European Heart Journal (1994) **15,** 815–822

The 'Ermenonville' classification of observations at coronary angioscopy — evaluation of intra- and inter-observer agreement

P. den Heijer*, D. P. Foley†, H. L. Hillege*, J. M. Lablanche‡, R. B. van Dijk*, D. Franzen§, M. C. Morice‖, A. Serra¶, I. K. de Scheerder**, P. W. Serruys† and K. I. Lie* On behalf of the European Working Group on Coronary Angioscopy

*Thoraxcenter, University Hospital Groningen, The Netherlands; †Thoraxcenter, Erasmus University, Rotterdam, The Netherlands; ‡Hopital Cardiologique, Lille, France; §Universitäts Klinikum, Köln, Germany; ‖C.N.N. Paris, Centre de Cardiologie du Nord, Saint-Denis, France; ¶Hospital Clinic I Provincial de Barcelona, Barcelona, Spain; **U.Z. Gasthuisberg, Leuven, Belgium*

KEY WORDS: Angioscopy, coronary artery disease, coronary angioplasty, PTCA, intra-observer agreement, inter-observer agreement.

A European coronary angioscopy working group has been established to create and evaluate a classification system for angioscopic observation. The 'Ermenonville' classification features items, graded in 3–5 categories, such as lumen diameter, shape of narrowing, colours of surface, atheroma, dissection, thrombus, etc. Inter- and intra-observer agreement on the interpretation of angioscopic images, using this classification system, was studied within the working group. Kappa values for chance-corrected intra-observer agreement of the diagnostic items were 0·51–0·67. The mean kappa values for inter-observer agreement were very low at 0·13–0·29. The important items, such as red thrombus and dissection were studied after recoding as either present or absent. These items proved to have a good intra-observer agreement, and an acceptable inter-observer agreement after recoding. Other angioscopic diagnoses should be made with caution. Multicentre angioscopy studies should make use of an angioscopy core laboratory. A set of definitions for coronary angioscopy is proposed, and this working group will re-evaluate observer agreements using these definitions.

Introduction

Coronary angioscopy is rapidly becoming established as an important diagnostic and scientific tool in interventional cardiology[1–6]. Since the introduction of a new, 4·5 French flexible coronary angioscope (Baxter, Edwards LIS Div., Irvine, CA, U.S.A.) in 1991, more than 800 coronary angioscopy procedures have been preformed in 44 European centres. Angioscopy is especially suitable for detecting thrombus[7–9], and intimal tears and flaps[9,10]. However, a common classification system to describe angioscopic observations is still lacking. Reliability of a classification system has important implications for the validity of conclusions that are drawn from using it. Reliability estimates are made on the basis of intra-observer agreement and/or inter-observer agreement. Although it is generally accepted that angioscopy provides more vascular surface detail than can be detected with angiography, it remains to be demonstrated whether different observers report similar visual findings from angioscopic images and further, whether the observations are similarly interpreted.

An angioscopy working group, consisting of European interventional cardiologists with ample and continually expanding experience in angioscopy, was established to address these questions. This report describes the first steps of this working group to create a classification system for angioscopic observations, and to evaluate intra- and inter-observer agreements using this classification system.

Methods

Through a process of discussion while reviewing high quality videotapes of angioscopic recordings, the working group has established a consensus on a first, preliminary classification system for angioscopic observations. Table 1 shows the classification system, which underwent various modifications during these discussions, to reach its present form. The classification system contains two parts. The first part consists purely of descriptive items without any clinical interpretation, such as image quality, estimated lumen diameter, and a description of the vessel surface and its colours. In the second part, diagnostic items such as atheroma, dissection, and thrombus are given as pathophysiological interpretations of the angioscopic observations. Each item in this classification can be given a score ranging from 0 to 2–5.

Two working group meetings were organized for the main purpose of evaluating intra- and inter-observer agreement on angioscopic observations. The first meeting took place in October 1992 in Ermenonville, France,

Submitted for publication on 27 September 1993, and in revised form 20 January 1994.

Correspondence: Peter den Heijer, MD, Department of Cardiology, Thoraxcenter, University Hospital, Oostersingel 59, 9713 EZ Groningen, The Netherlands.

0195-668X/94/060815+08 $08.00/0

Table 1 Coronary angioscopy classification system

Descriptive items	Categories
Image quality	1. Adequate
	2. Inadequate
Obtained image of target	1. No recognizable image
	2. Only vessel wall, lumen not visible
	3. Vessel wall, lumen incomplete
	4. Vessel wall, complete lumen
	5. Central lumen visualization
Lumen diameter	0. Not assessable, not applicable
	1. No or minimal narrowing
	2. Narrowing
	3. Total occlusion
Shape of narrowing	0. Not assessable, not applicable
	1. Round or elliptical
	2. Slit-like
	3. Complex shaped
Vessel surface description	0. Not assessable, not applicable
	1. Smooth
	2. Non-mobile protruding irregularities
	3. Mobile protruding irregularities
	4. Both mobile & non-mobile protr. irr.
Colours of surface	1. Homogeneous colour
	2. Patchy colours (mixed, multiple)
White	1. Present
Gray	1. Present
Yellow	1. Present
Brown	1. Present
Red	1. Present
Pink	1. Present
Diagnostic items	**Categories**
Atheroma	0. Not assessable, not applicable
	1. None
	2. Lining atheroma
	3. Protruding atheroma
Dissection	0. Not assessable, not applicable
	1. None
	2. Small surface disruptions (flaps)
	3. Large dissection
	4. Large mobile flap(s)
Red thrombus	0. Not assessable, not applicable
	1. None
	2. 1 lining thrombus
	3. Multiple lining thrombi
	4. Protruding thrombus
White thrombus	0 Not assessable, not applicable
	1. None
	2. 1 lining thrombus
	3. Multiple lining thrombi
	4. Protruding thrombus
Mixed red/white (pink) thrombus	0. Not assessable, not applicable
	1. None
	2.–1 lining thrombus
	3. Multiple lining thrombi
	4. Protruding thrombus

and the second meeting was held in January 1993 in Amsterdam. The working group decided to name its angioscopy classification system after Ermenonville, the place of their first meeting. During both meetings, discussions preceded the observer agreement scoring sessions. The discussions served to define angioscopic features of diseased coronary arteries before and after intervention, to reach consensus on these features. An additional goal was to assess the effect of group discussion and consequent modification of the Ermenonville classification system on the inter-observer variability.

For each of the two scoring sessions, a special videotape was made, which contained 30 recordings, each 0·5 to 1·5 min long, of angioscopic examinations. The angioscopic images were all recorded at the Department of Cardiology, University Hospital Groningen, The Netherlands. They were selected from images obtained from 62 patients (44 male, 18 female, mean age 60·4 years, range 40–76 years), who all underwent coronary angioscopy pre- and/or post-PTCA. The edited videotapes contained only angioscopic recordings with no angiographic images or additional clinical or procedural data. Each recording was repeated once in a pre-assigned random order to calculate intra-observer agreement. During each of the two sessions, the same six physicians, who were all participants of the angioscopy working group, individually reviewed the videotapes, and completed a single classification form for each recording (Table 1).

STATISTICS

The inter- and intra-observer variation were assessed as percentages of total agreement and with unweighted Cohen's Kappa coefficients[11]. Kappa is a parameter of the agreement between two or more observations in excess of the chance agreement. As this study deals with more than two observers, an unweighted group Kappa[12] was utilized as well, offering a measure of the average agreement between all pairs of observers in excess of the average chance agreement. An overview of the application of Kappa statistics to this study is given in the appendix. Additionally, the parameters, lumen diameter, atheroma, dissection, red thrombus, white thrombus, and mixed thrombus were re-categorized and dichotomized. The Kappa coefficients and percentages of positive and negative agreements of these parameters were re-computed after this operation.

Results

To assess the intra-observer agreement for each parameter 360 observations in each session (60 paired recordings and six observers) were available. Table 2 lists the intra-observer agreement on all angioscopic observation items for both scoring sessions in Ermenonville and Amsterdam. These agreements are presented first as uncorrected total agreement percentages, and secondly as chance corrected Kappa values. The lowest value of uncorrected intra-observer agreement was 51·4% for 'obtained image' in the Ermenonville scoring session, and the highest value was 95·0 for 'brown surface colour' in the Amsterdam session. The chance corrected Kappa values range from 0·29 for gray surface colour in the Ermenonville scoring session to 0·90 for

Table 2 Intra-observer total replication reliability and Kappa values

		Ermenonville	Amsterdam
Image quality	Agreement	78·1%	83·9%
	Kappa	0·57	0·60
Obtained image	Agreement	51·4%	61·7%
	Kappa	0·30	0·38
Lumen diameter	Agreement	66·5%	77·7%
	Kappa	0·50	0·58
Shape of narrowing	Agreement	80·1%	73·1%
	Kappa	0·62	0·58
Vessel surface description	Agreement	64·4%	72·2%
	Kappa	0·50	0·59
Colours of surface (homogeneous vs. patchy)	Agreement	80·0%	87·8%
	Kappa	0·60	0·60
Colours of surface, white	Agreement	75·0%	73·3%
	Kappa	0·48	0·43
Colours of surface, gray	Agreement	69·4%	72·8%
	Kappa	0·29	0·50
Colours of surface, yellow	Agreement	78·9%	90·0%
	Kappa	0·51	0·70
Colours of surface, brown	Agreement	94·4%	95·0%
	Kappa	0·82	0·90
Colours of surface, pink	Agreement	—	81·1%
	Kappa	—	0·71
Colours of surface, red	Agreement	84·4%	76·7%
	Kappa	0·57	0·65
Atheroma	Agreement	77·3%	77·0%
	Kappa	0·63	0·51
Dissection	Agreement	77·0%	76·3%
	Kappa	0·57	0·67
Red Thrombus	Agreement	82·2%	75·4%
	Kappa	0·57	0·60
White thrombus	Agreement	97·2%	84·2%
	Kappa	0·60	0·57
Mixed thrombus	Agreement	92·8%	83·1%
	Kappa	0·51	0·53

brown surface colour in the Amsterdam scoring session. However, these extreme values are not the most meaningful, since obtained image is a relatively unimportant item, and gray and brown surface colour have low incidences of occurrence. Generally, the intra-observer agreement was satisfactory, with most items scoring a total agreement exceeding 75%, and a Kappa value between 0·40 and 0·70. Among all possible angioscopic diagnoses, red thrombus and dissection are probably the most important items. In the two scoring sessions, we found a 75·4%–82·2% intra-observer agreement, and Kappa values of 0·57–0·60 for red thrombus. The intra-observer agreement on dissection was 76·3%–77·0%, with Kappa values of 0·57–0·67. These Kappa values represent a fair to good agreement beyond chance (see Appendix). In general, diagnostic parameters appear to be more reliable than the descriptive parameters. Similarly, comparison between the observers revealed the same difference in favour of the interpretive items (Table 3), although overall the Kappa's were substantially lower than the intra-observer values, ranging from poor to fair agreement beyond chance. For red thrombus we calculated an inter-observer agreement of 49·7%–71·1%, and for dissection of 48·6%–53·1%.

Table 4 shows the results of the Amsterdam scoring session, when a recategorized version of the six most important items of the classification system was used. For these calculations, the items were considered as being either present or absent, whereby cut-off points were applied with the intention of identifying clinically relevant positive findings as 'present'. Thus, for luminal diameter a cut-off point between grades 1 (no or minimal narrowing) and two (narrowing) was used. For dissection, the categories 1–2 (none and small flaps) and 3–4 (large dissection and large mobile flaps) were clustered. We had observed that there was wide variability in the interpretation of small, patchy surface discolourations, with frequent classification as actual thrombus by some observers and no abnormality by others. Therefore, in this Table, red, white and mixed thrombus was counted as present only in case of a grade 4 (protruding thrombus) score. The observed agreement, not corrected for chance, exceeded 85% for all observations on thrombi. However, the Kappa values are all in the category of poor chance-corrected agreement. The observed agreement is separated into P-pos and P-neg. These data indicate that high values of Kappa are associated with P-neg and P-pos values which are of

Table 3 Inter-observer total replication reliability and Kappa values

		Ermenonville	Amsterdam
Image quality	Agreement	61·3%	58·9%
	Kappa	0·28	0·09
Obtained image	Agreement	41·1%	41·8%
	Kappa	0·18	0·10
Lumen diameter	Agreement	43·8%	58·0%
	Kappa	0·19	0·24
Shape of narrowing	Agreement	64·6%	53·5%
	Kappa	0·32	0·27
Vessel surface description	Agreement	42·4%	52·1%
	Kappa	0·21	0·30
Colours of surface (homogeneous vs. patchy)	Agreement	62·3%	77·4%
	Kappa	0·23	0·22
Colours of surface, white	Agreement	49·6%	50·2%
	Kappa	− 0·04	0·05
Colours of surface, gray	Agreement	64·0%	43·1%
	Kappa	0·13	− 0·03
Colours of surface, yellow	Agreement	70·0%	75·8%
	Kappa	0·30	0·24
Colours of surface, brown	Agreement	94·9%	38·4%
	Kappa	0·06	− 0·14
Colours of surface, pink	Agreement	—	30·7%
	Kappa	—	− 0·02
Colours of surface, red	Agreement	77·3%	38·4%
	Kappa	0·52	0·07
Atheroma	Agreement	52·4%	59·8%
	Kappa	0·13	0·15
Dissection	Agreement	53·1%	48·6%
	Kappa	0·27	0·30
Red Thrombus	Agreement	71·1%	49·7%
	Kappa	0·29	0·21
White thrombus	Agreement	92·7%	66·4%
	Kappa	0·02	0·05
Mixed thrombus	Agreement	85·1%	69·7%
	Kappa	0·07	0·07

comparable magnitude to each other, and also the value of the observed agreement.

Discussion

It seems likely that coronary angioscopy will play an increasingly important role in interventional cardiology — directed towards either documentation, for the purpose of prospective research, or guidance during the intervention itself[13–15]. We have established a European coronary angioscopy working group in order to develop and evaluate a classification system that can be used to describe systematically angioscopic findings, as well as to encourage, guide and coordinate multicentre clinical trials. The classification system in its present form, as shown in Table 1, is suitable for routine clinical use and selected clinical angioscopy studies. However, it must be emphasized that this is merely a preliminary classification, and that the work of refining and adding other items is still in progress.

As our data show, a detailed interpretation of angioscopic images is subject to an acceptable intra-observer variability, but a large inter- observer variability. Although this finding may seem somewhat disconcerting, in our opinion it does not necessarily reflect negatively on the clinical value of coronary angioscopy, but should rather be regarded as an impetus to develop quantitative angioscopic measurement tools, for instance for luminal dimensions[16], or colorimetry, it is relevant to refer to well-known studies published in the 1970s, which revealed unexpectedly poor agreement in the interpretation of coronary angiograms[17,18]. These observations did not prevent the increasingly widespread acceptance of the clinical value of coronary angiography, but they certainly played an important role in the development of quantitative coronary angiography[19].

Our study demonstrates that the angioscopic diagnosis of thrombus and dissection is associated with the highest intra- and inter-observer agreements. This in itself is an important conclusion, in agreement with previous studies which have shown a high sensitivity of angioscopy, compared with angiography, for intravascular thrombus formation and intimal dissection[4,20,21].

There are several factors which may play a role in the low Kappa values for most of the other angioscopic features. In an effort to ensure objectivity and lack of potential bias in reporting by the observers, the videotaped angioscopic recordings were edited and presented

Table 4 Intra- and inter-observer Kappa values and other indices of agreement of the recoded items of the Amsterdam scoring session

		Intra-observer	Inter-observer
Lumen diameter	n	175	880
	Agreement	95%	91%
	Kappa	0·62	0·26
	P-neg	97%	095%
	P-pos	64%	30%
Atheroma	n	178	880
	Agreement	89%	84%
	Kappa	0·43	0·12
	P-neg	49%	20%
	P-pos	94%	91%
Dissection	n	177	880
	Agreement	87%	73%
	Kappa	0·73	0·43
	P-neg	89%	77%
	P-pos	84%	66%
Red thrombus	n	179	890
	Agreement	91%	81%
	Kappa	0·66	0·34
	P-neg	94%	88%
	P-pos	72%	45%
White thrombus	n	177	880
	Agreement	92%	84%
	Kappa	0·57	0·10
	P-neg	95%	91%
	P-pos	62%	18%
Mixed thrombus	n	177	890
	Agreement	90%	86%
	Kappa	0·53	0·16
	P-neg	95%	92%
	P-pos	59%	24%

n=number of compared observations; Agreement=Total percentage observed agreement; Kappa=kappa value; P-neg=Proportion of negative agreement; P-pos=Proportion of positive agreement.

without any angiographic data and without any knowledge of the patients, their clinical status, and the examined vessel and its 'angioscopic target lesion'. However, during everyday clinical angioscopy, the knowledge and use of such data may have a positive influence on the reliability of angioscopic data.

The absence of cine-angiograms during these sessions made it difficult for the observers to define the 'target' lesion as the substrate for description. Probably the most important factor contributing to the poor agreement both within, and especially between, observers would seem to be the lack of clear and unambiguous definitions of the various angioscopic features of interest. We consider this to be the single most important contributory factor to the poor inter-observer agreement. This is not surprising, as the likelihood of agreement in subjective impressions is clearly dependent on availability of tangible reference points. The National Heart Lung and Blood Institute of the U.S.A., the American Heart Association and American College of Cardiology classifications of angiographic features of dissection and lesion type provided such a reference for investigators in that field and thus an international language of angiography[22–24]. A similar approach is urgently required to coin an international angioscopic language. The working group now considers it its first priority to establish such definitions, and to re-evaluate observer agreements using these definitions. On the basis of the findings of the group, a 'simplified Ermenonville classification', with proposed definitions, is presented in Table 5. Figures 1–4 show typical examples of normal vessel wall, red thrombus, dissection, and atheroma. The drawback of such photographic still-frames is that not all features observed on the real-time, moving angioscopic images, can be adequately reproduced, since they lack the third dimension of depth. Software for a computerized atlas and training program with digital video recordings of angioscopic images has been developed for enhanced illustration of the Ermenonville classification system, and to help investigators use the classification system and definitions[25]. The hypothesis that training and learning effects play an important role in the optimal assessment of angioscopic images is supported by the finding that the intra-observer agreement was higher in all but five of the 17 factors studied at the Amsterdam session, compared to the Ermenonville session. The Ermenonville session was prior to the Amsterdam session, and considerable additional individual angioscopic experience was gained by all observers in the meantime, which apparently had a positive influence on the replication reliability.

Small intraluminal structures and minute discolorations may either be difficult to detect for some observers, or give rise to more differing and inconsistent interpetations. Therefore an additional analysis was performed for the items luminal narrowing, atheroma, dissection, and red, white and mixed thrombus, in which these items could only be scored as being either absent or present. For this purpose, the original gradings were recategorized such that only substantial abnormal findings were considered to give a positive score. Table 4 shows that protruding red thrombi and large dissection can be detected with acceptable intra- and inter-observer agreement, although the inter-observer Kappa values remain low. In this particular case, Kappa values are probably not the best choice for gaining insight into the measure of inter-observer agreement, as the uncorrected inter-observer agreement percentages illustrate: the exceptional distribution of the assessments exerts a major influence on the Kappa values[26,27]. In these cases, a factor which can be in part held responsible for the low Kappa values, the so-called 'Kappa- paradox', can be demonstrated. A typical example is given by the Kappa value for the inter-observer agreement on white thrombus in Table 4. The low Kappa value of 0·10 is in apparent contradiction to the high agreement score of 84%, and can partially be explained by the low prevalence of large white thrombi, causing an uneven distribution of the possible scores. In such cases, other indices like P-pos and P-neg provide better insight into the strength of the inter-observer agreement. P-neg and

Table 5 Coronary angioscopy simplified classification and definitions

Items		Categories and definitions
Normal vessel		Vessel wall appearing uniformly smooth in contour, and of a uniform gray colour. There are no abrupt changes in diameter. The lumen is free of intraluminal structures.
Atheroma	1.	None.
	2.	Lining atheroma. Non-elevated yellow or white discoloration.
	3.	Protruding atheroma. Non-mobile, elevated and/or protruding yellow or white discoloration. There may be focal or diffuse narrowing.
Dissection	0.	Not assessable, not applicable.
	1.	None
	2.	Small surface disruptions. Small, very mobile, structures, which are contiguous with the vessel wall. They do not impede the visualization of the lumen.
	3.	Large surface disruptions. Visible cracks or fissures on the luminal surface and/or large mobile or non-mobile protruding structures, which are contiguous with the vessel wall and of homogeneous appearance with the vessel wall. They impede visualization of (part of) the lumen.
Red thrombus	0.	Not assessable, not applicable.
	1.	None.
	2.	Lining thrombus. Red, predominantly mural, non-mobile, superficial mass, adherent to the vessel surface, but clearly a separate structure.
	3.	Protruding thrombus. Red, intraluminal, protruding, mobile or non-mobile mass, adherent to the vessel surface, but clearly a separate structure.
White thrombus	0.	Not assessable, not applicable.
	1.	None.
	2.	Lining thrombus. White, predominantly mural, non-mobile, superficial mass, adherent to the vessel surface, but clearly a separate structure.
	3.	Protruding thrombus. White, intraluminal, protruding, mobile or non-mobile mass, adherent to the vessel surface, but clearly a separate structure. The appearance is shaggy, irregular, and cotton-wool-like.
Mixed thrombus	0.	Not assessable, not applicable.
	1.	None.
	2.	Lining thrombus. Mixed red and white, predominantly mural, non-mobile, superficial mass, adherent to the vessel surface, but clearly a separate structure.
	3.	Protruding thrombus. Mixed red and white, intraluminal, protruding, mobile or non-mobile mass, adherent to the vessel surface, but clearly a separate structure.
Wall haemorrhage	0.	Not assessable, not applicable.
	1.	None.
	2.	Present. A distinct, demarcated, red, non-elevated discolouration, which is clearly within the vessel wall.

P-pos represent the probability of negative and positive agreements between random observers, and are thus comparable to the well-known concept of sensitivity and specificity. These figures reveal fair to good agreement for red thrombus and the absence of thrombus, and for dissection and the absence of dissection.

Conclusions

In coronary angioscopy, the intra-observer agreement for the presence of a substantial red thrombus or dissection is good. The inter-observer agreement for these diagnoses can be considered to be acceptable. Other angioscopic diagnoses should be made with caution. Descriptive items in a classification system reach lower observer agreements than pathophysiological interpretations. Due to the inherently subjective nature of angioscopy, for clinical purposes, an angioscopic classification system should be as simple as possible, containing 'diagnostic' rather than descriptive items. Multicentre angioscopy studies should make use of a central angioscopy core laboratory to minimize the problem of inter- observer variability. Quantitative angioscopy measurement tools should be developed to assist in clinical research. Finally, the next most urgent task in the development of this powerful diagnostic tool in interventional cardiology is to establish a set of clear and unambiguous definitions. A preliminary definition system is given in this paper, but refinement and additional analysis of angioscopic observer agreement using these definitions will be carried out, and reported, by this coronary angioscopy working group.

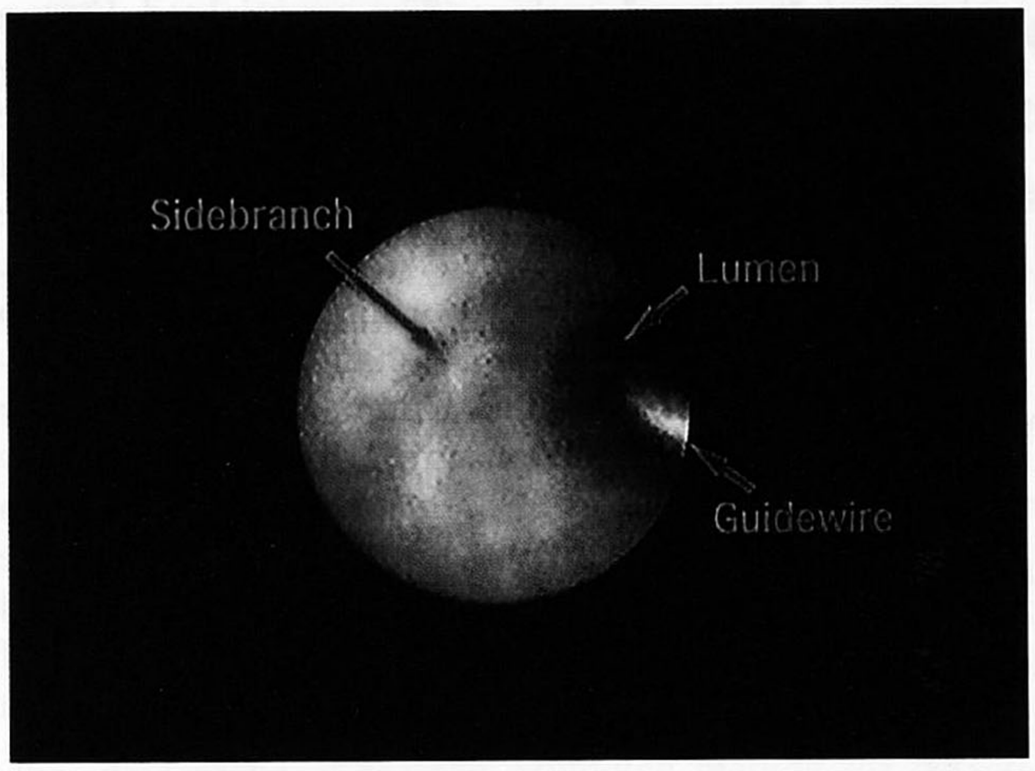

Figure 1 Normal vessel. The colour is uniformly gray, and there are no intraluminal structures, narrowings, or protrusions visible. The lumen is round and does not appear to be narrowed. The orifice of a side branch is seen at 11 o'clock.

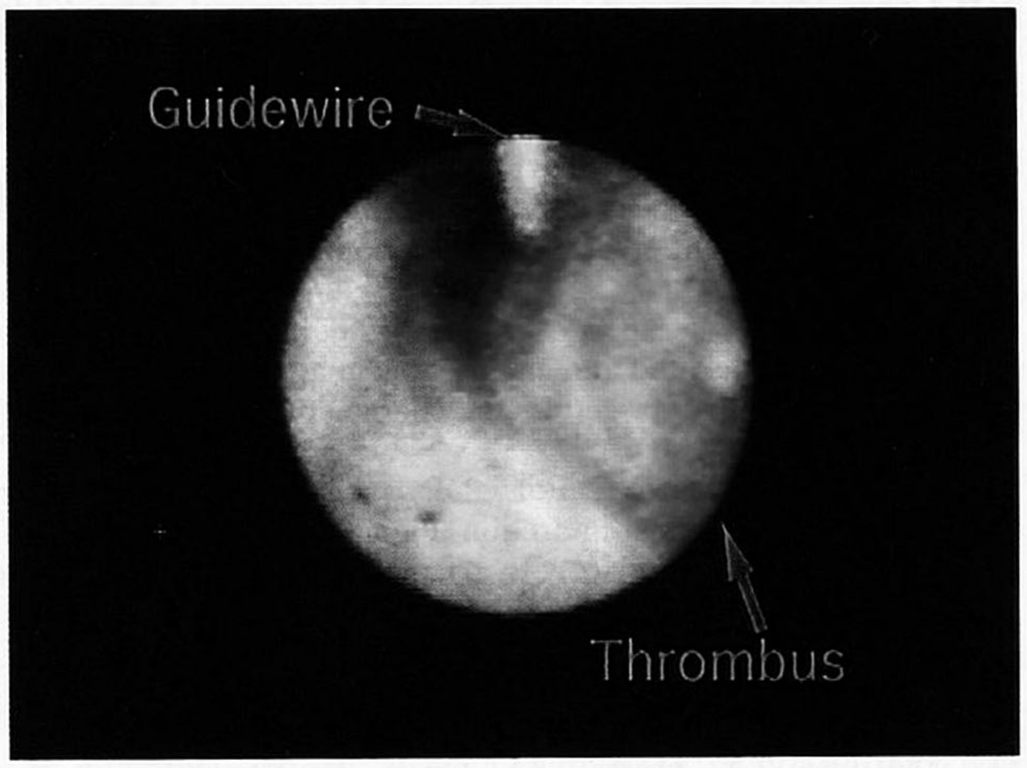

Figure 2 Protruding red thrombus, grade 3 according to the simplified Ermenonville classification (recorded by A. Serra, Barcelona, Spain).

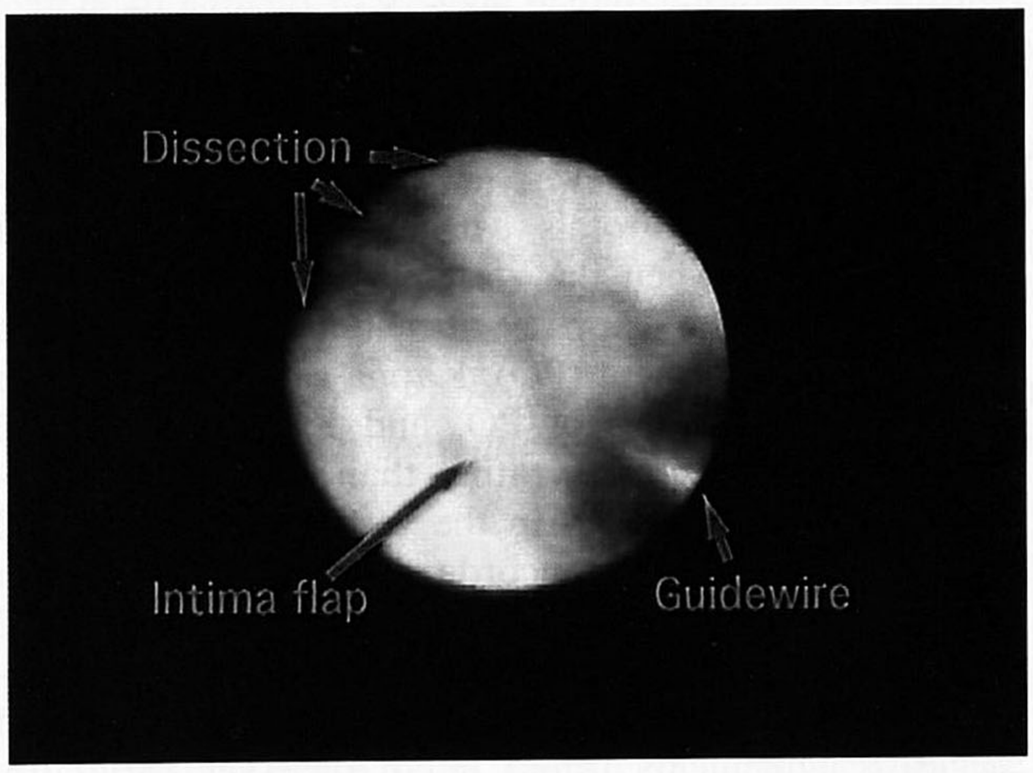

Figure 3 Grade 3 dissection. There is a large fissure at 10 o'clock, together with mobile intima flaps.

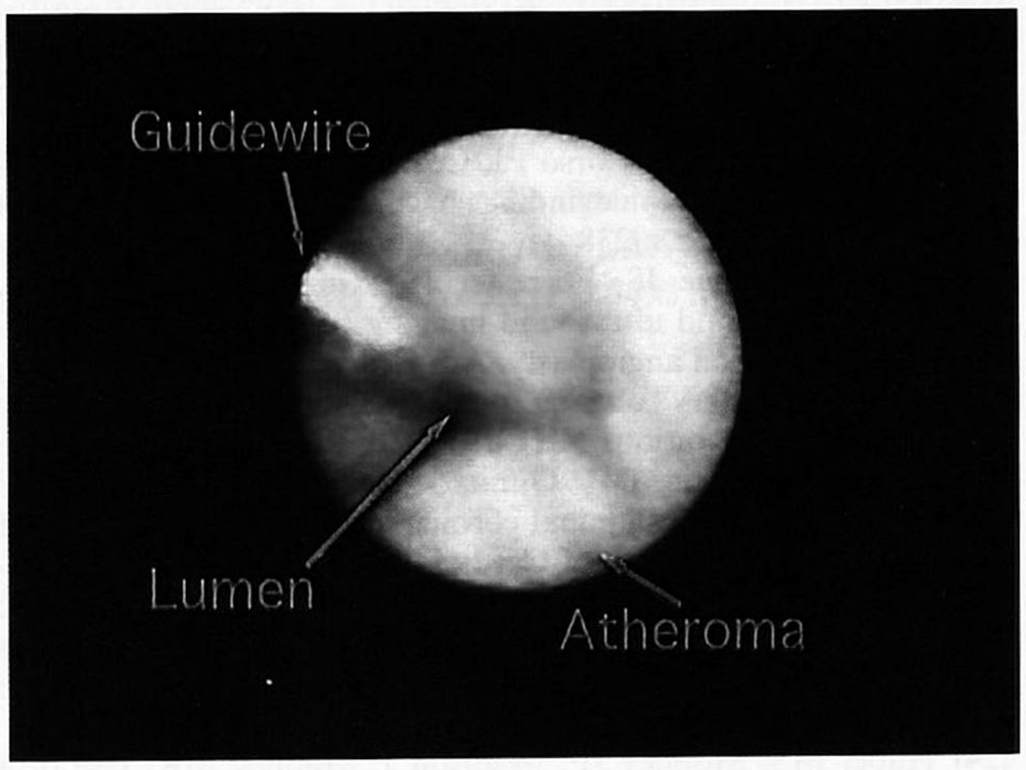

Figure 4 Protruding atheroma, grade 3 according to the simplified Ermenonville classification. The vessel wall is yellow, with a small red discolouration, which may be a sign of a ruptured plaque. The remaining lumen is small due to the severe protrusion of the ahteromata.

Baxter Healthcare Corporation, Interventional Cardiology Division supports the meetings of the European Working Group on Coronary Angioscopy. We gratefully acknowledge the continued cooperation, support and assistance of Astrid de Jager, Mark Price, Peter de Jong, Anna Marie Smith, Jonathan Strauss, David Smith, and Yves Guillemain d'Echon.

References

[1] Spears JR, Marais HJ, Serur J. In vivo coronary angioscopy. J Am Coll Cardiol 1983; 1: 1311–4.

[2] Litvack F, Grundfest WS, Lee ME. Angioscopic visualization of blood vessel interior in animals and humans. Clin Cardiol 1985; 8: 65–70.

[3] Uchida Y, Tomaru T, Nakamura F, Furuse A, Fujimori Y, Hasegawa K. Percutaneous coronary angioscopy in patients with ischemic heart disease. Am Heart J 1987; 114: 1216–22.

[4] Johnson C, Hansen D, Vracko R. Angioscopy — more sensitive for identifying thrombus, distal emboli, and subintimal dissection. J Am Coll Cardiol 1989; 13: 146 (Abstr).

[5] Ramee SR, White CJ, Collins TJ, Mesa JE, Murgo JP. Percutaneous angioscopy during coronary angioplasty using a steerable microangioscope. J Am Coll Cardiol 1991; 17: 100–5.

[6] Franzen D, Hopp HW, Korsten J, Hilger HH. A prospective study on percutaneous coronary angioscopy with different guiding techniques in patients with coronary heart disease. Eur Heart J 1992; 13: 655–60.

[7] Mizuno K, Miyamoto A, Satomura K *et al.* Angioscopic coronary macromorphology in patients with acute coronary disorders. Lancet 1991; 337: 809–12.

[8] Mizuno K, Satomura K, Miyamoto A *et al.* Angioscopic evaluation of coronary artery thrombi in acute coronary syndromes. N Engl J Med 1992; 326: 287–91.

[9] Siegel RJ, Ariani M, Fishbein MC *et al.* Histopathologic validation of angioscopy and intravascular ultrasound. Circulation 1991; 84: 109–17.

[10] Uchida Y, Hasegawa K, Kawamura K, Shibuya I. Angioscopic observation of the coronary luminal changes induced by percutaneous transluminal coronary angioplasty. Am Heart J 1989; 117: 769–76.

[11] Cohen JA. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960; 20: 37–46.
[12] Schouten HJA. Nominal scale agreement among observers. Psychometrika 1986; 51: 453–66.
[13] Knopf WD, Cates CU, Doby B, Langlois K. Coronary angioscopy influences intervention in patients with unstable angina and recent myocardial infarction. Circulation 1992; 86: 1–650 (Abstr).
[14] De Feyter PJ, Escaned J, DiMario C *et al.* Combined intracoronary ultrasound and angioscopic imaging in patients with unstable angina: target lesion characteristics. Eur Heart J 1993; 14: 25 (Abstr).
[15] Escaned J, Di Mario C, Foley DP *et al.* Coronary angioscopy in unstable and post-infarction angina pectoris. Eur Heart J 1993; 14: 476 (Abstr).
[16] Spears JR, Raza SJ, Ali M, Iyer GS, Cheong WF, Crilly RJ. Quantitative angioscopy: a new method for measurement of luminal dimensions by use of a 'lightwire'. JACC 1993; 21: 133A (Abstr).
[17] DeRouen TA, Murray JA, Owen W. Variability in the Analysis of Coronary Arteriograms. Circulation 1977; 55: 324–8.
[18] Zir LM, Miller SW, Dinsmore RE, Gilbert JP, Harthorne JW. Interobserver Variability in Coronary Angiography. Circulation 1976; 53: 627–32.
[19] Reiber JHC, Serruys PW, Kooijman CJ. Assessment of short-, medium-, and long-term variations in arterial dimensions from computer-assisted quantitation of coronary cineangiograms. Circulation 1985; 71: 280–8.
[20] Lee G, Garcia JM, Corso PJ. Correlation of coronary angioscopic to angiographic findings in coronary artery disease. Am J Cardiol 1986; 58: 238–41.
[21] Siegel RJ, Chae JS, Forrester JS, Ruiz CE. Angiography, Angioscopy, and ultrasound imaging before and after percutaneous balloon angioplasty. Am Heart J 1990; 120: 1086–90.
[22] Coronary artery angiographic changes after PTCA. In: Manual of operations NHLBI PTCA Registry 1985: 6–9.
[23] Ryan TJ, Faxon DP, Gunnar RM *et al.* Guidelines for percutaneous transluminal coronary angioplasty. A report of the American College of Cardiology/American Heart Association Task Force on assessment of diagnostic and therapeutic cardiovascular procedures (subcommittee on percutaneous transluminal coronary angioplasty). Circulation 1988; 78: 486–502.
[24] Huber MS, Mooney JF, Madison J, Mooney MR. Use of a morphologic classification to predict clinical outcome after dissection from coronary angioplasty. Am J Cardiol 1991; 68: 467–71.
[25] Den Heijer P, Van Dijk RB, Lie KI. Coronary Angioscopy: digital recording of real-time, moving color images — technique and applications. Eur Heart J 1993; 14: 116 (Abstr).
[26] Feinstein AR, Cicchetti DV. High agreement but low Kappa: I. The problem of two paradoxes. J Clin Epidemiol 1990; 43: 543–9.
[27] Cicchetti DV, Feinstein AR. High agreement but low Kappa: II. The problem of two paradoxes. J Clin Epidemiol 1990; 43: 551–8.
[28] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 671–9.

Appendix

Estimates of reliability in terms of reproducibility are provided by tests of concordance, such as Kappa[11] and correlations. The unweighted Kappa coefficient[28], which measures agreement beyond that expected by chance, was used to evaluate intra- and inter-observer reliability for nominal scaled parameters, such as the colours of the vessel surface. The unweighted Kappa was used because it represents a measure of the frequency of exact agreement rather than a measure of the degree of approximate agreement. The same coefficient was used for angioscopic parameters which could be addressed as ordinal scaled, such as the lumen diameter. The magnitude of weighted Kappa may be arbitrary, so to avoid this arbitrariness, unweighted Kappa coefficients were used. Unweighted Kappa statistics consider all differences of observer agreement to be equally important, and its values to be in the range from -1 to $+1$. The Kappa statistic is $+1$ if there is complete agreement between two observers, but 0 if observed agreement is equal to that expected by chance alone. If observed agreement is less than chance agreement then Kappa has a negative value.

The magnitude of Kappa can be divided in arbitrary categories: Kappa values below 0·40 represent a poor agreement beyond chance, values between 0·40 and 0·59 represent fair agreement, those between 0·60 and 0·74 good agreement, whereas values greater than 0·75 may be taken to represent excellent agreement beyond chance[28].

Kappa offers a single statistical summary for a study of concordance. However, imbalances in the distributions of marginal totals can produce two types of paradoxes[26,27] when the variability of observations is expressed with the Kappa coefficient and the proportion of observed agreement (PO). Kappa can be low despite relatively high values of PO, and Kappa values will sometimes be increased, rather than decreased, by departures from symmetry in the vertical and horizontal marginal total of the concordance table. This problem can be resolved by a recoding of the parameter in a 2×2 concordance table, and by citing separate indices, analogous to sensitivity and specificity, for the positive and negative agreements found in an observer variability study. The proportion of positive agreements in cases of positive diagnosis can be designated as P-pos, and the proportion of negative agreements as P-neg. P-pos and P-neg indicate consistency of the two observers when going in the opposite directions of positive and negative decisions. The distinction will help a reader decide about the persuasiveness of the individual results. If needed, it will also help the investigator to design further work in an effort to decrease the observers' disparities in the positive direction, the negative or both directions. A second contribution of the P-pos and P-neg value is that they can explain and 'eliminate' the Kappa paradoxes. These indices are demonstrated in the recategorized dichotomized parameters lumen diameter, atheroma, dissection, red thrombus, white thrombus, and mixed thrombus.

This study was also supported, in part, by a Biomedical and Health Research (Biomed I) grant, from the Commission of the European Communities, Directorate-General XII for Science, Research and Development.