Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures

Rolla Edward Park, Arlene Fink, Robert H. Brook,
Mark R. Chassin, Katherine L. Kahn, Nancy J. Merrick,
Jacqueline Kosecoff, David H. Solomon

Rand

The research described in this report was supported jointly by the Commonwealth Fund, the John A. Hartford Foundation, the Pew Memorial Trust, and the Robert Wood Johnson Foundation.

Library of Congress Cataloging in Publication Data

Physician ratings of appropriate indications for six medical and surgical procedures.

(Publications of the Rand/UCLA Health Services Utilization Study)
"This report presents results from one part of the Rand/UCLA Health Services Utilization Study"—Pref.
"July 1986."
"R-3280-CWF/HF/PMT/RWJ."
1. Clinical indications—Evaluation. 2. Surgical indications—Evaluation. I. Park, Rolla Edward. II. Commonwealth Fund. III. Rand Corporation. IV. Rand/UCLA Health Services Utilization Study. V. Series. [DNLM: 1. Health Services—utilization—United States. W 84 AA1 P5]
RC67.P49 1986 362.1'7 86-17830
ISBN 0-8330-0751-3

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

R-3280-CWF/HF/PMT/RWJ

Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures

Rolla Edward Park, Arlene Fink, Robert H. Brook,
Mark R. Chassin, Katherine L. Kahn, Nancy J. Merrick,
Jacqueline Kosecoff, David H. Solomon

July 1986

Prepared for
The Commonwealth Fund
The John A. Hartford Foundation
The Pew Memorial Trust
The Robert Wood Johnson Foundation

Rand
1700 MAIN STREET
P.O. BOX 2138
SANTA MONICA, CA 90406-2138

PREFACE

This report presents results from one part of the Rand/UCLA Health Services Utilization Study (HSUS). HSUS is exploring clinical reasons for geographic variations in the use of medical and surgical procedures. The study has four major stages:

- Stage I calculated use rates for a large number of procedures in 13 large areas of the United States. The patient population studied was Medicare elderly.
- Stage II reviewed the medical literature to summarize knowledge about efficacy, complications, and indications for six procedures.
- Stage III convened panels of physicians to rate the appropriateness of a large number of indications for performing the six procedures.
- Stage IV is abstracting medical records to determine the actual indications for performing three of the six procedures in several of the 13 geographic areas.

This report presents some results of the Stage III panel process. Without going into the clinical details of the indications that were rated, it describes the panel process, summarizes the panel ratings (with particular emphasis on the amount of agreement among the panelists), and describes how we plan to use the results in Stage IV. The report should interest physicians and health researchers concerned with establishing consensus on appropriate indications for performing procedures.

A shorter paper based on this report appears in *American Journal of Public Health*, July 1986. A complete list of HSUS publications follows.

PUBLICATIONS OF THE RAND/UCLA HEALTH SERVICES UTILIZATION STUDY

Brook R. H., M. R. Chassin, A. Fink, J. Kosecoff, R. E. Park, and D. H. Solomon, "A Method for the Detailed Assessment of the Appropriateness of Medical Technologies," *International Journal of Technology Assessment in Health Care*, 1986.

Brook, R. H., and K. N. Lohr, "Efficacy, Effectiveness, Variations, and Quality: Boundary-Crossing Research," *Medical Care*, 1985; 23(5):710–722.

Brook, R. H., K. N. Lohr, M. R. Chassin, J. Kosecoff, A. Fink, and D. H. Solomon, "Geographic Variations in the Use of Services: Do They Have Any Clinical Significance?" *Health Affairs*, 1984; 3(4):63–73.

Chassin, M. R., J. Kosecoff, R. E. Park, A. Fink, S. Rauchman, J. Keesey, M. F. Flynn, and R. H. Brook, *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Coronary Angiography*, The Rand Corporation, R-3204/1-CWF/HF/HCFA/PMT/RWJ, 1986.

Chassin, M. R., R. E. Park, A. Fink, S. Rauchman, J. Keesey, and R. H. Brook, *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Coronary Artery Bypass Graft Surgery*, The Rand Corporation, R-3204/2-CWF/HF/HCFA/PMT/RWJ, 1986.

Chassin, M. R., R. H. Brook, R. E. Park, J. Keesey, A. Fink, J. Kosecoff, K. Kahn, N. Merrick, and D. H. Solomon, "Variations in the Use of Medical and Surgical Services by the Medicare Population," *New England Journal of Medicine*, 1986; 314:285–290. (A longer version appears as Rand R-3288-CWF/HF/PMT/RWJ, forthcoming.)

Fink, A., J. Kosecoff, M. R. Chassin, and R. H. Brook, "Consensus Methods: Characteristics and Guidelines for Use," *American Journal of Public Health*, September 1984; 74:979–983.

Kahn, K. L., C. P. Roth, J. Kosecoff, R. E. Park, A. Fink, R. H. Brook, M. F. Flynn, J. Keesey, M. R. Chassin, G. Anderson, and D. H. Solomon, *Indications for Selected Medical and Surgical Procedures—A Review of the Literature and Ratings of Appropriateness: Diagnostic Gastrointestinal Endoscopy*, The Rand Corporation, R-3204/4-CWF/HF/HCFA/PMT/RWJ, forthcoming.

Kosecoff, J., A. Fink, R. H. Brook, and M. R. Chassin, "The Appropriateness of Using a Medical Procedure: Is Information in the Medical Record Valid?" *Medical Care*, forthcoming.

Merrick, N. J., R. E. Park, J. Kosecoff, A. Fink, R. H. Brook, C. Roth, L. McCloskey, J. Keesey, and D. H. Solomon, *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Carotid Endarterectomy*, The Rand Corporation, R-3204/6-CWF/HF/HCFA/PMT/RWJ, forthcoming.

Merrick, N., and R. H. Brook, "Use of Carotid Endarterectomy in Five Veterans Administration Medical Centers," *Journal of the American Medical Association*, forthcoming.

Merrick, N. J., A. Fink, R. E. Park, R. H. Brook, J. Kosecoff, M. Chassin, and D. H. Solomon, "Derivation of Clinical Indications for Carotid Endarterectomy by an Expert Panel," *American Journal of Public Health*, forthcoming.

Park, R. E., A. Fink, R. H. Brook, M. R. Chassin, K. Kahn, N. Merrick, J. Kosecoff, and D. H. Solomon, "Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures," *American Journal of Public Health*, July 1986 (a shortened version of the present report).

Roth, C. P., K. L. Kahn, A. Fink, R. E. Park, R. H. Brook, J. Kosecoff, J. Keesey, and D. H. Solomon, *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Fiberoptic Colonoscopy*, The Rand Corporation, R-3204/5-CWF/HF/PMT/RWJ, forthcoming.

Solomon, D. H., R. H. Brook, A. Fink, R. E. Park, and J. Keesey, *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Cholecystectomy*, The Rand Corporation, R-3204/3-CWF/HF/PMT/RJW, May 1986.

SUMMARY

As part of the Rand-UCLA Health Services Utilization Study, we convened three panels of physicians to rate the appropriateness of a large number of indications for performing a total of six medical and surgical procedures. The procedures were coronary angiography, coronary artery bypass graft surgery, cholecystectomy, upper gastrointestinal endoscopy, colonoscopy, and carotid endarterectomy. The ratings demonstrate that it is possible to obtain ratings of hundreds of indications for each procedure—a much finer level of detail than previous consensus processes (for example, the NIH consensus panels) have ever attempted.

There were three panels. A cardiovascular panel rated indications for coronary angiography and coronary artery bypass surgery; a gastroenterological panel rated indications for cholecystectomy, upper gastrointestinal endoscopy, and colonoscopy; a cerebrovascular panel rated indications for carotid endarterectomy. Each of the panels was organized and conducted according to the following general plan:

1. Project staff invited nine distinguished physicians to serve on the panel.
2. Project staff prepared initial lists of clinical indications for performing each procedure.
3. The panelists rated the appropriateness of each indication on a nine-point scale and mailed their ratings to Santa Monica, where project staff tabulated them for use at the panel meeting.
4. The panelists met in Santa Monica. Taking one procedure at a time, they discussed the indications, revised the indications list, and individually assigned final appropriateness ratings.
5. Project staff analyzed the ratings.

In broad outline, the panel results were:

- The panelists substantially revised the initial lists of indications. The final lists for most procedures included more indications than the initial lists.
- The average median rating differed substantially from procedure to procedure, and for some procedures changed substantially from the initial to the final ratings.
- The average dispersion around the median of the panelists' final ratings was substantially less than the initial dispersion.
- Representatives of different specialties rated differently, but the differences were not consistent across all procedures.
- The median ratings were widely distributed across the rating scale. For all procedures, there were more ratings at the ends of the scale than there were in the middle, and dispersion was less at the ends than in the middle.
- We defined a number of different, intuitively appealing, standards for "agreement" or "disagreement" among the panelists about the rating for individual indications. Based on our preferred definitions, the panelists agreed on the ratings for about half of the indications, and disagreed for 10 to 30 percent of them, for all except one procedure. For colonoscopy (which was rated without group discussion), the proportions were reversed.
- By lumping together into an "equivocal" category those indications with middle ratings and those indications on which the panelists disagreed, we can establish three useful categories of indications: clearly inappropriate, equivocal, and clearly appropriate.

- For all procedures except colonoscopy, the proportion of equivocal ratings varied from 23 to 33 percent.
- Most of the remaining indications were clearly appropriate for coronary angiography and coronary artery bypass surgery. Most were clearly inappropriate for endoscopy and carotid endarterectomy. (As explained below, this is not in any sense an indictment of these two procedures.) The remaining indications were evenly divided for cholecystectomy.

In the final stage of our larger study, we will abstract medical records to ascertain the actual indications for which three procedures were done during 1981. The three procedures are coronary angiography, endoscopy, and carotid endarterectomy. We will compare the indications found in areas of high use with those in areas of low use. The major question to be asked is whether a higher percentage of the procedures were done for equivocal and inappropriate reasons in high use areas than in low use areas.

We are currently abstracting a random sample of approximately 500 medical records for each procedure studied in several geographic areas to determine the exact indications for which the procedures were done during 1981. Using information abstracted from the medical records, we will determine the indication or indications for which the procedure was done. Each indication translates into an appropriateness category—appropriate, equivocal, or inappropriate—based on the panel ratings. Using the highest rated indication for each individual procedure, we will count up the number of procedures that fall in each category. We will then statistically compare the proportion of procedures done for equivocal and inappropriate reasons in the different geographic areas.

ACKNOWLEDGMENTS

We greatly appreciate the indispensable help of the technical panel members, the wise counsel of the policy advisory board, and thoughtful reviews by Dr. Noralou P. Roos (University of Manitoba) and the editor and two anonymous referees for *American Journal of Public Health.*

CONTENTS

PREFACE iii

PUBLICATIONS OF THE RAND/UCLA HEALTH SERVICES UTILIZATION STUDY v

SUMMARY vii

ACKNOWLEDGMENTS ix

FIGURES AND TABLES xiii

Section

I. INTRODUCTION 1

II. ORGANIZATION AND CONDUCT OF THE PANELS 2
- Panel Selection 2
- Initial Indications Lists 2
- Initial Ratings 4
- Panel Meetings 7
- Planned Analyses 10

III. PANEL RESULTS 11
- Measures Used 11
- Overall Results 12
- Comparison of Initial and Final Ratings 13
- Differences in Ratings by Speciality 15
- Distribution of Ratings and Deviations 16
- Defining Agreement and Disagreement 16
- A Useful Categorization of Rated Indications 20

IV. DISCUSSION 22
- What Did We Learn from the Panels? 22
- How Will We Use the Panel Results in a Medical Records Study? 23
- How Might Additional Ratings Based on Current Knowledge Be Used? 23
- Why Does There Appear To Be So Much Disagreement? 24

REFERENCES 27

Appendix

A. THE RAND/UCLA HEALTH SERVICES UTILIZATION STUDY (HSUS) 29

B. HSUS POLICY ADVISORY BOARD AND TECHNICAL PANEL MEMBERS 31

C. INITIAL INDICATIONS 37

D. FINAL INDICATIONS 40

E. NUMBER OF INDICATIONS WITH VARIOUS NUMBERS OF MISSING FINAL RATINGS, BY PROCEDURE 43

F. PERCENT OF INDICATIONS ON WHICH PANELISTS AGREED OR DISAGREED 44

FIGURES

2.1. Sample Initial Rating Sheet for Coronary Angiography 5
2.2. Sample Initial Rating Sheet for Carotid Endarterectomy 6
2.3. Sample Final Rating Sheet for Coronary Angiography 8
2.4. Sample Final Rating Sheet for Carotid Endarterectomy 9

TABLES

2.1. Panel Membership 3
2.2. Clinical Settings and Factors: A Framework for the Indications for Coronary Angiography 4
3.1. Overall Panel Results for Six Procedures 12
3.2. Panel Results for Indications that Were Not Changed During the Panel Process . . . 14
3.3. Panel Results for Indications that Split in Two 15
3.4. Average Deviation from Overall Median by Specialty 16
3.5. Distribution of Median Ratings and Dispersion 17
3.6. Percent of Indications on Which Panelists Agreed, Using Four Different Definitions of Agreement 18
3.7. Percent of Indications on Which Panelists Disagreed Using Four Different Definitions of Disagreement 20
3.8. Categorization of Final Indications by Appropriateness 21
F.1. Percent of Indications on Which Panelists Agreed, Using Four Different Definitions of Agreement (Median Rating 1 to 3) 44
F.2. Percent of Indications on Which Panelists Agreed, Using Four Different Definitions of Agreement (Median Rating 4 to 6) 44
F.3. Percent of Indications on Which Panelists Agreed, Using Four Different Definitions of Agreement (Median Rating 7 to 9) 45
F.4. Percent of Indications on Which Panelists Disagreed, Using Four Different Definitions of Disagreement (Median Rating 1 to 3) 45
F.5. Percent of Indications on Which Panelists Disagreed, Using Four Different Definitions of Disagreement (Median Rating 4 to 6) 46
F.6. Percent of Indications on Which Panelists Disagreed, Using Four Different Definitions of Disagreement (Median Rating 7 to 9) 46

I. INTRODUCTION

The Rand-UCLA Health Services Utilization Study (HSUS) is exploring clinical reasons for geographic variation in the use of medical and surgical procedures. For over a hundred procedures, we calculated use rates in Medicare elderly populations. When we compared the use rates in 13 large areas of the U.S., we found substantial variation for almost all of the procedures (Chassin, Brook, Park, et al., 1986), a finding that parallels the results of numerous small-area studies (see, for example, Wennberg and Gittelsohn, 1982). We will abstract medical records in high-, average-, and low-use areas to identify the indications for which procedures were performed. We hypothesize that the distribution of indications was different in high-use areas than it was in low-use areas. Specifically, we want to see if a larger fraction of procedures were performed for inappropriate indications in high-use areas. If so, then overuse in some areas would explain at least part of the variation in use rates. If not, then other explanations must be sought.[1]

In order to distinguish between appropriate and inappropriate indications, we convened three panels of physicians to rate the appropriateness of a large number of indications for performing a total of six procedures: coronary angiography, coronary artery bypass graft surgery, cholecystectomy, upper gastrointestinal endoscopy, fiberoptic colonoscopy, and carotid endarterectomy.[2] The main purpose of the panels was to produce appropriateness ratings that we can use in our comparison of high- and low-use areas. The ratings that came out of the panel meetings are well suited for that purpose, and we will use them in the larger study.

The panel ratings are also interesting in their own right. They demonstrate that it is possible to obtain ratings of hundreds of indications for each procedure—a much finer level of detail than previous consensus processes (for example, the NIH consensus panels) have ever attempted. At the same time, they reveal a perhaps surprisingly high degree of residual disagreement among the panelists about the appropriate use of these six procedures.

In this report, we present an overview of the panel ratings. Without going into the clinical details of the indications,[3] we describe the composition and working of the panels (Sec. II) and summarize the ratings, with particular emphasis on the amount of agreement or disagreement among the panelists (Sec. III). Finally (in Sec. IV), we discuss the results and sketch how we plan to use them in the final phase of the larger study.

---

[1]For a more complete description of HSUS, see App. A.

[2]We selected these six procedures because they are frequently performed, they use substantial medical resources, and they exhibit significant variation in use rates between large geographic areas of the United States.

[3]For detailed lists of the indications and their ratings, see Chassin, Kosecoff, Park, et al. (coronary angiography, 1986); Chassin, Park, Fink, et al. (coronary artery bypass graft surgery, 1986); Solomon, Brook, Fink, et al. (cholecystectomy, 1986); Kahn, Roth, Kosecoff, et al. (upper gastrointestinal endoscopy, 1986); Roth, Kahn, Fink, et al. (fiberoptic colonoscopy, 1986); and Merrick, Park, Kosecoff, et al. (carotid endarterectomy, 1986). Hereinafter we refer to these publications as "the six procedure reports."

II. ORGANIZATION AND CONDUCT OF THE PANELS

There were three panels. A cardiovascular panel rated indications for coronary angiography and coronary artery bypass graft surgery; a gastrointestinal panel rated indications for cholecystectomy, upper gastrointestinal endoscopy, and colonoscopy; a cerebrovascular panel rated indications for carotid endarterectomy. Each of the panels was organized and conducted according to the following general plan:[1]

1. Project staff invited nine distinguished physicians to serve on the panel.
2. Project staff prepared initial lists of clinical indications for performing each procedure.
3. The panelists rated the appropriateness of each indication and mailed their ratings to Santa Monica, where project staff tabulated them for use at the panel meeting.
4. The panelists met in Santa Monica. Taking one procedure at a time, they discussed the indications, revised the indications list, and individually assigned final appropriateness ratings.
5. Project staff analyzed the ratings.

The details varied somewhat among procedures, and some procedures departed slightly from this general plan, as indicated by the more detailed descriptions of each step that follow.

PANEL SELECTION

In selecting panelists, project staff relied heavily on the suggestions of members of the Health Services Utilization Study Policy Advisory Board (PAB). The Board consisted of selected leaders of American medicine, many representing prominent medical organizations.[2] Those nominated by the PAB were all distinguished physicians. In choosing nine of them to serve on each panel, we attempted to keep a rough balance among different parts of the country, different relevant specialties, and academic versus private practice. Each panel had members from all four U.S. census regions. The specialty composition of the panels is shown in Table 2.1.

INITIAL INDICATIONS LISTS

Physicians on the project staff compiled the initial indications lists, using as guides reviews of the medical literature on each procedure. (The literature reviews are included in the six procedure reports.) The indications categorized patients in terms of their symptoms, past medical history, and the results of previous diagnostic tests. We attempted to compile clinical indications lists that were detailed, comprehensive, and manageable—detailed so that patients presenting with an indication would be reasonably homogeneous, in the sense that doing the procedure would be equally appropriate (or inappropriate) for all of them; comprehensive enough to include all indications for doing the procedure that arise in practice; and yet short

[1]We decided on this plan following a careful review of formal consensus methods that have been used before by others; see Fink, Kosecoff, Chassin, and Brook (1984) for details.

[2]Appendix B lists the HSUS PAB members and the members of the technical panels.

Table 2.1

PANEL MEMBERSHIP

Specialty	Panel: Cardiovascular (ANGIO,CABS)	Panel: Gastroenterological (CH,END,CO)	Panel: Cerebrovascular (CE)
Medical generalists	1 family physician 2 internists	1 family physician 2 internists	1 family physician 1 internist
Medical specialists	3 cardiologists	3 gastroenterologists	2 neurologists
Surgeons	2 cardiothoracic surgeons	2 general surgeons	3 vascular surgeons 1 neurosurgeon
Radiologist	1 radiologist	1 radiologist	1 neuroradiologist

NOTE: ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

enough so that the panelists could rate all of the indications within a reasonable length of time.

The indications for each procedure were organized into "chapters," which in most cases corresponded to major symptoms or clinically related problems for which the procedure could be used. For example, Table 2.2 lists the eight clinical settings (chapters) for coronary angiography. We defined indications for the use of coronary angiography in asymptomatic patients, patients with chest pain of uncertain origin, and patients in six other categories. Under each chapter title are listed the factors that we used to define the indications in that clinical setting; each factor may appear at several levels and in combination with different levels of the other factors. In Chapter III (patients with chronic stable angina), for example, contraindications to coronary artery bypass graft surgery include dementia, severely impaired one-year life expectancy, and severe, irreversible functional incapacity. Medical management is classified as none, less than maximal, or maximal. Angina is class I, II, III, or IV according to the Canadian Cardiovascular Society classification. Test results may be negative, positive, or very positive, or the test may not have been performed. We used various combinations of these factors to define the indications for coronary angiography in patients with chronic stable angina. For example, one such indication is: patients with chronic stable angina (without strong contraindications to coronary artery bypass graft surgery) in whom angina occurs with mild exertion (class III or IV) and who have received no or less than maximal medical management and have a very positive exercise stress test, a negative exercise thallium scan, and no exercise MUGA (multiple gated cardiac scan).

Comorbidity was accounted for in the indications for all of the procedures. For the cardiovascular procedures, comorbid conditions figured in the definition of contraindications to coronary artery bypass graft surgery. For the gastrointestinal and cerebrovascular procedures, the panelists explicitly assigned adjustments to their ratings to account for varying levels of comorbidity.

Table 2.2

CLINICAL SETTINGS AND FACTORS: A FRAMEWORK FOR THE INDICATIONS FOR CORONARY ANGIOGRAPHY

1. Asymptomatic patients
   - Occupation
   - Exercise ECG test
   - Exercise thallium scan
   - Exercise MUGA test

2. Chest pain of uncertain origin
   - Chest pain with exertion
   - Exercise ECG test
   - Exercise thallium scan
   - Exercise MUGA test

3. Chronic stable angina
   - Contraindications to coronary artery bypass graft surgery
   - Medical management of angina
   - Degree of angina
   - Exercise ECG test
   - Exercise thallium scan
   - Exercise MUGA test

4. Unstable angina
   - Contraindications to coronary artery bypass graft surgery
   - Medical management of angina
   - Degree of angina
   - Exercise ECG test
   - Exercise thallium scan
   - Exercise MUGA test

5. During an acute MI
   - Contraindications to coronary artery bypass graft surgery
   - Complications of MI

6. Within six months of an acute MI
   - Contraindications to coronary artery bypass graft surgery
   - Type of MI
   - Presence of angina post MI
   - Medical management of angina
   - Degree of angina
   - Exercise ECG test
   - Exercise thallium scan
   - Exercise MUGA test

7. Sudden death survivors
   - Contraindications to coronary artery bypass graft surgery
   - Relation of sudden death episode to an MI
   - Presence of angina post MI
   - Medical management of angina
   - Degree of angina
   - Exercise ECG test
   - Exercise thallium scan
   - Exercise MUGA test

8. Following coronary artery bypass graft surgery
   - Timing of coronary angiography
   - Presence of angina post CABS
   - Medical management of angina
   - Degree of angina
   - Exercise ECG test
   - Exercise thallium scan
   - Exercise MUGA test

The chapter titles from the initial indications list for each procedure are in Appendix C, which also shows the number of indications in each chapter. The total number of initial indications for each procedure ranged from 192 for cholecystectomy to 1685 for endoscopy.

INITIAL RATINGS

We sent literature reviews, rating sheets, and instructions to the panelists. The literature reviews gave all panelists equal access to a central core of relevant literature. (The literature reviews are included in the six procedure reports.) The ratings sheets listed all of the indications for each procedure and provided space for an appropriateness rating on a scale of 1 to 9. As examples, Figs. 2.1 and 2.2 show one page each from the initial rating sheets for coronary angiography and for carotid endarterectomy. The initial rating sheets for the other four procedures took the same form as those for coronary angiography.

CHAPTER 1

INDICATIONS FOR CORONARY ANGIOGRAPHY

Appropriateness Scale 1 2 3 4 5 6 7 8 9 1 = extremely inappropriate 5 = equivocal (neither clearly appropriate nor clearly inappropriate) 9 = extremely appropriate	Frequency Ratings Check the 50-75 indications that account for the majority of angiographies performed in the USA in 1981.

	Rating of Appropriateness (Circle One)
I. Asymptomatic Patients	
A. Coronary angiography (CA) is indicated in patients in high risk occupations if:	
1. No exercise ECG, no exercise thallium scan, and no exercise MUGA	1 2 3 4 5 6 7 8 9
2. Negative exercise ECG and	
a. No or negative exercise thallium scan regardless of MUGA results, if any	1 2 3 4 5 6 7 8 9
b. Reversible defect on exercise thallium scan and	
(i) No exercise MUGA	1 2 3 4 5 6 7 8 9
(ii) Negative exercise MUGA	1 2 3 4 5 6 7 8 9
(iii) Positive exercise MUGA	1 2 3 4 5 6 7 8 9
3. Positive exercise ECG and	
a. No exercise thallium scan regardless of MUGA results, if any	1 2 3 4 5 6 7 8 9
b. Negative exercise thallium scan and	
(i) No exercise MUGA	1 2 3 4 5 6 7 8 9
(ii) Negative exercise MUGA	1 2 3 4 5 6 7 8 9

Fig. 2.1—Sample initial rating sheet for coronary angiography

CLINICAL PRESENTATION:
1. CAROTID TIA and/or AMAUROSIS FUGAX - Single Episode

APPROPRIATENESS OF OPERATING IPSILATERALLY IF ANGIOGRAPHY SHOWS: Ipsi: Degree of stenosis of ipsilateral artery Contra: Degree of stenosis of contralateral artery	Low Surgical Risk	Elevated Surgical Risk	High Surgical Risk
1. Ipsi: Occluded Contra: None or 1-49%	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
2. Ipsi: Occluded Contra: 50-99%	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
3. Ipsi: 50-99% Contra: None, 1-49%, or 50-99%	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
4. Ipsi: 50-99% Contra: Occluded	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
5. Ipsi: 1-49% Contra: None or 1-49%	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
6. Ipsi: 1-49% Contra: 50-99%	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
7. Ipsi: 1-49% Contra: Occluded	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
8. Ipsi: 1-49% with large ulcerative lesion Contra: None, 1-49% or 50-99%	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
9. Ipsi: 1-49% w/multicentric ulcerative lesion Contra: None, 1-49%, or 50-99%	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

Fig. 2.2—Sample initial rating sheet for carotid endarterectomy

The instructions asked the panelists to rate the appropriateness of each indication as of 1981 using their own best clinical judgment (rather than their perceptions of what other experts might say), and considering an average group of patients presenting to an average U.S. physician who performed the procedure during 1981.[3] "Appropriate" was defined to mean that the expected health benefit (i.e., increased life expectancy, relief of pain, reduction in anxiety, improved functional capacity, etc.) exceeded the expected negative consequences (i.e., mortality, morbidity, anxiety of anticipating the procedure, pain produced by the procedure, time lost from work, etc.) by a sufficiently wide margin that the procedure was worth doing. We deliberately excluded cost of the procedure from the list of negative consequences, because we wanted the panelists to rate the medical appropriateness, not economic appropriateness.

[3]We specified 1981 so that the ratings would correspond to the claims data and medical records sample that we are using in the larger study.

"Inappropriate" meant the opposite—the negative consequences outweighed the expected benefits. Extremely appropriate indications should be rated 9, equivocal indications (neither clearly appropriate nor clearly inappropriate) should be rated 5, and extremely inappropriate indications should be rated 1.

In some cases, the instructions also included definitions of important terms used in the indications lists. The panelists were encouraged to modify and supplement the indications lists to make them more complete and clinically relevant. The panelists suggested very few changes during this initial ratings step.

We know from informal conversation with the panelists that it required about 20 to 35 minutes per 100 indications to complete the initial ratings.

PANEL MEETINGS

The panels all met in Santa Monica during the first half of 1984. The cardiovascular panel met for two days. After brief preliminaries, they spent one day discussing and rerating coronary angiography, and one day on coronary artery bypass graft surgery.[4] The gastrointestinal panel met for two and a half days. They spent one half day on cholecystectomy and two days on endoscopy, running out of time before they got to colonoscopy. Rerating of colonoscopy was done by mail without group discussion two months after the meeting. The cerebrovascular panel met for one and a half days. They spent one day discussing and rerating carotid endarterectomy, followed by a half day providing input to a related decision analysis project. The discussion of each procedure was led by the physician who prepared the initial indications list for that procedure. He or she was assisted by other physicians and social scientists on the project staff.

The panelists discussed the indications for each procedure one chapter at a time. During the discussion, they had computer printouts in front of them that summarized their initial ratings for that chapter. For example, Figs. 2.3 and 2.4 show one page each from the printouts for angiography and for carotid endarterectomy. The printouts for the gastrointestinal procedures were similar to that for carotid endarterectomy.

By looking at the printouts, the panelists could see the distribution of initial ratings. The numbers above the 1-to-9 rating line show how many panelists assigned each rating. For example, all nine panelists assigned a rating of 1 to the first coronary angiography indication in Fig. 2.3. Each panelist received a different printout; the distribution of ratings was the same on all, but the caret below the rating line showed the particular panelist's own initial rating. For example, the panelist whose printout is shown in Fig. 2.3 rated the first two indications 1 and the third indication 6.

The printouts for coronary angiography, coronary artery bypass graft surgery, and carotid endarterectomy used the initial indications lists without modification. The printouts for the three gastrointestinal procedures used modified indications lists. The modifications reflected suggestions made by the panelists during telephone conversations with a project physician. We hoped that revising the lists prior to the meeting would expedite discussion at the meeting.

The modifications to the gastrointestinal indications lists were as follows: For cholecystectomy, the initial list repeated 48 indications for four patient categories (older and younger patients, with and without diabetes). The list for the printout used at the meeting was instead divided by four comorbidity categories (no, low, medium, and high). The printout

[4]The radiologist on this panel did not attend the second day, and thus did not participate in the discussion or rerating of bypass surgery.

INDICATIONS FOR CORONARY ANGIOGRAPHY

Appropriateness Scale 1 2 3 4 5 6 7 8 9 1 = extremely inappropriate 5 = equivocal (neither clearly appropriate nor clearly inappropriate) 9 = extremely appropriate	Frequency Ratings Check the 50-75 indications that account for the majority of angiographies performed in the USA in 1981.

	Rating of Appropriateness (Circle One)
I. Asymptomatic Patients	
A. Coronary angiography (CA) is indicated in patients in high risk occupations if:	
1. No exercise ECG, no exercise thallium scan, and no exercise MUGA	9 1 2 3 4 5 6 7 8 9
2. Negative exercise ECG and	
a. No or negative exercise thallium scan regardless of MUGA results, if any	9 1 2 3 4 5 6 7 8 9
b. Reversible defect on exercise thallium scan and	
(i) No exercise MUGA	1 1 2 2 2 1 1 2 3 4 5 6 7 8 9
(ii) Negative exercise MUGA	1 1 2 2 2 1 1 2 3 4 5 6 7 8 9
(iii) Positive exercise MUGA	2 1 6 1 2 3 4 5 6 7 8 9
3. Positive exercise ECG and	
a. No exercise thallium scan regardless of MUGA results, if any	2 3 3 1 1 2 3 4 5 6 7 8 9
b. Negative exercise thallium scan and	
(i) No exercise MUGA	1 3 2 1 2 1 2 3 4 5 6 7 8 9
(ii) Negative exercise MUGA	1 3 2 2 1 1 2 3 4 5 6 7 8 9

Fig. 2.3—Sample final rating sheet for coronary angiography

1

CLINICAL PRESENTATION:
1. CAROTID TIA and/or AMAUROSIS FUGAX - Single Episode

APPROPRIATENESS OF OPERATING IPSILATERALLY IF ANGIOGRAPHY SHOWS:	Low Surgical Risk	Elevated Surgical Risk	High Surgical Risk
Ipsi: Degree of stenosis of ipsilateral artery Contra: Degree of stenosis of contralateral artery	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
1. Ipsi: Occluded Contra: None or 1-49%	9 1	9 1	9 1
2. Ipsi: Occluded Contra: 50-99%	9 1	9 1	9 1
3. Ipsi: 50-99% Contra: None, 1-49%, or 50-99%	1 1 2 1 4 3 5 6 7 8 9	1 1 3 1 2 2 3 4 5 6 7 8 9	4 2 1 1 1 1 3 4 5 6 7 8 9
4. Ipsi: 50-99% Contra: Occluded	1 1 3 4 2 5 6 7 8 9	1 1 1 2 1 3 2 3 4 5 6 7 8 9	1 4 1 1 1 1 2 3 4 5 6 7 8 9
5. Ipsi: 1-49% Contra: None or 1-49%	4 1 2 1 1 1 3 4 5 6 7 8 9	6 1 1 1 2 3 4 5 6 7 8 9	8 1 1 2 3 4 5 6 7 8 9
6. Ipsi: 1-49% Contra: 50-99%	3 1 1 1 1 1 1 2 3 4 5 6 7 8 9	5 2 1 1 1 2 3 4 5 6 7 8 9	7 1 1 1 2 3 4 5 6 7 8 9
7. Ipsi: 1-49% Contra: Occluded	3 1 1 2 1 1 2 3 4 5 6 7 8 9	5 3 1 1 2 3 4 5 6 7 8 9	8 1 1 2 3 4 5 6 7 8 9
8. Ipsi: 1-49% with large ulcerative lesion Contra: None, 1-49% or 50-99%	1 1 1 1 3 3 2 3 4 5 6 7 8 9	1 1 1 4 2 1 2 3 4 5 6 7 8 9	4 1 1 2 1 1 2 3 4 5 6 7 8 9
9. Ipsi: 1-49% w/multicentric ulcerative lesion Contra: None, 1-49%, or 50-99%	1 1 1 1 5 2 3 4 5 6 7 8 9	1 1 1 2 2 2 2 3 4 5 6 7 8 9	4 1 1 2 1 1 2 3 4 5 6 7 8 9

Fig. 2.4—Sample final rating sheet for carotid endarterectomy

showed the distribution of initial ratings for nondiabetic patients under 55 years old together with blank rating lines for the four comorbidity categories. For endoscopy and colonoscopy, a project physician substantially restructured the initial indications lists following the telephone conversations with the panelists. Most of the indications were changed. The printouts for these two procedures showed the distribution of initial ratings for the unchanged indications, together with blank rating lines for the new indications.

For all procedures except colonoscopy (which was not discussed at the panel meeting), the indications lists were substantially revised during discussions at the meetings. The changes were meant to tailor the indications so that they better described clinically homogeneous categories. Many of the changes simply split one indication into two. For example, indications for carotid endarterectomy that described a 50 through 99 percent occlusion of the ipsilateral artery were divided into separate indications for 50 through 69 percent and 70 through 99

percent occlusion. Other changes adjusted the boundaries between indications. For example, many indications on the initial list for coronary artery bypass graft surgery distinguished between ejection fraction under 30 percent and ejection fraction 30 percent or greater. For many of the final indications, the corresponding distinction is threefold: ejection fraction under 20 percent, 20 to 49 percent, and 50 percent or greater.

In other cases, some indications were dropped and others were added. Some initial indications were divided into multiple final indications. Other groups of initial indications were merged into a single final indication. Many of the changes were even more complex transformations. In some cases, whole chapters were divided, combined, or eliminated.

As one example of the changes that were made, in coronary angiography Chapter III (chronic stable angina), the panelists split some of the indications by age. After the split, all of the indications for patients with Class I or II angina were rated separately for patients under 65 and for patients 65 or older.

The final chapter titles and the number of indications in each chapter are in Appendix D. The total number of final indications ranged from 196 for cholecystectomy to 2862 for colonoscopy.

The printouts showing distributions of initial ratings also served as rating sheets. After discussion of each chapter, the panelists marked their final ratings directly on the printouts.

For colonoscopy, rating sheets were mailed to the panelists eight weeks after the meeting and were returned within twelve weeks of the meeting.

PLANNED ANALYSES

The next section presents an overview of the panel ratings; it treats the six procedures separately, but does not go into any finer level of clinical detail. Project staff members are now preparing several more-detailed analyses. These include regression analyses that link ratings to the clinical elements of the indications for several of the procedures.[5]

---

[5]See Merrick, Fink, Park, et al. (1986) for carotid endarterectomy; others are in progress.

III. PANEL RESULTS

We summarize the panel results in this section. In broad outline:

- The panelists substantially revised the initial lists of indications. The final lists for most procedures included more indications than the initial lists.
- The average median rating varied substantially from procedure to procedure, and for some procedures changed substantially from the initial to the final ratings.
- The average dispersion around the median of the panelists' final ratings was substantially less than the initial dispersion.
- Representatives of different specialties rated differently, but the differences were not consistent across all procedures.
- The median ratings were widely distributed across the rating scale. For all procedures, there were more ratings at the ends of the scale than there were in the middle, and dispersion was less at the ends than in the middle.
- We defined a number of different, intuitively appealing, standards for "agreement" or "disagreement" among the panelists about the rating for individual indications. Based on our preferred definitions, the panelists agreed on the ratings for about half of the indications, and disagreed for 10 to 30 percent of them, for all except one procedure. For colonoscopy (which was rated without group discussion), the proportions were reversed.
- By lumping together into an "equivocal" category those indications with middle ratings and those on which the panelists disagreed, we can establish three useful categories of indications: clearly inappropriate, equivocal, and clearly appropriate.
- For all procedures except colonoscopy, the proportion of equivocal ratings varied from 23 to 33 percent.
- Most of the remaining indications were clearly appropriate for coronary angiography and coronary artery bypass graft surgery. Most were clearly inappropriate for endoscopy and carotid endarterectomy. (As explained below, this is not in any sense an indictment of these two procedures.) The remaining indications were evenly divided for cholecystectomy.

MEASURES USED

For each indication, we used the median to measure the central tendency of the nine panelists' ratings and the mean absolute distance from the median to measure the dispersion of the ratings. We chose these measures because we think they are well suited to our appropriateness scale.

Our 1 through 9 scale is, strictly speaking, an ordinal scale. It ranks the excess or deficiency of benefit compared with risk. A 9 is always more appropriate than an 8, and an 8 is more appropriate than a 7. But because risk-benefit levels are not specified for each point on the scale, we cannot assert that the difference between a 9 and an 8 is necessarily the same as the difference between an 8 and a 7. This suggests that we should avoid such measures as means and standard deviations that treat the intervals as though they were equal.

For three reasons, however, we have not slavishly restricted ourselves to purely ordinal measures. First, our scale has some of the characteristics of an interval scale. The center of the scale (5) is well anchored at the point where risk equals benefit. Although the ends (1 and 9) are not precisely specified, they are anchored to some degree. At 1, risks exceed benefits by a sufficiently wide margin that the procedure should definitely not be done. At 9, it definitely should be done. Second, it is now well established that using interval measures on ordinal scales seldom has much effect on the results.[1] Third, to shun interval measures entirely would throw away information. A 4-point difference on our scale may not represent precisely four times as big a difference in the excess of benefit over risk as does a 1-point difference, but it almost certainly represents a bigger difference. A strictly ordinal measure would not distinguish between them.

OVERALL RESULTS

Table 3.1 summarizes the overall panel results. The indication counts on the first two lines refer to a basic indications list that excludes any explicit reference to comorbid conditions. In some cases, the panelists actually rated more items than the number shown. For cholecystectomy, they rated four times as many, including three replications of each primary indication for patients with higher levels of comorbidity. For the endoscopy and colonoscopy final ratings, the rating sheets included a few dozen items to allow the panelists to adjust their ratings for medium or high comorbidity (chapter by chapter rather than indication by indication). For carotid endarterectomy final ratings, the panelists rated 17 items for each chapter,

Table 3.1

OVERALL PANEL RESUTS FOR SIX PROCEDURES

Item	Procedure					
	ANGIO	CABS	CH	END	CO	CE
Number of initial indications	205	370	48	1685	1086	675
Number of final indications	300	480	49	1069	2862	864
Average initial MEDIAN	6.09	5.77	6.28	6.02	4.37	2.27
Standard deviation	(2.46)	(2.54)	(2.78)	(2.42)	(2.37)	(2.23)
Average final MEDIAN	6.65	5.56	4.76	3.77	5.44	2.66
Standard deviation	(2.51)	(2.65)	(3.23)	(3.07)	(2.73)	(2.50)
Average initial MAD	1.39	1.50	1.19	1.51	2.03	0.98
Standard deviation	(0.57)	(0.67)	(0.76)	(0.74)	(0.63)	(0.96)
Average final MAD	0.97	1.12	1.05	1.28	1.93	0.93
Standard deviation	(0.55)	(0.73)	(0.87)	(0.85)	(0.70)	(0.88)

NOTE: MEDIAN is the median of the nine panelists ratings.
MAD is the mean absolute deviation from the median.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

[1]See, for example, Moses, Emerson, and Hosseini (1984).

to be used to adjust their appropriateness ratings for differences in patient quality of life and surgeon experience in performing the procedure.

The average median ratings differed substantially among the procedures and between some initial and final ratings. We caution the reader, though, not to attribute too much meaning to these numbers. They are meant to be broadly descriptive, but they do not support comparisons. The low ratings for carotid endarterectomy, for example, do not necessarily mean that this procedure is somehow less appropriate than the others. Nor does the substantial decrease from initial to final ratings for endoscopy necessarily mean that the panelists changed their minds about the appropriateness of doing that procedure.

The reason for these strong cautions is that the average median ratings depend on the makeup of the indications list. At the extreme, adding a large number of unusual indications for which the procedure is almost never done in practice could greatly lower the average median rating. Yet this was precisely what was required for the purposes of the larger study. Also, splitting a batch of low-rated indications in two would double the number of low ratings and hence lower the average. To make valid comparisons between procedures or between initial and final ratings, we would have to know the relative frequency with which the indications occur in practice. We could then use the relative frequencies as weights in calculating comparable averages.

The average dispersion measures (mean absolute deviations from the median, or MAD in the tables) were mostly in the 1 to 2 rating point range. A dispersion of a little less than 1 point would occur, for example, if four panelists assigned ratings of 4, one assigned a 5, and four assigned 6s. Alternatively, it would occur if the ratings were eight 1s and one 9. A dispersion of almost 2 would occur with ratings like four 3s, one 5, and four 7s, or like seven 1s and two 9s.

The caution against making comparisons applies to some extent to the dispersion measures, as well as to the medians. Nevertheless, we think it is worth noting that the dispersion decreased from the initial to the final ratings for all six procedures. Certainly the dispersion measures, like the medians, depend on the makeup of the indications list. But the revisions in the indications lists between the initial and final ratings were *intended* in part to increase agreement. To judge from the dispersion measures in Table 3.1, the revisions were successful.

COMPARISON OF INITIAL AND FINAL RATINGS

We can legitimately compare initial and final ratings for indications that did not change during the panel meetings. As shown in Table 3.2, these indications constituted a fairly small fraction of the final indications for most of the procedures. The average medians for these indications decreased significantly for all of the procedures except coronary angiography. The dispersion decreased significantly for all six procedures. For most of the procedures, the decrease was about one third of the original amount of dispersion. This suggests that our panel process was successful as a method to bring physicians closer to agreement. But all parts of the panel process may not have been essential. The dispersion decreased for colonoscopy without any group discussion of the list of indications.

We present a slightly more complicated, but still legitimate, comparison of initial and final ratings in Table 3.3. The indications in this table were split in two between the initial and final ratings. For example, 83 coronary angiography indications were split to make 166 indications on the final list.

Table 3.2

PANEL RESULTS FOR INDICATIONS THAT WERE NOT CHANGED DURING THE PANEL PROCESS

Item	Procedure					
	ANGIO	CABS	CH	END	CO	CE
Number of unchanged indications	115	90	33	98	187	246
Average initial MEDIAN	5.95	5.52	5.79	3.77	4.34	1.28
Average final MEDIAN	6.46	4.88	4.88	3.48	3.68	1.20
Average change in MEDIAN	0.509	−0.639	−0.909	−0.286	−0.660	−0.077
Standard error	(0.076)	(0.157)	(0.223)	(0.132)	(0.101)	(0.025)
Average initial MAD	1.37	1.58	1.29	1.39	1.91	0.52
Average final MAD	0.86	1.06	0.98	0.83	1.36	0.34
Average change in MAD	−0.517	−0.639	−0.314	−0.559	−0.550	−0.173
Standard error	(0.031)	(0.157)	(0.115)	(0.066)	(0.046)	(0.018)

NOTE: MEDIAN is the median of the nine panelists ratings.
MAD is the mean absolute deviation from the median.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

We divided the final indications (after the split) into those we expected to have the higher rating (branch H) and those we expected to have the lower rating (branch L).[2] Unless the panelists changed their views on appropriateness of these items, the average median for branch H should be at least as high as the average inital rating, and the average for branch L should be at least as low. The expected pattern prevailed for all of the procedures except endoscopy. The changes for branch L were all negative or insignificantly different from zero. The changes for branch H were all positive or insignificantly different from zero except for endoscopy. Apparently, the discussion did lead the panelists to conclude that these indications for endoscopy were substantially less appropriate than they originally thought.

The changes in dispersion were negative or insignificantly different from zero for both branches of the split, indicating an improvement in agreement. The improvement was, however, generally somewhat smaller than it was for indications that stayed the same. The panelists had distributions of initial ratings for the indications that split in front of them when they rated the two branches of the split. This was apparently sufficient to improve agreement on both branches of the split, but not as much as if the indications had not changed.

[2]It was almost always obvious which branch should have the higher rating. For example, when indications for carotid endarterectomy that described a 50 through 99 percent stenosis of the ipsilateral artery were divided into separate indications for 50 through 69 percent and 70 through 99 percent, it was clear that the 70 through 99 percent branch should have the higher rating.

Table 3.3

PANEL RESULTS FOR INDICATIONS THAT SPLIT IN TWO

Item	Procedure					
	ANGIO	CABS	CH	END	CO	CE
Number of indications that split in two	83	19	6	120	133	249
Average initial MEDIAN	6.21	6.61	5.67	6.17	4.99	3.33
Average final MEDIANH	7.48	6.76	5.50	4.57	5.56	3.95
Average final MEDIANL	6.37	5.74	3.00	3.80	4.70	2.76
Average change in MEDIANH	1.27	0.158	−0.167	−1.60	0.571	0.618
Standard error	(0.126)	(0.230)	(0.307)	(0.240)	(0.188)	(0.071)
Average change in MEDIANL	0.157	−0.868	−2.67	−2.37	−0.293	−0.570
Standard error	(0.112)	(0.335)	(0.760)	(0.246)	(0.178)	(0.067)
Average initial MAD	1.41	1.19	0.98	1.83	2.19	1.40
Average final MADH	0.99	1.05	0.98	1.48	1.84	1.20
Average final MADL	1.05	0.87	1.46	1.22	1.95	0.96
Average change in MADH	−0.423	−0.139	0.000	−0.355	−0.348	−0.193
Standard error	(0.066)	(0.889)	(0.099)	(0.078)	(0.067)	(0.033)
Average change in MADL	−0.360	−0.327	0.481	−0.615	−0.239	−0.437
Standard error	(0.040)	(0.151)	(0.549)	(0.085)	(0.067)	(0.032)

NOTE: MEDIAN is the median of the nine panelists ratings.
MAD is the mean absolute deviation from the median.
H suffix refers to the branch of the split expected to have the higher rating.
L suffix refers to the branch of the split expected to have the lower rating.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

DIFFERENCES IN RATINGS BY SPECIALITY

Table 3.4 reports on median ratings by three specialty groupings: medical generalists, medical specialists, and surgeons. (See Table 2.1 for more information on the composition of the specialty groups.) The radiologist on each panel constituted a fourth group of one; we do not report his ratings so as not to disclose information about individuals. (The radiologists' ratings are included in the overall median, of course.)

The ratings are shown as average deviations of the median rating for each speciality from the overall median for all the panelists. Thus, for example, the median initial rating by medical generalists was, on average, 1.49 points lower than the overall median for the initial coronary angiography indications.

The ratings by specialty group were often significantly different from the overall ratings, but consistent patterns are hard to discern. Two observations stand out, however: Final ratings by the medical specialists were never more than half a rating point from the overall median, and the surgeons rated all three surgical procedures significantly higher than the overall median.

Table 3.4

AVERAGE DEVIATION FROM OVERALL MEDIAN BY SPECIALTY

Specialty	Procedure					
	ANGIO	CABS	CH	END	CO	CE
Medical generalists						
Initial rating	−1.49	−1.07	−0.66	−0.42	0.90	0.26
Standard error	(0.093)	(0.078)	(0.150)	(0.035)	(0.057)	(0.056)
Final rating	−0.60	−0.40	−0.27	0.26	0.94	−0.43
Standard error	(0.050)	(0.044)	(0.136)	(0.042)	(0.029)	(0.037)
Medical specialists						
Initial rating	0.65	0.01	0.16	0.47	−0.86	−0.39
Standard error	(0.834)	(0.068)	(0.172)	(0.043)	(0.049)	(0.064)
Final rating	0.11	−0.02	0.18	0.32	−0.10	−0.48
Standard error	(0.055)	(0.040)	(0.159)	(0.040)	(0.032)	(0.049)
Surgeons						
Initial rating	−0.50	0.93	0.49	0.38	0.44	0.45
Standard error	(0.078)	(0.087)	(0.159)	(0.036)	(0.053)	(0.036)
Final rating	0.09	0.65	1.02	−0.21	−1.94	0.46
Standard error	(0.066)	(0.070)	(0.251)	(0.044)	(0.043)	(0.030)

NOTE: ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

DISTRIBUTION OF RATINGS AND DEVIATIONS

Table 3.5 shows the distribution of indications by median rating. For each procedure, all of the possible medians were represented. For the two cardiology procedures, the ratings were skewed toward the appropriate end of the scale. For endoscopy and carotid endarterectomy, they were skewed toward the inappropriate end.

The table also shows the average dispersion measure for the indications with each median rating. Dispersion was lower at the ends of the scale than in the middle. This may not be surprising, since, for example, we cannot get a median of 1 unless at least five panelists agreed on a 1 rating. It is, however, not simply an artifact of the way we measure dispersion; the maximum possible mean absolute deviation is just the same for a median of 1 as it is for a median of 5.[3]

DEFINING AGREEMENT AND DISAGREEMENT

Mean absolute deviation from the median conveniently measures the dispersion of individual panelists' ratings for a particular indication, but it does not easily help us to decide whether there is "agreement" or "disagreement" on that indication. One problem is that it is a continuous measure; any dividing line between agreement and disagreement is necessarily

[3]For a median of 1, maximal dispersion occurs if five panelists assign a 1 rating and four assign a 9. That results in a dispersion measure (mean absolute deviation from the median) of $(4 \times 8)/9$, or about 3.6. For a median of 5, maximal dispersion occurs if only one panelist assigns a 5 rating while four assign a 1 and four assign a 9. That results in a dispersion measure of $[(4 \times 4) + (4 \times 4)]/9$, which again is 3.6.

Table 3.5

DISTRIBUTION OF MEDIAN RATINGS AND DISPERSION

Procedure		Median Rating 1	2	3	4	5	6	7	8	9
ANGIO	Number	18	19	10	11	15	35	46	50	96
	Average MAD	0.26	1.07	1.42	1.57	1.33	1.55	1.32	1.01	0.51
	Standard error	(0.072)	(0.114)	(0.108)	(0.114)	(0.107)	(0.051)	(0.054)	(0.046)	(0.035)
CABS	Number	37	43	45	40	45	54	55	58	103
	Average MAD	0.48	0.96	1.35	1.69	1.71	1.65	1.46	1.11	0.38
	Standard error	(0.085)	(0.091)	(0.066)	(0.114)	(0.101)	(0.060)	(0.062)	(0.061)	(0.041)
CH	Number	14	2	6	3	4	2	3	3	12
	Average MAD	0.34	1.94	1.54	2.33	2.06	2.44	1.52	1.22	0.44
	Standard error	(0.118)	(0.228)	(0.168)	(0.128)	(0.167)	(0.333)	(0.296)	(0.321)	(0.142)
END	Number	430	114	77	39	72	39	67	92	128
	Average MAD	0.85	1.61	2.08	2.11	2.15	2.12	1.80	1.48	0.55
	Standard error	(0.037)	(0.049)	(0.057)	(0.076)	(0.058)	(0.074)	(0.073)	(0.062)	(0.051)
CO	Number	210	430	266	165	279	214	334	471	493
	Average MAD	1.04	1.84	2.09	2.41	2.33	2.43	2.24	2.04	1.37
	Standard error	(0.036)	(0.025)	(0.031)	(0.033)	(0.032)	(0.036)	(0.030)	(0.029)	(0.033)
CE	Number	536	30	53	27	78	33	39	37	31
	Average MAD	0.40	1.57	1.92	1.85	2.04	2.21	1.74	1.47	0.98
	Standard error	(0.025)	(0.084)	(0.058)	(0.062)	(0.052)	(0.082)	(0.079)	(0.097)	(0.112)

NOTE: ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

arbitrary. Another problem is that it does not have much intuitive content; it is hard to know where is a reasonable place to draw that line. Here we suggest and apply four definitions of agreement and disagreement that have more intuitive content than the mean absolute deviation.

In defining agreement and disagreement, we adopted the following premises. (1) We seek definitions that most people would accept as reasonable. Consequently, we try hard not to include too much in the definitions. Some people might favor broader definitions, but we hope that few would advocate stricter ones. (2) Widely acceptable definitions of agreement and disagreement will leave a middle grey area that is not obviously one or the other. (3) Our 9-point scale meaningfully divides into three 3-point regions. Ratings from 1 to 3 say that risks outweigh benefits and the procedure should not be done. Ratings from 4 to 6 say that risks and benefits are roughly equal and doing the procedure is questionable. Ratings from 7 to 9 say that benefits outweigh risks and the procedure should be done. (4) In an imperfect world, we may need to be content with less than perfect unanimity. Although we base half of our definitions on the ratings of all nine panelists, for the other half we discard one rating at either extreme and use only the remaining seven.

Four Definitions of Agreement

We propose two "conceptions" of agreement. Both can be applied using either nine ratings or seven ratings, yielding a total of four definitions.

The first, stricter, conception is: The raters agreed if all of the ratings were within one of the above-specified 3-point regions—1 to 3, 4 to 6, or 7 to 9. We would interpret this to mean that all of the raters agreed to one of the following statements: The procedure should not be done, doing it is questionable, or it should be done.

The second conception is somewhat more relaxed: The raters agreed if all of the ratings were within *any* 3-point range, even if that range straddled the boundary between two of the regions specified above. If all of the ratings were within the range 3 to 5, for example, there was agreement according to this second conception, but not according to the first.

Four definitions of agreement result:[4]

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

A9R: All nine of the ratings fell within any 3-point range.

A7S: After discarding one extreme high and one extreme low rating, the remaining seven ratings all fell within a single 3-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 3-point range.

(The codes for each definition are mnemonic: "A" for agreement, "9" or "7" for the number of ratings used, "S" or "R" for the stricter or more relaxed conception.)

Table 3.6 shows the percent of indications on which the panelists agreed, using our four definitions of agreement. On any definition, colonoscopy showed much lower agreement than

Table 3.6

PERCENT OF INDICATIONS ON WHICH PANELISTS AGREED, USING FOUR DIFFERENT DEFINITIONS OF AGREEMENT

		Procedure											
		ANGIO		CABS		CH		END		CO		CE	
Definition of Agreement		I	F	I	F	I	F	I	F	I	F	I	F
A9S:	9 ratings, strict	12.2	28.0	7.8	21.7	29.2	38.8	16.6	25.4	1.8	2.6	39.9	40.9
	Standard error	(2.3)	(2.6)	(1.4)	(1.9)	(6.6)	(7.0)	(0.9)	(1.3)	(0.4)	(0.3)	(1.9)	(1.7)
A9R:	9 ratings, relaxed	12.7	28.7	7.8	23.1	29.2	38.8	16.6	25.4	1.8	2.6	39.9	40.9
	Standard error	(2.3)	(2.6)	(1.4)	(1.9)	(6.6)	(7.0)	(0.9)	(1.3)	(0.4)	(0.3)	(1.9)	(1.7)
A7S:	7 ratings, strict	21.0	50.0	21.4	35.4	43.8	49.0	27.8	41.3	10.6	14.6	55.6	53.4
	Standard error	(2.9)	(2.9)	(2.1)	(2.2)	(7.2)	(7.2)	(1.1)	(1.5)	(0.9)	(0.7)	(1.9)	(1.7)
A7R:	7 ratings, relaxed	22.0	56.3	23.2	41.9	43.8	53.1	28.6	41.6	11.1	14.9	55.6	53.8
	Standard error	(2.9)	(2.9)	(2.2)	(2.2)	(7.2)	(7.2)	(1.1)	(1.5)	(1.0)	(0.7)	(1.9)	(1.7)

NOTE: See text for full definitions of agreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

[4]All four definitions are slightly complicated by the problem of missing ratings. As shown in App. E for all six procedures, some panelists failed to rate a few of the indications. The fewer the ratings, the easier it is to satisfy any of our definitions of agreement. To avoid artificially inflating the agreement counts, we counted only indications rated by at least eight of the panelists.

any other procedure. There are several possible reasons for this. Colonoscopy was the only procedure for which final ratings were assigned without group discussion. It had by far the largest number of final indications, possibly producing less carefully considered ratings. Or there may simply have been a lower level of consensus on appropriate indications for use of colonoscopy than there was for the other procedures. In any event, since the panel process differed substantially for colonoscopy, our discussion disregards colonoscopy and focuses on the other five procedures.

Based on the definitions that use all nine ratings (A9S and A9R), the panelists agreed on the ratings for 20 to 40 percent of the indications, depending on the procedure. Using A9R rather than A9S makes surprisingly little difference; almost all of the ratings that clustered in any 3-point range fell within a single one of our regions (1 to 3, 4 to 6, or 7 to 9).

If we do not require unanimity (definitions A7S and A7R), agreement jumps by 16 percentage points on average, reaching 50 percent or better for three of the procedures.

Although we are focusing here on the final ratings, we note that a comparison of initial and final levels of agreement reinforces the conclusion that we reached above by comparing dispersion measures: The panel process resulted in increased agreement.

The results shown in Table 3.5 suggest that agreement may have been better on indications with high or low ratings. Results shown in Appendix Tables F.1 through F.3 confirm that that is true with one exception: For carotid endarterectomy, there was *less* agreement on indications with high ratings.

Four Definitions of Disagreement

As with agreement, we propose two conceptions of disagreement that can be used with either nine or seven ratings to yield four definitions.

The first conception is: The raters disagreed if at least one assigned a rating of 1 and at least one assigned a rating of 9. Surely no one would deny that this extreme polarization represents disagreement.

The second conception is more relaxed, but still, we think, compelling: The raters disagreed if at least one rating fell in the lowest 3-point region (1 to 3) and at least one in the highest (7 to 9). We would interpret this to mean that one thought the procedure should not be done for this indication and one thought that it should be.

Four definitions of disagreement result:[5]

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 3-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 3-point region (1 to 3) and at least one fell in the highest (7 to 9).

Table 3.7 shows the percent of indications that satisfied each of the four definitions of disagreement. Colonoscopy showed a high level of disagreement by any of the definitions, corresponding to the low level of agreement noted above. Again, our discussion focuses on the other five procedures.

[5]In much the same way that missing values make it easier to find agreement, they make it harder to find disagreement. Thus we do not need to worry about artificial inflation of the disagreement counts. If anything, our measures will slightly understate the true amount of disagreement.

Table 3.7

PERCENT OF INDICATIONS ON WHICH PANELISTS DISAGREED, USING FOUR DIFFERENT DEFINITIONS OF DISAGREEMENT

		Procedure											
		ANGIO		CABS		CH		END		CO		CE	
Definition of Disagreement		I	F	I	F	I	F	I	F	I	F	I	F
D9S:	9 ratings, strict	9.3	2.0	20.5	9.0	16.7	20.4	37.9	30.2	49.2	60.9	28.4	14.9
	Standard error	(2.0)	(0.8)	(2.1)	(1.3)	(5.4)	(5.8)	(1.2)	(1.4)	(1.5)	(0.9)	(1.7)	(1.2)
D9R:	9 ratings, relaxed	60.0	30.0	59.5	40.2	47.9	42.9	63.8	48.5	82.0	80.8	40.9	34.0
	Standard error	(3.4)	(2.7)	(2.6)	(2.2)	(7.3)	(7.1)	(1.2)	(1.5)	(1.2)	(0.7)	(1.9)	(1.6)
D7S:	7 ratings, strict	0.0	0.3	5.4	2.3	0.0	2.0	2.8	7.0	9.9	21.7	4.7	2.3
	Standard error	(0.0)	(0.3)	(1.2)	(0.7)	(0.0)	(2.0)	(0.4)	(0.8)	(0.9)	(0.8)	(0.8)	(0.5)
D7R:	7 ratings, relaxed	31.7	11.0	29.7	16.9	33.3	20.4	36.0	28.9	55.2	48.1	21.6	18.1
	Standard error	(3.3)	(1.8)	(2.4)	(1.7)	(6.9)	(5.8)	(1.2)	(1.4)	(1.5)	(0.9)	(1.6)	(1.3)

NOTE: See text for full definitions of agreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

Disregarding colonoscopy, the percentage of indications with at least one 1 and at least one 9 (D9S) ranged from 2 to 30; depending on the procedure, requiring at least two ratings at each extreme (D7S) reduced these percentages substantially, to a maximum of 7 for endoscopy. Between 30 and 50 percent of all indications had at least one 1 to 3 rating and at least one 7 to 9 (D9R). Requiring two ratings at each extreme (D7R) reduced the range to between 10 and 30 percent.

As we have come to expect by now, there was less disagreement in the final ratings than in the initial ratings, and there was usually less disagreement about low- and high-rated indications than about those with middle ratings (Appendix Tables F.4 through F.6).

A USEFUL CATEGORIZATION OF RATED INDICATIONS

How can we use the panel results to decide if a procedure is done appropriately for a particular indication? We propose to use the following scheme. Each indication can fall in one of three categories: clearly appropriate, equivocal, or clearly inappropriate. We classify the indication as "equivocal" for either of two reasons: The benefits and risks of doing the procedure are roughly the same (a median rating of 4 to 6),[6] or the panelists disagreed on the proper rating (according to one of the definitions discussed above). An indication is "clearly appropriate" if the panelists assigned a median rating in the 7 to 9 range without disagreement, and it is "clearly inappropriate" if they assigned a 1 to 3 rating without disagreement.[7]

[6]Non-integer values for medians may arise when the number of ratings is even. For purposes of this categorization, we use medians rounded to the nearest integer.

[7]A variant of this scheme is possible. Instead of assigning disagreed-upon indications to the equivocal category, we could assign indications that were *not* agreed upon. That is, we could treat indications in the grey area between agreement and disagreement as equivocal.

Table 3.8 shows the percentages of indications that fall in each of the three appropriateness categories when we use the fourth definition of disagreement (D7R: at least two 1 to 3 ratings and two 7 to 9 ratings).

Equivocal indications constituted over half of the total for colonoscopy. For the other five procedures, they constituted one quarter to one third of the total. For the cardiovascular procedures, equivocal indications were evenly divided between those with 4 to 6 ratings and those on which the panelists disagreed. For the other procedures, disagreement among the panelists was the usual reason for designation as equivocal.

The fractions of clearly appropriate and clearly inappropriate indications varied widely across procedures. There were more that were clearly appropriate for the two cardiology procedures, more that were clearly inappropriate for endoscopy and carotid endarterectomy, and they were evenly divided for cholecystectomy. Until we know the frequencies with which these procedures are actually performed for the various indications, however, comparisons across procedures are not legitimate; see the discussion at the beginning of this section.

Table 3.8

CATEGORIZATION OF FINAL INDICATIONS BY APPROPRIATENESS
(In percent)

Category	Procedure					
	ANGIO	CABS	CH	END	CO	CE
Clearly inappropriate						
Median 1-3,						
without disagreement[a]	14.0	24.6	38.8	45.1	22.1	67.0
Standard error	(2.0)	(2.0)	(7.0)	(1.5)	(0.8)	(1.6)
Equivocal						
Median 4-6,						
without disagreement	12.7	15.8	4.1	3.8	4.3	5.2
Any median,						
with disagreement	11.0	16.9	20.4	29.2	48.1	18.1
Total	23.7	32.7	24.5	33.0	52.4	23.3
Standard error	(2.5)	(2.1)	(6.1)	(1.4)	(0.9)	(1.4)
Clearly appropriate						
Median 7-9,						
without disagreement	62.3	42.7	36.7	21.9	25.4	9.7
Standard error	(2.8)	(2.2)	(6.9)	(1.3)	(0.8)	(1.0)

NOTE: ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

[a]"Disagreement" means at least two 1-3 ratings and at least two 7-9 ratings.

IV. DISCUSSION

We deal with several questions about the meaning and use of the panel results in this final section.

WHAT DID WE LEARN FROM THE PANELS?

Our panels demonstrated that physicians can rate the appropriateness of large numbers of indications for performing medical and surgical procedures. The lists of indications that they rated were much more detailed and comprehensive than any that have been attempted in the past.

The detailed lists of rated indications in the six procedure reports may interest researchers, but they are not a valid guide to current medical practice. Because we plan to use the ratings to investigate procedures done during 1981, we instructed the panelists to rate the indications in light of 1981 knowledge. Rapid development of some treatments (for example, intracoronary streptokinase in heart attacks and balloon angioplasty for obstruction of coronary arteries) has surely altered the appropriateness of some of the indications since 1981.

The panel process clearly brought the individual panelists' ratings closer together. Any way we look at it—comparing average dispersion, percent agreement, or percent disagreement—the final ratings tended to be closer together than the initial ratings. This was true whether we made the comparisons for all indications, or only for those that stayed the same or simply split in two. What is not clear is what caused the increased agreement. Was it careful reevaluation of initial opinions, or simply group-dynamic pressures to agree?[1]

Even in the final ratings, the panelists at best agreed on a little over half of the indications, and they clearly disagreed on up to 30 percent of them. These statements disregard colonoscopy, which had a much higher level of disagreement. They are based on the following relatively relaxed definitions of "agreement" and "disagreement": First, discard one extreme high and one extreme low rating for each indication, in order to reduce the influence of outliers. If the remaining seven ratings are clustered within any 3-point range of our 9-point scale, they agree. If at least one of them is in the 1 to 3 range (meaning the procedure should not be done) and at least one of them is in the 7 to 9 range (meaning it should be done), they disagree.

The average level of the ratings does *not* tell us anything about how any of the procedures are actually used. For example, the low average rating for carotid endarterectomy (Table 3.1) does not in any way suggest that this is a "bad" procedure. It may well be true that it is used almost exclusively for highly appropriate indications, and rarely or never for the lower-rated indications on our list. The larger study will provide data on this point.

On the other hand, the disagreement suggests that these procedures are sometimes done for equivocal indications, if, as seems likely, patients do in fact present with disagreed-upon indications. The argument is simple: A disagreed-upon indication is by definition an equivocal indication. It is also by definition an indication for which some physicians think the procedure should be done. If a patient presents to such a physician with such an indication, that physician will likely do the procedure or recommend that it be done.

[1]If it was the latter, then one could argue about whether the initial or the final ratings better reflect the range of clinical opinion about appropriate procedure use. Adopting the initial ratings as a standard would substantially increase the number of equivocal indications.

The final, medical-records, stage of our larger project is designed to measure the extent to which procedures are actually done for indications that the panel results suggest are inappropriate or equivocal. This use of the panel results is outlined below.

HOW WILL WE USE THE PANEL RESULTS IN A MEDICAL RECORDS STUDY?

The central question for our Health Services Utilization Study is: To what extent does greater inappropriate or equivocal use of procedures in high-use areas explain geographic differences in use rates? To answer the question, we need to know both the indications for which procedures were done in different geographic areas, and the appropriateness of doing procedures for those indications. The panel ratings provide a way to measure appropriateness. Abstracts of medical records will determine the indications for which procedures were done.

In the final stage of HSUS, we are studying three procedures (coronary angiography, upper gastrointestinal endoscopy, and carotid endarectomy) in high-, medium-, and low-use areas.[2] We are currently abstracting a random sample of approximately 500 medical records for each procedure studied in each area to determine the exact indications for which the procedures were done during 1981. We may find more than one indication in a single medical record. If so, the highest-rated indication is the best justification for doing the procedure. Using the highest-rated indication in each record, we will count up the number of procedures that fall in each appropriateness category: appropriate, equivocal, or inappropriate. We will then statistically compare the proportion of procedures done for equivocal and inappropriate reasons in the different areas.

Several variants of the analysis are possible. Less-than-appropriate indications can include only those rated inappropriate by the panelists, or they can include those rated equivocal as well. As we noted in Sec. III, indications can be assigned to the equivocal category based on any of our definitions of disagreement. Some sensitivity analysis is in order, and we will do it; however, we do not expect that the results of the comparisons will be much affected by the particular choice of definition of a less-than-appropriate indication.

HOW MIGHT ADDITIONAL RATINGS BASED ON CURRENT KNOWLEDGE BE USED?

Physicians today face mounting pressure to use procedures only when clinically valid criteria indicate that they are appropriate. Much of the pressure is due to the government's limits on health care financing (Aaron and Schwartz, 1984), businesses' efforts to curb health insurance costs (Thurow, 1985), increasing concern over large regional variations in procedure use (Chassin, Brook, Park, et al., 1986), and the establishment of PROs (Lohr, 1985).

Where can physicians turn for help in deciding what indications for using a procedure are generally accepted as appropriate? The medical literature is only a partial guide; expensive, time-consuming, randomized, clinical trials can never investigate the effectiveness of procedures in more than a small fraction of the situations in which they might be used. The NIH consensus statements offer some guidance for some procedures, but are generally too broadly stated to be of much help in individual cases (Rennie, 1981).

---

[2]Cost constraints precluded studying all six procedures.

Our method could be used to generate ratings based on current knowledge and practice (rather than 1981). Individual physicians might use such panel ratings as an aid in their own practices. If the panel rated an indication inappropriate, the physician might seriously question the advisability of doing a procedure for that indication. If the rating were equivocal, the physician might pause to ask what unusual circumstances justify use of a procedure in this particular case.

The panel ratings could never absolutely proscribe (or prescribe) procedures for individual patients. Each patient is unique in ways that might sometimes legitimately influence the physician's decision to use a procedure. However, patients also differ in ways that should not, but in fact do, influence treatment decisions (Phibbs, 1979). Because the panel ratings are based solely on clinically relevant factors, they could serve as a valuable check on decisions in individual cases.

Furthermore, indications about which the expert panelists disagree highlight areas about which present knowledge is inconclusive. If these are indications that occur frequently in practice, they become promising candidates for clinical research.

WHY DOES THERE APPEAR TO BE SO MUCH DISAGREEMENT?

Several explanations are possible for the amount of disagreement recorded:

1. Perhaps the appearance reflects reality; there may really be substantial disagreement about the value of these procedures for many indications. This is a plausible reason in light of the large number of indications rated and the paucity of controlled clinical trials to establish scientifically the value of the six procedures in different circumstances. (See the six procedure reports.)

2. Perhaps panelists had in mind different "average groups of patients presenting to an average U.S. physician." This is plausible, since their perceptions of what constitutes an average group presumably depend to some exent on what the panelists see in their own practices. It is possible that what is appropriate for one group of patients is inappropriate for another group with the same indication.

3. Perhaps the indications were not properly framed—for example, they did not match clinically relevant categories, or were not tightly enough defined to avoid the "different group" problem. However, the fact that the panelists substantially revised the initial lists and unanimously accepted the final lists suggests that the final lists should define indications that are well suited for rating their appropriateness.

4. Perhaps panelists differed in their understanding of what the indications meant. But the definitions were supplied both in writing and orally at the meetings, and there was plenty of opportunity to discuss questions and clarify the definitions.

5. Perhaps the indications with disagreement do not arise in practice. They may be hypothetical situations with which the panelists have had no experience. There may be a much higher degree of agreement on the indications that do occur frequently in practice. However, the panelists did drop some indications from the initial lists on the grounds that they never occur in practice, suggesting that most of the final indications do occur at least occasionally.

Our tentative conclusion is that there really is disagreement about the value of these procedures for many indications, and that the disagreement reflects the lack of detailed evidence about the circumstances in which these procedures are efficacious. This conclusion is consistent with the view that uncertainty underlies much of the geographic variation in procedure

use (Eddy, 1984). Uncertainty about the efficacy of a procedure would allow different regional practice styles to develop, some of which use the procedure more than others (Wennberg, 1984). In that case, it seems likely that some physicians are doing procedures for indications that others would rate as inappropriate. If many procedures are done for equivocal or inappropriate reasons, their curtailment might save health care resources without compromising quality of care.

REFERENCES

Aaron, J. H., and W. B. Schwartz, *The Painful Prescription: Rationing Health Care*, Brookings Institution, Washington, D.C., 1984.

Brook, R. H., K. N. Lohr, M. R. Chassin et al., "Geographic Variations in the Use of Services: Do They Have Any Clinical Significance?" *Health Affairs*, 1984, 3(4):63.

Chassin, M. R., R. H. Brook, R. E. Park et al., "Variations in the Use of Medical and Surgical Services by the Medicare Population," *New England Journal of Medicine*, 1986, 314:285–290.

Chassin, M. R., J. Kosecoff, R. E. Park et al., *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Coronary Angiography*, The Rand Corporation, R-3204/1-CWF/HF/HCFA/PMT/RWJ, 1986.

Chassin, M. R., R. E. Park, A. Fink et al., *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Coronary Artery Bypass Graft Surgery*, The Rand Corporation, R-3204/2-CWF/HF/HCFA/PMT/RWJ, 1986.

Eddy, D. M., "Variations in Physician Practice: The Role of Uncertainty," *Health Affairs*, 1984, 3(4):74–89.

Fink, A., J. Kosecoff, M. Chassin, and R. H. Brook, "Consensus Methods: Characteristics and Guidelines for Use," *American Journal of Public Health*, 1984, 74(9):979–983.

Kahn, K. L., C. P. Roth, J. Kosecoff et al., *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Diagnostic Gastronintestinal Endoscopy*, The Rand Corporation, R-3204/4-CWF/HF/HCFA/PMT/RWJ, forthcoming.

Lohr, K., *Peer Review Organizations: Quality Assurance in Medicare*, The Rand Corporation, P-7125, 1985.

Merrick, N. J., R. E. Park, J. Kosecoff, et al., *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Carotid Endarterectomy*, The Rand Corporation, R-3204/6-CWF/HF/HCFA/PMT/RWJ, forthcoming.

Merrick, N. J., A. Fink, R. E. Park et al., "Derivation of Clinical Indications for Carotid Endarterectomy by an Expert Panel," *American Journal of Public Health*, 1986, forthcoming.

Moses, L. E., J. D. Emerson, and H. Hosseini, "Analyzing Data From Ordered Categories," *New England Journal of Medicine*, 1984, 311:442–448.

Park, R. E., A. Fink, R. H. Brook et al., "Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures," *American Journal of Public Health*, July 1986.

Phibbs, B., "The Abuse of Coronary Angiography," *New England Journal of Medicine*, 1979, 301:1394–1396.

Rennie, D., "Consensus Statements," *New England Journal of Medicine*, 1981, 304: 665–666.

Roth, C. P., K. L. Kahn, A. Fink et al., *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Fiberoptic Colonoscopy*, The Rand Corporation, R-3204/5-CWF/HF/PMT/RWJ, forthcoming.

Solomon, D. H., R. H. Brook, A. Fink et al., *Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Cholecystectomy*, The Rand Corporation, R-3204/3/CWF/HF/PMT/RWJ, 1986.

Thurow L. C., "Medicine Versus Economics," *New England Journal of Medicine,* 1985, 313:611–614.

Wennberg, J. E., "Dealing with Medical Practice Variations: A Proposal for Action," *Health Affairs*, 1984, 3(4):6–32.

—— and A. M. Gittelsohn, "Variations in Medical Care Among Small Areas," *Scientific American*, 1982, 246:120–134.

Appendix A

THE RAND/UCLA HEALTH SERVICES UTILIZATION STUDY (HSUS)

Geographic areas differ sharply in their use of health services, a finding repeatedly documented by researchers analyzing data from areas as large as major U.S. census regions to areas as small as New England towns. The differences occur across a wide range of health services, from hospital days to physician office visits, and include virtually all surgical procedures.

These data have been used by legislators and policymakers to accuse the medical profession of overuse of services; their assumption is that the high-use areas use too many health services. Up to now, however, research has been unable to explain the significance of these geographic variations. We simply do not know whether high-use areas provide more services than low-use areas because more patients with the relevant illnesses reside there, whether low-use areas provide too few services and high-use areas are just right, or whether there is some overuse in high-use areas.

While randomized, controlled trials might be the most scientifically correct way to study the circumstances in which health services ought to be employed, the cost and time required to perform the large number of trials needed to answer all the important questions in this area do not make this approach feasible. The Health Services Utilization Study combines epidemiologic data, clinical information, and a consensus of expert physicians to illuminate how the use of health services differs among geographic areas. Four major foundations and the Health Care Financing Administration have awarded grants to The Rand Corporation and UCLA for the purpose of carrying out the study. Several major professional organizations and individual academic and community doctors have endorsed the study's purposes and methods.

HSUS is being conducted in four stages. In the first, we collected health claims data from Medicare Part B insurance carriers in eight states. From these data, we calculated age- and sex-adjusted, population-based, rates of use for a wide variety of health services. In the second stage, for a selected number of services (coronary angiography, upper gastrointestinal endoscopy, colonoscopy, cholecystectomy, coronary bypass surgery, and carotid endarterectomy), we performed reviews of the literature to summarize what is known concerning their efficacy, complication, indications, use, and cost. These procedures were selected because they are costly and commonly done, and because our data showed large geographic differences in their use.

In the third stage, we convened panels of expert physicians to review the data assembled in the first two stages. The principal task of the panels was to rate the appropriateness of every indication for performing each of the six procedures for which we have done literature reviews.

For the fourth stage of the study, we have selected a subset of the six procedures analyzed through stage III. These are: coronary angiography, upper gastrointestinal endoscopy, and carotid endarterectomy. We have selected five sites from among the thirteen geographic areas for which we have calculated population-based rates of use. The areas differ in their utilization of the three procedures, with some being relatively high, average, or low. Using medical-records abstracting techniques, we will review a representative sample of medical records in

two areas for all three procedures, and in one additional area for each procedure. The primary goal of this part of HSUS is to identify the indications for which these procedures were performed.

Our plan is to compare the distribution of indications for use of procedures in high-, average-, and low-use areas. Using the appropriateness ratings provided by the expert panels in stage III, we will be able to assess the extent to which high-use areas employ these procedures more often for the lowest-rated indications. If they do, then overuse is a likely explanation for at least some of the geographic differences we observe in rates of health service use. If the distributions are similar across areas, the study results may lend support to the argument that variations in disease incidence can explain a large part of the differences. Whatever the specific results of the study, it will provide for the first time a clinical information base that can be used to study specifically how patients who receive services in high-use areas differ from those who receive services in low-use areas.

Appendix B

HSUS POLICY ADVISORY BOARD AND TECHNICAL PANEL MEMBERS

POLICY ADVISORY BOARD

David H. Solomon, M.D., Chairman
Professor of Medicine
University of California, Los Angeles
Co-Principal Investigator

Organizational Representatives

Gerald E. Bisbee, Ph.D. (1983–1984)
Former President, Hospital Research and Educational Trust
(Representative of the American Hospital Association)

Linda K. Demlo, Ph.D. (1984–1985)
Former Director of the Center for Health Services Research
Former Associate Professor of the Graduate Program in
Hospital & Health Administration, University of Iowa
(Representative of the American Hospital Association)

John M. Lowe, M.D. (1985–)
Vice President, Hospital Research and Educational Trust
(Representative of the American Hospital Association)

Ernie J. Chaney, M.D.
Past President of the American Academy of Family Physicians
(Representative of the American Academy of Family Physicians)

John H. Dawson, Jr., M.D.
Trustee, American Medical Association
(Representative of the American Medical Association)

Ronald G. Evens, M.D.
Director of the Mallinckrodt Institute of Radiology
Washington University School of Medicine
(Representative of the American College of Radiology)

Paul F. Griner, M.D.
Associate Chairman
Department of Medicine, University of Rochester
Governor, Upstate New York
American College of Physicians
(Representative of the American College of Physicians)

Anthony M. Imparato, M.D. (1985–)
Professor of Surgery
Director, Division of Vascular Surgery
New York University Medical Center
President, Society for Vascular Surgery
(Representative, Joint Council of the Society for Vascular Surgery and the North American Chapter of the International Society for Cardiovascular Surgery)

Douglas B. McGill, M.D.
Chairman of the Division of Gastroenterology
Mayo Clinic
(Representative of the American Gastroenterological Association)

Melvin Schapiro, M.D. (1985–)
Associate Professor, Clinical Medicine
University of California, Los Angeles
President, American Society for Gastrointestinal Endoscopy
(Representative of the American Society of Gastrointestinal Endoscopy)

Individual Members

Robert J. Adolph, M.D.
Chairman of Cardiovascular Norms Committee
American College of Cardiology
Professor of Medicine
Director, Cardiac Research Laboratory
University of Cincinnati

John D. Alexander, M.D.
Director of Cardiology and Coronary Care
Downey Community Hospital
Executive Committee, Council on Clinical Cardiology
American Heart Association

Richard Berman, M.P.H. (1984–)
Executive Vice President
New York University Medical Center

Neil J. Elgee, M.D.
Clinical Professor of Medicine
University of Washington School of Medicine

Donald S. Fredrickson, M.D.
Past Director, National Institutes of Health
President, Howard Hughes Medical Institute

Francis D. Moore, M.D.
Professor of Surgery, Emeritus
Harvard Medical School

Marvin J. Shapiro, M.D.
Past Chairman of the Board
Blue Shield Association

Thomas L. Stern, M.D.
Corporate Development
Family Health Foundation of America

Richard S. Wilbur, M.D.
Executive Vice President
Council of Medical Specialty Societies

HSUS CARDIOVASCULAR PANEL

Mortimer Buckley, M.D.
Chief of Cardiovascular Surgical Unit
Massachusetts General Hospital
Boston, Massachusetts

Ernie Chaney, M.D.
Associate Professor
University of Kansas School of Medicine
Wichita, Kansas

Gottlieb Friesinger, M.D.
Director, Division of Cardiology
Vanderbilt University School of Medicine
Nashville, Tennessee

J. Ward Kennedy, M.D.
Director, Division of Cardiology
University of Washington School of Medicine
Seattle, Washington

John Kirklin, M.D.
Professor and Director
Division of Cardiothoracic Surgery
University of Alabama Medical Center
Birmingham, Alabama

Ceylon Lewis, M.D.
Clinical Professor of Medicine
University of Oklahoma Medical College
Tulsa, Oklahoma

Joe Noble, M.D.
Associate Clinical Professor of Medicine
Indiana University School of Medicine
Consulting Cardiologist, St. Vincent's Hospital
Indianapolis, Indiana

Harold Sox, Jr., M.D.
Chief of General Internal Medicine Ambulatory Care
Palo Alto VA Medical Center
Palo Alto, California

William Wilson, M.D.
Director of Radiology
Long Beach Memorial Medical Center
Long Beach, California

HSUS GASTROINTESTINAL PANEL

John R. Amberg, M.D.
Professor of Radiology
University of California
San Diego, California

Arthur Feinberg, M.D.
Professor of Clinical Medicine
Cornell University Medical College
New York City, New York
Associate Director
Department of Medicine
North Shore University Hospital
Manhasset, New York

Martin Greene, M.D.
Clinical Associate Professor of Medicine
University of Washington School of Medicine
Seattle, Washington
Chairman, Gastroenterology
Swedish Hospital Medical Center
Seattle, Washington

Thomas Hendrix, M.D.
Paulson Professor of Gastroenterology
Director, Division of Gastroenterology
Johns Hopkins University School of Medicine
Baltimore, Maryland

Edwin Maynard, III, M.D.
Assistant Clinical Professor of Medicine
Harvard Medical School
Cambridge, Massachusetts

Robert McCurdy, M.D.
Associate Clinical Professor of Surgery
University of Colorado School of Medicine
Denver, Colorado

John Morrissey, M.D.
Vice-Chairman and Professor,
Head of Section of Gastroenterology
Department of Medicine
University of Wisconsin Medical School
Madison, Wisconsin

Robert Tyson, M.D.
Professor of Surgery
Past Chairman, Department of Surgery
Temple University Health Science Center
Philadelphia, Pennsylvania

James Weber, M.D.
Clinical Professor of Family Medicine
University of Arkansas Medical Science Center
Little Rock, Arkansas

HSUS CAROTID ENDARTERECTOMY PANEL

William H. Baker, M.D.
Professor and Chief
Section of Periphral Vascular Surgery
Loyola University Medical Center
Maywood, Illinois

Richard R. Brummett, M.D.
Clinical Assistant Professor of Family Practice
University of Kansas School of Medicine
Wichita, Kansas

Randall M. Cebul, M.D.
Assistant Professor of Medicine
Henry J. Kaiser Family Foundation Faculty Scholar
Department of Medicine, Section of General Medicine
University of Pennsylvania Medical School
Philadelphia, Pennsylvania

John Paul Conomy, M.D.
Chairman, Department of Neurology
Cleveland Clinic Foundation
Associate Professor of Medicine (Neurology)
Case Western Reserve University
Cleveland, Ohio

Robert W. Hobson II, M.D.
Professor of Surgery
Vice-Chairman, Department of Surgery
Chief, Section of Vascular Surgery
University of Medicine and Dentistry of New Jersey
New Jersey Medical School
Newark, New Jersey

James T. Robertson, M.D.
Professor and Chairman
Department of Neurosurgey
University of Tennessee Center for the Health Sciences
Memphis, Tennessee

John O. F. Roehm, Jr., M.D.
Clinical Professor of Radiology
Baylor College of Medicine
Chief, Section of Cardiovascular Radiology
Methodist Hospital
Houston, Texas

Hugh Henry Trout III, M.D.
Associate Professor of Surgery
George Washington University Medical School
Washington, D.C.

Jack Page Whisnant, M.D.
Professor of Neurology
Mayo Clinic and Mayo Medical School
Rochester, Minnesota

Appendix C

INITIAL INDICATIONS

Chapter Number	Chapter Title	Number of Indications
	INITIAL INDICATIONS FOR CORONARY ANGIOGRAPHY	
1.	Asymptomatic patients	28
2.	Chest pain of uncertain origin	15
3.	Chronic stable angina	68
4.	Unstable angina	10
5.	During acute myocardial infarction	3
6.	Within six months of an acute myocardial infarction	50
7.	Sudden death survivors	11
8.	Following coronary artery bypass graft surgery	16
9.	Other	4
	Total	205
	INITIAL INDICATIONS FOR CORONARY ARTERY BYPASS GRAFT SURGERY	
1.	Chronic stable angina	30
2.	Unstable angina	66
3.	During acute myocardial infarction	25
4.	Within six months of an acute myocardial infarction	41
5.	Following coronary artery bypass graft surgery	91
6.	Survivors of sudden cardiac death	81
7.	Intractable ventricular arrhythmias	9
8.	Chest pain of uncertain origin	9
9.	Asymptomatic coronary artery disease	9
10.	Patients undergoing primary valvular surgery	9
	Total	370
	INITIAL INDICATIONS FOR CHOLECYSTECTOMY	
1.	Patients less than 55 and nondiabetic	48
2.	Patients less than 55 and diabetic	48
3.	Patients 55 or older and nondiabetic	48
4.	Patients 55 or older and diabetic	48
	Total	192

INITIAL INDICATIONS—continued

Chapter Number	Chapter Title	Number of Indications
	INITIAL INDICATIONS FOR ENDOSCOPY IN PATIENTS 16 AND OLDER	
1.	Asymptomatic patients with no history of upper gastrointestinal cancer	31
2.	Endoscopy follow-up to prior abnormal endoscopy	114
3.	Endoscopy follow-up to prior upper gastrointestinal series	28
4.	Upper gastrointestinal bleeding in patient with no or minimal alcohol intake, not using ulcerogenic medications, with or without evidence for Mallory-Weiss tear	38
5.	Upper gastrointestinal bleeding in patient with moderate or heavy alcohol intake, or with use of ulcerogenic medications, and/or known portal hypertension, with or without evidence for Mallory-Weiss tear	38
6.	Melena or stool positive for occult blood	66
7.	Hematochezia	132
8.	Classic peptic symptoms with no evidence for bleeding or obstruction	576
9.	Peptic symptoms with evidence for obstruction	10
10.	Atypical peptic symptoms with no evidence for bleeding or obstruction	576
11.	Dysphagia	10
12.	Odynophagia	18
13.	Pregnancy	3
14.	Unexplained chest pains, persistent after a negative cardiovascular work-up	20
15.	Therapeutic indications	15
16.	Miscellaneous indications	10
	Total	1685
	INITIAL INDICATIONS FOR COLONOSCOPY IN PATIENTS 16 AND OLDER	
1.	Asymptomatic patients with screening sigmoidoscopy	21
2.	Asymptomatic patients undergoing cancer surveillance	34
3.	Lower gastrointestinal bleeding (includes hematochezia, melena, or stool positive for occult blood) but no evidence for upper gastrointestinal source of bleeding	169
4.	Lower gastrointestinal bleeding with possible upper gastrointestinal source	507
5.	Lower gastrointestinal symptoms and signs (nonbleeding)	120
6.	Colonoscopy follow-up to abnormal barium enema	67
7.	Inflammatory bowel disease	60
8.	Miscellaneous conditions	23
9.	Colonoscopy follow-up to prior colonoscopy	72
10.	Therapeutic indications for colonoscopy	13
	Total	1086

INITIAL INDICATIONS—continued

Chapter Number	Chapter Title	Number of Indications
	INITIAL INDICATIONS FOR CAROTID ENDARTERECTOMY	
1.	Carotid transient ischemic attacks (TIA) and/or amaurosis fugax—single episode	45
2.	Carotid TIAs and/or amaurosis fugax—multiple episodes, never tried on medical therapy	45
3.	Carotid TIAs and/or amaurosis fugax—multiple episodes with at least one recurrence since initiation of medical therapy	45
4.	Carotid TIAs and/or amaurosis fugax—multiple episodes; no recurrence since initiation of medical therapy	45
5.	Vertebrobasilar TIA(s)	45
6.	Postatherothrombotic stroke	45
7.	Stroke-in-evolution	45
8.	Crescendo TIAs	45
9.	Asymptomatic	90
	9A. Asymptomatic, normal stroke risk	
	9B. Asymptomatic, high stroke risk	
10.	Asymptomatic, patient to undergo surgery	180
	10A1. Asymptomatic, patient to undergo intra-abdominal or intrathoracic surgery; normal stroke risk	
	10A2. Asymptomatic, patient to undergo intra-abdominal or intrathoracic surgery; high stroke risk	
	10B1. Asymptomatic, patient to undergo extra-abdominal, extrathoracic surgery; normal stroke risk	
	10B2. Asymptomatic, patient to undergo extra-abdominal, extrathoracic surgery; high stroke risk	
11.	Dementia of vascular origin	45
	Total	675

Appendix D

FINAL INDICATIONS

Chapter Number	Chapter Title	Number of Indications
	FINAL INDICATIONS FOR CORONARY ANGIOGRAPHY	
1.	Asymptomatic patients	28
2.	Chest pain of uncertain origin	30
	2A. Chest pain of uncertain origin in males of any age or females over 50	
	2B. Chest pain of uncertain origin in females 50 or under	
3.	Chronic stable angina	108
4.	Unstable angina	28
	4A. Unstable angina, age less than 65	
	4B. Unstable angina, age 65 or more	
5.	During acute myocardial infarction	5
6.	Within six months of an acute myocardial infarction	68
7.	Sudden death survivors	6
8.	Following coronary artery bypass graft surgery	16
9.	Other	11
	Total	300
	FINAL INDICATIONS FOR CORONARY ARTERY BYPASS GRAFT SURGERY	
1.	Chronic stable angina	73
2.	Unstable angina	117
3.	During acute myocardial infarction	25
4.	Following myocardial infarction	144
	4A. Following transmural myocardial infarction with or without sudden death (within 6 months)	
	4B. Following subendocardial myocardial infarction with or without sudden death (within 6 months)	
5.	Following coronary artery bypass graft surgery	91
6.	Survivors of sudden cardiac death without a myocardial infarction	4
7.	Intractable ventricular arrhythmias	3
8.	(Note: Chapter 8 merged with Chapter 9)	
9.	Asymptomatic coronary artery disease or chest pain of uncertain origin	14
10.	Patients undergoing primary valvular surgery	9
	Total	480
	FINAL INDICATIONS FOR CHOLECYSTECTOMY	
1.	No comorbidity	49
2.	Low comorbidity	49
3.	Medium comorbidity	49
4.	High comorbidity	49
	Total	196

Final Indications—continued

Chapter Number	Chapter Title	Number of Indications
	FINAL INDICATIONS FOR ENDOSCOPY	
1.	Endoscopy is indicated in the asymptomatic patient	54
2.	Endoscopy is indicated as a follow-up to prior endoscopy	186
3.	Endoscopy is indicated as a follow-up to a prior upper gastrointestinal series	63
4.	Endoscopy is indicated in patient with upper gastrointestinal bleeding but with no evidence for portal hypertension or stigmata of liver disease. Includes melena regardless of evidence for Mallory-Weiss tear	49
5.	Endoscopy is indicated in patient with upper gastrointestinal bleeding (includes melena), portal hypertension, or stigmata of liver disease (includes melena) or labelled alcoholic regardless of acute ingestion	12
6.	Endoscopy is indicated in patients with stool positive for occult blood	288
7.	Endoscopy is indicated in patients with hematochezia, i.e., bright red bleeding per rectum requiring more than two transfusions	152
8&10.	Endoscopy is indicated in patients with peptic symptoms	120
9.	Endoscopy is indicated in patients with evidence for obstruction, i.e., vomiting or positive saline loading test (no evidence for metabolic or infectious cause)	44
11.	Endosocopy is indicated in normal host patients with dysphagia or odynophagia	18
12.	Endoscopy is indicated in immunocompromised or diabetic patients with dysphagia or odynophagia	5
14.	Endoscopy is indicated in patient with unexplained chest pains persistent after negative cardiovascular work-up	51
15.	Endoscopy is indicated for therapeutic indications	9
16.	Endoscopy is indicated for miscellaneous indications	18
	Total	1069
	FINAL INDICATIONS FOR COLONOSCOPY IN PATIENTS 16 AND OLDER	
1.	Colonoscopy is indicated in asymptomatic adults with screening sigmoidoscopy	43
2.	Colonoscopy is indicated in asymptomatic patients undergoing cancer surveillance	339
3.	Colonoscopy is indicated in patient with lower gastrointestinal bleeding: hematochezia	864
4.	Colonoscopy is indicated in patient with lower GI bleeding: stool positive for occult blood	672
5.	Colonoscopy is indicated in patient with melena	324
6.	Colonoscopy is indicated with lower GI symptoms and signs (nonbleeding)	192
7.	Colonoscopy is indicated if barium enema is abnormal	148
8.	Colonoscopy is indicated when diagnosis of inflammatory bowel disease is newly made	96
9.	Colonoscopy is indicated for miscellaneous indications	27
10.	Colonoscopy is indicated as follow-up to a prior colonoscopy	144
11.	Colonoscopy is indicated for therapeutic indications	13
	Total	2862

Final Indications—continued

Chapter Number	Chapter Title	Number of Indications
	FINAL INDICATIONS FOR CAROTID ENDARTERECTOMY	
1.	Carotid transient ischemic attack (TIA) and/or amaurosis fugax—single episode	66
2.	Carotid TIAs and/or amaurosis fugax—multiple episodes, never tried on medical therapy	66
3.	Carotid TIAs and/or amaurosis fugax—multiple episodes with at least one recurrence since initiation of medical therapy	66
4.	Carotid TIAs and/or amaurosis fugax—multiple episodes; no recurrence since initiation of medical therapy (at least 3 months of therapy)	66
5.	Vertebrobasilar TIA(s)	66
6.	Postatherothrombotic stroke	66
7.	Stroke-in-evolution	66
8.	Crescendo TIAs—carotid TIAs	66
9.	Asymptomatic	90
	9A. Asymptomatic, normal stroke risk	
	9B. Asymptomatic, high stroke risk	
10.	Asymptomatic, patient to undergo surgery	180
	10A1. Asymptomatic, patient to undergo intra-abdominal or intrathoracic surgery (except coronary bypass graft surgery), normal stroke risk	
	10A2. Asymptomatic, patient to undergo intra-abdominal or intrathoracic surgery (except coronary artery bypass graft surgery), high stroke risk	
	10B1. Asymptomatic, patient to undergo coronary artery bypass graft surgery, normal stroke risk	
	10B2. Asymptomatic, patient to undergo coronary artery bypass graft surgery, high stroke risk	
11.	Dementia of vascular origin	66
	Total	864

Appendix E

NUMBER OF INDICATIONS WITH VARIOUS NUMBERS OF MISSING FINAL RATINGS, BY PROCEDURE

Number of Missing Ratings	Procedure					
	ANGIO	CABS[a]	CH	END	CO	CE
0	265	65	48	1007	1937	848
1	32	359	1	43	674	16
2	3	46	0	4	194	0
3	0	6	0	0	32	0
4	0	4	0	0	21	0
5	0	0	0	0	0	0
6	0	0	0	0	4	0
7	0	0	0	0	0	0
8	0	0	0	4	0	0
9	0	0	0	11	0	0

[a]One panelist did not assign final ratings for CABS. The 65 indications with no missing ratings were in a chapter for which the other panelists decided to leave their initial ratings unchanged.

Appendix F

PERCENT OF INDICATIONS ON WHICH PANELISTS AGREED OR DISAGREED

Table F.1

PERCENT OF INDICATIONS ON WHICH PANELISTS AGREED, USING FOUR DIFFERENT DEFINITIONS OF AGREEMENT
(Median rating 1 to 3)

		Procedure											
		ANGIO		CABS		CH		END		CO		CE	
Definition of Agreement		I	F	I	F	I	F	I	F	I	F	I	F
A9S:	9 ratings, strict	23.7	34.0	18.3	20.0	18.2	45.5	2.5	30.9	0.4	3.3	51.1	55.3
	Standard error	(7.0)	(7.0)	(4.3)	(3.5)	(12.2)	(10.9)	(0.8)	(1.9)	(0.3)	(0.6)	(2.2)	(2.0)
A9R:	9 ratings, relaxed	23.7	34.0	18.3	24.8	18.2	45.5	2.5	30.9	0.4	3.3	51.1	55.3
	Standard error	(7.0)	(7.0)	(4.3)	(3.8)	(12.2)	(10.9)	(0.8)	(1.9)	(0.3)	(0.6)	(2.2)	(2.0)
A7S:	7 ratings, strict	34.2	53.2	26.8	38.4	36.4	59.1	12.6	48.1	12.5	14.9	70.5	70.1
	Standard error	(7.8)	(7.4)	(4.9)	(4.2)	(15.2)	(10.7)	(1.7)	(2.0)	(1.6)	(1.2)	(2.0)	(1.8)
A7R:	7 ratings, relaxed	34.2	55.3	28.0	44.8	36.4	63.6	12.6	48.1	12.5	15.0	70.5	70.1
	Standard error	(7.8)	(7.3)	(5.0)	(4.4)	(15.2)	(10.5)	(1.7)	(2.0)	(1.6)	(1.2)	(2.0)	(1.8)

NOTE: See text for full definitions of agreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

Table F.2

PERCENT OF INDICATIONS ON WHICH PANELISTS AGREED, USING FOUR DIFFERENT DEFINITIONS OF AGREEMENT
(Median rating 4 to 6)

		Procedure											
		ANGIO		CABS		CH		END		CO		CE	
Definition of Agreement		I	F	I	F	I	F	I	F	I	F	I	F
A9S:	9 ratings, strict	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Standard error	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
A9R:	9 ratings, relaxed	1.6	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Standard error	(1.6)	(1.6)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
A7S:	7 ratings, strict	0.0	6.6	0.8	1.4	0.0	0.0	0.4	0.7	0.0	0.2	0.0	0.0
	Standard error	(0.0)	(3.2)	(0.8)	(1.0)	(0.0)	(0.0)	(0.3)	(0.7)	(0.0)	(0.2)	(0.0)	(0.0)
A7R:	7 ratings, relaxed	1.6	19.7	4.0	9.4	0.0	0.0	0.6	1.3	1.6	0.9	0.0	1.4
	Standard error	(1.6)	(5.1)	(1.7)	(2.4)	(0.0)	(0.0)	(0.4)	(0.9)	(0.7)	(0.4)	(0.0)	(1.0)

NOTE: See text for full definitions of agreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

Table F.3

PERCENT OF INDICATIONS ON WHICH PANELISTS AGREED,
USING FOUR DIFFERENT DEFINITIONS OF AGREEMENT
(Median rating 7 to 9)

	Procedure											
	ANGIO		CABS		CH		END		CO		CE	
Definition of Agreement	I	F	I	F	I	F	I	F	I	F	I	F
A9S: 9 ratings, strict	15.1	35.4	8.6	36.6	44.4	50.0	33.3	27.9	6.6	3.3	0.0	10.3
Standard error	(3.5)	(3.5)	(2.2)	(3.3)	(9.7)	(12.1)	(1.6)	(2.7)	(1.5)	(0.5)	(0.0)	(2.9)
A9R: 9 ratings, relaxed	15.1	35.9	8.6	37.0	44.4	50.0	33.3	27.9	6.6	3.3	0.0	10.3
Standard error	(3.5)	(3.5)	(2.2)	(3.3)	(9.7)	(12.1)	(1.6)	(2.7)	(1.5)	(0.5)	(0.0)	(2.9)
A7S: 7 ratings, strict	28.3	63.0	34.6	55.6	63.0	61.1	50.4	49.5	21.7	21.6	7.0	25.2
Standard error	(4.4)	(3.5)	(3.7)	(3.4)	(9.5)	(11.8)	(1.7)	(3.0)	(2.5)	(1.1)	(3.4)	(4.2)
A7R: 7 ratings, relaxed	29.2	68.2	35.8	61.1	63.0	66.7	51.8	50.2	21.7	21.9	7.0	27.1
Standard error	(4.4)	(3.4)	(3.8)	(3.3)	(9.5)	(11.4)	(1.7)	(3.0)	(2.5)	(1.1)	(3.4)	(4.3)

NOTE: See text for full definitions of agreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

Table F.4

PERCENT OF INDICATIONS ON WHICH PANELISTS DISAGREED,
USING FOUR DIFFERENT DEFINITIONS OF DISAGREEMENT
(Median rating 1 to 3)

	Procedure											
	ANGIO		CABS		CH		END		CO		CE	
Definition of Disagreement	I	F	I	F	I	F	I	F	I	F	I	F
D9S: 9 ratings, strict	0.0	4.3	18.3	7.2	9.1	9.1	31.2	25.9	32.3	41.9	17.3	7.8
Standard error	(0.0)	(3.0)	(4.3)	(2.2)	(9.1)	(6.3)	(2.4)	(1.8)	(2.2)	(1.6)	(1.7)	(1.1)
D9R: 9 ratings, relaxed	36.8	17.0	47.6	32.8	54.5	27.3	66.3	41.4	75.1	77.5	26.8	19.4
Standard error	(7.9)	(5.5)	(5.5)	(4.1)	(15.7)	(9.7)	(2.5)	(2.0)	(2.0)	(1.4)	(1.9)	(1.6)
D7S: 7 ratings, strict	0.0	2.1	2.4	0.0	0.0	0.0	2.5	6.1	5.8	7.1	1.1	0.2
Standard error	(0.0)	(2.1)	(1.7)	(0.0)	(0.0)	(0.0)	(0.8)	(1.0)	(1.1)	(0.9)	(0.5)	(0.2)
D7R: 7 ratings, relaxed	5.3	10.6	15.9	5.6	36.4	13.6	27.7	23.2	41.6	30.1	8.7	6.4
Standard error	(3.7)	(4.5)	(4.1)	(2.0)	(15.2)	(7.5)	(2.3)	(1.7)	(2.3)	(1.5)	(1.2)	(1.0)

NOTE: See text for full definitions of disagreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

Table F.5

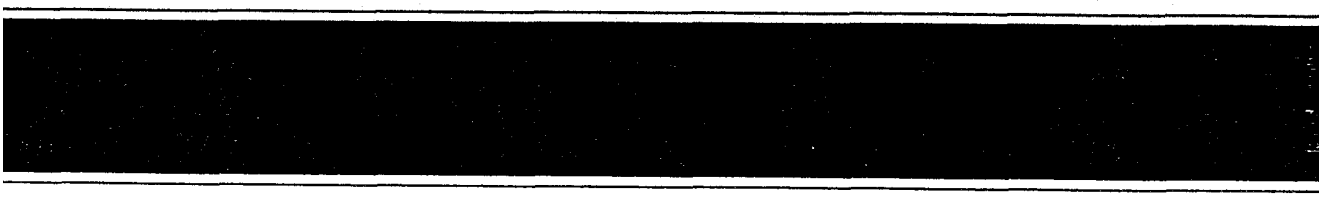
PERCENT OF INDICATIONS ON WHICH PANELISTS DISAGREED, USING FOUR DIFFERENT DEFINITIONS OF DISAGREEMENT
(Median rating 4 to 6)

		Procedure											
		ANGIO		CABS		CH		END		CO		CE	
Definition of Disagreement		I	F	I	F	I	F	I	F	I	F	I	F
D9S:	9 ratings, strict	18.0	6.6	33.3	20.9	40.0	66.7	69.0	55.3	60.0	76.1	83.7	37.0
	Standard error	(5.0)	(3.2)	(4.2)	(3.4)	(16.3)	(16.7)	(2.1)	(4.1)	(2.6)	(1.7)	(3.9)	(4.1)
D9R:	9 ratings, relaxed	98.4	80.3	90.5	79.1	100.0	100.0	96.3	90.0	96.4	97.4	100.0	89.1
	Standard error	(1.6)	(5.1)	(2.6)	(3.5)	(0.0)	(0.0)	(0.9)	(2.5)	(1.0)	(0.6)	(0.0)	(2.7)
D7S:	7 ratings, strict	0.0	0.0	8.7	5.0	0.0	11.1	5.2	12.7	12.1	33.4	16.3	8.0
	Standard error	(0.0)	(0.0)	(2.5)	(1.8)	(0.0)	(11.1)	(1.0)	(2.7)	(1.7)	(1.8)	(3.9)	(2.3)
D7R:	7 ratings, relaxed	77.0	37.7	59.5	45.3	80.0	77.8	76.2	73.3	80.5	81.3	81.5	67.4
	Standard error	(5.4)	(6.3)	(4.4)	(4.2)	(13.3)	(14.7)	(1.9)	(3.6)	(2.1)	(1.5)	(4.1)	(4.0)

NOTE: See text for full definitions of disagreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.

Table F.6

PERCENT OF INDICATIONS ON WHICH PANELISTS DISAGREED, USING FOUR DIFFERENT DEFINITIONS OF DISAGREEMENT
(Median rating 7 to 9)

		Procedure											
		ANGIO		CABS		CH		END		CO		CE	
Definition of Disagreement		I	F	I	F	I	F	I	F	I	F	I	F
D9S:	9 ratings, strict	7.5	0.0	11.7	2.3	11.1	11.1	22.8	27.5	62.5	66.4	42.1	28.0
	Standard error	(2.6)	(0.0)	(2.5)	(1.0)	(6.2)	(7.6)	(1.5)	(2.6)	(2.9)	(1.3)	(6.6)	(4.4)
D9R:	9 ratings, relaxed	46.2	17.2	41.4	19.4	25.9	33.3	43.9	43.9	73.9	74.7	75.4	47.7
	Standard error	(4.9)	(2.7)	(3.9)	(2.7)	(8.6)	(11.4)	(1.7)	(2.9)	(2.7)	(1.2)	(5.8)	(4.9)
D7S:	7 ratings, strict	0.0	0.0	4.3	1.9	0.0	0.0	1.7	6.3	14.0	25.9	19.3	7.5
	Standard error	(0.0)	(0.0)	(1.6)	(0.9)	(0.0)	(0.0)	(0.4)	(1.4)	(2.1)	(1.2)	(5.3)	(2.6)
D7R:	7 ratings, relaxed	15.1	2.6	13.6	5.1	14.8	0.0	16.4	19.2	43.4	43.9	43.9	21.5
	Standard error	(3.5)	(1.2)	(2.7)	(1.5)	(7.0)	(0.0)	(1.3)	(2.3)	(3.0)	(1.4)	(6.6)	(4.0)

NOTE: See text for full definitions of disagreement.
I indicates initial ratings, F indicates final ratings.
ANGIO: coronary angiography.
CABS: coronary artery bypass graft surgery.
CH: cholecystectomy.
END: upper gastrointestinal endoscopy.
CO: fiberoptic colonoscopy.
CE: carotid endarterectomy.