Interobserver Variability in Coronary Angiography

LEONARD M. ZIR, M.D., STEPHEN W. MILLER, M.D., ROBERT E. DINSMORE, M.D., JOHN P. GILBERT, PH.D., AND J. WARREN HARTHORNE, M.D.

SUMMARY Four experienced coronary angiographers (two radiologists and two cardiologists) independently assessed the location and degree of coronary artery stenosis, and the location and degree of left ventricular wall motion abnormalities in 20 coronary angiograms.

Marked interobserver variability was noted in quantifying percent coronary artery stenosis and degree of left ventricular wall motion abnormalities. For example, in only 13/20 (65%) of the coronary angiograms did all observers agree as to the significance of a stenosis (defined as greater than 50% in diameter luminal narrowing) in the proximal or mid left anterior descending coronary artery.

In 3/20 (15%) angiograms there was disagreement by at least one observer about the significance of lesions noted in the main left coronary artery.

The ventricle was divided into five segments and the degree of wall motion abnormality graded into six categories of increasing severity from normal to dyskinesis. There was a 42% mean disagreement among all four observers where a disagreement between observers was defined as any difference in grading wall motion abnormalities.

Interobserver variability reveals a significant limitation of coronary angiography.

Methods

Four experienced coronary angiographers from the Massachusetts General Hospital independently evaluated 20 coronary angiograms for the presence and location of coronary artery stenosis and location and degree of left ventricular wall motion abnormality. Although the background and experience of the four angiographers differed in some ways, each angiographer has performed or interpreted at least 1500 coronary angiograms. Two observers, JWH, a cardiologist, and RED, a radiologist, introduced coronary angiography to the Massachusetts General Hospital in 1967. The other two observers, LMZ, a cardiologist, and SWM, a radiologist, are senior members of the Catheterization Laboratory staff of the Massachusetts General Hospital. The 20 good quality coronary angiograms used in this study were selected by a senior X-ray technician from the 600 coronary angiograms performed at the Massachusetts General Hospital from January to June, 1974. Starting with the first coronary angiogram performed in January, 1974, approximately every twentieth coronary angiogram was selected for analysis. The senior X-ray technician, who is in a large part responsible for angiographic quality control in our laboratory, was instructed to select only angiograms of good quality. In this study good quality was based upon the visual recognition of an acceptable diagnostic coronary angiogram. The optimal resolution of our cineangiographic system was measured by placing a star phantom on the image intensifier screen and exposing the phantom image on 35 mm Dupont SF-2 film. The cine of the star phantom was projected on a Tajarno-35 projector and demonstrated a resolution of 2 line pairs per mm.

None of the observers participated in the selection of the angiograms and none of the angiograms were of patients followed clinically by the observers. As one or more of the observers may have interpreted the angiograms prior to the present study, the angiograms were blinded as to patient

From the Cardiac Unit and the Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts, and the Office of Information Technology, Harvard University, Cambridge, Massachusetts.


Address for reprints: Leonard M. Zir, M.D., Cardiac Catheterization Laboratory, Massachusetts General Hospital, Boston, Massachusetts 02114.

Received July 28, 1975; revision accepted for publication November 4, 1975.
name, date of study, and diagnosis. Furthermore, angiograms were interpreted in the present study six months to one year after they had been performed.

Coronary angiography and ventriculography were performed by the Judkins or Sones technique. Coronary angiograms were recorded at 30 frames/second on 35 mm Dupont SF-2 film utilizing a Philips 6-inch Cesium iodide image intensifier with submaximal overframing (100 mm lens). Multiple transverse projections of the right coronary artery and left coronary artery were recorded until all lesions were satisfactorily delineated. Cineangiograms were simultaneously recorded on video tape and the tapes were reviewed for adequacy of the diagnostic study prior to withdrawal of the catheters. All cineangiograms were independently interpreted by all four observers on the same projector (Tajarno-35). The percent arterial stenosis was graded from 0–100% and based on the maximal diameter narrowing seen in all projections of the coronary artery. A significant coronary artery lesion was defined as a 50% or greater reduction in diameter.

The coronary arteries were divided into the following segments to assist in the analysis of the interobserver variability (fig. 1). The main left coronary artery was defined as that portion of the left coronary artery from the left coronary ostium to the bifurcation of the left anterior descending (LAD) and the left circumflex (CFX) coronary arteries, or trifurcation when an intermediate artery was noted. The proximal LAD extends from the origin of the LAD up to and including the first septal perforating branch. The mid-LAD includes the middle third of the anterior descending after the origin of the first septal perforating branch. The proximal circumflex extends from the bifurcation of the LAD and CFX to the first obtuse marginal branch, and the distal CFX from the first obtuse marginal branch to its termination. The proximal right coronary artery (RCA) extends from the ostium and includes the first third of the vessel, the mid-RCA and the middle third of the RCA.

Analysis of suitability of distal vessels for bypass grafts was not included in this study. It was felt that assessment of distal vessels, which includes distal vessel size, runoff, and the presence and morphology of obstruction (single lesion or diffuse disease), is highly subjective. In the present study analysis of interobserver variability was limited to percent narrowing of arterial lesions, presumably the most objective aspect of coronary angiography.

Interobserver variability in the interpretation of coronary artery lesions was analyzed in several ways. First, analysis was directed at the presence or absence of a significant coronary lesion in the various coronary segments. A significant coronary lesion was defined as a 50% or greater narrowing in diameter.

Each coronary segment in all 20 coronary angiograms was analyzed for the number of times there was agreement by all four observers. Percent agreement for a coronary segment was defined as the number of times all four observers agreed divided by 20, the total number of possible comparisons × 100. In addition, for any segment analyzed, the number of disagreements in which one observer differed from the other three, or two observers differed with the other two was also tallied.

Second, an analysis of variance was performed comparing the four observers, three vessels (RCA, LAD, CFX) and 19 angiograms. (Angiogram B showed all normal vessels for all observers and was not included in this analysis.) To minimize location errors the maximum percent stenosis seen in the proximal or mid-LAD and RCA and proximal or distal CFX was used.

Third, the degree of agreement between observers was also measured for the LAD by computing the Pearson product-moment correlation coefficient for each pair of observers. For this calculation, the highest percent stenosis in the proximal or mid-LAD was taken for each angiogram.

One factor influencing interobserver variability is the number of positive findings in a study. Observers would be expected to agree more if there were a high percentage of negative findings and less if there were a high percentage of possible positive findings. Therefore, the percent of "positive" findings was determined for each coronary segment and was calculated as the number of significant lesions counted for a segment over all possible 80 counts (4 observers × 20 angiograms).

The right anterior oblique ventriculogram was divided into five segments (fig. 2): anterobasal, anterolateral, apical, diaphragmatic, and posterobasal. Wall motion abnormalities were classified as follows: normal, mild, moderate and severe hypokinesis, akinesis, and dyskinesis. Hypokinesis was defined as a reduction in systolic contraction, akinesis as absent systolic contraction, and dyskinesis as paradoxical systolic bulging.

Percent disagreement for each segment was calculated as follows: for each ventriculographic segment the total number of disagreements between each observer and the

---

**Figure 1.** Coronary artery segments used for comparison of observer evaluation. LMCA is the left main coronary artery. LAD is the left anterior descending. CFX is the left circumflex and RCA is the right coronary artery.

**Figure 2.** Ventricular segments used for comparison of observer evaluation.
other three observers was tallied and expressed as a percentage of 60, the total number of possible comparisons. For example, to calculate the percent disagreement for observer 1 for the 20 apical segments, the number of times observer 1 disagreed with observers 2, 3, and 4 would be tallied and expressed as a percent of 60, the total number of comparisons. The mean percent disagreement for that segment would be the average percent disagreement for all four readers. The over-all percent disagreement was the average percent disagreement of all four observers for all five segments.

A disagreement in ventriculography between observers was defined in two ways in order to test the sensitivity of the ventricular classification system. First, a disagreement between two observers was defined as a difference by more than one class of severity. For example, the difference between two observers, one who read a segment as mild hypokinesis and one who read the same segment as moderate hypokinesis would be an agreement. Whereas, if one segment were labeled mild hypokinesis and one segment as severe hypokinesis, a disagreement would be noted. Second, disagreement was defined as any difference between two observers. If one observer read a segment as mild hypokinesis and one read the same segment as moderate hypokinesis a disagreement would be tallied.

Results

A striking degree of interobserver variability was noted in the interpretation of coronary angiography.

In the left main coronary artery, there was agreement by all four observers in 17/20 (85%) coronary angiograms. There was disagreement by one observer with the other three observers in two coronary angiograms and disagreement by two observers in one coronary angiogram. The number of "positive" findings was 4/80 (5%).

In only 9/20 (45%) of the coronary angiograms did all four observers agree about the significance of a lesion in the proximal LAD. There was disagreement by only one observer in 7/20 angiograms and disagreement by two observers in 4/20. Examination of table 1, which shows the percent stenosis of the proximal and mid-LAD for all 20 angiograms by all four observers, shows that part of the variability may be due to location errors. One observer's proximal lesion may be defined by another observer as a mid lesion. In three cases, angiograms H, M and S, a single significant lesion was noted in the proximal or mid LAD by all four observers and differences in these angiograms might be attributable in part to identification of the location of the lesion. If one assumes that in these three angiograms there is a location error and removes that factor as a disagreement, then there is an improvement in the number of cases of agreement in the proximal LAD from 9/20 (45%) to 12/20 (60%). The number of positive findings in the proximal LAD was 35/80 (44%).

In only 9/20 (45%) of the angiograms did all four observers agree about the significance of a lesion in the mid-LAD. In eight cases there was disagreement by one observer with the other three and in three cases disagreement by at least two observers. If one assumes that in patients H, M, and S there is a location error, then agreement is improved to 12/20 (60%). The number of positive findings in the mid LAD was 33/80 (41%).

If one asks how frequently did all observers agree about a lesion in the proximal or mid-LAD then in only 13/20 cases (65%) was there total agreement. There was disagreement by two observers in two cases.

The range of percent stenosis of the LAD in some of the coronary angiograms was striking. For example, in angiogram A (table 1) the range of percent stenosis was from 0–90% in the proximal LAD and from 60–99% in the mid-LAD. In patient C, two readers thought there was 100% stenosis of the mid-LAD and two observers thought that the LAD was completely normal.

In the proximal CFX, there was agreement by all four observers in 15/20 (75%) of the angiograms with three cases of disagreement by one observer and two cases of disagreement by two observers. In only one angiogram, S, was there a suggestion of location error, i.e., a single lesion noted by all four observers in the proximal or distal CFX. If one assumes the discrepancy in angiogram S is secondary to location rather than severity of lesion total agreement for the proximal CFX improves to 16/20 (80%). The number of positive findings for the proximal CFX was 18/80 (23%). In the distal CFX, there was total agreement in only 11/20 (55%). There was disagreement by one observer with the other three in eight angiograms and disagreement by two observers in one angiogram. The number of positive findings in the distal CFX was 19/80 (24%).

In 13/20 (65%) of the angiograms there was complete agreement about the significance of lesions in the proximal RCA with five cases of disagreement by one observer and two with disagreement by two observers. To circumvent potential location error, complete agreement improved to 17/20 (85%) when agreement was analyzed for the proximal or mid-RCA. In all three cases of disagreement, the disagreement was always by one observer with the other three. The number of positive findings in the proximal RCA was 29/80 (36%) and in the mid RCA 40/80 (50%).
The analysis of variance comparing observer, vessel, and angiogram effects found that angiograms, vessels, and angiogram-by-vessel interaction significantly affected the variability found and that consistent bias on the part of one or more of the observers was not a significant factor. Angiograms with 18 and 108 degrees of freedom yielded $F = 11.97$; vessels with 2 and 108 degrees of freedom, $F = 33.92$; and the angiogram by vessel interaction term with 36 and 108 degrees of freedom, $F = 8.30$. All three $F$ values relate to a $P$ of less than 0.001. This indicates that some patients have more disease than others, some vessels are more apt to be diseased than others, and (the interaction term) that the pattern of disease varies from patient to patient. In contrast, the value of the $F$ statistic for observers, 0.88, with 3 and 108 degrees of freedom, was not significant. Thus, the disagreement found between readers was due to one or two readers being consistently biased either in the direction of overestimating or underestimating the amount of stenosis.

The correlation coefficients for each pair of observers for the proximal or mid LAD are given in Table 2. The highest of these is 0.857 and a value which indicates that even for this pair of readers (No. 1 and No. 4) only about three quarters of the variability of one reader could be eliminated by utilizing the finding of the other, while the pair with the most disagreements (readers No. 2 and No. 4; $r = 0.448$) suggests that only 20% of the variance could be accounted for by knowing the value of the other reader.

A different method of determining if one reader’s interpretations were skewing the results was an analysis of the “odd man out.” In analysis of the left main, proximal and mid LAD, proximal and mid RCA, and proximal and distal CFX there were 18 disagreements where one observer differed from the other three. Observers 1, 2, and 4 differed with the other three observers in five cases and observer 3 in three cases. Hence, no one observer appeared to account for all the interobserver variability noted in this study.

In Table 3A disagreement was defined as any difference in interpretation of wall motion between observers. There was a mean percent disagreement of 42% by all four observers. There was a marked reduction in mean percent disagreement to 16% when a difference between observers is defined as more than one class of wall motion abnormality (Table 3B). There was least disagreement about anterobasal wall motion and approximately equal disagreement about the remaining four segments. Of interest was the range of variability between some observers in some segments where wall motion was variously called normal to dyskinetic. Of 100 LV wall segments analyzed (five segments per 20 coronary angiograms) there were 11 segments in which disagreement was by three or more classes of severity. Although marked interobserver variability was noted, the percent disagreement for each observer was remarkably similar, indicating that no one observer contributed to all the interobserver variability (Table 3A, B).

Discussion

This study demonstrates a substantial degree of interobserver variability among four experienced coronary angiographers in the interpretation of coronary angiography and ventriculography. This variability is surprising in that all four angiographers are from the same institution and have frequently participated as a group in the interpretation of coronary angiograms for the diagnosis of coronary artery disease and the advisability of coronary artery bypass surgery. No one observer accounted for a substantial portion of the variability.

One function of interobserver variability is the number of positive findings. A greater percentage of positive findings leads to a greater observer disagreement and a greater percentage of negative findings to greater observer agreement. The lesser interobserver variability in interpretation of the main left coronary artery when compared to the interpretation of the LAD coronary artery is undoubtedly explained by the higher percentage of insignificant lesions in the main left coronary artery. In the present study, observer agreement would have been different if a higher percentage of angiograms with “normal” coronary arteries or single vessel disease had been analyzed. The 20 angiograms in the present study included a fairly typical spectrum of coronary angiographic findings seen at our institution. The films included one normal angiogram (as judged normal by all four observers) as well as a sampling of single, double, and triple vessel disease.

The problem of interobserver variability is not new to radiology. Birkelo et al. in 1947 reported significant inter-

| Table 3A: Disagreement Among Four Observers on Ventriculograms |
|-------------------|--------|--------|--------|--------|
| Segment           | 1     | 2     | 3     | 4     |
| Anterobasal       | 5     | 6     | 8     | 6     |
| Anterolateral     | 55    | 50    | 48    | 63    |
| Apical            | 52    | 45    | 48    | 52    |
| Diaphragmatic     | 53    | 42    | 43    | 57    |
| Posterobasal      | 45    | 52    | 47    | 60    |
| Mean % observer disagreement | 42 | 39 | 39 | 48 | 42 |

| Table 3B: Disagreement Among Four Observers on Ventriculograms |
|-------------------|--------|--------|--------|--------|
| Segment           | 1     | 2     | 3     | 4     |
| Anterobasal       | 0     | 2     | 0     | 2     |
| Anterolateral     | 18    | 22    | 25    | 25    |
| Apical            | 22    | 13    | 15    | 13    |
| Diaphragmatic     | 20    | 18    | 20    | 22    |
| Posterobasal      | 18    | 17    | 17    | 19    |
| Mean % observer disagreement | 16 | 14 | 15 | 17 | 16 |

The numbers are percent disagreement. In 3A, disagreement is defined as any difference between observers. In 3B, disagreement is defined as differences between observers of more than one class of wall motion abnormality.
observer and intraobserver variability in the interpretation of chest films for tuberculosis by five experienced radiologists. In some cases 20% of the films called positive for tuberculosis by one observer were called negative by another observer. Bjork et al. allude to considerable interobserver and intraobserver differences in assessing coronary arterial stenosis in a study comparing the diagnostic value of the 70 mm camera and 16 cine in coronary angiography.

The reasons for interobserver variability in coronary angiography may be classified into three broad categories: those related to the performance of angiography; the angiographic equipment; and the observer. It is well recognized that the coronary vessels should be selectively catheterized for optimal resolution. Multiple transverse projections are necessary to fully delineate stenotic lesions as most atherosclerotic plaques are eccentric rather than concentric and lesions that appear insignificant in one projection may be quite significant in another projection. Failure to obtain multiple projections may leave stenotic vessels overlapped by normal vessels. Individual readings of the degree of stenosis in vessels that are overlapped or not clearly delineated vary because a certain amount of subjective visualization is involved to make quantitative assessment of these lesions. In our experience difficulties in demonstrating stenotic lesions occur most often with the left main and left anterior descending coronary arteries. To delineate lesions in these vessels it is frequently necessary to use transaxial views as well as multiple transverse projections.

The technical quality of the final product, the coronary cineangiogram, is the sum of many factors including the X-ray tube and generator, quality of image intensifier, the camera and lens system, the type of film and film processing and type of projector used. The entire angiographic system requires continuous attention by experienced personnel to maintain optimum resolution. What a technically acceptable coronary angiogram is has not been fully defined. Our experience with viewing films from many other institutions suggests that the factors affecting quality of coronary angiograms such as background “noise,” magnification, and degree of film contrast (as opposed to resolution) may vary widely among catheterization laboratories and yet films of different quality may still be of equal diagnostic value. However, such differences in “film quality” may lead to differences in interpretation of coronary artery lesions.

Differences in reader perception undoubtedly contribute to interobserver variability. The type of training of the angiographer, whether radiologist or cardiologist, and the level of experience and ability are both important factors in interobserver variability. Reader fatigue, pertinent at our institution where 10-12 coronary angiograms are performed daily, and “reader set” (prior knowledge of a patient’s clinical history) all may be important contributing factors. For example, a single vessel with a borderline stenosis might be more likely to be called significant by a reader who knows the patient has definite angina pectoris associated with ECG changes in the area of the borderline lesion.

Discussion of the results of the study among the four observers revealed that one of the most important factors contributing to variability was which portion of the vessel to use in calculating the percent coronary arterial stenosis. Frequently no part of a vessel was normal and there was no standard to judge the degree of coronary stenosis. Coronary stenosis is infrequently a simple and concentric lesion. There may or may not be poststenotic dilatation, lesions may be multiple, lesions may be smooth and tapering, vessels may become suddenly narrowed and remain narrowed in their distal course. Clearly, a simple definition of percent stenosis as the maximal diameter narrowing divided by the diameter of the immediate proximal or immediate distal artery was not always appropriate.

Our study suggests that one method of reducing the error introduced by interobserver variability is “group opinion” based on independent reading, a method suggested to reduce error in the mass screening of chest X-rays for tuberculosis. If a significant lesion is defined as a diameter narrowing greater than 50% seen by the majority of observers (in this case three of four observers), then there is a marked reduction in the effects of interobserver variability in our study. Defined in this way the number of disagreements about the left main was reduced from 3/20 to 1/20, the proximal LAD from 11/20 to 4/20, the proximal or mid-LAD from 7/20 to 2/20, the proximal CFX from 6/20 to 1/20 and the proximal or mid-RCA from 3/20 to no disagreement. The accuracy of “group opinion” remains to be verified by other means of quantifying lesions such as autopsy findings.

An alternative approach is “consensus opinion” based on several observers reading the coronary angiogram at the same time. This is the method frequently used at our institution where the cardiologist, radiologist, and surgeon examine the angiograms together and decide upon the necessity for bypass surgery. One of the difficulties with consensus opinion is that valid disagreements may be overlooked especially if one member of the group is dominant or overly persistent. Nevertheless, obvious errors of individual members of the group may be corrected by other members.

The marked interobserver variability in the subjective interpretation of left ventricular wall motion abnormalities is more to be expected than the variability noted in the interpretation of coronary lesions. The differences between observers has been noted previously and many methods have been suggested to quantify left ventricular wall motion abnormalities. Nevertheless, the extreme differences in interpretation seen in some segments between the four observers was striking. In some cases, one man’s dyskinesis was another man’s normal motion. The results of this study underscore the need for an objective system of quantifying left ventricular wall motion abnormalities so that data among several institutions will be based on similar scales of interpretation and studies from several institutions can be compared in a meaningful way.

The clinical implications of marked interobserver variability in coronary angiography are disturbing. Although the decision for coronary artery bypass surgery is not based solely on the presence of a significant lesion, variability in the interpretation of the significance of a coronary lesion could lead to variability in the recommendations for coronary artery bypass surgery. For example, the demonstration of a significant lesion in the main left coronary artery is an indication in many centers for coronary artery bypass surgery even in the absence of significant angina. In the present study, in 3/20 angiograms at least one observer’s interpretation of significance of a left main coronary artery le-
sion could have changed a decision for immediate coronary surgery. Interobserver variability in the interpretation of lesions in other coronary vessels might be similarly translated into different decisions about the necessity for coronary artery bypass surgery, or if coronary artery bypass surgery is to be performed, which vessels are bypassable. Interobserver variability is a significant limitation of coronary angiography and clearly requires further study.

References

---

**Intervention Ventriculography**

**Comparative Value of Nitroglycerin, Post-extrasystolic Potentiation and Nitroglycerin Plus Post-extrasystolic Potentiation**

**Vidya S. Banka, M.D., Monty M. Bodenheimer, M.D., Rajnikant Shah, M.D., and Richard H. Helfant, M.D.**

**SUMMARY**

The comparative value of nitroglycerin (TNG), post-extrasystolic potentiation (PESP), and their combination (TNG + PESP) to unmask asynergic residual contraction was examined, each patient serving as his own control. Twelve of 13 hypokinetic zones improved both with TNG and PESP. One remained unchanged with either. Of 15 akinetic zones, four improved with both TNG and PESP, while ten remained unchanged. One akinetic zone, although improved with TNG, remained unchanged with PESP. Four dyskinetic zones did not change with either. Six asynergic zones responding to TNG alone demonstrated further augmentation with TNG + PESP. However, none of 13 TNG unresponsive zones improved with TNG + PESP.

Thus, TNG, PESP, and TNG + PESP are each equally capable of unmasking asynergic residual contractile ability.

**RECENT STUDIES** have shown that the residual contractile ability of asynergic zones can be assessed ventriculographically using nitroglycerin, post-extrasystolic potentiation, or catecholamine (epinephrine) infusion. These interventions involve very different mechanisms for unmasking contractile reserve. Nitroglycerin acts presumably by improving the balance between oxygen demand and supply to the chronically ischemic zone either due to its unloading effect and/or to increased regional coronary blood flow. Post-extrasystolic potentiation has been found superior to epinephrine as an intervention which increases contractility.

A comparison of the value of nitroglycerin and post-extrasystolic potentiation in assessing contractile reserve has not been previously examined. In addition, the potential utility of a combined intervention which would both decrease ischemia (i.e., nitroglycerin) and increase contractility (i.e., post-extrasystolic potentiation) in unmasking reversible asynergy in a zone unresponsive to a single intervention is unknown. The present study was, therefore, undertaken to compare the ventriculographic changes induced by both nitroglycerin and post-extrasystolic potentiation in patients with coronary artery disease and asynergy. These responses were also compared to a nitroglycerin plus post-extrasystolic potentiation intervention.

**Material and Methods**

Thirty-six patients undergoing cardiac catheterization for evaluation of coronary heart disease were selected for study based on the following criteria: 1) asynergy on ventriculography (defined as a localized abnormality of left ventricular contraction); 2) appearance of one to three premature ventricular beats during injection of contrast material into the left ventricle during the initial ventriculogram and/or nitroglycerin ventriculogram; 3) significant (≥ 75% decrease in diameter) obstruction of one or more of the three major