

Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs— An Extension of the Modified Airlie House Classification

ETDRS Report Number 10

EARLY TREATMENT DIABETIC RETINOPATHY STUDY RESEARCH GROUP*

Abstract: The modified Airlie House classification of diabetic retinopathy has been extended for use in the Early Treatment Diabetic Retinopathy Study (ETDRS). The revised classification provides additional steps in the grading scale for some characteristics, separates other characteristics previously combined, expands the section on macular edema, and adds several characteristics not previously graded. The classification is described and illustrated and its reproducibility between graders is assessed by calculating percentages of agreement and kappa statistics for duplicate gradings of baseline color non-simultaneous stereoscopic fundus photographs. For retinal hemorrhages and/or microaneurysms, hard exudates, new vessels, fibrous proliferations, and macular edema, agreement was substantial (weighted kappa, 0.61 to 0.80). For soft exudates, intraretinal microvascular abnormalities, and venous beading, agreement was moderate (weighted kappa, 0.41 to 0.60). A double grading system, with adjudication of disagreements of two or more steps between duplicate gradings, led to some improvement in reproducibility for most characteristics. *Ophthalmology* 1991; 98:786–806

The modified Airlie House classification of diabetic retinopathy used in the Diabetic Retinopathy Study (DRS)¹ was extended for use in the Early Treatment Diabetic Retinopathy Study (ETDRS). The same seven standard photographic fields were used (Fig 1), and the

original standard color nonsimultaneous stereoscopic photographs were retained. Several new photographs were added and some existing photographs were used to define additional grades for certain characteristics. An additional step was added to the grading scale for the following five characteristics: hard exudates, soft exudates, arteriovenous nicking, retinal elevation, and vitreous hemorrhage (VH). The component characteristics previously pooled as venous abnormalities and arterial abnormalities were separated and graded individually. The single DRS question regarding macular edema was replaced by separate items assessing location, extent, and severity of retinal thickening in the posterior pole (field 2). Items were added to

Originally received: July 19, 1990.
Revision accepted: January 24, 1991.

* A list of the ETDRS Research Group Investigators appears at the end of ETDRS report number 7 in this supplement to *Ophthalmology*.

Reprint requests to the Biometry and Epidemiology Program, National Eye Institute, Bldg 31, Room 6A-24, 9000 Rockville Pike, Bethesda, MD 20892.

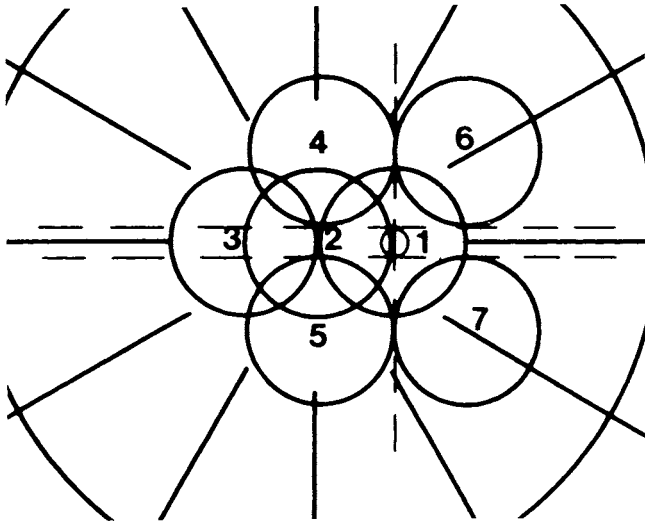


Fig 1. Seven standard fields of the modified Airlie House classification (shown for a right eye). Field 1 is centered on the optic disc, field 2 on the macula. Field 3 is temporal to the macula. Fields 4 to 7 are tangential to horizontal lines passing through the upper and lower poles of the disc and to a vertical line passing through its center.

assess microaneurysms (considered separately from retinal hemorrhages), drusen, papillary swelling, hard exudate rings, macular hole, dragged macula, subretinal hemorrhage, and deep pale spot at the center of the macula. An item allowing the grader to record any other abnormality not specified on the grading form also was added.

This report describes the classification, the methods used to process and summarize the gradings, and the reproducibility of the system. To use the classification for grading fundus photographs, it is necessary to have copies of the stereoscopic color slides, the descriptions that accompany them, and the grading form (available at cost from the Fundus Photograph Reading Center, University of Wisconsin, PO Box 5240, Madison, WI 53705).

METHODS

GRADING PROCEDURES

As in the DRS, nonsimultaneous stereoscopic pairs of the seven standard fields (Fig 1) taken with 30° fundus cameras and mounted in plastic sheets are viewed against light boards using Donaldson stereo viewers (5× magnification). The graders are nonophthalmologists who have been selected for normal depth perception and trained to assess diabetic retinopathy from photographs. Sets of photographs from the two eyes of each patient, and from different visits, are graded independently, and the results are entered on a form.

Each abnormality is graded separately, some of them in a single field and some in each of 5 to 7 fields. Information obtained from overlapping parts of adjoining fields is used whenever it is helpful to determine the nature or severity of an abnormality within the field being graded.

The area of retina included in the photograph being graded is assessed, without making any adjustment for photographs that deviate from the field definitions specified in the protocol. An optional field 8 (or two such fields, 8A and 8B, if necessary) is used during follow-up to document new vessels, fibrous proliferations, plane of proliferations, preretinal hemorrhage (PRH), and/or VH outside the seven standard fields.

A grade of “questionable” is generally recorded for a characteristic when the grader is 50 to 90% certain that it is present. When an abnormality is definitely present but its nature is uncertain, the grader assigns the grade “questionable” for the characteristic considered to be most likely and “absent” for the one(s) considered less likely. For example, when an abnormality is believed to be either new vessels or intraretinal microvascular abnormalities (IRMA), but is more likely the former, the grader assigns the grade of questionable for new vessels and absent for IRMA. Only if the grader favors neither characteristic over the other is a grade of questionable assigned to both. A different definition of “questionable” is used when grading involvement of the center of the macula by retinal thickening, hard exudate, or other abnormality. In this case, a grade of “questionable” is recorded when involvement of the center cannot be confidently recorded as neither definitely present nor definitely absent.

“Cannot grade” is recorded when inadequate photographic quality or obscuration from VH or other abnormality makes it impossible to determine whether the characteristic being graded is present. If one or more disc areas (DA) of retina are visible in a photographic field and that area is free of the characteristic, it is graded as “absent,” rather than “cannot grade.” If the characteristic is present in the unobscured part of the field, its extent is estimated for the field as a whole, as though it were absent in the obscured area.

CHARACTERISTICS GRADED IN MULTIPLE PHOTOGRAPHIC FIELDS

Microaneurysms (fields 1–7). This item was added so that eyes with microaneurysms only could be distinguished from those with hemorrhages and/or microaneurysms. A red spot that is less than $\frac{1}{12}$ the diameter of an average optic disc, or 125 μm in its longest dimension (approximately the width of an average major vein at the disc margin) and that has sharp margins is considered to be a microaneurysm. A red spot of this size with irregular margins and/or uneven density, particularly when it surrounds a smaller central punctate lesion considered to be a microaneurysm, is considered to be a hemorrhage. Any red spot equal to or greater than 125 μm in its longest dimension is considered to be a hemorrhage unless features such as round shape, smooth margins, and a central light reflex suggest that it is probably a microaneurysm. The longstanding clinical convention of considering the diameter of the average optic disc to be 1500 μm is used in this classification, although 1800 to 1900 μm may be a better estimate.^{2,3} Because one or more retinal hemorrhages alone (in the absence of microaneurysms) is less



Fig 2. Standard photograph 1, mildest standard for hemorrhages/microaneurysms and drusen.



Fig 3. Standard photograph 2A, intermediate standard for hemorrhages/microaneurysms.



Fig 4. Standard photograph 2B, severe standard for hemorrhages/microaneurysms.

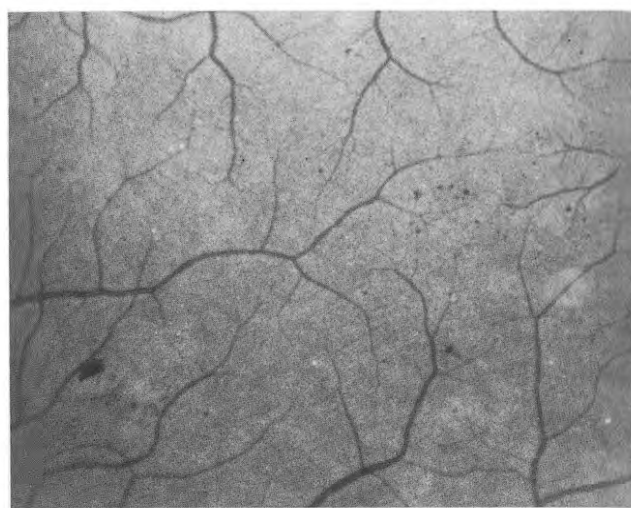


Fig 5. Part of standard photograph 1, enlarged; approximately 20 microaneurysms, one small hemorrhage, and 10 drusen can be seen.

suggestive of diabetic retinopathy than are one or more microaneurysms alone, a step in the scale less severe than that of one microaneurysm is provided for eyes with hemorrhage(s) only. The grading scale is as follows: grade 0 = no hemorrhages or microaneurysms; grade 1 = questionable hemorrhages and/or microaneurysms; grade 2 = hemorrhages only (one or more); grade 3 = one microaneurysm (only); grade 4 = two or more microaneurysms (only); grade 5 = one or more hemorrhages and one or more microaneurysms; grade 8 = cannot grade.

Hemorrhages and/or microaneurysms (fields 2-7). All punctate, blot, or linear hemorrhages and all microaneurysms (whether the color of venous blood or paler, and with or without a recognizable wall) are included. Microaneurysms that appear as white dots with no blood visible in a central lumen are graded as hard exudates. Hemorrhages that are clearly preretinal are excluded. The grading

is based on the area of retina covered by hemorrhages and/or microaneurysms, using standard photographs 1, 2A, and 2B (Figs 2-5) for comparison. The grading scale is as follows: grade 0 = no hemorrhages and/or microaneurysms; grade 1 = questionable hemorrhages and/or microaneurysms; grade 2 = definite hemorrhages and/or microaneurysms less than standard photograph 1; grade 3 = hemorrhages and/or microaneurysms greater than or equal to standard photograph 1 but less than standard photograph 2A; grade 4 = hemorrhages and/or microaneurysms greater than or equal to standard photograph 2A but less than standard photograph 2B; grade 5 = hemorrhages and/or microaneurysms greater than or equal to standard photograph 2B; grade 8 = cannot grade.

Drusen (fields 2-7). Drusen (deposits at or anterior to Bruch's membrane associated with thinning or hypopigmentation of the overlying retinal pigment epithelium)

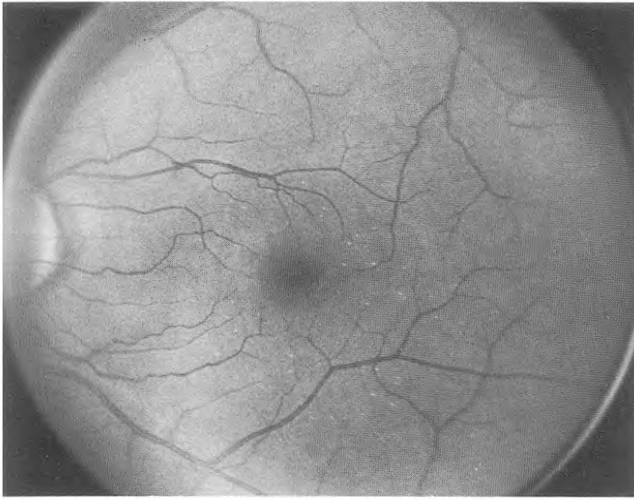


Fig 6. Standard photograph 20, intermediate standard for drusen.

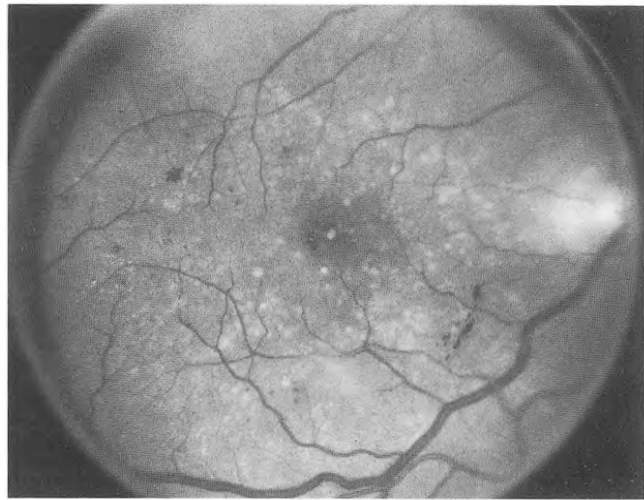


Fig 7. Standard photograph 21, severe standard for drusen.

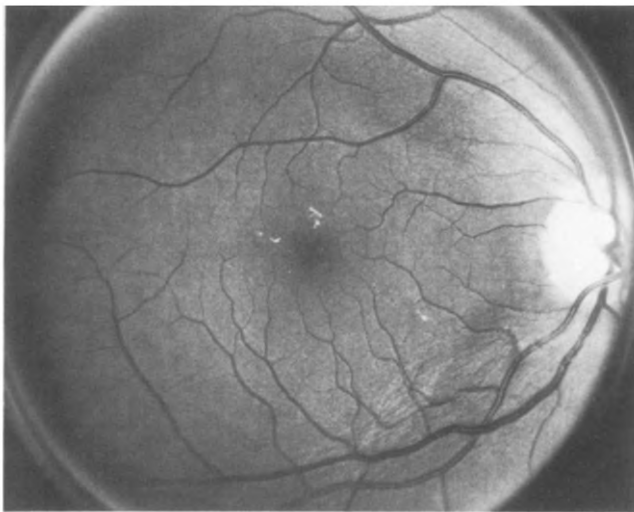


Fig 8. Standard photograph 3, mildest standard for hard exudates.

appear as deep, yellowish-white dots, sometimes circumscribed by a thin line of pigment. Whereas large drusen are easily recognized, small drusen may be confused with small punctate hard exudates. In distinguishing between these alternatives, good stereoscopic effect is important, since drusen appear very deep, at the level of the retinal pigment epithelium, while hard exudates usually appear slightly more superficial, within the outer or middle layers of the retina (or sometimes quite superficial if the retina is thickened).

Several other features are used in distinguishing between drusen and hard exudates: (a) drusen are usually scattered diffusely or clustered near the center of the macula, while hard exudates are usually located near prominent microaneurysms or at the edges of zones of retinal edema; (b) drusen are usually round while hard exudates are often irregular in shape; (c) drusen usually have a dull appearance while hard exudates often appear more shiny or

waxy; (d) drusen often have a faint border of pigment while hard exudates do not.

The grading is based on the area covered by drusen, using standard photographs 1, 20, and 21 (Figs 2, 5, 6, 7) for comparison, as follows: grade 0 = no drusen; grade 1 = drusen questionable or less than standard photograph 1; grade 2 = drusen greater than or equal to standard photograph 1 but less than standard photograph 20; grade 3 = drusen greater than or equal to standard photograph 20 but less than standard photograph 21; grade 4 = drusen greater than or equal to standard photograph 21; grade 8 = cannot grade.

Hard exudates (fields 2-7). Hard exudates are small white or yellowish-white deposits with sharp margins. Often they have a slightly waxy or glistening appearance. They are usually located in the outer layers of the retina, but may be more superficial, particularly when retinal thickening (edema) is present. Hard exudates may be arranged as individual dots, as confluent patches, or in partial or complete rings surrounding prominent microaneurysms or zones of retinal edema. At areas of localized retinal elevation, confluent exudates may be seen beneath or in the outer layers of the retina. The grading of hard exudates is based on the area of retina involved, using standard photographs 3, 4, and 5 (Figs 8-10) for comparison, as follows: grade 0 = no hard exudate; grade 1 = questionable hard exudate; grade 2 = definite hard exudate, less than standard photograph 3; grade 3 = hard exudate greater than or equal to standard photograph 3 but less than standard photograph 5; grade 4 = hard exudate greater than or equal to standard photograph 5 but less than standard photograph 4; grade 5 = hard exudate greater than or equal to standard photograph 4; grade 8 = cannot grade.

Exudates that lie immediately adjacent to retinal veins, simulating venous sheathing, are designated "perivenous exudates" and excluded from consideration in the hard exudate category. They are graded separately in the section dealing with venous abnormalities.

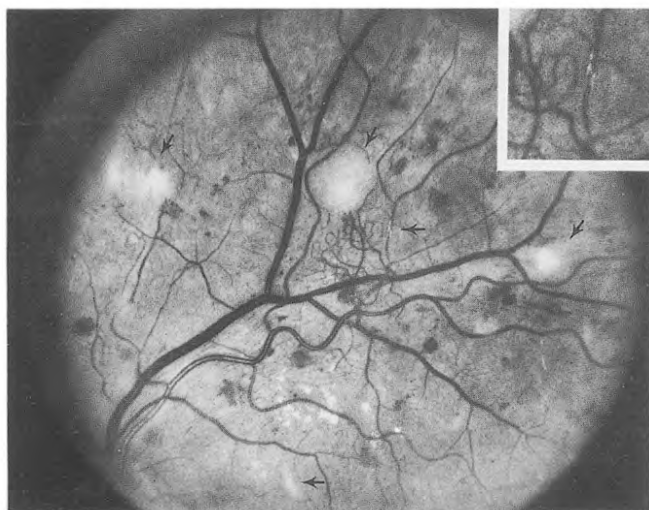


Fig 9. Standard photograph 5, intermediate standard for hard exudates and the more severe of two standards for soft exudates. Hard exudates are present below the center of the photograph and occupy about 4 to 5 times the area of those in Figure 8. Soft exudates are present in three locations in the upper half of the photograph (oblique arrows), with an additional smaller one near its lower edge (lower horizontal arrow). They occupy about 3 times the area of the soft exudates in Figure 11. The small segment of arteriolar sheathing (upper horizontal arrow and inset) defines the lower boundary of grade 3 (moderate) arteriolar sheathing. Part of the abnormal vessels at the center of the photo are IRMA and part are new vessels (the two parts can be distinguished clearly only in the stereo transparency, by slight elevation of the new vessels).

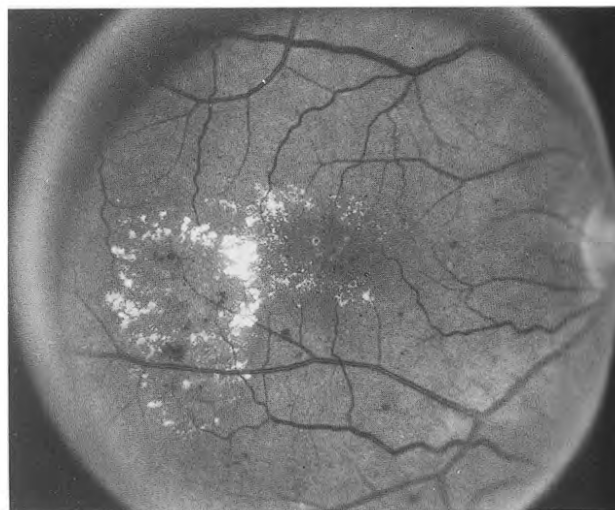


Fig 10. Standard photograph 4, severe standard for hard exudates.

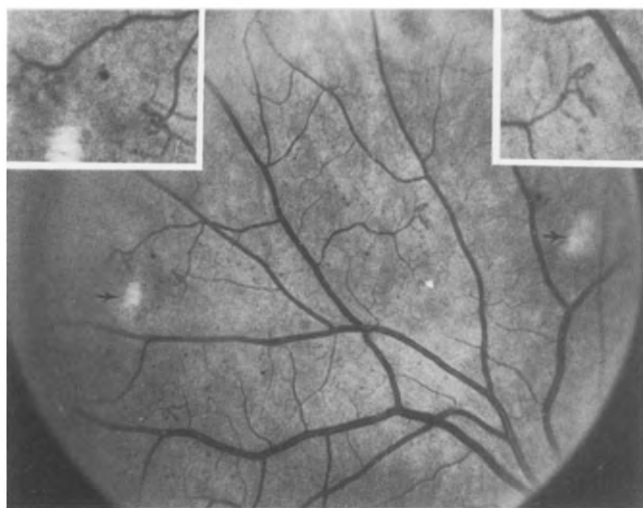


Fig 11. Standard photograph 8A, less severe of two standards for grading soft exudates (horizontal arrows) and IRMA, which are present in three areas (two shown in insets).

Soft exudates (cotton-wool patches; fields 1–7). Soft exudates are localized superficial swellings in the nerve fiber layer. They are round or oval in shape, white, pale yellow–white, or greyish–white in color, and have ill-defined (feathery) edges, frequently with striations parallel to the nerve fibers. The grading is based on the area of

retina involved, using standard photographs 8A and 5 (Figs 9 and 11) for comparison, as follows: grade 0 = no soft exudate; grade 1 = questionable soft exudate; grade 2 = definite soft exudate, less than standard photograph 8A; grade 3 = soft exudate greater than or equal to standard photograph 8A but less than standard photograph 5; grade 4 = soft exudate greater than or equal to standard photograph 5; grade 8 = cannot grade.

Intraretinal microvascular abnormalities (fields 2–7).

Tortuous intraretinal vascular segments, varying in caliber from barely visible to 31 μm (approximately one fourth the width of a major vein at the disc margin) or occasionally larger, are the abnormalities considered in this category. Punctate microaneurysms in the retina and new vessels located on the surface of the retina are excluded. The grading is based on the area of the retina covered by IRMA, using standard photographs 8A and 8B (Figs 11 and 12) for comparison: grade 0 = no IRMA; grade 1 = questionable IRMA; grade 2 = definite IRMA, less than standard photograph 8A; grade 3 = IRMA greater than or equal to standard photograph 8A but less than standard photograph 8B; grade 4 = IRMA greater than or equal to standard photograph 8B; grade 8 = cannot grade.

Venous abnormalities. In the ETDRS, three abnormalities that were combined in the DRS grading protocol are assessed separately: (a) venous beading, (b) venous narrowing, and (c) venous loops and/or reduplication. Venous sheathing and perivenous exudates are graded as in the DRS. Diffuse increase in venous caliber is not graded, because it is considered too difficult to evaluate consistently. Abnormalities within one half disc diameter (DD) of the disc margin are ignored.

Venous beading (fields 3–7). Localized increases in venous caliber, which sometimes resemble a string of beads and are typical of diabetic retinopathy, are the abnormalities assessed. The grade is based on the total length of vein involved and the severity of beading, using standard photographs 6A and 6B (Figs 13 and 14) for comparison, as follows: grade 0 = no beading; grade 1 = ques-

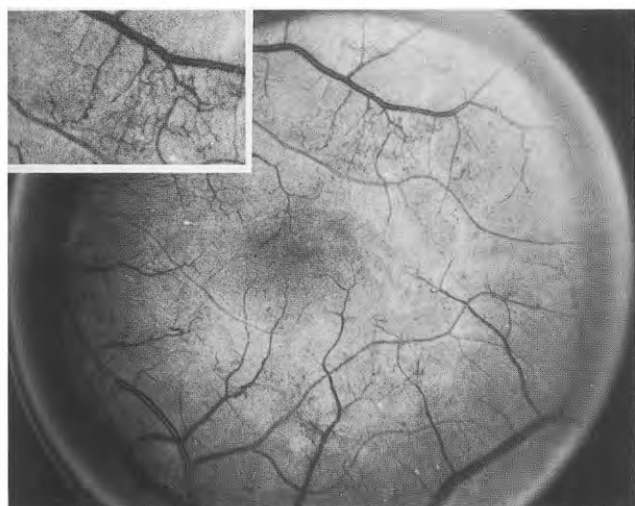


Fig 12. Standard photograph 8B, more severe standard for IRMA, which are present in all quadrants. Inset shows IRMA above and temporal to center of macula.



Fig 13. Standard photograph 6A, less severe of two standards for venous beading. Two main branches of the superior temporal vein show beading that is definite, but not severe.

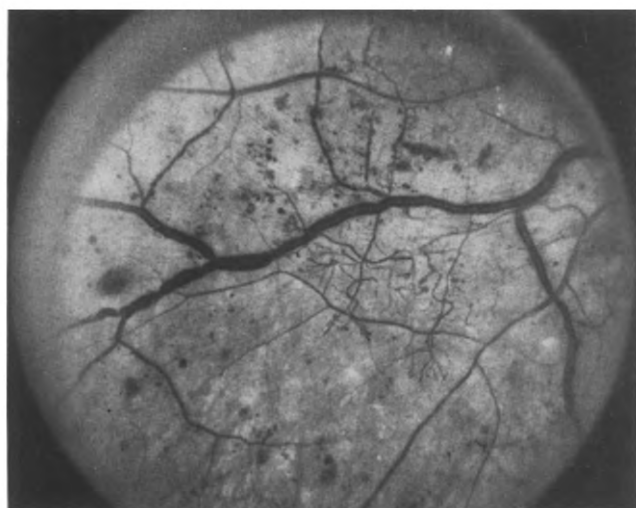


Fig 14. Standard photograph 6B, more severe standard for venous beading. Most venous branches, both large and small, are involved by severe beading.

tionable beading; grade 2 = definite beading, less than standard photograph 6A; grade 3 = beading greater than or equal to standard photograph 6A but less than standard photograph 6B; grade 4 = beading greater than or equal to standard photograph 6B; grade 8 = cannot grade.

Venous narrowing (fields 3–7). Localized narrowing of venous caliber is the abnormality considered. The total length of all narrowed segments within the field being graded is estimated and compared with the width of an average major vein at the disc margin (considered to be 125 μm) and the diameter of an average disc (considered to be 1500 μm), as follows: grade 0 = no narrowing; grade 1 = narrowing(s) questionable or total less than 125 μm ; grade 2 = narrowing(s) total greater than or equal to 125 μm but less than 1500 μm ; grade 3 = narrowing(s) total

greater than or equal to 1500 μm ; grade 4 = veins thread-like throughout field; grade 8 = cannot grade.

Venous loops and/or reduplication (fields 3–7). A venous loop is an abrupt, curving deviation of a vein from its normal path. Reduplication of a vein is the dilation of a pre-existing channel or the proliferation of a new channel of similar caliber adjacent to the original vein. The grading is based on the caliber of the veins involved, as shown below: grade 0 = no loops or reduplications; grade 1 = questionable loop(s) or definite loop(s) in a branch less than 31 μm wide if the loop is less than 50% wider than vein from which it arises; grade 2 = loop(s) in a branch greater than or equal to 31 μm wide or loop(s) in a branch less than 31 μm wide if the loop is greater than or equal to 50% wider than vein from which it arises; grade 3 = reduplication of a small vein (<31 μm wide); grade 4 = reduplication of a larger vein (≥ 31 μm wide); grade 8 = cannot grade.

Venous sheathing (fields 3–7). White lines along one or both sides of the venous blood column (sheathing) and/or complete opacification of the venous wall (white threads) are the abnormalities graded. The total length of vein having either appearance is assessed, as with venous narrowing: grade 0 = no sheathing or white threads; grade 1 = sheathing or white threads questionable or total less than 125 μm ; grade 2 = sheathing and/or white threads total greater than or equal to 125 μm but less than 1500 μm ; grade 3 = sheathing and/or white threads total greater than or equal to 1500 μm ; grade 8 = cannot grade.

Perivenous exudate (fields 3–7). Hard exudates occurring immediately adjacent to one or both sides of retinal veins, simulating venous sheathing, are classified as perivenous exudates and graded separately. The total length of vein involved is assessed: grade 0 = no perivenous exudate; grade 1 = perivenous exudate questionable or involves venous segments totaling less than $\frac{1}{8}$ DD; grade 2 = perivenous exudate (total of all segments) greater than or equal to $\frac{1}{8}$ DD; grade 8 = cannot grade.

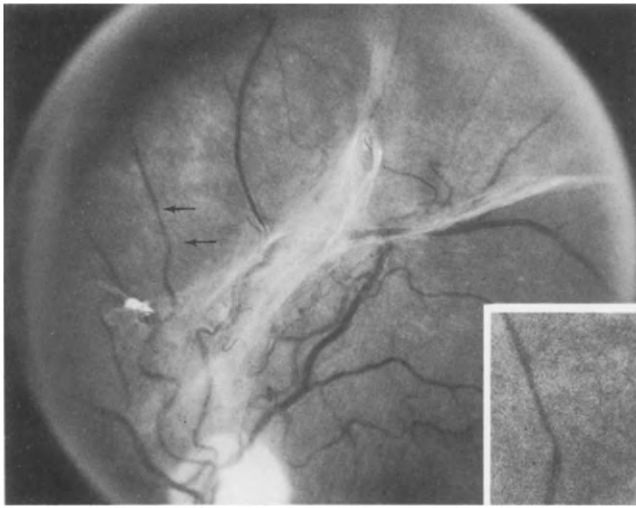


Fig 15. Standard photograph 11, milder of two standards for arteriolar narrowing and the standard defining the lower boundary of grade 4 (severe) FPE. The superior nasal arteriole has a subtle but definite narrowing about 500 μm long (between the horizontal arrows and in inset). In using this standard for grading FPE, the fibrous tissue in it that is within 1 DD of the disc is excluded (and graded as FPD), leaving FPE covering about 2.5 times the area of the standard disc (2.5 DA) for comparison with the photo being graded. The standard disc is slightly smaller than the disc in this eye, which was recognized to have a larger-than-usual disc only after its use as a standard had been established.

Arteriolar abnormalities. Arteriolar narrowing and opacification of arteriolar walls (sheathing and white threads), which were combined in the DRS grading protocol, are graded separately. Abnormalities within $\frac{1}{2}$ DD of the disc margin are ignored.

Arteriolar narrowing (fields 3–7). Localized irregularities in arteriolar caliber are the abnormalities graded. Because it is considered too difficult to grade consistently, generalized narrowing is not evaluated, except to indicate narrowing so extensive that the arterioles appear thread-like throughout the field. Sheathed arterioles and white threads are excluded when grading arteriolar narrowing. The grading is based on the length of arteriolar segments involved, using standard photographs 11 and 7 (Figs 15 and 16) for comparison: grade 0 = no arteriolar narrowing; grade 1 = questionable arteriolar narrowing; grade 2 = definite arteriolar narrowing, less than standard photograph 11; grade 3 = arteriolar narrowing greater than or equal to standard photograph 11 but less than standard photograph 7; grade 4 = arteriolar narrowing greater than or equal to standard photograph 7; grade 5 = arterioles thread-like throughout field; grade 8 = cannot grade.

Arteriolar sheathing (fields 3–7). Arteriolar walls that are partially opaque (a ribbon of red blood can still be seen with white lines on one or both sides of it) or completely opaque (no visible red blood column) are the abnormalities graded. The total length of arteriole that is either partially or fully opaque is assessed, using as cut points on the grading scale 125 μm and standard photographs 5 and 7 (Figs 9 and 16): grade 0 = no sheathing or white threads; grade 1 = sheathing/white threads ques-

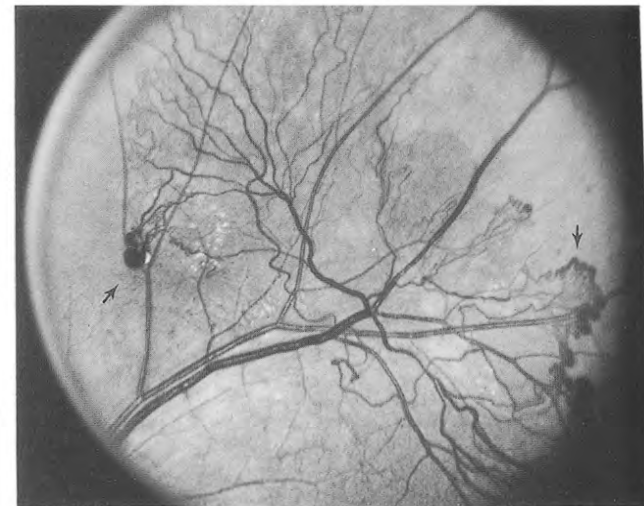
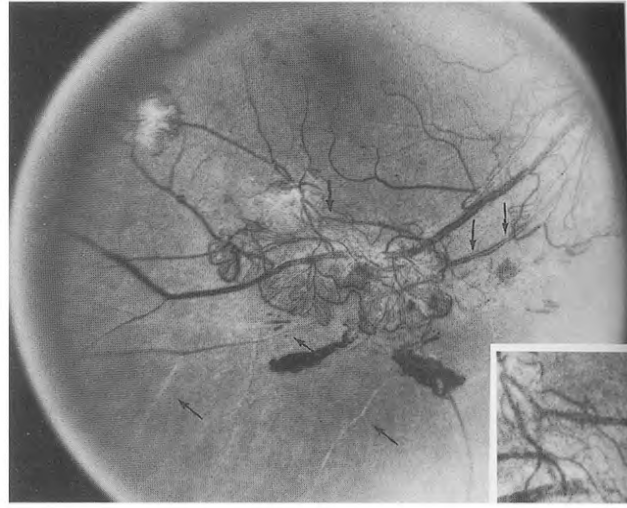


Fig 16. Top, standard photograph 7, arteriolar abnormalities and NVE, more severe of two standards for arteriolar narrowing and sheathing and the standard defining the lower boundary of grade 4 (severe) NVE. To the right of the center, where the inferior temporal arteriole is clearly visible between two patches of new vessels, two obvious focal narrowings are present (vertical arrows). A more subtle focal narrowing is present near the center of the photo (vertical arrow and inset). Arteriolar sheathing/white threads involve a total of about 4 DD, all in the lower part of the photograph (oblique arrows). In using this standard for grading NVE, most of the new vessels at the upper right part of the photograph are excluded, because they are within 1 DD of the disc and therefore are considered to be NVD. The NVE in this standard occupy an area about 2.5 times that of an average disc (2.5 DA). Bottom, example of NVE with dilated tips definitely present (arrows).

tionable or less than 125 μm in length; grade 2 = sheathing and/or white threads greater than or equal to 125 μm but less than standard photograph 5; grade 3 = sheathing and/or white threads greater than or equal to standard photograph 5 but less than standard photograph 7; grade 4 = sheathing and/or white threads greater than or equal to standard photograph 7; grade 8 = cannot grade.

Arteriovenous nicking (fields 3–7). Although arteriovenous nicking is not a typical feature of diabetic reti-

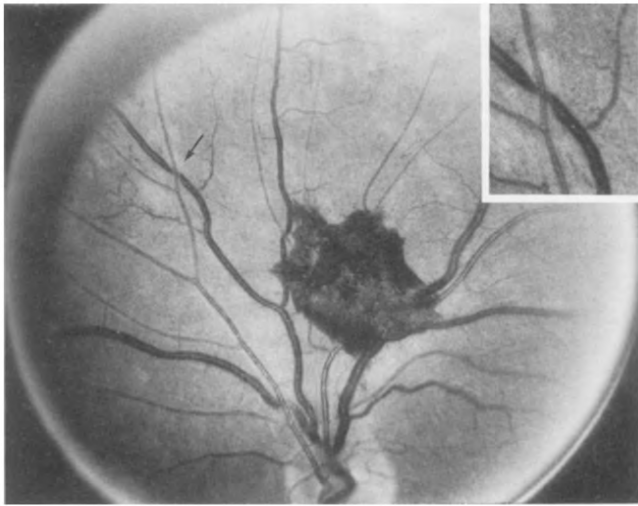


Fig 17. Standard photograph 9, one of two standards used (interchangeably) to define the lower boundary of grade 3 (moderate) preretinal hemorrhage, and used as well to define the lower boundary of grade 3 (moderate/severe) A-V nicking (arrow and inset). Preretinal hemorrhage without a fluid level occupies an area about 1.5 times that of the standard disc.

nopathy, evaluation of this characteristic is included as a possible measure of longstanding vascular hypertension. Arteriovenous crossings within $\frac{1}{2}$ DD of the disc margin are excluded. Tapering of both sides of the venous blood column on both sides of the crossing is required. Neither an obscuration of a narrow zone adjacent to the arterial blood column nor tapering of only the side of the venous column adjacent to the artery when the crossing is far from a right angle is sufficient for the designation “definite nicking.” Standard photograph 9 (Fig 17) provides an example of arteriovenous nicking that defines the lower boundary of grade 3 (moderate-to-severe). The grades are: grade 0 = no arteriovenous nicking; grade 1 = questionable arteriovenous nicking; grade 2 = definite nicking, less than standard photograph 9; grade 3 = nicking greater than or equal to standard photograph 9; grade 8 = cannot grade.

New vessels elsewhere (fields 1, 3–8). New vessels that are clearly *on the surface* of the retina (that is, *not within* the retina) or further forward in the vitreous cavity are considered to be new vessels elsewhere (NVE), except for those on the disc or within 1 DD of its margin (or in the vitreous anterior to this area), which are designated new vessels of the disc (NVD). However, if new vessels located mostly elsewhere extend into the area between $\frac{1}{2}$ and 1 DD from the disc margin, *and no other new vessels are present closer to or on the disc*, all new vessels are included in the NVE category. These same criteria are used to distinguish between fibrous proliferations elsewhere (FPE) and fibrous proliferations of the disc (FPD). The grading is based on the area of retina covered by the new vessels, using as cut points on the grading scale $\frac{1}{2}$ DA and the NVE in standard photograph 7 (Fig 16): grade 0 = no NVE; grade 1 = questionable NVE; grade 2 = definite NVE, less than $\frac{1}{2}$ DA; grade 3 = NVE greater than or

equal to $\frac{1}{2}$ DA, less than standard photograph 7; grade 4 = NVE greater than or equal to standard photograph 7; grade 8 = cannot grade.

New vessels may form networks less dense than those in standard photograph 7 or from no networks. When grading such NVE, the grader mentally compresses the new vessels to the density of the network seen in standard photograph 7 and then estimates the area that would be covered.

Dilated tips of new vessels elsewhere. New vessels may have dilated tips, which are defined as saccular or fusiform dilations of the peripheral parts of neovascular loops. Often such dilated segments are 2 to 4 times the caliber of the new vessel segments proximal to them (Fig 16, bottom). They are graded as follows: grade 0 = no dilated tips; grade 1 = dilated tips questionable, or caliber less than $2\times$ proximal NVE; grade 2 = caliber of dilated tips greater than or equal to $2\times$ proximal NVE; grade 8 = cannot grade.

Fibrous proliferations elsewhere (fields 1, 3–8). Fibrous tissue opaque enough to be definitely seen, with or without accompanying new vessels, is the abnormality considered, including fibrous strands or sheets that comprise a thickened posterior hyaloid surface, as well as completely atrophic new vessels (no visible red blood column present). Fibrous proliferations are designated FPE or FPD by the same criteria described previously for new vessels. In grading FPE, the area of retina covered is the main consideration, not the density or whiteness of the proliferations, using as cut points on the scale $\frac{1}{2}$ DA and the FPE in standard photograph 11 (Fig 15): grade 0 = no FPE; grade 1 = questionable FPE; grade 2 = definite FPE, less than $\frac{1}{2}$ DA; grade 3 = FPE greater than or equal to $\frac{1}{2}$ DA, less than standard photograph 11; grade 4 = FPE greater than or equal to standard photograph 11; grade 8 = cannot grade.

Occasionally, fibrous tissue occurs in fine strands rather than broad sheets. In this case, the grader mentally compresses the fine strands into a network of approximately the same density as the NVE in standard photograph 7 and then estimates the area that would be covered.

Plane of proliferation elsewhere (fields 1, 3–8). As a measure of the degree of contraction of vitreous and/or fibrovascular proliferations, the distance of these proliferations from the normal position of the *attached* retina is estimated. When the retina is detached, the degree of elevation of the fibrovascular proliferations from the normal position of the retina, not its detached position, is judged. The highest point of elevation determines the grade, as described below. If the plane of all neovascular and fibrous proliferations is on the surface of the attached retina, and the proliferations appear to be in contact with the retina throughout the involved area, the grade is 0. If proliferations appear not to be in contact with the retina in some areas but are not elevated as much as $\frac{1}{4}$ DD anywhere, and if no edge of the patch of proliferations is definitely elevated from the retina, the grade is 1. Proliferations definitely anterior to the retina by $\frac{1}{4}$ DD or more but less than 1 DD, or patches with one or more edges definitely elevated away from the retina (even if less than

$\frac{1}{4}$ DD), so that the anterior surface of the patch appears concave in this area, are graded 2. Elevations of at least 1 but less than 2 DD are graded 3; those of at least 2 DD are graded 4. If the stereoscopic effect is too poor to allow assessment of the degree of elevation, the designation "cannot grade," is indicated. The grading scale is: grade 0 = no elevation; grade 1 = elevation questionable or less than $\frac{1}{4}$ DD; grade 2 = elevation greater than or equal to $\frac{1}{4}$ DD, less than 1 DD; grade 3 = elevation greater than or equal to 1 DD, less than 2 DD; grade 4 = elevation greater than or equal to 2 DD; grade 8 = cannot grade.

Preretinal hemorrhage (fields 1–8). Both boat-shaped hemorrhages with a fluid level (standard photograph 13) (Fig 18) and round, oval, or linear patches of hemorrhage just anterior to the retina or under its internal limiting membrane (standard photograph 9) (Fig 17) are included. Hemorrhage on the surface of detached retina is also considered to be PRH. The grading is based on the area of retina covered by hemorrhage, using standard photographs 9 and 13 (Figs 17 and 18) and the area of one half of the field for comparison: grade 0 = no PRH; grade 1 = questionable PRH; grade 2 = definite PRH, less than standard photographs 9 or 13; grade 3 = PRH greater than or equal to standard photographs 9 or 13, less than $\frac{1}{2}$ field; grade 4 = PRH greater than or equal to $\frac{1}{2}$ field; grade 8 = cannot grade.

Vitreous hemorrhage (fields 1–8). Hemorrhage further forward in the vitreous cavity than PRH, including hemorrhage on or within fibrovascular proliferations, is considered to be VH. When the haziness present could be due to hemorrhage, lens opacity, or poor focus, grade 1 (questionable) is assigned. If definite hemorrhage is present, but whether it is PRH or VH is uncertain, the grade questionable may not be used for both. The grader must decide which definite grade is the best choice. When VH is definitely present but interferes with grading of less than 1 DA of the photographic field, the grade is 2. The grade of 2 is also assigned to diffuse VH (so thin that it does not prevent the grading of fundus abnormalities) covering but not obscuring 1 DA or more of a field (occasionally all of a field). When VH is dense enough to obscure a major retinal vessel (to the extent that beading, narrowing, and so forth, cannot be graded) and extensive enough to obscure at least 1 DA but less than one half of the photographic field, the grade is 3. Grade 4 is assigned when VH obscures one half of the field or more but some detail can still be graded (such as elevated fibrovascular proliferations). Grade 5 is assigned for VH only when *all* fundus details are obscured by VH, in which case "cannot grade" is assigned to all other characteristics in the field, with two exceptions (both in field 2 only): (1) obscuration of the center of the macula by VH is graded 2, definitely present, and (2) it might be possible to grade "posterior vitreous detachment" (PVD) by inference from adjacent fields (see below). The grading scale is: grade 0 = no VH; grade 1 = questionable VH; grade 2 = definite VH, obscuring less than 1 DA; grade 3 = VH obscures 1 DA or more but less than one half of field; grade 4 = VH obscures one half of field or more but some characteristics still able to be graded; grade 5 = VH obscures entire field (all char-

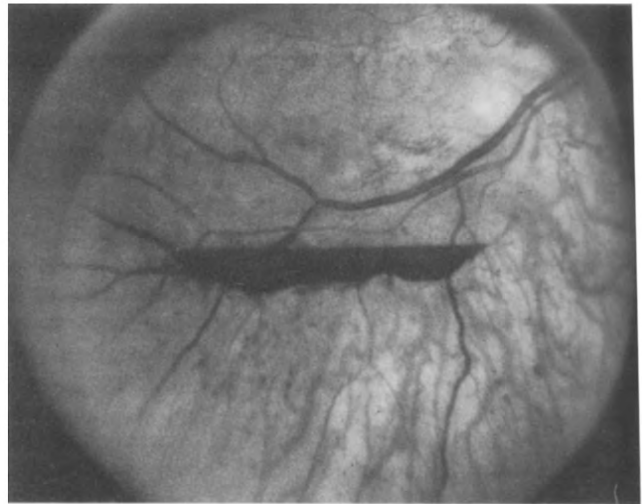


Fig 18. Standard photograph 13, one of two standards used (interchangeably) to define the lower boundary of grade 3 (moderate) preretinal hemorrhage. Preretinal hemorrhage of about the same amount as that in Figure 17 has settled to form a fluid level.

acteristics ungradable); grade 8 = cannot grade for reasons other than VH.

Retinal elevation (fields 2–7). The term "retinal elevation" includes retinal detachment, retinoschisis, or the elevation of a large retinal vessel (pulled forward in a narrow fold or avulsed from the retina). The *area* of retinal elevation rather than its height is judged, using 1 DA and standard photograph 12 (Fig 19) for comparison: grade 0 = no retinal elevation; grade 1 = questionable retinal elevation; grade 2 = retinal vessel and/or less than 1 DA of retina elevated; grade 3 = retinal elevation greater than or equal to 1 DA, less than standard photograph 12; grade 4 = retinal elevation greater than or equal to standard photograph 12; grade 8 = cannot grade.

Scars of prior photocoagulation (fields 1–7). The area of retina covered by scars of past photocoagulation or other scars of similar appearance is estimated. In a subsidiary item, the grader evaluates the area of retina obscured (e.g., by hemorrhage or fibrous proliferations) to such a degree that scars (as opposed to more subtle lesions) could not be discerned if present. For both scars and obscurity, the area of retina covered is estimated as a percentage of the entire field. If a portion of the retina is obscured by hemorrhage or fibrous proliferations, the grader estimates the area of scars seen, expressing the proportion scarred as a percentage of the *entire* field, both obscured and unobscured. For example, when 50% of a given field is obscured and 50% of the remaining part is occupied by scars, the grader records scars as 25% and obscurity as 50%. When only a small area of retina is obscured and the complete absence of scars in all other photographic fields indicates that no photocoagulation treatment has been applied, the grader estimates 0 for both scars and obscurity. Thus, "obscurity" in this item pertains only to obscurity sufficient to prevent estimation of the area involved by scars.

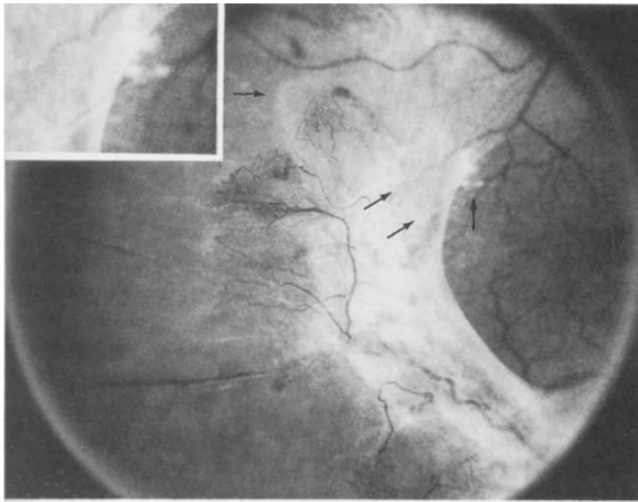


Fig 19. Standard photograph 12, defining the lower boundary of grade 4 (severe) retinal elevation, and providing a reference area for judging the thickness of the detached posterior vitreous surface. Inferiorly, retinal elevation is present, extending from the center of the photograph to its periphery between the 5-o'clock and 8-o'clock meridians (not easily appreciated in this nonstereoscopic photograph). The left (temporal) edge of a smaller area of retinal elevation in the superior temporal quadrant is indicated by the horizontal arrow. Retinal elevation totals about 6 to 8 disc areas in this photograph. The reference area used in judging the thickness of the detached posterior vitreous surface is an area along the 1:30 meridian about 1 DD from the center of the photograph (inset). At the right edge of this area, an E-shaped white structure (either hard exudate or regressed new vessels) lies on the surface of the retina (vertical arrow). The upper and lower boundaries of the reference area are formed by two small retinal vessels (oblique arrows) that course to the left and down from the upper and lower ends of the vertical bar of the E. The left edge of the reference area is the fibrous tissue and elevated retina to which the detached posterior vitreous surface is adherent (oblique arrows).

To determine the area of retina scarred or obscured, the grader makes use of either of two methods. The first, used for fields in which spacing of scars is fairly uniform, uses standard photographs (Fig 20). When only a few scars are present, or when their spacing is not uniform, the grader uses a second method, mentally moving the scars together and judging the area occupied. A transparent sheet containing a circle the size of the average disc and various segments and sectors of a circle is used in making such judgments.

CHARACTERISTICS GRADED ONLY IN FIELD 1

New vessels on or within 1 DD of the disc. New vessels were defined in the section on NVE. The grading is based on the area involved by NVD, using standard photographs 10A and 10C (Figs 21 and 22) for comparison: grade 0 = no NVD; grade 1 = questionable NVD; grade 2 = definite NVD, less than standard photograph 10A; grade 3 = NVD greater than or equal to standard photograph 10A but less than standard photograph 10C; grade 4 = NVD greater than or equal to standard photograph 10C; grade 8 = cannot grade.

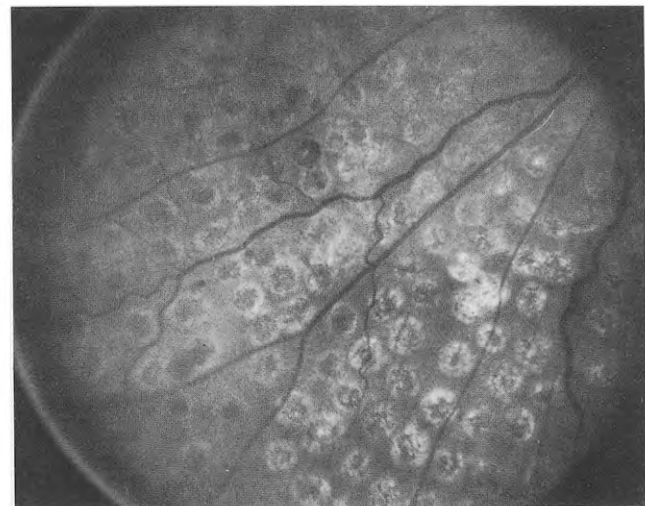
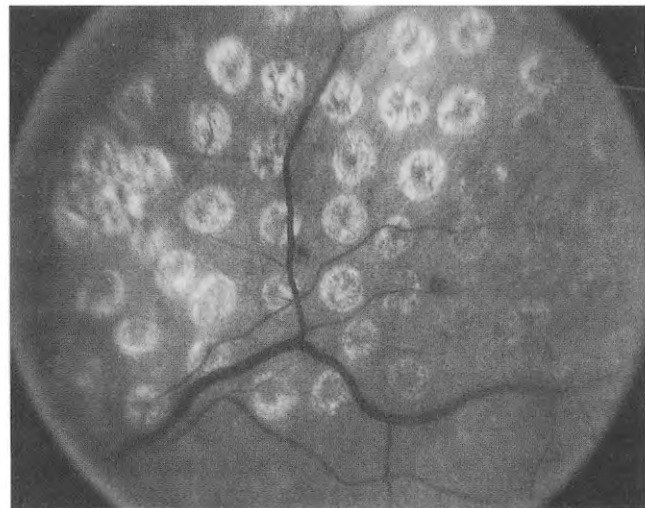
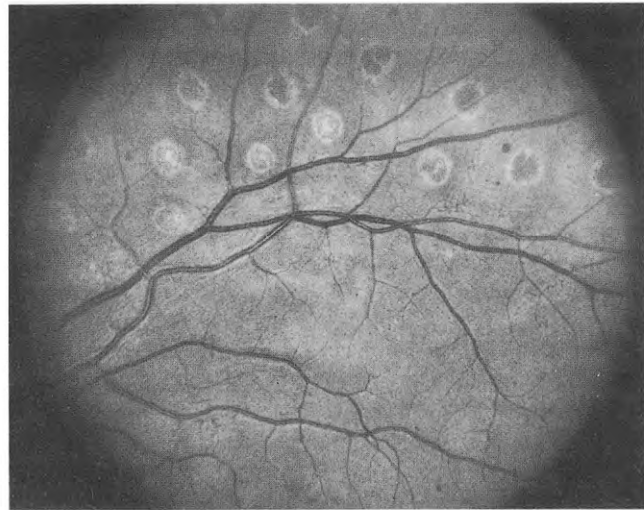


Fig 20. Examples of standard photographs used in judging percent of field covered by photocoagulation scars. *Top*, widely spaced scars cover 5% of this field 4. *Center*, more closely spaced scars over 25% of this field 4. *Bottom*, closely spaced scars cover 50% of this field 6.



Fig 21. Standard photograph 10A, defining the lower boundary of grade 3 (moderate) NVD. NVD cover approximately one third the area of the standard disc (4.7-mm diameter on the original transparency) and about one fifth the area of this unusually large disc (6.0-mm diameter on the original transparency). The unusual size of this disc was not taken into consideration when it was chosen as a standard, and is disregarded in grading (i.e., the area of NVD in the photograph being graded is compared with that in the standard, without regard to the size of the discs).



Fig 22. Standard photograph 10C, defining the lower boundary of grade 4 (severe) NVD. NVD cover about 1.3 to 1.5 times the area of the disc (which is about equal in size to the standard disc). All new vessels within 1 DD of the disc margin are included as NVD.

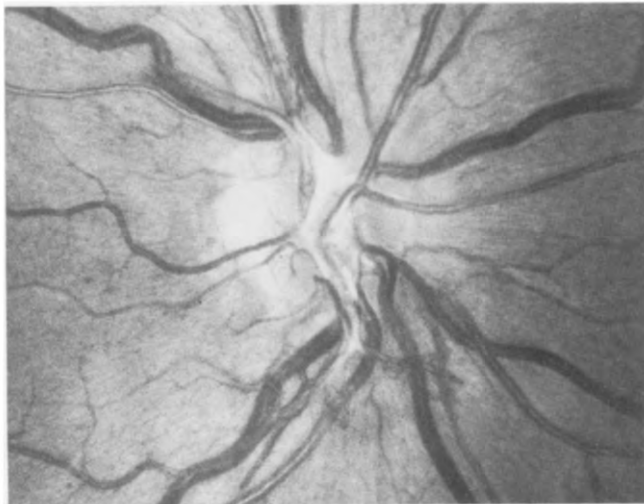


Fig 23. Standard photograph 10B, defining the lower boundary of grade 3 (moderate) FPD. FPD cover approximately one fourth the area of the standard disc, but a somewhat larger percentage of this rather small disc.

When grading NVD that are long and stringy (as opposed to those forming a network), the grader estimates the area they would cover if they were to be compressed into a network of the same density as those in standard photographs 10A and 10C (Figs 21 and 22).

Dilated tips of new vessels of the disc. Considerations for grading dilated tips of NVD are the same as those for evaluation of dilated tips of NVE: grade 0 = no dilated tips; grade 1 = dilated tips questionable, or caliber less than 2 times proximal NVD; grade 2 = caliber of dilated

tips greater than or equal to 2 times proximal NVD; grade 8 = cannot grade.

Fibrous proliferations on or within 1 DD of the disc.

Fibrous proliferations of the disc have been defined in the section on NVE. The grading is based on the area involved by FPD, not their thickness or whiteness, using as cut points on the scale 2 DA and standard photograph 10B (Fig 23): grade 0 = no FPD; grade 1 = questionable FPD; grade 2 = definite FPD, less than standard photograph 10B; grade 3 = FPD greater than or equal to standard photograph 10B, less than 2 DA; grade 4 = FPD greater than or equal to 2 DA; grade 8 = cannot grade.

If the FPD occur in fine strands, the grader mentally compresses them into a network of density similar to that of the NVE in standard photograph 7 and estimates the area that would be covered.

Plane of proliferation on or within 1 DD of the disc.

Considerations for grading plane of proliferation of NVD/FPD are the same as those given above for plane of proliferation elsewhere: grade 0 = no elevation; grade 1 = elevation questionable or less than $\frac{1}{4}$ DD; grade 2 = elevation greater than or equal to $\frac{1}{4}$ DD, less than 1 DD; grade 3 = elevation greater than or equal to 1 DD, less than 2 DD; grade 4 = elevation greater than or equal to 2 DD; grade 8 = cannot grade.

Papillary swelling. The maximum severity of swelling and the percentage of the disc margin involved are considered. If the disc margin is blurred for 270° or more, even though swelling is mild, no more than example C (Fig 24), the grade is 3 (moderate-to-severe). If swelling is as severe as that in example E (Fig 25), the grade is also 3, even if the swelling is confined to only one quadrant of the disc. If neither of the above conditions is met, but swelling is considered to be definitely present, the grade is 2. Grading is as follows: grade 0 = no papillary swelling; grade 1 = questionable papillary swelling; grade 2 = severity less than example E and extent less than 270° ; grade



Fig 24. Example photograph C, disc swelling involving $\geq 270^\circ$ of the disc margin (4- to 2-o'clock position in this photograph), even though mild, is assigned grade 3 (moderate/severe).

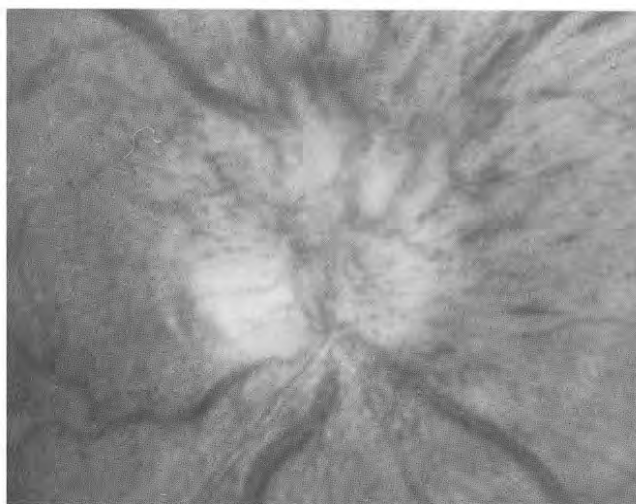


Fig 25. Example photograph E, disc swelling of this severity, even if $<270^\circ$ of the disc margin is involved, is assigned grade 3 (moderate/severe).

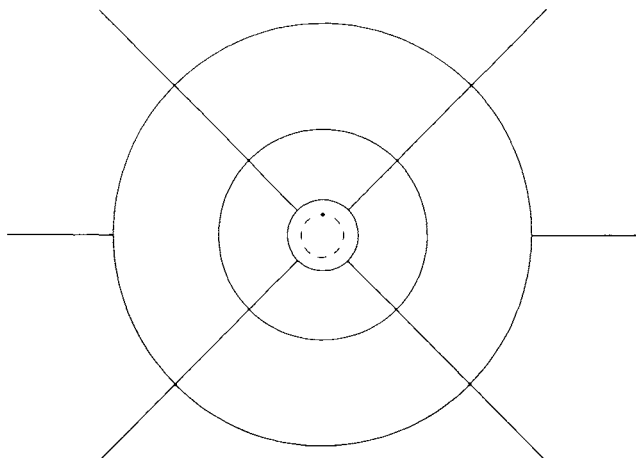


Fig 26. Grid centered on the macula and used for judging distances of 500 μm (inner solid circle) and 1 DD (middle solid circle) from the center. (The dashed circle has a radius of 300 μm , and the outer solid circle has a radius of 2 DD.)

3 = severity greater than or equal to example E *or* extent greater than or equal to 270° ; grade 8 = cannot grade.

CHARACTERISTICS GRADED ONLY IN FIELD 2

Grid used in grading retinal thickening and hard exudate. The grid designed for grading fluorescein angiograms (Fig 26), which provides circles with radii of $\frac{1}{3}$, 1, and 2 DD (500, 1500, and 3000 μm), is used to define the borders of areas within 500 and 1500 μm of the center of the macula when assessing retinal thickening and hard exudates.

Hard exudate rings. Rings of hard exudate usually surround zones of retinal thickening (edema), within which prominent microaneurysms are usually visible. When

hard exudates are definitely present in field 2 (grade ≥ 2), the percentage of the total forming complete or partial rings, defined as arcs of 180° to 360° , is estimated. Arcs less than 180° are not included in the estimate; if only such arcs are present, the grade assigned is 1 (questionable). The grades are: grade 0 = no hard exudate rings; grade 1 = rings questionable, or less than 180° ; grade 2 = definite ring(s), less than 10% of hard exudates form ring(s); grade 3 = greater than or equal to 10% but less than 50% of hard exudates form ring(s); grade 4 = greater than or equal to 50% but less than 90% of hard exudates form ring(s); grade 5 = greater than or equal to 90% of hard exudates form ring(s); grade 8 = cannot grade.

In Figure 10, three prominent hard exudate rings can be seen. Almost all of the hard exudates in the field appear to form part of one of the rings ($\geq 90\%$, grade 5). In the lower part of Figure 9, hard exudates form an incomplete oval surrounding a microaneurysm that lies just above an arteriole. Of the total hard exudates in the field, at least 50% but less than 90% are part of a ring (grade 4). In Figure 8, none of the hard exudates appear to be part of a ring (grade 0).

Posterior vitreous detachment. Both the presence of posterior vitreous detachment and the thickness of the detached posterior vitreous surface (the degree to which it is visible) are considered. Although this characteristic is graded in field 2 only, information from adjacent fields also is used. When there is no evidence of PVD, the grade is 0. Detachment of the posterior vitreous is considered questionable (grade 1) when a visible vitreous surface is elevated less than $\frac{1}{4}$ DD from the retina, when new vessels or fibrous proliferation are elevated less than $\frac{1}{4}$ DD from the retina, or when the upper edge of a boat-shaped preretinal hemorrhage is less than $\frac{1}{4}$ DD in its greatest anterior-posterior dimension. If the grader does not see a posterior vitreous surface but does see new vessels or fibrous proliferations elevated $\frac{1}{4}$ DD or more away from

the retina in field 2, PVD is graded "definitely present, inferred from other lesions," grade 2. Likewise, grade 2 is assigned when, in an adjacent field, the orientation of the fibrovascular proliferations indicates that posterior vitreous detachment must also involve field 2. Grade 2 is also assigned when the maximum anterior-posterior dimension of the upper edge of a boat-shaped hemorrhage is at least $\frac{1}{4}$ DD. When the posterior vitreous surface is visible in field 2 and is $\frac{1}{4}$ DD or more anterior to the retina, the grade is 3 or 4: grade 3 if the thickest area visible in field 2 is thinner than the reference area in standard photograph 12 (Fig 19); grade 4 if the thickness equals or exceeds this area. In summary: grade 0 = no PVD; grade 1 = questionable PVD; grade 2 = definite PVD, inferred from other lesions; grade 3 = PVD, surface visible, thin ($<$ standard photograph 12); grade 4 = PVD, surface visible, thick (\geq standard photograph 12); grade 8 = cannot grade.

Retinal thickening (edema). Thickening of the retina (with or without partial loss of transparency) is the characteristic considered. Two aspects are graded: the *size* of the thickened area and its maximum *thickness*. Judgments are made separately (1) for all of field 2, (2) for the macular region only (that area of retina within 1 DD of the center of the macula), and (3) for the center of the macula (where, because the center is defined as a point, only thickness need be graded). These three locations are not mutually exclusive, but rather are included within one another. Thus, if thickening is recorded at the center, it must also be recorded in the macula and in field 2 (and thickness must be at least as great). On the other hand, thickening recorded in field 2 need not necessarily be present in the macula, and thickening in the macula need not necessarily be present at its center.

Size of thickened area. The size of the thickened area in all of field 2 is recorded in the item designated "field thickening size" and the size of the thickened area within 1 DD of the center of the macula is recorded in the item designated "macular thickening size," using the scale shown below. In either case, when several patches of thickening are present, they are added together and the total area is estimated.

Field thickening size (all of field 2) is graded as: grade 0 = no retinal thickening; grade 1 = questionable thickening; grade 2 = thickening definite less than $\frac{1}{2}$ DA; grade 3 = thickening greater than or equal to $\frac{1}{2}$ DA, less than 1 DA; grade 4 = thickening greater than or equal to 1 DA, less than 2 DA; grade 5 = thickening greater than or equal to 2 DA, less than $\frac{1}{4}$ field; grade 6 = thickening greater than or equal to $\frac{1}{4}$ field, less than $\frac{1}{2}$ field; grade 7 = thickening greater than or equal to $\frac{1}{2}$ field; grade 8 = cannot grade.

Macular thickening size (area $<$ 1 DD from center of macula) is graded as: grade 0 = no retinal thickening; grade 1 = questionable thickening; grade 2 = thickening definite, less than $\frac{1}{2}$ DA; grade 3 = thickening greater than or equal to $\frac{1}{2}$ DA, less than 1 DA; grade 4 = thickening greater than or equal to 1 DA, less than 2 DA; grade 5 = thickening greater than or equal to 2 DA; grade 8 = cannot grade.

Maximum thickness of thickened area. When estimating maximum thickness, the grader evaluates the distance from the retinal pigment epithelium to the surface of the retina. In stereo photographs with good stereoscopic effect, the maximum thickness of the normal retina (in the area $\frac{1}{2}$ to 1 DD from the center of the macula) *appears to be* about twice the diameter of a major retinal vein at the disc margin. This is used as the *reference thickness* in defining the retinal thickness scale. In stereo photographs of normal eyes, the retinal surface at the foveal depression appears to rest on the pigment epithelium without visible thickness.

When thickening in field 2 is at least questionably present (grade 1 or greater), the maximum thickness is estimated and recorded in the item "field thickness," using the scale shown below. The lowest step in the scale, less than reference thickness, is seldom used, except when thickening is limited to the normally thin retina at and near the center of the macula or when thickening is only questionably present. When thickening in the macular area is at least questionably present (grade 1 or greater), its maximum thickness is estimated and recorded in the item "macular thickness," and the presence (or absence) and degree of thickening at the center of the macula is recorded in the item "center thickness."

Evaluation of center thickness may be difficult, either because thickening is subtle or because the location of the center is uncertain. In identifying the center of the macula consideration is given first to the increased pigmentation that normally surrounds it and second to the pattern of the smallest visible perifoveal blood vessels, which approach but do not reach it. In some eyes, neither of these methods allows a confident estimate. For this reason, a different definition is used for grade 1 (questionable) for this characteristic, namely that involvement of the center cannot be confidently recorded as either definitely absent or definitely present (rather than 50 to 90% certainty that it is present).

Field thickness (all of field 2) is graded as: grade 2 = thickening questionable, or thickness less than reference; grade 3 = thickness greater than or equal to $1\times$, less than $2\times$ reference; grade 4 = thickness greater than or equal to $2\times$ reference, less than $\frac{1}{2}$ DD; grade 5 = thickness greater than or equal to $\frac{1}{2}$ DD; grade 7 = cannot grade, poor stereo; grade 8 = cannot grade, other.

Macular thickness (area within 1 DD from center of macula) is graded as: grade 2 = thickening questionable, or thickness less than reference; grade 3 = thickness greater than or equal to $1\times$, less than $2\times$ reference; grade 4 = thickness greater than or equal to $2\times$ reference, less than $\frac{1}{2}$ DD; grade 5 = thickness greater than or equal to $\frac{1}{2}$ DD; grade 7 = cannot grade, poor stereo; grade 8 = cannot grade, other.

Center thickness (center of the macula, a point) is graded as: grade 0 = no retinal thickening; grade 1 = thickening of center uncertain; grade 2 = definite thickening, thickness less than reference; grade 3 = thickness greater than or equal to $1\times$, less than $2\times$ reference; grade 4 = thickness greater than or equal to $2\times$ reference, less than $\frac{1}{2}$ DD; grade 5 = thickness greater than or equal

to ½ DD; grade 7 = cannot grade, poor stereo; grade 8 = cannot grade, other.

Hard exudate in the macular area. The presence and amount of hard exudate within 1 DD of the center of the macula is assessed. If hard exudate is present in this area, its presence at the center of the macula also is assessed.

Hard exudate within 1 DD of the center of the macula. The grading is based on the area of retina involved, using standard photographs 3, 5, and 4 (Figs 8–10) for comparison. All of the hard exudate in each standard photograph is compared with the hard exudate within 1 DD of the center of the macula in the photograph being graded. The grades are: grade 0 = no hard exudate within 1 DD of center; grade 1 = questionable hard exudate within 1 DD; grade 2 = definite hard exudate within 1 DD, less than standard photograph 3 (Fig 8); grade 3 = hard exudate within 1 DD, greater than or equal to standard photograph 3, less than standard photograph 5 (Fig 9); grade 4 = hard exudate within 1 DD greater than or equal to standard photograph 5, less than standard photograph 4 (Fig 10); grade 5 = hard exudate within 1 DD, greater than or equal to standard photograph 4; grade 8 = cannot grade.

Hard exudate at the center of the macula. For this characteristic, the questionable grade is defined as it is for involvement of the center of the macula by retinal thickening, i.e., involvement of the center cannot be confidently recorded as either definitely absent or definitely present (rather than 50 to 90% certainty that it is present). Grading is as follows: grade 0 = no hard exudate at center of macula; grade 1 = involvement of center by hard exudate uncertain; grade 2 = definite involvement of center by hard exudate; grade 8 = cannot grade.

Cystoid spaces. Cystoid spaces are seen in color photographs as areas of increased retinal transparency occurring at or near the center of the macula, usually in retinas that appear thickened. Cystoid spaces are usually round or oval with diameters 1 to 4 times that of an average major retinal vein at the disc margin. Reflexes from the anterior surface of the retina and/or areas of mild depigmentation in the retinal pigment epithelium may easily be mistaken for cystoid spaces, as may areas of retina that have normal transparency when they are bordered by an arc of hard exudate or retinal hemorrhage. For cystoid spaces to be considered definitely present a clearly visible edge at which retinal transparency increases (not because of a transition from adjacent hard exudate or hemorrhage) must be traced around at least one half of the circumference of at least one cyst. If such an edge is visible for less than one half the circumference of a suspected cyst, grade 1 (questionable) is recorded. The grades for cystoid spaces are: grade 0 = no cystoid spaces; grade 1 = questionable cystoid spaces, or edge visible less than 180°; grade 2 = definite cystoid spaces, edge visible greater than or equal to 180°; grade 8 = cannot grade.

Clinically significant macular edema. The term clinically significant macular edema (CSME) was chosen to define macular edema involving or threatening the center of the macula. In the ETDRS, the presence of CSME was used to mandate follow-up treatment in eyes assigned to

early photocoagulation. Clinically significant macular edema is defined as either (1) retinal thickening at least 1 DA in extent, part of which is within 1 DD of the center of the macula, or (2) retinal thickening, or hard exudates adjacent to thickened retina, extending to within 500 µm of the center of the macula. The latter category may be subdivided on the basis of involvement of the center by thickening, as noted previously. The grades for CSME are: grade 0 = no CSME; grade 1 = questionable CSME; grade 2 = thickening greater than or equal to 1 DA, part within 1 DD of center; grade 3 = thickening or adjacent hard exudate within 500 µm of center; grade 8 = cannot grade.

Abnormalities within 1 DD of the center of the macula.

A number of abnormalities assume particular importance when they are located at or near the center of the macula. The characteristics listed below are divided into two groups on the basis of the area of the retina to be assessed. In each case, the grades are absent (0), questionable (1), definitely present (2), and cannot grade (8).

Abnormalities in the area within 1 DD of the center of the macula include: (1) new vessels on the retinal surface; (2) fibrous proliferations on the retinal surface; (3) new vessels on the detached posterior hyaloid surface; (4) fibrous proliferations on the detached posterior hyaloid surface; (5) pigment disturbance (i.e., increased or decreased pigmentation); (6) tension lines in the retina; (7) macular hole.

Abnormalities at the center of the macula. For these abnormalities, the questionable grade is defined as it is for involvement of the center of the macula by retinal thickening or hard exudate, i.e., involvement of the center cannot be confidently recorded as either definitely absent or definitely present. The characteristics graded are: (1) new vessels; (2) fibrous proliferations; (3) retinal detachment; (4) retinal distortion from tension lines; (5) dragged macula; (6) retinal hemorrhage; (7) preretinal hemorrhage; (8) subretinal hemorrhage; (9) confluent plaque of hard exudate, organized exudate, or subretinal fibrous scar; (10) deep white spot (the ill-defined yellowish-white spot, usually about 50 to 100 µm in diameter, sometimes seen in eyes with retinal thickening at the center of the macula and often *appearing to be* very deep in the retinal pigment epithelium or choroid); (11) center of macula obscured by VH.

OTHER (NONDIABETIC) ABNORMALITIES GRADED FOR THE EYE AS A WHOLE

When an abnormality not specified on the grading form is observed in any field, the grader enters the appropriate code letters from the following list: (A) questionable abnormality, type unspecified; (B) asteroid hyalosis; (C) central vein occlusion; (D) branch vein occlusion; (E) central artery occlusion; (F) branch artery occlusion; (G) disciform macular degeneration; (H) chorioretinal scar other than photocoagulation scar; (I) other abnormality (written on form); (J) choroidal nevus; (K) subretinal fibrous tissue (as with demarcation lines associated with retinal detachment); (L) coloboma or staphyloma.

PROCESSING OF GRADINGS AND ASSESSMENT OF REPRODUCIBILITY

Summary grade for the eye. For characteristics graded only in field 1 or field 2, the grading code can be used directly to express the severity of that characteristic for the eye, or the grades can be converted to descriptive terms, as follows: grade 0 = absent (A); grade 1 = questionable (Q); grade 2 = definitely present (D); grade 3 = moderate (M); grade 4 = severe (S); grade 5 = very severe (VS).

For characteristics graded in multiple fields, a summary of the grades assigned in the various fields is necessary, such as the maximum grade in any field, the number of fields in which the characteristic is definitely present, or some other measure. One convenient scale combines maximum severity in any field with the number of non-overlapping fields (i.e., fields 3–7) in which that maximum is present. Any field in which a characteristic is graded can establish the maximum grade, but for two (or more) fields to be recorded as having the maximum, both (or all) must be among fields 3 to 7. This convention produces five possible steps in the scale for each severity grade, which have been pooled to produce three steps: maximum in 1 field, maximum in 2 or 3 fields, and maximum in 4 or 5 fields.

Double grading. Baseline photographs were graded independently by two graders, with pairs chosen so that recently trained graders were paired with more experienced graders. The initial field-by-field gradings of each member of the grader pair were entered into a computer and compared. Disagreements by one step on the scale were averaged. When there were disagreements of two steps or more, these fields were regraded for these characteristics by each grader, without knowledge of the initial grades, and the grades again compared. When disagreement of two or more steps remained, they were resolved by a senior grader, who was also masked to the original grades. The senior grader, or adjudicator, had the option of choosing two adjacent steps as equally appropriate in borderline cases. Averaging adjacent steps in the scale produced new, intermediate steps, thus lengthening the grading scale. This expanded scale was convenient for characteristics graded in a single field or for characteristics graded in multiple fields when the maximum grade alone was used. However, the expanded scale was too complex for the characteristics graded in multiple fields when the maximum grade was combined with the number of fields in which it was present. Therefore, in this latter situation, the additional steps of the double-grading scale were pooled with the existing single-grading steps, as shown in Table 1.

Quality control. During grading of the baseline photographs, the duplicate gradings allowed comparisons between each of the original grades and the final "grade of record" attained by the double-grading system. Each grader was provided with cross-tabulations of his or her original grades and the grade of record, along with a summary of proportions of agreement and disagreement. The senior grader performing the adjudications reviewed these

Table 1. Scale for Characteristics Graded in Multiple Photographic Fields

Grade	Single Grading*		Additional Steps, Double Grading†	
	Maximum Grade	No. Fields	Maximum Grade	No. Fields
A	A	1–5‡	AQ	1
Q/1	Q	1	AQ	2–5
Q/2–3	Q	2–3		
Q/4–5	Q	4–5	QD	1
D/1	D	1	QD	2–5
D/2–3	D	2–3		
D/4–5	D	4–5	DM	1
M/1	M	1	DM	2–5
M/2–3	M	2–3		
M/4–5	M	4–5	MS	1
S/1	S	1	MS	2–5
S/2–3	S	2–3		
S/4–5	S	4–5	SVS	1
VS/1	VS	1	SVS	2–5
VS/2–3	VS	2–3		
VS/4–5	VS	4–5		

* Severity categories are of the form (maximum severity/extent), where maximum severity can be absent (A), questionable (Q), definite (D), moderate (M), severe (S), or very severe (VS), and extent is the number of fields at that severity level. For example, M/2–3 means there are 2 or 3 fields from fields 3 to 7 with moderate severity, and none with higher severity.

† The notations AQ, QD, and so forth, indicate intermediate steps derived from averaging two gradings that disagreed by one step. These are pooled as shown.

‡ If the maximum grade is A but this appears in less than five fields, the remaining fields were missing or ungradable.

statistics, along with the grading history of eyes with frequent disagreements, and then conducted further training sessions, focusing on particular characteristics as needed.

During follow-up, when single grading rather than double grading was used, variability of the gradings was monitored with haphazardly selected "reading lists" (randomized lists of ten eyes each, arranged so that the two eyes of a patient did not appear on the same list). Each grader graded two of these lists each calendar quarter, interspersed with routine work (i.e., graders were masked). The senior grader evaluated these photographs as well, although not being a routine grader, she could not be masked. Once everyone had graded the sample, each grader was provided with two kinds of information. First, two summary tables for each characteristic were prepared, with a row for each grader and columns showing proportions of all fields graded with agreement and all observed degrees of disagreement between each grader and (1) the senior grader or (2) the modal score determined from the corps of graders excluding the senior grader. Second, the grades chosen by each grader for each field of each eye in the sample were listed, along with the modal grade and that of the senior grader, grouped by characteristic graded. The graders were instructed to examine the summary tables to identify items for which their grades appeared dis-

Table 2. Percent of Eyes with Specified Levels of Agreement between First and Second Gradings of Baseline Color Fundus Photographs for Characteristics Graded Only in Fields 1 or 2, with Kappa Statistics (N = 7422)

Characteristic	Complete Agreement		Agree within 1 Step	Kappa (Standard Error)	
	A/A*	Other†		Unweighted	Weighted
NVD	86.1	3.4	98.0	0.388 (0.016)	0.595 (0.018)
Dilated tips of NVD	99.1	0.0	99.9	0.117 (0.046)	0.184 (0.066)
FPD	97.1	0.3	99.3	0.185 (0.032)	0.312 (0.048)
PPD	99.0	0.1	99.9	0.123 (0.051)	0.208 (0.088)
Papillary swelling	97.4	0.1	99.1	0.135 (0.028)	0.225 (0.044)
PVD	97.2	0.2	99.3	0.141 (0.031)	0.208 (0.047)
Hard exudate rings	41.3	20.0	83.4	0.423 (0.007)	0.604 (0.007)
Hard exudate $\leq 1500 \mu\text{m}\ddagger$	20.6	45.9	94.2	0.559 (0.007)	0.767 (0.005)
Hard exudate at center	91.6	1.6	99.1	0.354 (0.020)	0.507 (0.024)
Cystoid spaces	80.5	4.0	97.0	0.329 (0.014)	0.469 (0.017)
Area of thickening, field 2	24.9	26.1	82.9	0.396 (0.007)	0.621 (0.006)
Area of thickening $\leq 1500 \mu\text{m}\ddagger$	35.0	21.9	86.5	0.425 (0.007)	0.654 (0.006)
Maximum thickness field 2	33.5§	44.9	99.3	0.591 (0.009)	0.629 (0.008)
Maximum thickness $\leq 1500 \mu\text{m}\ddagger$	44.2§	35.0	99.2	0.606 (0.009)	0.641 (0.008)
Maximum thickness center	51.0	18.3	91.5	0.492 (0.007)	0.693 (0.006)
CSME	46.1	30.9	88.4	0.608 (0.007)	0.713 (0.007)

NVD = new vessels on or within 1 DD of disc; FPD = fibrous proliferations on or within 1 DD of disc; PPD = plane of proliferation on or within 1 DD of disc; PVD = posterior vitreous detachment; CSME = clinically significant macular edema.

* A/A = graders agreed on grade of absent.

† Other = graders agreed exactly on a grade other than absent.

‡ $\leq 1500 \mu\text{m}$ from the center of the macula.

§ For maximum thickness, field 2, and maximum thickness $\leq 1500 \mu\text{m}$ from the center of the macula, the lowest grade on the scale (2) represented pooled grades of absent, questionable, and thickening less than 1 \times reference. Agreement on this lowest step is shown here.

crepant, and then to review the relevant eyes, using the list of grades as a guide. The identity of each grader was scrambled so that only he/she among the grader corps could identify his/her gradings.

The senior grader examined the summary tables (unmasked to grader identity) to determine which items might be causing difficulties for which graders. On the basis of this review and problems referred by graders during the course of routine grading, the senior grader conducted tutorial sessions with individual graders as needed. For grading problems of general concern (e.g., evaluation of retinal thickening, which can be quite difficult), the senior grader held review meetings, for which graders were required to prepare by grading a set of eyes selected for the characteristics under consideration.

Reproducibility of gradings. Since many graders (23) participated in the ETDRS, intergrader agreement was considered more important than intragrader agreement and all analyses of reproducibility considered the former comparison. Reproducibility of single gradings was assessed by comparing the duplicate gradings of all baseline photographs. Reproducibility of the double grading system was assessed in a randomly selected sample of 454 eyes in which a repeat duplicate grading of the baseline photographs was carried out (by different graders from those doing the initial gradings, but with the same adjudicator). Reproducibility was assessed for each characteristic graded by tabulating percentages of agreement and selected degrees of disagreement, and by calculating unweighted and weighted kappa statistics.^{4,5} Weights were chosen so that

disagreements between absence and definite presence of a characteristic would receive 0 weight and those between absence and questionable presence of a weight of approximately 0.75. Thus, for single gradings of individual fields or for the maximum grade in any of the multiple fields in which a characteristic was graded, the weights used were 1.0 for exact agreement, 0.75 for one-step disagreement, and 0 for all other disagreements. For double gradings, either of individual fields or of the maximum grade across multiple fields (without regard to the number in which the maximum grade is present), the number of steps in the grading scale between absent and definitely present was four (A, AQ, Q, QD, D) rather than two (A, Q, D), and the weights chosen were 1.0, 0.875, 0.75, 0.5, and 0, respectively, for exact agreement, disagreement by one, two, or three steps, and all greater disagreements. These same weights were used for the scale using maximum grade/number of fields with the maximum, since it also has four steps between absent and definitely present (A, Q/1, Q/2–3, Q/4–5, D/1).

RESULTS

Agreement for single gradings of characteristics graded in one field only (field 1 or 2) is shown in Table 2. The first column of the table shows agreement on a grade of absent, the second agreement on any other grade. The third column presents the sum of the first two, plus one-

step disagreements. The remaining columns present unweighted and weighted kappa statistics and their standard errors. The first characteristic listed, NVD, was infrequent, and graders agreed on the grade of absent in 86.1% of eyes. Largely because of this, exact agreement was present for 89.5% of eyes (sum of columns 1 and 2). In nearly all of the remaining eyes, disagreements were of one step only, giving a total for agreement within one step of 98% (column 3). Unweighted kappa statistic was 0.388 and weighted kappa statistic was 0.595.

For the next five characteristics listed in Table 2, which occurred very infrequently, there was agreement on "absent" in at least 97% of eyes, and kappa statistics were low. The next four characteristics, three expressing aspects of hard exudate and the fourth cystoid changes at the center of the macula, were less frequently graded absent, and agreement within one step ranged from 83 to 99%, with weighted kappa statistics of 0.47 to 0.77. The remaining six characteristics, all expressing aspects of retinal thickening, were usually present, and agreement within one step ranged upward from 82.9%, with weighted kappa statistics from 0.62 to 0.71. Abnormalities within 1 DD of the center of the macula were rarely present and are not included in Table 2.

Table 3 presents a similar analysis for characteristics graded in multiple fields. Part A presents information for the maximum grade in any field, using a format identical to that of Table 2. Part B presents similar information for the scale that uses maximum grade and number of fields with that maximum (Table 1). Because this expanded scale has three steps within each maximum grade (except 0), each step of disagreement on this scale is designated as one third of one full step. The third column in Part B includes only complete agreement and disagreements of one third step. The fourth column, designated agreement within one full step, includes disagreements of up to three of these smaller steps.

For microaneurysms, the graders agreed on the highest step in the maximum grade scale (one or more retinal hemorrhages and one or more microaneurysms) in 97.1% of eyes (and on some other grade in 0.7%, for a total with complete agreement of 97.8%). For drusen, there was 84.6% agreement on the next to lowest step (drusen, questionable or less than those in standard photograph 1, which has about 10 small drusen). Kappa statistics for these characteristics were relatively low, even with weighting (0.390 and 0.331, respectively).

The next five characteristics, hemorrhages and/or microaneurysms, hard exudates, soft exudates, IRMA, and venous beading, were important in the ETDRS because their severity was one factor in determining the photocoagulation treatment strategy used.⁶ For these characteristics, all steps in the severity scale were present among the 7422 eyes represented in Table 3, except for the absent grade for hemorrhages and/or microaneurysms. On the maximum grade scale, complete agreement (the sum of the first two columns) ranged from 51.5% for IRMA to 70.7% for hemorrhages and/or microaneurysms and agreement within one step from 90.1 to 99.5%. Weighted kappa statistics ranged from 0.479 for IRMA to 0.783 for

hard exudates. In part B of Table 3, agreement within one third step ranged from 57.2% for IRMA to 78.1% for hemorrhages and/or microaneurysms. As expected, agreement within one full step and weighted kappa statistics were virtually the same for the two scales for each of these characteristics. Unweighted kappas were lower with the second, longer scale.

The next seven rows of Table 3 consider other characteristics of the retinal veins and arterioles, most of which were less prevalent than venous beading. These seven characteristics may be divided into two groups on the basis of weighted kappa statistics above and below 0.40. The most frequent, and least reproducible, was arteriolar narrowing, for which (in part A of Table 3) complete agreement was 35.2%, agreement within one step was 81.0%, and weighted kappa was 0.270. Agreement was better (largely because of more agreements on the grade of absent) for venous narrowing and arteriovenous nicking, but weighted kappas were low (0.276 and 0.339, respectively, on the maximum grade scale). Arteriolar sheathing, venous sheathing, venous loops/reduplication, and perivenous exudate, although infrequent, were graded more reproducibly, with weighted kappas above 0.40.

The next four rows of Table 3 deal with features of new vessels and FPE. For extent of NVE and FPE, weighted kappas exceeded 0.7. Dilated tips of NVE and elevated proliferations were infrequent, and weighted kappas were lower (0.528 and 0.385, respectively in part A). The last three rows of Table 3 deal with characteristics that were rarely present at the baseline ETDRS visit, because of eligibility criteria.⁶ A grade of questionable or greater by at least one grader was recorded for preretinal hemorrhage in only 88 eyes, for VH in only 24 eyes, and for retinal elevation in only 12 eyes. Weighted kappas for preretinal hemorrhage were above 0.40, but for the remaining two characteristics kappas became technically incalculable or had a very large standard error.

Similar analyses limited to 4718 eyes selected in such a way that each grader is always classified as either more or less experienced gave almost identical results. In those analyses, for example, if grader 5 was paired with grader 7 for some gradings, thus defining grader 5 as more experienced, any gradings for which grader 5 was paired with a lower numbered (more experienced) grader were excluded. Graders received numbers consecutively as they were trained, so that lower numbered graders were generally more experienced.

For selected characteristics, Tables 4 and 5 present similar analyses for 454 eyes selected randomly for a repeat double grading of baseline photographs. There are two lines for each characteristic, the first providing comparisons of single gradings (a randomly selected grading from the first pair versus a randomly selected grading from the second pair) and the second comparisons of the two adjudicated double gradings.

Table 4 presents characteristics graded only in fields 1 or 2, using the expanded scale that results from the averaging of one-step disagreements in the double grading process. The additional steps are designated half steps and differences between a full step and an adjacent half step

Table 3. Percent of Eyes with Specified Levels of Agreement between First and Second Gratings of Baseline Color Fundus Photographs for Characteristics Graded in Multiple Photographic Fields, with Kappa Statistics (n = 7422)

Characteristic	A. Maximum Grade in Any Field				B. Maximum Grade/Number of Fields with Maximum						
	Complete Agreement		Agree within 1 Step	Kappa (SE)		Complete Agreement		Agree within 1/3 Step	Agree within 1 Full Step	Kappa (SE)	
	A/A*	Other†		Unweighted	Weighted	A/A	Other			Unweighted	Weighted
Ma	0	97.8	99.7	0.404 (0.036)	0.390 (0.039)	0	83.3	96.4	99.2	0.483 (0.011)	0.578 (0.012)
Drusen	0.2	85.9	98.1	0.198 (0.013)	0.331 (0.019)	0.2	57.0	86.4	97.9	0.042 (0.008)	0.179 (0.013)
H/Ma	0	70.7	99.5	0.463 (0.009)	0.631 (0.008)	0	41.4	78.1	99.3	0.262 (0.007)	0.612 (0.006)
HE	2.5	64.3	97.2	0.566 (0.007)	0.783 (0.004)	2.5	45.9	69.5	93.9	0.419 (0.006)	0.736 (0.004)
SE	2.1	53.0	94.9	0.334 (0.008)	0.585 (0.007)	2.1	29.2	61.8	92.9	0.213 (0.006)	0.563 (0.006)
IRMA	7.1	44.4	90.1	0.268 (0.008)	0.479 (0.009)	7.1	20.1	57.2	87.1	0.152 (0.006)	0.455 (0.008)
VB	33.7	21.1	93.7	0.277 (0.009)	0.508 (0.010)	33.7	10.4	69.4	92.4	0.192 (0.007)	0.495 (0.009)
VN	55.9	6.6	92.4	0.146 (0.009)	0.276 (0.013)	55.9	3.5	76.6	97.9	0.119 (0.008)	0.243 (0.013)
VLR	61.8	11.5	97.2	0.365 (0.011)	0.469 (0.012)	61.8	6.6	86.6	96.8	0.290 (0.008)	0.432 (0.012)
VS	93.7	0.8	98.3	0.279 (0.022)	0.436 (0.029)	93.7	0.6	97.3	98.3	0.249 (0.020)	0.413 (0.031)
PVEX	91.9	1.6	98.8	0.366 (0.021)	0.572 (0.024)	91.9	0.8	96.1	98.7	0.293 (0.017)	0.545 (0.024)
AN	17.2	18.0	81.0	0.092 (0.008)	0.270 (0.010)	17.2	7.1	42.8	77.7	0.058 (0.005)	0.224 (0.009)
AS	81.5	2.9	95.4	0.322 (0.013)	0.521 (0.015)	81.5	2.0	90.7	95.1	0.288 (0.011)	0.491 (0.015)
AVN	71.6	4.2	95.1	0.223 (0.011)	0.339 (0.015)	71.6	2.9	89.3	94.7	0.195 (0.010)	0.309 (0.015)
NVE	73.0	10.7	96.1	0.551 (0.010)	0.745 (0.009)	72.8	7.7	89.5	95.7	0.480 (0.010)	0.725 (0.009)
DLTE	93.0	1.3	99.4	0.343 (0.022)	0.528 (0.027)	93.0	0.9	97.7	99.3	0.302 (0.020)	0.517 (0.030)
FPE	90.1	3.5	98.6	0.537 (0.017)	0.724 (0.016)	90.1	2.8	96.9	98.5	0.486 (0.017)	0.723 (0.016)
PPE	94.9	0.8	99.3	0.288 (0.026)	0.385 (0.035)	94.9	0.6	98.0	99.2	0.266 (0.025)	0.353 (0.038)
PRH	98.8	0.2	99.7	0.315 (0.052)	0.489 (0.065)	98.8	0.1	99.5	99.9	0.306 (0.050)	0.479 (0.071)
VH	99.7	0.0	99.9	0.147 (0.077)	0.280 (0.131)	99.7	0.0	99.9	99.9	0.147 (0.077)	0.313 (0.169)
REL	99.8	0.0	99.9	0.555 (0.176)	0.280 (0.131)	99.8	0.0	99.8	99.8	0.312 (0.110)	0.487 (0.158)

Ma = microaneurysms; H/Ma = hemorrhages/microaneurysms; HE = hard exudates; IRMA = intraretinal microvascular abnormalities; VB = venous beading; VN = venous narrowing; VLR = venous loops/reduplication; VS = venous sheathing; PVEX = perivenous exudate; AN = arteriolar narrowing; AS = arteriolar sheathing; AVN = arterio-venous nicking; NVE = new vessels elsewhere (not on or within 1 DD of disc); DLTE = dilated tips elsewhere; FPE = fibrous proliferations elsewhere; PPE = plane of proliferation elsewhere; PRH = preretinal hemorrhage; VH = vitreous hemorrhage; REL = retinal elevation.

* A/A = graders agreed on grade of absent.

† Other = graders agreed exactly on a grade other than absent.

Table 4. Percent of Eyes with Specified Levels of Agreement between First and Second Single and Double Gratings of Baseline Color Fundus Photographs for Characteristics Graded Only in Fields 1 or 2, with Kappa Statistics (N = 454)

Characteristic	Type of Grading	Complete Agreement		Agree within ½ Step	Agree within 1 Full Step	Kappa (SE)	
		A/A*	Other†			Unweighted	Weighted
NVD‡	S§	85.2	2.6		96.6	0.307 (0.063)	0.478 (0.078)
	D§	78.9	3.7	95.8	98.2	0.308 (0.052)	0.673 (0.056)
Hard exudate rings	S	44.3	20.0		82.6	0.451 (0.029)	0.600 (0.030)
	D	39.4	14.5	81.0	85.8	0.363 (0.026)	0.681 (0.026)
Hard exudate < 1500 µm from center	S	20.3	44.5		94.1	0.530 (0.029)	0.753 (0.022)
	D	17.2	37.0	87.9	96.9	0.456 (0.027)	0.838 (0.013)
Hard exudate center	S	93.8	1.5		99.3	0.444 (0.090)	0.629 (0.097)
	D	90.3	1.8	98.7	99.6	0.409 (0.063)	0.710 (0.061)
Cystoid spaces	S	82.8	4.0		97.6	0.357 (0.061)	0.553 (0.070)
	D	73.6	5.9	95.1	98.8	0.376 (0.044)	0.653 (0.049)
Area of thickening field 2	S	23.3	27.1		83.0	0.390 (0.027)	0.619 (0.025)
	D	18.5	22.7	73.4	84.2	0.336 (0.024)	0.738 (0.018)
Area of thickening < 1500 µm from center	S	33.5	23.1		86.3	0.426 (0.028)	0.652 (0.026)
	D	27.5	21.1	78.1	90.0	0.388 (0.025)	0.780 (0.016)
Maximum thickness field 2	S	39.5	43.8		99.1	0.624 (0.034)	0.663 (0.032)
	D	28.2	41.8	96.0	99.3	0.558 (0.029)	0.723 (0.022)
Maximum thickness < 1500 µm from center	S	44.1	33.5		99.4	0.578 (0.035)	0.639 (0.032)
	D	37.4	31.1	96.9	99.8	0.534 (0.030)	0.722 (0.021)
Center thickness	S	50.4	19.2		92.1	0.507 (0.030)	0.699 (0.028)
	D	45.8	14.3	87.2	95.6	0.421 (0.026)	0.807 (0.165)
Clinically significant macular edema	S	37.0	28.2		76.7	0.560 (0.032)	0.667 (0.033)
	D	32.4	29.7	76.4	79.9	0.554 (0.029)	0.765 (0.027)

* A/A = graders agreed on grade of absent.

† Other = graders agreed exactly on a grade other than absent.

‡ NVD = new vessels on or within 1 DD of the disc.

§ S = single grading; D = double grading.

^{||} For maximum thickness field 2 and maximum thickness < 1500 µm, the lowest grade on the scale (grade 2) represented pooled grades of absent, questionable, and thickening less than 1× reference. Agreement on this lowest step is shown here.

are classified as half step disagreement. In Table 4, the improvement in reproducibility resulting from double grading is difficult to quantify because the scales used are different. If it is postulated that agreement within one half step on the double-grading scale is almost as satisfactory as complete agreement on the single-grading scale, then the sum of the first two columns for single grading may be compared with the third column for double grading for each characteristic. For hard exudate rings, this comparison was 64.3% versus 81.0%; for area of thickening within 1 DD of the center of the macula, 56.6% versus 78.1%; for center thickness, 69.6% versus 87.2%, and for CSME, 65.2% versus 76.4%. For all of the characteristics in Table 4, unweighted kappas decreased by less than 0.1 between single and double gradings. Since unweighted kappa is expected to decrease as the scale becomes longer, these small or absent decreases suggest improved reproducibility with double grading.

Double and single grading are more readily compared for characteristics graded in multiple fields, where pooling has maintained identical scales and weighting schemes (Table 5). Only the scale considering number of fields with the maximum grade is presented, with the format like that of part B of Table 3. The improvement in reproducibility resulting from double grading is demonstrated by comparing the single and double grading entries for a given characteristic in the columns showing agree-

ment within one third step and weighted kappa. For each characteristic, agreement within one third step increased by about 15 percentage points with double grading, except for NVE and its elevated proliferations, which already showed 89.0% and 98.0% agreement within one third step with single grading. With double grading, weighted kappas increased by about 0.1 to 0.2 units for most characteristics.

DISCUSSION

In assessing reproducibility of a grading system, percent of exact agreement is the simplest measure, but this percentage may be misleadingly high when a large majority of observations fall within a single step on the scale, as is the case for agreement on the absence of a rare characteristic. Conversely, when the severity of the characteristic being graded is close to the cutoff between adjacent steps of the scale, frequent disagreements, often trivial, are inevitable. Pooling one-step disagreement with exact agreement (agreement within one step) removes the trivial disagreements but also pools other, less trivial differences.

The unweighted kappa statistic provides an estimate of the proportion of exact agreement achieved more than that expected by chance, thus compensating for the problem of agreement that looks better than it really is because

Table 5. Percent of Eyes with Specified Levels of Agreement between First and Second Single and Double Gradings of Baseline Color Fundus Photographs for Characteristics Graded in Multiple Photographic Fields, with Kappa Statistics (N = 454)

Characteristic	Type of Grading	Complete Agreement		Agree within 1/3 Step	Agree within 1 Full Step	Kappa (SE)	
		A/A*	Other†			Unweighted	Weighted
H/Ma	S‡	0.0	43.6	74.4	99.5	0.307 (0.028)	0.597 (0.024)
	D‡	0.0	46.5	86.8	100.0	0.343 (0.028)	0.700 (0.019)
HE	S	1.5	42.5	66.0	95.5	0.371 (0.026)	0.705 (0.019)
	D	2.2	40.5	79.5	98.6	0.363 (0.025)	0.789 (0.013)
SE	S	3.7	30.6	57.4	90.2	0.244 (0.025)	0.522 (0.028)
	D	3.1	36.6	74.1	97.6	0.312 (0.026)	0.714 (0.019)
IRMA	S	8.6	19.4	55.6	87.5	0.159 (0.024)	0.457 (0.031)
	D	6.4	24.9	71.4	96.5	0.191 (0.025)	0.621 (0.023)
VB	S	38.1	8.8	68.8	91.7	0.220 (0.027)	0.476 (0.037)
	D	39.4	14.3	83.4	98.4	0.312 (0.030)	0.688 (0.027)
VN	S	50.2	3.5	70.2	92.0	0.078 (0.030)	0.171 (0.049)
	D	47.1	6.6	87.2	96.9	0.089 (0.036)	0.355 (0.053)
AN	S	17.2	8.8	46.5	74.5	0.075 (0.023)	0.243 (0.036)
	D	12.8	15.0	60.6	87.3	0.099 (0.025)	0.404 (0.033)
NVE	S	73.8	7.7	89.0	93.6	0.512 (0.035)	0.742 (0.033)
	D	74.4	6.6	91.8	96.6	0.492 (0.034)	0.812 (0.025)
PPE	S	93.6	0.9	98.0	98.4	0.266 (0.094)	0.426 (0.136)
	D	96.5	0.4	99.5	99.9	0.382 (0.094)	0.794 (0.070)

H/Ma = hemorrhages/ microaneurysms; HE = hard exudates; SE = soft exudates; IRMA = intraretinal microvascular abnormalities; VB = venous beading; VN = venous narrowing; AN = arteriolar narrowing; NVE = new vessels elsewhere (not on or within 1 DD of disc); PPE = plane of proliferation elsewhere.

* A/A = graders agreed on grade of absent.

† Other = graders agreed on a grade other than absent.

‡ S = single grading; D = double grading.

the characteristic is rare. However, unweighted kappa is not well suited to ordinal scales because all disagreements, large and small, are treated equally. Weighted kappa is more suitable for such scales, although the values obtained are highly dependent on the weighting scheme used and must be interpreted and compared with caution.

Our goal in choosing weights was to facilitate comparison of the reproducibility of longer scales with that of shorter ones. Given the weights chosen, the similarity of weighted kappas between the two sections of Table 4 for most characteristics was predictable. Finding the expected perhaps provides some reassurance that no great loss in reproducibility resulted from the longer scales, which are desirable because they provide more choices for defining subgroups of convenient size. This was particularly important for hemorrhages and/or microaneurysm, IRMA, and soft exudates, since for each of these characteristics, approximately 50% of eyes were classified in the same step of the maximum grade scale.

The ETDRS eligibility criteria excluded eyes with very mild retinopathy and those with more than early proliferative retinopathy. Thus, the baseline photographs do not provide a good sample for assessing the reproducibility of the classification for some characteristics. In Table 2, there were enough eyes with questionable or definitely present NVD (several hundred) to test the lower part of the scale for this characteristic; kappas for those gradings were 0.388 unweighted and 0.595 weighted. But other characteristics associated with the disc and posterior vit-

reous detachment were rare, and therefore their kappas are difficult to interpret. A similar problem exists for PRH and VH and for retinal elevation, as shown in Table 3. However, similar analyses (unpublished photographic gradings from the DRS) in which these characteristics were more frequently present have shown satisfactory reproducibility.

A comparable problem occurs for microaneurysms because nearly all of the 97.8% of eyes for which graders agreed were placed in the same step of the maximum grade scale, its highest (at least one retinal hemorrhage and at least one microaneurysm). The increase in kappa for gradings of microaneurysms between the maximum grade scale and the scale specifying maximum grade/number of fields with the maximum (Table 3) may be the result of expanding the scale at its upper end, which redistributed the 97.1% of eyes in which graders agreed on the maximum grade into its three subdivisions and left only 76.0% of eyes in which the graders agreed on the highest step in the expanded scale (one or more hemorrhages and one or more microaneurysms in four or five fields), thus presumably allowing the substantial agreement for the two steps below this to increase the kappa.

Landis and Koch⁷ have suggested the following general levels for interpretation of unweighted kappa statistics: 0 to 0.20, slight agreement; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and more than 0.81, almost perfect agreement. It seems reasonable to apply these same guidelines to weighted kappas when, as de-

scribed, weights are selected to give 0 weight for absent versus present disagreements. If so, then it seems fair to characterize agreement with single grading as substantial for NVE, FPE, hemorrhages and/or microaneurysms, hard exudates, the several individual characteristics of retinal thickening, and CSME; as moderate for NVD, soft exudates, IRMA, venous beading and other venous abnormalities (except narrowing), arteriolar sheathing, and dilated tips of NVE; and as fair for all other characteristics sufficiently prevalent to allow assessment (except drusen, which did not have high priority in our ongoing quality control procedures).

Double grading improved reproducibility but did not move any characteristics into the "almost perfect" category described by Landis and Koch. For soft exudates, IRMA, venous beading, perivenous exudate, arteriolar sheathing, dilated tips of NVE, and plane of proliferation elsewhere, the improvement with double grading was sufficient for movement to the "substantial agreement" category. The satisfactory reproducibility achieved with the single grading system for the most important characteristics was recognized during the early years of the ETDRS and provided support for a decision to discontinue duplicate grading of follow-up visit photographs, with con-

sequent savings of time and expense. The ETDRS experience supports the conclusion that this classification system has satisfactory reproducibility for most characteristics.

REFERENCES

1. The Diabetic Retinopathy Study Research Group. A modification of the Airle House classification of diabetic retinopathy. DRS report #7. *Invest Ophthalmol Vis Sci* 1981; 21:210-26.
2. Jonas JB, Gusek GC, Guggenmoos-Holzmann I, Naumann GOH. Variability of the real dimensions of normal human optic discs. *Graefes Arch Clin Exp Ophthalmol* 1988; 226:332-6.
3. Mansour AM, Walsh JB, Henkind P. Optic disc size in central retinal vein occlusion. *Ophthalmology* 1990; 97:165-6.
4. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20:37-46.
5. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70:213-20.
6. Early Treatment Diabetic Retinopathy Study Research Group. Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics: ETDRS report number 7. *Ophthalmology* 1991; 98: 741-56.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.