Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures

Lucian L. Leape, MD,[a,b] Rolla Edward Park, PhD,[a] Thomas M. Bashore, MD,[c] J. Kevin Harrison, MD,[c] Charles J. Davidson, MD,[d] and Robert H. Brook, MD, ScD[a,e] ***Santa Monica, Calif; Boston, Mass; Durham, NC; Chicago, Ill; and Los Angeles, Calif***

Background Evidence from numerous studies of coronary angiography show differences between observers' assessments of 15% to 45%. The implication of this variation is serious: If readings are erroneous, some patients will undergo revascularization procedures unnecessarily and others will be denied an essential treatment. We evaluated the variation in interpretation of angiograms and its potential effect on appropriateness of use of revascularization procedures.

Methods and Results Angiograms of 308 randomly selected patients previously studied for appropriateness of angiography, coronary artery bypass grafting (CABG), and percutaneous transluminal coronary angioplasty (PTCA) were interpreted by a blinded panel of 3 experienced angiographers and compared with the original interpretations. The potential effect on differences on the appropriateness of revascularization was assessed by use of the RAND criteria. Technical deficiencies were found in 52% of cases. Panel readings tended to show less significant disease (none in 16% of vessels previously read as showing significant disease), less severity of stenosis (43% lower, 6% higher), and lower extent of disease (23% less, 6% more). The classification of CABG changed from necessary/appropriate to uncertain/inappropriate for 17% to 33% of cases when individual ratings were replaced by panel readings.

Conclusions The general level of technical quality of coronary angiography is unsatisfactory. Variation in the interpretation of angiograms was substantial in all measures and tended to be higher in individual than in panel readings. The effect was to lead to a potential overestimation of appropriateness of use of CABG by 17% and of PTCA by 10%. These findings indicate the need for increased attention to the technical quality of studies and an independent second reading for angiograms before recommending revascularization. (Am Heart J 2000;139:106-13.)

Recent studies of the appropriateness of the use of coronary artery bypass graft surgery (CABG) and percutaneous transluminal coronary angioplasty (PTCA) in New York State showed low rates of inappropriate use of these procedures.[1,2] The validity of those assessments depends on the credibility of the data that were used to determine appropriateness. These data include clinical symptoms, response to medical therapy, results of laboratory studies such as the exercise stress test, findings on coronary angiography, and the presence of risk factors such as hypertension, diabetes, and peripheral vascular disease. Of those data, the most important for determining feasibility of revascularization is the extent of disease of the coronary arteries. The extent of coronary artery disease is measured by the number of vessels involved and the severity of their obstructions. Patients with significant left main disease or 3-vessel disease, for example, are often appropriate candidates for bypass surgery because controlled trials have shown CABG improves long-term survival in these patients.[3] By contrast, CABG offers little survival advantage for most patients with single-vessel disease, who are often more appropriately treated medically or by angioplasty.[4]

Both the number of vessels affected and the severity of their stenoses are determined by coronary angiography. The American College of Cardiology has defined significant coronary artery disease as evidence by angiography of at least 50% narrowing of the lumen of the artery.[5] Clearly, much depends on the accuracy of the angiographic assessment of the extent of stenosis. Unfortunately, the reproducibility of interpretations of angiograms is far from perfect. Evidence from a number of studies indicates that differences between observers' assessments of the presence of significant disease range from 15% to 45%.[6-18] When individual readings are compared with a "standard," such as interpretation by a group of experts or quantitative coronary angiography, 28% to 31% are found to be erroneous.[8,10,15-18]

With the exception of the Coronary Artery Surgical

From [a]RAND; [b]Harvard School of Public Health; [c]Duke University Medical Center; [d]Northwestern University Medical Center; and the [e]Schools of Medicine and Public Health, UCLA.

Guest Editor for this manuscript was Blase A. Carabello, MD, Medical University of South Carolina, Charleston.

Submitted May 13, 1998; accepted October 26, 1998.

Reprint requests: Lucian L. Leape, MD, Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115.

0002-8703/2000/$12.00 + 0 **4/1/96001**

Study (CASS), these studies of rater reliability have been characterized by small numbers of observers and small numbers of angiograms, often in a single institution. Although the CASS study included patients from many regions of the country, most were treated in academic medical centers. Thus none of the previous studies is truly representative of the standard of community practice.

The implication of variation in interpretation of angiograms is serious: If readings are erroneous, some patients will undergo revascularization procedures unnecessarily and others will be denied an essential treatment. If, in the previous study in New York State, variation in angiographic interpretation was substantial, the findings of little inappropriate use of revascularization must be reconsidered because the judgments depended so much on the number of vessels with significant disease.

Because of these concerns, we assessed the interpretation of angiograms in a sample of patients from our previous study of appropriateness of CABG, PTCA, and coronary angiography in New York[1,2,19] by comparing the hospital cardiologists' interpretations with those of a blinded panel of cardiologists. We asked 3 questions: (1) What is the overall reproducibility of interpretation of coronary angiograms? (2) Is there evidence of upward bias in readings? (3) To the extent that variation is present, how would that alter the previous evaluations of appropriateness of use of CABG and PTCA?

Methods

Sample

We selected 3 samples of angiograms for review from our previously reported study of the appropriateness of coronary angiography, CABG, and PTCA in 1990 in nonfederal hospitals in New York State.

To evaluate variation in assessing the extent of disease, we sampled angiograms from the study of appropriateness of angiography.[13] For that study, patients were selected by a 2-step sampling process. Hospitals were randomly sampled from several strata: geographic location, number of angiographies performed, and whether they also performed CABG. Then medical records were randomly selected within the 15 hospitals chosen, a total of 1335 patients. For the current study, we provided each of these sample hospitals with a randomized listing of the angiograms that had been used in the appropriateness study. Each then located the studies consecutively on the list, replacing missing studies until the target number of angiograms was obtained or the list was exhausted. From a desired sample of 120, cardiac catheterization laboratories provided 119 angiograms for our review ("angiography sample").

To assess the effect of variations on the rates of appropriateness, we added studies of "threshold" cases (3-vessel disease for CABG and single-vessel disease for PTCA) from our previous studies of appropriateness in patients who received CABG[1] (1338 patients) and PTCA[2] (1306 patients). These patients were selected by a sampling plan similar to that for the study of coronary angiography. From patients in the CABG study diagnosed with 3-vessel disease (56% of all patients with CABG), hospital cardiac catheterization laboratories provided 91 of the requested 100 angiograms ("CABG sample"). From the PTCA study, we requested 100 randomly selected angiograms of patients diagnosed with single-vessel disease (46% of all PTCAs). Catheterization laboratories provided 98 of these sample angiograms ("PTCA sample"). Thus the total number of angiograms evaluated was 308 (119 + 91 + 98).

For the evaluation of variation and bias in the assessment of the presence of disease and its severity in individual vessels, we combined the samples.

Table I. Types of technical deficiencies in 295 cineangiograms

	No.	%
No technical deficiencies	153	52%
No reference segment	32	11%
Inadequate separation from background	35	12%
Inadequate lesion/vessel separation	67	23%
Inadequate opacification flow	48	16%
Inadequate opacification technique	68	23%
Inadequate radiographic procedure	10	3%

Interpretation of angiograms

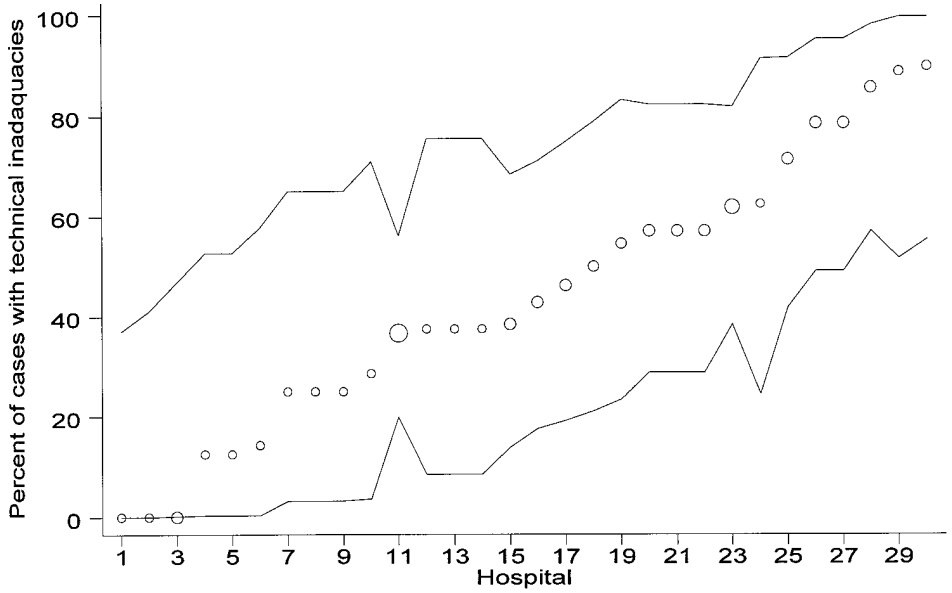
Cineangiogram films were sent to the Core Laboratory at Duke University Medical Center. The Core Laboratory was unaware of the clinical histories, the interpretations of the angiograms by cardiologists at the New York sites, or the appropriateness ratings of the cases. All films were reviewed visually by a panel of 3 experienced angiographers, and a consensus was reached regarding disease severity and adequacy of the cinefilms for interpretation. The previous New York cardiologists' interpretations were then compared with this panel reading to assess variation.

The quality of the coronary angiographic study was evaluated during the visual reading sessions. Each vessel was assessed as to adequacy of (1) a reference segment to which stenoses could be compared (Was the reference segment seen adequately on the frames wherein the lesions were present?), (2) separation of vessels and lesions from background (Was the lesion in question separated from background by adequate angulation?), (3) lesions/vessel separation (Was the lesion in question separated from other vessels by adequate angulation?), (4) flow for coronary opacification (Was the injection rate adequate to ensure appropriate contrast mixing?), (5) technical opacification and panning of the image (Were images of adequate quality to ensure that the relevant lesions were not lost in the lung field and all were seen by appropriate panning?), and (6) radiographic exposure (Was the film underexposed or overexposed?). A single study could have more than 1 technical deficiency. Every effort was made to interpret the lesions in question. Film quality was deemed inadequate only if lesions could not be interpreted based on the sum of all of these quality factors.

The presence or absence of each coronary segment and its relative size was first encoded.[20] Lesions were then identified within these coronary segments. Lesions were encoded in 2 levels of nonsignificant disease: 1% to 25% and 26% to 50%, and 4 levels of significant disease: 51% to 75%, 76% to 95%, 96% to 99%, and 100%.[21]

Coronary lesions identified by New York cardiologists had

Figure 1



Sizes of circles are proportional to sample sizes, which varied from 7 to 30. *Solid lines* bound binomial 95% CIs.

been previously coded for the study of appropriateness according to the most severe lesion in each vessel also into 2 levels of nonsignificant disease: <30% and 30% to 49%, and 4 levels of significant disease: 50% to 69%, 70% to 89%, 90% to 99%, and occluded. When the Duke panel reported multiple lesions in a single vessel, we chose the most severe lesion for comparison of readings. Because of variation in the definition of whether a lesion was in the proximal or distal left anterior descending artery or within a specific diagonal or marginal branch, we collapsed lesions from all locations within the major epicardial vessels into a single reading of the most severe stenosis in each vessel for the comparison of the Duke panel and New York cardiologists' interpretations of individual vessels. Thus for the comparison of severity and the presence of significant disease we defined 4 vessels: left main, left anterior descending, circumflex, and right coronary arteries. The distinction between proximal and distal lesions in the left anterior descending artery was maintained, however, for the analyses of the extent of disease and its effect on appropriateness judgments.

Rating appropriateness

The appropriateness of CABG or PTCA was determined for each patient by applying the ratings from the RAND expert panel process used for the previous studies of appropriateness.[1,2] In this process, each panelist rates each indication for CABG and PTCA for appropriateness. An "indication" is a unique combination of factors that are considered before recommending surgery. These include findings on coronary angiography, clinical symptoms, response to medical therapy, results of laboratory studies such as the exercise stress test, and the presence of risk factors such as hypertension, diabetes, and peripheral vascular disease. The indications are rated for appropriateness by each expert on a 9-point scale. The final appropriateness rating for an indication is the median of the 9 experts' ratings. An example of a typical indication is: CABG is appropriate within 21 days of an acute myocardial infarction for a patient who has continuing pain, a low operative risk, an ejection fraction of 15% to 35%, and in whom coronary angiography has demonstrated significant 3-vessel disease.

To apply appropriateness ratings to patients, clinical and laboratory data (listed above) are abstracted from the medical records and used to define the indication for each patient. The patient is then assigned the appropriateness rating for that indication. Thus these ratings of appropriateness derive from application of expert panel ratings to clinical and angiographic data. They are not judgments of appropriateness made by either the New York or the Duke cardiologists.

"Significant disease" was defined as a reduction in the luminal diameter of 50% or more. To determine extent of disease (single-vessel, 2-vessel, etc), we used the RAND expert panel criteria that required in addition to 50% reduction in the lumen of all affected vessels, a reduction of 70% or more in the lumen of at least 1 vessel (except for the left main coronary artery). Note that this results in fewer patients being identified as having "significant disease" than when the 70% stenosis is not required. Scores derived by use of the New York cardiologists' vessel stenosis interpretations were compared with those derived by use of the Duke panel interpretations.

Analysis

We analyzed the data by using cross-tabulations of New York and Duke assessments. We assessed the predominant direction of disagreement ("imbalance") in each table by counting the number of cases for which the New York assessment was higher (for example, more severe disease) and sub-

Table II. Presence of significant disease by vessel

New York cardiologists	Duke panel reading					
	NSD	LM	LAD	LCX	RCA	Total
NSD	647	3	2	4	8	664
LM	7	5	0	0	0	12
LAD	31	0	155	0	0	186
LCX	27	0	0	121	0	148
RCA	19	0	0	0	151	170
Total	731	8	157	125	159	1180

Imbalance (84 – 17) = 67 (90% CI 51 to 83). For this comparison we combined proximal LAD disease with middle, distal, or unspecified LAD disease because the definition of "proximal" varied among cardiologists.
NSD, No significant difference (<50% narrowing of the vessel); *LM,* left main coronary artery; *LAD,* left anterior descending coronary artery; *LCX,* left circumflex coronary artery; *RCA,* right coronary artery.

Table III. Assessment of severity of stenosis by vessel (percent narrowing)*

New York cardiologists	Duke panel reading						
	NRD	26-50	51-75	76-95	96-99	Occl	Total
NRD	451	10	5	0	0	3	469
30-49	149	37	5	1	0	3	195
50-69	21	14	10	2	0	2	49
70-89	8	22	71	29	0	6	136
90-99	6	3	61	91	1	19	181
Occl	8	2	4	12	0	124	150
Total	643	88	156	135	1	157	1180

Imbalance (310 – 46) = 264 (90% CI 235 to 290). Ratings were considered to be in agreement when New York and Duke categories were a close match (eg, 30 to 49 and 26 to 50, 50 to 69 and 51 to 75, 70 to 89 and 76 to 95, 90 to 99 and occluded) as well as when they overlapped (eg, NRD and 26 to 50, 70 to 89 and 51 to 75, 90 to 99 and 76 to 95).
NRD, No reportable disease (Duke, 0% to 25%; New York, 0% to 29%); *Occl,* occluded vessel (100% narrowing).
*Comparison of most severe lesion in each of 4 vessels (left main, circumflex, left anterior descending, and right coronary) in 332 patients.

tracting the number of cases for which the Duke assessment was higher. We also calculated approximate 90% confidence intervals (CIs) for the imbalance by using the 5th and 95th percentiles from 1000 bootstrap replications. All calculations were done with the use of Stata version 4.0.

Results

Adequacy of angiograms

Although in most cases it was possible for Duke angiographers to visually interpret lesions despite technical deficiencies, in 7 cases the technical quality of the study was so poor that adequate visual assessment of the lesions was impossible. In another 5 cases, major epicardial vessels were not injected. These cases were excluded. In 1 previous angiography case, clinical data were inadequate to permit classification for appropriateness of revascularization; therefore it was also excluded, leaving 295 cases for analysis.

Of the 295 cineangiographic studies examined, 142 (48%) exhibited 1 or more technical deficiencies. The types of technical deficiencies are shown in Table I. We found 260 technical deficiencies in these 142 studies. The percentage of inadequate studies varied markedly by hospital. Figure 1 shows the percentage of studies from each hospital that had 1 or more technical deficiencies. In 12 of the 29 hospitals, 50% or more of the studies submitted had technical deficiencies. Six of these 12 hospitals are teaching hospitals.

Assessment of extent of disease

Presence of significant disease by vessel. Of the 1180 vessels evaluated (4 vessels for 295 cases), the Duke panel assessments agreed with the New York hospitals cardiologists' assessments in 1079 (91%) vessels (Table II). However, among the 533 vessels in which either New York or Duke cardiologists found significant disease (50% narrowing or more), agreement dropped to 81% (432 of 533). (Significant disease is defined by most cardiologists as narrowing of the lumen by at least 50%. Because the extent of stenosis is estimated by visual examination of the radiologic image, precision is at best ±5% to 10%; therefore cardiologists often report readings in ranges. Although the RAND convention is to classify

Table IV. Assessment of extent of disease: Patients who had angiography

New York cardiologists	Duke panel reading							
	LM	3VD	2VDP	2VD	1VDP	1VD	NSD	Total
LM	4	1	2	3	0	1	0	11
3VD	0	5	2	4	0	2	1	14
2VDP	0	2	0	0	1	1	0	4
2VD	0	0	2	13	0	6	1	22
1VDP	0	0	0	0	1	0	0	1
1VD	0	0	0	1	0	12	1	14
NSD	0	0	0	1	0	1	47	49
Total	4	8	6	22	2	23	50	115

Imbalance (26 – 7) = 19 (90% CI 11 to 28).
LM, Left main disease; *3VD,* 3-vessel disease; *2VDP,* 2-vessel disease including proximal LAD; *2VD,* 2-vessel disease without proximal LAD; *1VDP,* 1-vessel disease including proximal *LAD;* 1VD, 1-vessel disease not including proximal LAD; *NSD,* no significant disease.

Table V. Appropriateness of CABG: Patients who had angiography

New York cardiologists	Duke panel reading				
	Necessary	Appropriate	Uncertain	Inappropriate	Total
Necessary	15	2	5	4	26
Appropriate	1	9	0	4	14
Uncertain	1	1	7	3	12
Inappropriate	0	0	1	62	63
Total	17	12	13	73	115

Imbalance (13 – 2) = 11 (90% CI 5 to 17).

the lowest level of significant disease as 50% to 69%, the Duke convention is to classify the lowest level of significant disease as 51% to 75%. Because of the limits of visual assessment, there is no significant difference between the readings.) The Duke panel found no significant disease in 84 (16%) of 516 vessels that New York cardiologists read as showing significant disease, including 7 of 12 patients previously diagnosed with left main disease. New York cardiologists found no significant disease in 17 (4%) of 449 vessels interpreted by the Duke panel as having significant disease, including 3 patients with left main disease (imbalance 84 – 17 = 67; 90% CI 51 to 83).

Severity of stenosis. Overall, New York and Duke cardiologists agreed on the severity of stenosis in 824 (70%) of 1180 of vessels (Table III). If the 451 vessels that both New York and Duke cardiologists agreed had no disease are excluded, comparison of the assessments of the most severe lesion in each of the remaining 729 vessels revealed agreement in 373 (51%). New York cardiologists rated 310 (43%) vessels as having a higher degree of stenosis than the Duke panel, whereas Duke cardiologists found 46 (6%) vessels to have more severe narrowing than the New York cardiologists (imbalance 310 – 46 = 264; 90% CI 235 to 290).

Extent of disease. We compared assessment of the extent of disease (3-vessel, 2-vessel, etc) according to the New York and Duke readings in the population-based angiography sample by using the RAND expert panel criteria of 50%/70%. Of the 119 cases in the sample, 115 were suitable for analysis. The appraisals agreed in 82 (71%) cases (Table IV). In 26 (23%) cases, the New York cardiologists' readings resulted in an assessment of more extensive disease (eg, 3-vessel vs 2-vessel), whereas in 7 cases (6%) the Duke panel readings resulted in a higher assessment (imbalance 26 – 7 = 19; 90% CI 11 to 28).

Effect of variation in angiographic interpretation on appropriateness classification

Potential appropriateness of the use of CABG was computed both for the 115 angiography cases (possible candidates for CABG) and for the 85 patients who had received CABG by use of the RAND expert panel ratings of appropriateness applied to the clinical and angiographic data for these cases. Note: These ratings were not the judgments of either the New York or Duke cardiologists.

In the angiography sample, appropriateness ratings calculated with the Duke angiographic interpretations agreed with those obtained by using the New York

Table VI. Appropriateness of CABG: Patients who had CABG

New York cardiologists	Duke panel reading				Total
	Necessary	Appropriate	Uncertain	Inappropriate	
Necessary	60	6	9	5	80
Appropriate	0	3	0	0	3
Uncertain	1	0	0	0	1
Inappropriate	0	0	1	0	1
Total	61	9	10	5	85

Imbalance (14 – 1) = 13 (90% CI 8 to 19).

readings in 93 (81%) of the 115 cases (Table V). Of 40 (35%) cases classified as necessary or appropriate for CABG when New York cardiologists' interpretations were used, 13 (33%) were classified as inappropriate or uncertain when Duke panel readings were used. In contrast, of 29 (25%) cases rated as necessary or appropriate when Duke angiographic interpretations were used, only 2 (7%) were rated as uncertain according to New York readings, and none were rated as inappropriate (imbalance 13 – 2 = 11; 90% CI 5 to 17).

When the Duke angiographic interpretations were applied to the 85 patients who had undergone CABG for 3-vessel disease, 14 (17%) of the 83 cases that had previously been rated as necessary or appropriate by use of the New York interpretations were classified as inappropriate or uncertain, whereas 1 patient was reclassified from uncertain to necessary (Table VI) (imbalance 14 – 1 = 13; 90% CI 8 to 19).

Application of the Duke angiographic interpretations to the 95 patients who had undergone PTCA for single-vessel disease resulted in reclassification of 6 (10%) of 63 patients from necessary/appropriate to inappropriate/uncertain and 1 patient from inappropriate to appropriate (Table VII) (imbalance 6 – 1 = 5, 90% CI 1 to 9).

Discussion

Cardiologists have wrestled with the problem of reproducibility of visual angiographic interpretation of lesion severity for more than 2 decades. The lack of agreement between readers has been disturbingly high as well as remarkably consistent. In 1975, Detre et al[6] reported the results of a study in which 22 physicians interpreted 13 angiograms on 2 separate occasions. They found interobserver agreement on extent of disease to be approximately halfway between 100% and chance, and intraobserver agreement on the extent of disease in each vessel varied from 72% to 91%. Zir et al[7] compared readings of 20 good-quality angiograms by 4 experienced angiographers who worked together but interpreted the studies independently. Lack of agreement among all 4 observers as to the presence or absence of significant disease (defined as >50% stenosis) varied from 15% (3 of 20) for the left main coronary artery to 65% (11 of 20) for the proximal or mid left anterior descending artery.

In a study of independent readings of 10 angiograms by 11 observers compared with interpretations by an expert panel of 3, DeRouen et al[8] found that the expert panel disagreed with the readers about the number of major vessels with 70% stenosis in 31% of cases. The reproducibility of panel interpretations was examined by Sanmarco et al,[9] who compared the consensus readings of 14 films by a panel of 4 on 2 occasions 7 months apart. The panel agreed with itself on the presence or absence of a 70% or greater lesion among 186 vessel sites 95% of the time.

Fisher et al[10] studied 870 randomly selected abnormal angiograms from 14 clinical sites in the CASS study. They compared local readings with those at 1 of 4 quality control sites. In 28% of cases, the readings differed concerning the number of vessels with 50% or greater stenosis. In 18.6% of cases read as having left main disease, no lesion was found by the second reader.[10] Although Topol and Nissen[22] caution clinicians to not rely excessively on angiography for clinical decision-making, assessment of the extent of arterial stenosis remains an important factor in therapeutic decisions.

Our study differed from these previous evaluations in several important ways. First, it was a population-based sample, sampling angiograms randomly from 29 hospitals in New York State, providing coronary angiography, including teaching and nonteaching hospitals, public and private ownership, and located in large and small cities. Thus it is representative of overall practice in New York State. Second, all angiograms in the sample were reviewed regardless of the technical quality of the study or the severity of the findings. Most previous studies have selected films of good quality or ones showing significant stenosis. However, in practice, treatment decisions are often based on interpretation of less than satisfactory studies. We found many in this sample. Third, we estimated the impact of variation in interpretation on possible outcomes, that is, judgments of the appropriateness of use of CABG or PTCA for these patients.

Table VII. Appropriateness of PTCA: Patients who had PTCA

New York cardiologists	Duke panel reading				
	Necessary	Appropriate	Uncertain	Inappropriate	Total
Necessary	39	1	0	3	43
Appropriate	1	16	2	1	20
Uncertain	0	0	21	2	23
Inappropriate	0	1	3	5	9
Total	40	18	26	11	95

Imbalance (6 – 1) = 5 (90% CI 1 to 9).

The first significant finding was that the technical quality of angiograms across the board was poor: 12 (4%) were unsuitable for evaluation, and 48% of the remaining studies had significant technical deficiencies that made interpretation difficult. Thus the total defect rate was 52%. This is disturbingly high for a procedure that has been in use for more than 30 years. Moreover, we found striking differences in the quality of angiograms among the study hospitals. Although some hospitals had few poor-quality studies, in 12 of the 29 study hospitals, 6 of them university hospitals, more than half of the angiograms had some deficiencies. Poor quality affects the validity of revascularization decisions. Clearly, these technical quality problems need to be addressed.

The second significant finding was that there are major differences in interpretations of the angiograms, both in whether disease was present and in the severity of stenosis. The panel found no significant disease in 16% of vessels that New York cardiologists read as showing significant disease, whereas the New York cardiologists found no significant disease in 3% of those read by the Duke panel as showing significant disease. Agreement on the severity of stenosis was even less: 70% overall and 51% when restricted to vessels that either group of cardiologists rated as having some stenosis.

When these differences were applied to the determination of the extent of disease (3-vessel, 2-vessel, etc) in the population-based "angiography" sample, the New York cardiologists rated 23% of cases as having more extensive disease than the Duke panel, whereas in 6% of cases the Duke panel rated cases as having more extensive disease.

In the absence of data on the reproducibility of readings, the Duke panel cannot be considered a "gold standard." However, studies by DeRouen et al[8] and Sanmarco et al[9] indicate that panel interpretations have substantially less variability than individual readings. Thus it seems likely that the blinded panel readings more nearly approximate the true state of the coronary arteries than individual physician readings.

Compared with the Duke panel interpretation, we found substantial evidence of overassessment in the New York readings. Some overreading is to be expected because the New York cardiologists had complete clinical information about the patient and the Duke interpreters did not. Prior knowledge is known to affect what one "sees." Because of the known inaccuracy of cineangiograms in defining lesion severity precisely in many instances, knowing that the patient had severe symptoms or a positive functional study in the area of distribution of the vessel in question, for example, could easily have led the examiner to interpret significant disease in less severe lesions.

The impact of the variations of interpretation of angiograms on the potential appropriateness of revascularization was substantial. In the sample of 115 patients who had angiography, 33% of those rated as necessary or appropriate for CABG by use of New York angiography readings were classified as inappropriate or uncertain when Duke panel angiography readings were used. Among patients who had undergone CABG for 3-vessel disease, 17% of those rated necessary or appropriate by use of the New York cardiologists' readings were classified as inappropriate or uncertain when Duke panel readings were used. Of the single-vessel disease patients who had undergone PTCA, 10% of those rated necessary or appropriate by New York cardiologists' readings were classified as inappropriate or uncertain when Duke panel readings were used.

Policy significance

The high rate of nonreproducibility that we found in angiographic studies performed in both community and teaching hospitals suggests that our previous studies may have underestimated the inappropriate use of both CABG and PTCA. If the panel ratings were applied to the cases from the previous studies of appropriateness, the estimate of the percentage of CABG cases that were classified as inappropriate changes from 2% to 8% and those classified as uncertain from 7% to 17%. For PTCA, the percentage classified inappropriate changes from 4% to 8% and uncertain changes from 38% to 40%.

The finding that despite improvements in radiographic equipment over the years, reproducibility rates are not

significantly improved from those of studies performed 20 years ago indicates that catheterization laboratories need to reevaluate their techniques and procedures. Several remedial measures are urgently needed.

First, most laboratories need to significantly improve the technical quality of their studies. Improving technical quality may require attention to multiple factors, such as training of professional and technical personnel, establishing standards of performance and enforcing compliance, revamping procedures, and, in some cases, replacing outdated and inferior equipment.

In the next decade, most cardiac catheterization laboratories will probably convert to digital angiography. Although every innovation presents its own unique potential for errors, overall this change should improve image quality. However, the operator must still select views appropriately, inject the vessel with sufficient contrast material, and perform the other technical maneuvers appropriately. If adequate images are not produced, the value of digital imaging will be limited.

Second, cardiologists need to reevaluate the methods used for interpretation of the angiograms. Our findings, as well as those of many others dating back 20 years and more,[6,7] indicate that reliance on angiographic interpretation by a single individual results in an unacceptable variation rate. Even highly qualified cardiologists have significant rates of nonreproducibility. Therefore concurrence of at least a second independent opinion should be required before a final diagnosis is assigned. A group reading of 3 or 4 cardiologists is better, as are multiple independent readings with resolution of differences. In all centers, periodic outside monitoring of randomly selected studies would almost certainly lead to improved quality of the readings.

References

1. Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York state. JAMA 1993;269:753-60.
2. Hilborne LH, Leape LL, Bernstein SJ, et al. The appropriateness of use of percutaneous transluminal coronary angioplasty in New York state. JAMA 1993;269:761-5.
3. Alderman E, Bourassa M, Cohen L, et al. Ten-year follow-up of survival and myocardial infarction in the randomized Coronary Artery Surgery Study. Circulation 1990;82:1-18.
4. Hilborne LH, Leape LL, Kahan JP, et al. Percutaneous transluminal coronary angioplasty: a literature review and ratings of appropriateness and necessity. Publication JRA01. Santa Monica, Calif: the RAND Corporation; 1991.
5. Ryan T, Faxon D, Gunnar R, et al. Guidelines for percutaneous transluminal coronary angioplasty. J Am Coll Cardiol 1988;12:529-45.
6. Detre K, Wright E, Murphy M, et al. Observer agreement in evaluating coronary angiograms. Circulation 1975;52:979-86.
7. Zir L, Miller S, Dinsmore R, et al. Interobserver variability in coronary angiography. Circulation 1976;53:627-32.
8. DeRouen T, Murray J, Owen W. Variability in the analysis of coronary arteriograms. Circulation 1977;55:324-8.
9. Sanmarco ME, Brooks SH, Blankenhorn DH. Reproducibility of a consensus panel in the interpretation of coronary angiograms. Am Heart J 1978;96:430-7.
10. Fisher L, Judkins M, Lesperance J, et al. Reproducibility of coronary arteriographic reading in the coronary artery surgery study (CASS). Cathet Cardiovasc Diagn 1982;8:565-75.
11. Galbraith JE, Murphy ML, de Soya N. Coronary angiogram interpretation. JAMA 1978;240:2053-6.
12. Beauman GJ, Vogel RA. Accuracy of individual and panel visual interpretations of coronary arteriograms: implications for clinical decisions. J Am Coll Cardiol 1990;16:108-13.
13. White CW, Wright CB, et al. Does visual interpretation of the coronary arteriogram predict the physiologic importance of a coronary stenosis? N Engl J Med 1984;310:819-24.
14. Gould KL. Functional and anatomic assessment of coronary artery stenosis. In: Heintzen PH, Bursch JH, editors. Progress in digital angiography. Boston, Mass: Kluwer Academic Publishers; 1988. p. 308-24.
15. Kalbfleisch SJ, McGillem MJ, et al. Comparison of automated quantitative coronary angiography with caliper measurements of percent diameter stenosis. Am J Cardiol 1990;55:329-37.
16. Brown BG, Bolson EL, Dodge HT. Arterographic assessment of coronary arteriosclerosis: review of current methods, their limitations, and clinical applications. Arteriosclerosis. 1982;2:1-15.
17. Goldberg R, Kleiman N, Minor ST, et al. Comparison of quantitative coronary angiography to visual estimates of lesion severity pre and post PTCA. Am Heart J. 1990;119:178-84.
18. Fleming R, Kirkeeide R, Smalling RW, et al. Patterns in visual interpretation of coronary arteriography as detected by quantitative angiography. J Am Coll Cardiol 1991;18:945-51.
19. Bernstein SJ, Hilborne LH, Leape L, et al. The appropriateness of use of coronary angiography in New York state. JAMA 1993;269:766-9.
20. Spero LA, Hanemann JD, Cusma JT, et al. A multiuser networked system for the large scale study of coronary artery stenosis using quantitative and qualitative coronary angiography. IEEE Comput Cardiol 1990;195-8.
21. Hartman CW, Kong Y, Margolis JR, et al. Aortocoronary bypass surgery: correlation of angiographic, symptomatic, and functional improvement at 1 year. Am J Cardiol 1976;37:352-7.
22. Topol EJ, Nissen SE. Our preoccupation with coronary luminology. Circulation 1995;92:2333-42.