Evaluation of Human Panelists in Assessing Coronary Atherosclerosis

Stanley P. Azen, Linda Cashin-Hemphill, Janice Pogoda, Wendy J. Mack, Miguel E. Sanmarco, Emily Wickham, and David H. Blankenhorn

The Cholesterol Lowering Atherosclerosis Study, a randomized, angiographic clinical trial, has demonstrated the beneficial effect of niacin/colestipol therapy on coronary and femoral atherosclerosis. The primary outcome was a panel-determined consensus score evaluating global coronary changes determined angiographically at 2 years. This article presents an evaluation of interreader agreement in independently assessing the status of native coronary arteries and overall coronary condition. Parameters include 1) identification of the presence of lesions and lesion changes; 2) estimation of lesion severity (percent stenosis) and amount of change in lesion severity; and 3) global assessment of change in coronary status. Readers independently agreed on 1) presence of lesions (82%) and change in lesions (51%); 2) percent stenosis±10% (76%) and change in stenosis±10% (81%); and 3) global assessment of change in coronary status within one step (96%). Results of these analyses may be useful in effectively designing angiographic trials that use a panel of human evaluators as well as computerized methods for angiographic interpretation. (Arteriosclerosis and Thrombosis 1991;11:385–394)

Valid and reliable information about the status of coronary lesions is needed for evaluation of interventional therapy in controlled angiographic trials. When patients are randomized to test or control, the required information should lead to analysis on a per-patient basis to determine trial outcome. For patients with multiple lesions, results for each lesion must be integrated into a single score for that patient.

Coronary lesions have been evaluated for clinical research in three general ways: 1) human readers viewing angiographic films alone1-3 or as panel members4-6; 2) human readers measuring films with calipers7 or tracing vessel edges by hand for later computer analysis8; and 3) computerized edge-finding after human designation of the search areas.8-10 The reliability of film evaluation by humans for changes in coronary status is influenced by 1) technical features, such as film quality, the conditions for film reading, the number of films a reader has recently evaluated, and order of film presentation and 2) coronary pathological and anatomic characteristics, such as lesion size, degree and direction of change in lesions, degree and direction of change in per-patient disease status, degree to which branches overlap, and frequency of lesion formation at branch junctions.11-13

The Cholesterol Lowering Atherosclerosis Study (CLAS) was a randomized, placebo-controlled clinical trial testing the effects of colestipol plus niacin therapy.4 Coronary, carotid, and femoral angiograms, selected to sample treatment effects as widely as possible, were obtained at baseline and at 2 years. CLAS was monitored by an external advisory committee, which mandated an analysis of coronary vessels at 2 years as a requirement for a proposed trial extension (CLAS-II), in which each patient would remain on his assigned treatment for 4 years and then have a third angiogram. Because CLAS involved high-risk patients, ethical considerations placed a premium on rapid and timely analysis of interim data. Therefore, the interim analysis and subsequent trial management were based on angiogram reading by a human panel rather than computerized image analysis, which required extensive processing time.

This article examines factors affecting the reliability of the coronary panel process in evaluating coronary films. We present the results of a detailed evaluation of 1) interreader agreement on native artery lesion identification and degree of stenosis, 2) interreader agreement on native artery lesion change and amount of change in stenosis, and 3) interreader and interpanel agreement on the overall evaluation.
of coronary status (including bypass grafts). In this situation, as in other controlled studies dealing with humans, analyses regarding the validity of the angiographic interpretation (by comparisons with a known “true” value) are not technically or ethically possible. Results of these analyses may be useful in designing angiographic trials that use all methods of angiographic interpretation most effectively.

Methods
Study Groups and Experimental Design

The CLAS design has been described elsewhere. Briefly, CLAS was a randomized, placebo-controlled angiographic trial testing combined therapy of colestipol hydrochloride and niacin plus diet in 162 (of 188 randomized) nonsmoking men aged 40-59 years with progressive atherosclerosis and previous coronary artery bypass surgery (80 men in the drug group, 82 in the placebo group). Both groups received diet intervention, with a different diet composition to enhance the differential in blood cholesterol responses between the two groups (a 30-40% drop in the drug group, and a 7-8% drop in the placebo group). Entry fasting blood cholesterol levels were in the range of 185-335 mg/dl. Age at entry (mean±SEM) of all patients was 54.2±0.5 years, and mean blood pressure was 124.0±1.5/80.6±1.0 mm Hg. Subjects were evaluated angiographically before randomization to obtain baseline data on the atherosclerotic disease of the coronary, femoral, and carotid arteries, as well as of the coronary artery bypass grafts. Subjects were then followed up at specified intervals for 2 years, at which time a repeat angiogram was performed.

The coronary end point at 2 years, as evaluated by panels of angiographers, revealed a significant effect due to treatment.14 This led to a truncation of CLAS-II, the 2-year trial extension that entailed 4 years of therapy (drug or placebo) for each patient. In addition, the femoral end point, as evaluated by computerized analyses, demonstrated a significant treatment effect on both a per-lesion and a per-patient basis.15 We have previously published the precision and reproducibility of computerized quantification in a random sample of CLAS coronary angiograms when automated edge tracking and multiple frame averaging were used. This article evaluates data derived from the panel evaluation of coronary end points at 2 years, which were used for the decision to truncate CLAS-II.

Panel Reading Procedures

A panel meeting involved three to six readers and a moderator (not one of the readers). To remove panelists from workday distractions and to allow adequate rest between film readings, panelists were convened for 3 days over a weekend at a vacation resort near Los Angeles or near the annual meeting of the Association of University Cardiologists.

A film-reading panel consisted of two cardiologists and an independent moderator. On average, 45 minutes were required to evaluate the film pair, and panelists were usually scheduled to two consecutive reading sessions and then rotated to a rest period. The procedure for film reading was as follows. Film pairs were mounted by the moderator for side-by-side projection on two 35-mm cineangiographic projectors (Vanguard Instruments, Inc., Melville, N.Y.) that magnified the image sixfold. All angiograms had been exposed to show identical coronary artery views in the same sequence and had been edited and spliced to eliminate extra footage. The baseline and 2-year angiograms from a drug- or placebo-treated patient were randomly selected, and the order of projecting the early or late member of a film pair was randomized to the two projectors. Each panel member was provided with an identical sketch of the patient's coronary tree (Figure 1). These were individually drawn for each patient to display a blank outline of the right anterior oblique view of the coronary arteries and bypass grafts. Panel members had no clinical data regarding patients and were masked as to treatment assignments and temporal ordering of films. Panel members were not shown ventriculograms.

Working independently, each film reader examined the first film (film 1) of a pair to sketch and record native artery stenosis estimates on a diagram without revealing these to the other film reader. Readers were instructed to note only those lesions with 20% or more diameter stenosis. In evaluating bypass grafts, each graft was divided into proximal, middle, and distal thirds. Percent stenosis as well as length of the segment involved were recorded for each third of each graft. Each film reader's diagram was then photocopied to record the individual film reader's identification of lesions and stenosis estimates. The first film was then replayed and discussed openly by the two film readers to arrive at a consensus estimate of the location and degree of all stenotic lesions. During this discussion, consensus estimates were diagrammed on a separate record by the moderator and photocopied for the record and for use by each film reader in evaluating the second member of the film pair (Figure 2).

Film readers then viewed the second film (film 2) for the same patient and independently marked the consensus diagram. All diagrammed lesions were evaluated for change in percent stenosis in native arteries and bypass grafts and for lengths of stenosis in bypass grafts. If there was a change, then the current percent stenosis was estimated. Any lesion not previously noted was recorded with an estimate of percent stenosis, provided that it was at least 20% and was labeled as “new.” In the situation in which a lesion had disappeared, the percent stenosis was noted as 0%.

Each reader working independently recorded a global evaluation of the films on a four-point scale (0-3). Global evaluation of films combined changes in both grafts and native coronary arteries. A predesignated weighting scheme was not used, each panelist
relying on subjective clinical judgment and prior experience to arrive at a score. A score of 0 indicated no demonstrable change; a score of 1, definitely discernible change; a score of 2, intermediate change; and a score of 3, extreme change. Direction of change was assigned by panelists according to the mounting of films on projectors. This score, which was masked for temporal order, is designated as the independent summary score. The panelist was also permitted to compare film 2 with film 1 in evaluating lesion change to determine the global coronary change.

After each reader completed scoring film 2, the second film was replayed and discussed openly to arrive at a consensus estimate of the location and degree of all lesion changes, as well as a consensus summary score of global coronary change, the consensus summary score. Agreed-on estimates were diagramed on the consensus drawing by the moderator (Figure 3).

Determination of Trial Outcome at 2 Years

After all film pairs had been evaluated, the code for temporal order was broken by the project statistician. The resulting decoded consensus summary score, designated as the global coronary change score, ranged from −3 (regression) to +3 (progression). As previously reported, deterioration in overall coronary status was significantly less in drug-treated patients than in placebo-treated patients ($p<0.007$). Atherosclerosis regression, as indicated

---

**Figure 1.** Coronary anatomy sketches for two Cholesterol Lowering Atherosclerosis Study subjects A and B. Bypass grafts are numbered and divided into thirds.

**Figure 2.** Moderator's diagram of coronary anatomy sketches for two Cholesterol Lowering Atherosclerosis Study subjects A and B after film 1 had been viewed by both panelists. Bypass grafts showed no lesions in either patient. Native coronary artery lesions have been drawn and assigned a consensus severity estimate expressed as percent diameter stenosis. The letter "D" preceding a severity estimate indicates that the lesion is diffuse.
by perceptible improvement in overall coronary status, occurred in 16.2% of colesteplol/niacin-treated patients versus 3.6% of placebo-treated patients.

Panel Database

In all, there were 11 readers who were paired in 23 different panels. Readers were not represented equally in the panels; for example, one reader was a member of nine different panels, while a second reader was a member of only two. Panels evaluated 190 pairs of angiograms (162 patients plus 28 randomly selected duplicates). On average, 15 native coronary artery segments were evaluated per patient, and these contained an average of 10.6 lesions. On average, 2.6 bypass grafts per patient were evaluated.

Data on native coronary arteries were abstracted by a statistician from the diagrams used by the readers to record their independent evaluations. (Bypass graft lesions were not abstracted for this article.) The statistician abstracted the recorded percent stenosis of each identified lesion (determined from film 1) and whether a lesion had changed, and if so, the percent stenosis of the changed lesion (from film 2). The amount of change was defined to be the absolute difference in percent stenosis values. In addition, the independent summary and the consensus summary scores were abstracted. Although data on bypass grafts were not abstracted for this study, the two summary scores did reflect changes in the bypass grafts as well as in the native coronary arteries.

A total of 2,135 native coronary artery lesions were identified, forming a data set of 8,540 observations (two observations per lesion for each of film 1 and film 2). Of the 8,540 observations, 94.8% were used for this study; 2% were excluded because of minor problems in abstracting the data (e.g., a panel reader noted total occlusion but failed to write 100%); 3.2% of the data was excluded as not comparable (e.g., final consensus indicated two contiguous lesions, but one reader recorded a single lesion).

Statistical Analysis

Interreader agreement was calculated as the ratio of joint agreement to consensus agreement. Joint agreement was defined as agreement on lesions independently identified by both readers before they shared opinions, while consensus agreement was defined as agreement independently identified by at least one reader and eventually agreed on by both readers after they shared opinions. The degree of interreader agreement was assessed (when possible) using a stratified $\kappa$ coefficient (stratum=panel). Comparisons of proportions used $\chi^2$ methods. Significance was established at the 0.05 level.

Results

Distribution of Lesions and Lesion Change

In total, there were 1,830 film 1 lesions: 12% (221/1,830) with stenosis equal to 20%; 28% (511/1,830) with stenosis in the range 25–45%; 23% (428/1,830) with stenosis in the range 50–70%; 21% (379/1,830) with stenosis in the range 75–95%; and 16% (291/1,830) with stenosis greater than 95%. Of these lesions, 86% (1,579/1,830) were identified on film 2 by both readers. Of these 29% (456/1,579) were regarded as changed lesions according to consensus opinion: 16% (73/456) with initial stenosis equal to 20%; 27% (121/456) with initial stenosis in the range 25–45%; 30% (136/456) with initial stenosis in the range 50–70%; 23% (107/456) with initial stenosis in the range 75–95%; and 4% (19/456) with...
Interreader Agreement of Lesions and Lesion Change was greater for diameter stenoses of 20% and greater before a patent graft, proximal to a very-high-grade lesion, or located in a very small vessel.

<table>
<thead>
<tr>
<th>Presence of lesion</th>
<th>Consensus agreement (n)</th>
<th>Joint agreement (n)</th>
<th>Interreader agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity (% stenosis)</td>
<td>1,830</td>
<td>1,498</td>
<td>82</td>
</tr>
<tr>
<td>Perfect</td>
<td>1,498</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>≤10%</td>
<td>1,142</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>&gt;10%</td>
<td>356</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Lesion change</td>
<td>456</td>
<td>232</td>
<td>51</td>
</tr>
<tr>
<td>Perfect</td>
<td>92</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>≤10%</td>
<td>187</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>&gt;10%</td>
<td>45</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

Consensus agreement, determined by one or both readers and agreed on by consensus; joint agreement, independently determined by both readers and agreed on by consensus; interreader agreement, joint agreement or consensus agreement; amount of change, absolute difference in percent stenosis.

Initial stenosis greater than 95%. The distribution of the amount of change (e.g., the absolute difference in percent stenosis values) according to consensus opinion was 20% (92/456) less than 10%; 29% (133/456) in the range 10–14%; 28% (128/456) in the range 15–20%; and 23% (103/456) greater than 20%.

Interreader Agreement of Coronary Status

Table 1 and Figure 4 (upper panels) summarize interreader agreement in assessing the presence of lesions and original lesion severity. Table 1 and Figure 4 (lower panels) summarize interreader agreement in assessing the presence of lesion change and the amount of change.

Interreader agreement in identifying the presence of lesions was 82% (95% confidence interval [CI]=79–85%, Table 1). This agreement increased with increasing lesion severity (p<0.001, Figure 4, upper left panel). $\chi^2$ testing indicated significantly greater interreader agreement for lesions greater than or equal to 30% than that for lesions less than 30% (p<0.001).

A 50% random sample was selected from the 116 lesions that had at least 50% stenosis (according to consensus) and were not independently identified by one of the two readers. All 11 readers were represented in this sample. Seventy-six percent of the nonidentified lesions were either high-grade lesions before a patent graft, proximal to a very-high-grade lesion, or located in a very small vessel.

For those lesions that were jointly identified, interreader agreement on lesion severity was perfect for 38% of the lesions (95% CI=34–43%) and was within 10% for 76% of the lesions (95% CI=72–80%, Table 1). Interreader agreement (both perfect and within 10%) in assessing lesion severity was greater for diameter stenoses of 20% and greater than 90% than it was for stenoses in the range 30–80% (Figure 4, upper right panel). $\chi^2$ analyses indicated that perfect interreader agreement was greatest (89%) when total occlusion was present (p<0.001).

Interreader agreement on the presence of lesion change was 51% (95% CI=43–58%, Table 1). This corresponded to a stratified $\kappa$ (±SE) of 0.44±0.02, p<0.001. Stratification by stenosis decile revealed that interreader agreement on the presence of lesion change did not depend on the original severity of the lesion (Figure 4, lower left panel). However, agreement on the presence of lesion change varied with the amount of change: 17% (16/92) for changes less than 10%; 39% (52/133) for changes in the range 10–14%; 66% (85/128) for changes in the range 15–20%; and 77% (79/103) for changes in the range greater than 20%. All jointly identified changes were in the same direction; there were no cases for which one reader saw increased stenosis while the other reader saw decreased stenosis.

Of those changed lesions that were jointly identified, interreader agreement on the amount of change was perfect for 40% of the lesions and was within 10% for 81% of the lesions. This corresponded to a stratified $\kappa$ (±SE) of 0.42±0.02, p<0.001. Interreader agreement on the amount of lesion change depended on the absolute amount of change (p<0.001, Figure 4, lower right panel).

Interpanel Agreement of Coronary Status

Table 2 summarizes interreader agreement in assessing global coronary status. One of the 190 angiograms had a missing global coronary change score by one reader, reducing the sample size to 189. Of these, there was perfect agreement on the global coronary change score in 54% of cases (95% CI=46–63%) and agreement within one step (e.g., ±1 versus +2 or −1 versus 0) in 96% of cases (95% CI=92–100%). This corresponds to a $\kappa$ (±SE) of 0.54±0.05, p<0.001.

There were eight occurrences in which readers differed by more than one step. In five of these instances, the readers disagreed on the direction of change. Although readers did not disagree on the direction of changed lesions (see previous section), disagreement on the direction of the global score was possible since the global score also integrated information regarding change in bypass grafts. In all eight occurrences, the readers compromised at consensus to the average of the steps.

Interreader agreement in identifying lesion change increased with increasing global coronary change: 43% agreement for patients scored 0 (no change), 50% for patients scored ±1 (slight change), and 62% for patients scored ±2 or ±3.

Interpanel Agreement of Coronary Status

Twenty-eight angiogram pairs were independently evaluated by two panels. For 20 panels (71%), there was perfect agreement in coronary status as determined by the global coronary change score. For all but one angiogram pair, panel agreement was within one step (96%). In this situation, all four panelists judged the film differently: one
FIGURE 4. Bar graphs of interreader agreement (%) in lesion identification by deciles of lesion severity (upper left panel), with numbers above bars indicating number of lesions in each decile identified by consensus. Upper right panel: Interreader agreement (%) on degree of stenosis by deciles of lesion severity among 1,498 lesions, with numbers above bars indicating number of lesions in each decile identified by consensus. Lower left panel: Interreader agreement (%) on lesion change by deciles of lesion severity, with numbers above bars indicating number of lesions in each decile identified by consensus. Lower right panel: Interreader agreement (%) on amount of lesion change by amount of change among 456 changed lesions. ■, in left panels, lesion severity as % stenosis; ■, in upper right panel, perfect; ■, in upper right panel, within 10%; ■, in lower right panel, amount of lesion change as % stenosis.
Table 2. Agreement in Reaching Global Coronary Change Score
(No. of Consensus Agreements = 189)

<table>
<thead>
<tr>
<th>Level of agreement</th>
<th>Joint agreement (n)</th>
<th>Interreader agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect</td>
<td>103</td>
<td>54</td>
</tr>
<tr>
<td>≤ 1 step</td>
<td>181</td>
<td>96</td>
</tr>
<tr>
<td>&gt; 1 step</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

Consensus agreement, determined by one or both readers and agreed on by consensus; joint agreement, independently determined by both readers and agreed on by consensus; interreader agreement, joint agreement/consensus agreement.

consensus global coronary change score was $+2$, and the second score was 0. Interreader disagreement was due to differences in subjective clinical judgment in weighting lesion changes to arrive at a global coronary change score.

Reader Bias in Evaluating Coronary Status

We evaluated whether readers were biased in favor of identifying progression over regression. To this end, we calculated individual reader sensitivities in identifying each outcome using the global coronary change score as the best estimate of outcome (Table 3). Readers were as sensitive in identifying progression as they were in identifying regression (83% versus 78%, $\chi^2=0.27, p=NS$).

In addition, we evaluated whether readers were more sensitive in detecting outcome if the films were presented in the actual order of examination, namely pre/post, compared with the reverse post/pre, as set by the randomization used to mask film sequence. As shown in Table 3, the sensitivity rates between pre/post and post/pre pairs of films were similar (78% for pre/post pairs, 87% for post/pre pairs, $\chi^2=2.8, p=NS$).

Interpanel Variability

We evaluated the degree of interpanel variability in identifying lesions and lesion change and in assessing lesion severity and the amount of lesion change. To this end, we calculated the mean, SD, and coefficient of variation ($CV=100 \, SD/mean$) of interreader agreement across the 28 panels (Table 4). Interpanel variability in identifying lesion change (53%) was fivefold greater than that in identifying the original lesion (10%). Once the lesion (or lesion change) was agreed on, there was relatively small interpanel variability in assessing the severity of the lesion (13%) or the amount of lesion change (20%).

Reader Dominance

We evaluated reader effectiveness in exerting control in the consensus opinion. We compared the individual summary score with the global coronary change score: if the two scores were the same, both panelists scored a “win”; if one reader’s score was used as the consensus, he scored a win while the other reader scored a “loss”; in the case where neither score was used as the consensus, each panelist scored a “compromise.” Data for six of the 11 panelists who read 10 or more angiograms is summarized in Table 5.

The percentage of wins ranged from 15% to 48%, with a mean of 34.0%. The win ratio, defined as the ratio of the number of wins to the number of losses and compromises, ranged from 0.18 to 0.92.

Discussion

Limitations of Coronary Imaging Technology

In choosing end-point measurements for clinical trials testing treatment effects at the level of the arterial wall, it is important to realize the limitations of current technology. Angiographic trials test for therapy effects on the basis of change in coronary anatomy without knowledge (for technical and ethical reasons) of the true anatomy of coronary atherosclerosis. In fact, no current coronary imaging method has been shown to reveal true coronary anatomy by adequate calibration against arterial specimens, but selective angiography most closely approaches this ideal. Selective arteriography has been calibrated against peripheral arterial specimens in intact cadavers under conditions that realistically simulate clinical radiography, but such studies cannot reproduce cardiac motion to test coronary angiography. Except for the study of Weiner et al (discussed in the next section), quantitative coronary angiographic procedures have been calibrated by procedures employing either radiographic test objects, measurements of coronary flow past single constrictions, or arteries, or measurements of total myocardial flow.

Table 4. Interpanel Variability (No. of Panels = 28)

<table>
<thead>
<tr>
<th>Determination</th>
<th>Mean±SD (%)</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of lesion</td>
<td>79.8±8.0</td>
<td>10</td>
</tr>
<tr>
<td>Severity (∆10%)</td>
<td>73.4±10.0</td>
<td>13</td>
</tr>
<tr>
<td>Lesion change</td>
<td>44.4±23.3</td>
<td>53</td>
</tr>
<tr>
<td>Amount of change (∆10%)</td>
<td>83.6±17.0</td>
<td>20</td>
</tr>
</tbody>
</table>

*Mean±SD=percent joint agreement across all panels.
†CV, coefficient of variation=100 SD/mean.
Human Evaluation of Coronary Atherosclerosis

They judge lesion changes directly or when they identify lesions for subsequent computerized analysis. In the National Heart, Lung, and Blood Institute (NHLBI) Type II Intervention Trial and CLAS, lesions were identified by human panel readers who also estimated lesion severity and change in stenosis to determine trial outcome. In other angiographic trials, computerized image processing has been added with the goal of reducing the size and duration of the trial by detecting smaller therapeutic benefits through increased precision of lesion measurement. In these studies, humans have either located and traced the lesions for computer processing or have selected segments with lesions to be measured by computer. However, because lesions within a coronary tree show divergent responses under therapy, gains from increased lesion measurement precision are offset by potential sampling biases if all lesions in each patient are not measured. Since current computer procedures are applicable only to unbranched segments of the coronary tree and will not operate in small branches, human oversight is required to designate which lesions or segments should be measured.

Panel Performance in the Cholesterol Lowering Atherosclerosis Study

Early clinical trials employed film reading by humans under routine or slightly modified clinical conditions. CLAS reading procedures embodied features suggested by results obtained in the NHLBI Type II trial. As in the NHLBI trial, CLAS film readers were angiographers drawn nationwide from universities or large medical care facilities. In addition, CLAS introduced the following features: 1) readings were conducted in a situation that minimized distracting influences and minimized reader fatigue; 2) uniformly high-quality films were used, in which coronary views were specifically matched and uniformly high-quality films were used, in which coronary views were specifically matched; 3) record-keeping procedures were introduced, which included prepared sketches individually characterizing the anatomy of each patient and resultant photocopies of these sketches tracking each stage of the procedure. Finally, the evaluation of human panel–reading performance in CLAS films was conducted over a longer time and included more films than any previous study.

Ideally, panel performance would be assessed by measuring the reliability of panelist evaluations of the presence and size of lesions and lesion changes in vivo images at sites of focal narrowing but also found larger lumens where vessels were normal. This latter finding was attributed to shrinkage during histology processing, despite perfusion fixation at physiological pressures. Weiner et al also found that diffuse wall thickening without focal stenosis was not well demonstrated by angiography.

Calibration Study in Coronary Atherosclerosis

Weiner et al produced coronary atherosclerosis in pigs (where coronary anatomy closely mimics that of humans), obtained coronary angiograms, opened the chest to repeat the angiograms, and measured coronary dimensions in vivo with vernier calipers. Coronary arteries were fixed with pressure perfusion and reangiogramed, and paraffin-sectioned histological specimens were projected, traced, and measured with digital image analysis. Weiner et al found excellent correlation between histological specimens and in vivo images at sites of focal narrowing but also found larger lumens where vessels were normal. This latter finding was attributed to shrinkage during histology processing, despite perfusion fixation at physiological pressures. Weiner et al also found that diffuse wall thickening without focal stenosis was not well demonstrated by angiography.

The findings of Weiner et al that focal stenoses are accurately represented has been corroborated in some but not all comparisons of clinical coronary angiograms performed shortly before death. There is general agreement that selective angiography detects focal areas where plaques intrude into the vessel lumen but underestimates diffuse wall thickening. However, intrusive lesions are an appropriate (and possibly the prime) end-point measurement because 1) enumeration of focal lesions in coronary angiograms significantly predicts long-term prognosis and 2) the extent of raised coronary lesions is the single most significant anatomic finding predictive of the presence of ischemic heart disease at autopsy.

The results reported here for human panel readers, taken with our published estimates of the reproducibility of image-processed lesion measurements in the same films, lead to the conclusion that the capabilities of selective angiography are optimized when films are first examined by human panel readers and next with image processing.

Human Evaluation of Coronary Atherosclerosis

Film evaluation by humans plays a critical role in the outcome of coronary angiographic trials when they judge lesion changes directly or when they identify lesions for subsequent computerized analysis. In the National Heart, Lung, and Blood Institute (NHLBI) Type II Intervention Trial and CLAS, lesions were identified by human panel readers who also estimated lesion severity and change in stenosis to determine trial outcome. In other angiographic trials, computerized image processing has been added with the goal of reducing the size and duration of the trial by detecting smaller therapeutic benefits through increased precision of lesion measurement. In these studies, humans have either located and traced the lesions for computer processing or have selected segments with lesions to be measured by computer. However, because lesions within a coronary tree show divergent responses under therapy, gains from increased lesion measurement precision are offset by potential sampling biases if all lesions in each patient are not measured. Since current computer procedures are applicable only to unbranched segments of the coronary tree and will not operate in small branches, human oversight is required to designate which lesions or segments should be measured.

Panel Performance in the Cholesterol Lowering Atherosclerosis Study

Early clinical trials employed film reading by humans under routine or slightly modified clinical conditions. CLAS reading procedures embodied features suggested by results obtained in the NHLBI Type II trial. As in the NHLBI trial, CLAS film readers were angiographers drawn nationwide from universities or large medical care facilities. In addition, CLAS introduced the following features: 1) readings were conducted in a situation that minimized distracting influences and minimized reader fatigue; 2) uniformly high-quality films were used, in which coronary views were specifically matched and recorded in identical order; and 3) record-keeping procedures were introduced, which included prepared sketches individually characterizing the anatomy of each patient and resultant photocopies of these sketches tracking each stage of the procedure. Finally, the evaluation of human panel–reading performance in CLAS films was conducted over a longer time and included more films than any previous study.

Ideally, panel performance would be assessed by measuring the reliability of panelist evaluations of the presence and size of lesions and lesion changes.

<table>
<thead>
<tr>
<th>Reader</th>
<th>No. panels</th>
<th>Wins</th>
<th>Losses</th>
<th>Compromises</th>
<th>Win ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>20</td>
<td>40</td>
<td>40</td>
<td>0.25</td>
</tr>
<tr>
<td>B</td>
<td>60</td>
<td>48</td>
<td>37</td>
<td>15</td>
<td>0.92</td>
</tr>
<tr>
<td>C</td>
<td>35</td>
<td>34</td>
<td>43</td>
<td>23</td>
<td>0.52</td>
</tr>
<tr>
<td>D</td>
<td>25</td>
<td>48</td>
<td>28</td>
<td>24</td>
<td>0.92</td>
</tr>
<tr>
<td>E</td>
<td>13</td>
<td>15</td>
<td>39</td>
<td>46</td>
<td>0.18</td>
</tr>
<tr>
<td>F</td>
<td>37</td>
<td>38</td>
<td>46</td>
<td>16</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Win, reader's score used as consensus score; loss, other reader's score used as consensus score; compromise, neither reader's score used as consensus score; win ratio, wins/(losses + compromises).
and of overall coronary status. This study design did not permit formal reliability analyses of panelist evaluations of the presence and size of lesions. To do so, one would need to know the frequency of both panelists independently seeing no lesions. However, it was possible to formally assess the reliability of identification of lesion changes, amounts of changes, and overall coronary status; consequently, $\kappa$ statistics are reported for these factors.

Overall, panel performance in CLAS was superior to that reported in the literature,\textsuperscript{11-13} possibly due to conditions under which films were read and/or to improved film quality. Specifically, CLAS results indicate good interreader agreement in identifying the presence of lesions (82%, Table 1) that did not vary greatly from one panel to the next (CV=10%, Table 4). Interreader agreement in lesion identification increased with increasing percent stenosis (Figure 4, upper left panel).

Agreement on severity (percent stenosis) was within 10% in 76% of instances when panelists jointly identified a lesion (Table 1) and did not vary from panel to panel (CV=13%, Table 4). Agreement was best for very small (20%) or very large (>90%) stenoses (Figure 4, upper right panel).

On the other hand, panelists did not perform as well in the identification of lesion change. On average, only 51% of changed lesions were independently identified (Table 1), with large interpanel variability (CV=53%, Table 4). Although identification of the presence of lesion change did not depend on the initial size of the lesion (Figure 4, lower left panel), it did depend on the size of the change. In fact, agreement on the presence of lesion change was quite good (>77%) for changes in the range greater than 20%. It is in the situation of small changes, where the interreader agreement was as small as 17%, that the consensus approach to angiographic interpretations compensates for the limitation of the human eye in independently assessing lesion change.

For those lesion changes that were independently and jointly agreed on, there was good agreement in evaluating the amount of change (81% within 10%; Table 1), with small variability across panels (CV=20%; Table 4). Again, interreader agreement on the amount of change increased with increasing size of the change (Figure 4, lower right panel).

There was excellent agreement in independently assessing overall coronary status within one step in a four-step scale both for readers (96%; Table 2) and for panels (96%; see text). No bias was observed in a reader's favoring progression over regression nor in assuming that films are presented in the usual pre/post order. Panel reading provided a reasonable balance of current opinion on what constitutes overall progression or regression in coronary status. However, there was evidence that certain panel readers dominated the designation of the global coronary change score. To reduce potential bias, it is essential in angiographic trial design that readers be randomly assigned to panels and that all readers grade equal numbers of film pairs.

Human readers are a primary means to obtain comprehensive lesion counts before and after therapy, but as indicated by this study, carefully controlled two-reader panels are required. Panel readings offer the advantage of rapid, comprehensive, clinically relevant information that can be used for online trial management and detection of large therapeutic effects. Also, panel evaluation can reliably provide a single score for each patient, which can accommodate a heterogeneous response of lesions within the same patient. The interreader consistency of the CLAS global coronary change score suggests the merit of using human judgment to bring lesion change, however measured, into a single score for each patient.

Computerized Evaluation of Coronary Atherosclerosis

As a corollary, our results suggest that computer-reading procedures can be expected to underestimate lesion counts unless applied to search areas designated by two-reader panels. Risk of lesion underestimation by computer reading will be particularly high in the proximity of patent grafts and major occlusive lesions. In addition, since detection of lesion change by humans is done less well than the identification of lesions, this may be a prime area in which image processing can augment human performance. In small trials and trials where expected therapeutic effects are small, the extra precision of computerized measurement may be critical, but in these trials it is particularly important to assure that all lesions are identified and measured to avoid sampling bias. We are currently evaluating the computerized coronary end points for CLAS patients, both in relation to the panel determinations and also in relation to treatment effect on computerized measures of stenosis, length of involvement, and other coronary artery dimensions.

Combining Lesion Changes

An analysis of how human panelists arrive at a global coronary change score, under reading conditions similar to those in CLAS, has been published by Long and coworkers.\textsuperscript{9} When elements of human decisions are better known, this information may be useful in bringing computer measures of lesion change into a single score. At present, the accepted means is to evaluate consistency of change across lesions.\textsuperscript{5,15} In larger trials and in those exhibiting strong and consistent therapeutic effects visible to humans,\textsuperscript{14} it will be of interest to learn whether additional information about the mechanisms of progression and regression can be extracted by computer. For example, in the NHLBI Type II Trial, when manual tracings of panel-read films (which did not show a significant therapeutic effect) were evaluated by computer, it was found that progressing coronary lesions grew more in length than in degree of luminal narrowing.
Acknowledgments

The authors acknowledge the CLAS Clinical and Biostatistical Staff and the Coronary Film Panelist physicians: George G. Rowe, Peter R. Mahler, Ivan L. Bunnell, William J. French, C. Richard Conte, J. Michael Criley, Harold T. Dodge, David G. Greene, W. David Johnston, K. Ramaswamy, Douglas K. Stewart, Laurence DeBoer, and Bonnie H. Weiner, and C.J. Darnall who organized the panel readings.

References


Key Words • coronary angiography • clinical trials • reliability • angiographic interpretation
S P Azen, L Cashin-Hemphill, J Pogoda, W J Mack, M E Sanmarco, E Wickham and D H Blankenhorn

doi: 10.1161/01.ATV.11.2.385

Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1991 American Heart Association, Inc. All rights reserved.
Print ISSN: 1079-5642. Online ISSN: 1524-4636

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://atvb.ahajournals.org/content/11/2/385

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Arteriosclerosis, Thrombosis, and Vascular Biology can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Arteriosclerosis, Thrombosis, and Vascular Biology is online at:
http://atvb.ahajournals.org//subscriptions/