiac valvular surgery) accounted for 18 per cent of all cases. (See reference 3 for other frequently occurring indications.) The eight most frequently occurring indications (2.7 per cent of the 300 indications on the theoretical list) accounted for over half of the cases. The other two procedures exhibited similar concentration. We found only 20 to 40 per cent of the theoretical indications to occur at all in our samples of about 1,500 cases for each procedure.

Table 1 shows the percent of agreed-upon indications using each of the four definitions of agreement. The comparison of agreement on theoretical indications with agreement

TABLE 1—Percent of Indications and Cases on Which Panelists Agreed Using Four Different Definitions of Agreement

Definition of Agreement	Coronary Angiography		Endoscopy		Carotid Endarterectomy	
	Indic	Cases	Indic	Cases	Indic	Cases
A9S: 9 ratings, strict	28.0	29.5	25.4	39.5	40.9	16.4
Standard error	(2.6)	(1.1)	(1.3)	(1.2)	(1.7)	(1.0)
A9R: 9 ratings, relaxed	28.7	30.6	25.4	39.5	40.9	16.4
Standard error	(2.6)	(1.1)	(1.3)	(1.2)	(1.7)	(1.0)
A7S: 7 ratings, strict	50.0	55.5	41.3	53.4	53.4	27.3
Standard error	(2.9)	(1.2)	(1.5)	(1.3)	(1.7)	(1.2)
A7R: 7 ratings, relaxed	56.3	63.4	41.6	53.4	53.8	27.9
Standard error	(2.9)	(1.2)	(1.5)	(1.3)	(1.7)	(1.2)

Notes: See text for full definitions of agreement.
 "Indic" denotes the theoretical indications on the entire list rated by the panel; "Cases" denotes actual cases to which indications were assigned in a field study of medical records.

on indications assigned to actual cases is different for the three procedures.

- For angiography, the per cent of agreed-upon indications is about the same for the theoretical list as it is for the actual cases.
- For endoscopy, agreement on cases is substantially higher than is agreement on theoretical indications.
- For endarterectomy, agreement on cases is substantially lower than is agreement on theoretical indications.

As defined by A7R, agreement on the appropriateness of indications assigned to actual cases ranged from over 60 per cent for angiography to under 30 per cent for endarterectomy.

Table 2 shows disagreed-upon indications using each of the five definitions of disagreement. Disagreement on cases, compared with disagreement on the theoretical indications, was somewhat lower for angiography, about the same for endoscopy, and substantially higher for endarterectomy. As defined by D7R, disagreement on indications for actual cases was 5 per cent for angiography, but was 25 per cent for endoscopy and 32 per cent for endarterectomy.

Table 3 categorizes theoretical indications and indications assigned to actual cases by level of appropriateness. For all three procedures, the per cent of appropriate cases was higher than the per cent of appropriate theoretical indications—substantially higher for endoscopy and endarterectomy. The per cent of inappropriate cases was about the same

APPENDIX

TABLE A—Per Cent of Cases Done for the N Most Frequently Occurring Indications

Per Cent of Cases	Number and Per Cent of Indications					
	Angiography		Endoscopy		Endarterectomy	
	N	%	N	%	N	%
10	1	0.3	1	0.1	2	0.2
25	2	0.7	4	0.4	6	0.7
50	8	2.7	14	1.3	23	2.7
75	26	8.7	43	4.0	80	9.3
100	119	39.7	225	21.1	280	32.4

Note: N = number of indications
 % = per cent of total theoretical indications:
 300 for angiography,
 1069 for endoscopy,
 864 for endarterectomy

TABLE 2—Percent of Indications on Which Panelists Disagreed Using Five Different Definitions of Disagreement

Definition of Disagreement	Coronary Angiography		Endoscopy		Carotid Endarterectomy	
	Indic	Cases	Indic	Cases	Indic	Cases
D9S: 9 ratings, strict	2.0	0.1	30.2	26.6	14.9	21.7
Standard error	(0.8)	(0.1)	(1.4)	(1.1)	(1.2)	(1.1)
D9R: 9 ratings, relaxed	30.0	22.6	48.5	40.9	34.0	50.4
Standard error	(2.7)	(1.0)	(1.5)	(1.2)	(1.6)	(1.4)
D7S: 7 ratings, strict	0.3	0.0	7.0	6.3	2.3	2.7
Standard error	(0.3)	(0.0)	(0.8)	(0.6)	(0.5)	(0.4)
D7R: 7 ratings, relaxed	11.0	5.3	28.9	25.0	18.1	32.0
Standard error	(1.8)	(0.5)	(1.4)	(1.1)	(1.3)	(1.3)
D5R: 5 ratings, relaxed	1.3	0.9	11.4	5.2	7.0	6.6
Standard error	(0.7)	(0.2)	(1.0)	(0.6)	(0.9)	(0.7)

Notes: See text for full definitions of disagreement.
 "Indic" denotes the theoretical indications on the entire list related by the panel; "Cases" denotes actual cases to which indications were assigned in a field study of medical records.

TABLE 3—Categorization of Indications by Appropriateness (%)

Appropriateness	Coronary Angiography		Endoscopy		Carotid Endarterectomy	
	Indic	Cases	Indic	Cases	Indic	Cases
Clearly inappropriate						
Median 1-3, without disagreement	14.0	17.3	45.1	10.5	67.0	28.5
Standard error	(2.0)	(0.9)	(1.5)	(0.8)	(1.6)	(1.3)
Equivocal						
Median 4-6, without disagreement	12.7	3.4	3.8	3.1	5.2	6.8
Any median, with disagreement	11.0	5.3	29.2	25.0	18.1	32.0
Total	23.7	8.7	33.0	28.1	23.3	38.8
Standard error	(2.5)	(0.7)	(1.4)	(1.1)	(1.4)	(1.4)
Clearly appropriate						
Median 7-9, without disagreement	62.3	74.0	21.9	61.3	9.7	32.7
Standard error	(2.8)	(1.1)	(1.3)	(1.2)	(1.0)	(1.3)

Notes: "Disagreement" means at least two 1-3 ratings and at least two 7-9 ratings. "Indic" denotes the theoretical indications on the entire list rated by the panel; "Cases" denotes actual cases to which indications were assigned in a field study of medical records.

as the per cent of inappropriate theoretical indications for angiography, and much lower for endoscopy and endarterectomy. The fraction of cases done for indications that we categorized as clearly inappropriate based on the panel ratings ranged from 10 per cent for endoscopy, through 17 per cent for angiography, up to nearly 30 per cent for endarterectomy.

The percentages in Table 3 differ somewhat from those that we have reported elsewhere.³ The reason is that Table 3 is based on disagreement defined by D7R, whereas the earlier report used D5R. We adopt the more easily satisfied definition here in order to provide a new perspective on the appropriateness results previously reported.

Discussion

Disagreement among expert physicians about appropriate indications for doing angiography, endoscopy, and endarterectomy is not limited to theoretical indications that are never done in practice. We found that there was generally almost as much or even more disagreement about the appropriateness of indications assigned to actual cases than there was about the complete lists of theoretical indications. Disagreement about either indications or cases was minimal for coronary angiography on most of our definitions. In contrast, there was substantial disagreement about both the theoretical indications and the indications actually used for endoscopy and carotid endarterectomy. For example, 27 per cent of endoscopies were done for reasons that at least one

out of nine panelists rated 1 (extremely inappropriate) and at least one rated 9 (extremely appropriate), and thus count as disagreement on definition D9S. Also, 32 per cent of endarterectomies were done for indications that at least two panelists rated 3 or lower and at least two panelists rated 7 or higher (disagreement on definition D7R).

There is generally more difference between the theoretical indications and the indications assigned to actual cases when they are categorized by appropriateness (Table 3), than when they are categorized by agreement or disagreement (Tables 1 and 2). Particularly for endoscopy and endarterectomy, inappropriate indications occur frequently on the theoretical lists, but most of them are not used in practice.

The high percentages of disagreement about the appropriateness of doing endoscopies and endarterectomies for indications that were assigned to actual cases reflect the real uncertainty about the efficacy of these procedures in many situations.^{6,7} Clinical trials may help to lessen the uncertainty in the future, but can never eliminate it entirely. Physicians should carefully weigh the advisability of doing a procedure for indications that lack firm scientific justification. If they do believe that the procedure is advisable in such circumstances, they should make sure that their patients understand that other physicians might disagree. Also, patients should attempt to ascertain and to understand the reasons for their doctor's recommendations before submitting themselves to expensive, invasive procedures.

ACKNOWLEDGMENTS

This work was supported in part by the Commonwealth Fund (Grant No. 6137), the John A. Hartford Foundation (Grant No. 83172-2H), the Pew Memorial Trust (Glenmede Trust), the Robert Wood Johnson Foundation (Grant No. 7309), and the Health Care Finance Administration (Cooperative Agreement No. 17-C-98442/9-01). Any opinions, conclusions, and proposals in the text are those of the authors and do not necessarily represent the views of any of these sponsors or of the Rand Corporation.

REFERENCES

1. Park RE, Fink A, Brook RH, *et al*: Physician ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health* 1986; 76:766-772.
2. Park RE, Fink A, Brook RH, *et al*: Physician ratings of appropriate indications for six medical and surgical procedures. R-3280-CWF/HF/PMT/RWJ. Santa Monica, CA: Rand Corporation, July 1986.
3. Chassin MR, Kosecoff J, Park RE, *et al*: Does inappropriate use explain geographic variations in the use of health services?: A study of three procedures. *JAMA* 1987; 258:2533-2537.
4. Kosecoff J, Chassin MR, Fink A, *et al*: Obtaining clinical data on the appropriateness of medical care in community practice. *JAMA* 1987; 258:2538-2542.
5. Chassin MR, Kosecoff J, Solomon DH, Brook RH: How coronary angiography is used: Clinical determinants of appropriateness. *JAMA* 1987; 258:2543-2547.
6. Wennberg JE: The paradox of appropriate care. *JAMA* 1987; 258:2568-2569.
7. Dawson JH: Practice variations: A challenge for physicians. *JAMA* 1987; 258:2570.