Automated Analysis of Retinal Images for Detection of Referable Diabetic Retinopathy

Michael D. Abràmoff, MD, PhD; James C. Folk, MD; Dennis P. Han, MD; Jonathan D. Walker, MD; David F. Williams, MD, MBA; Stephen R. Russell, MD; Pascale Massin, MD, PhD; Beatrice Cochener, MD, PhD; Philippe Gain, MD, PhD; Li Tang, PhD; Mathieu Lamard, PhD; Daniela C. Moga, MD, PhD; Gwénoële Quellec, PhD; Meindert Niemeijer, PhD

**Importance:** The diagnostic accuracy of computer detection programs has been reported to be comparable to that of specialists and expert readers, but no computer detection programs have been validated in an independent cohort using an internationally recognized diabetic retinopathy (DR) standard.

**Objective:** To determine the sensitivity and specificity of the Iowa Detection Program (IDP) to detect referable diabetic retinopathy (RDR).

**Design and Setting:** In primary care DR clinics in France, from January 1, 2005, through December 31, 2010, patients were photographed consecutively, and retinal color images were graded for retinopathy severity according to the International Clinical Diabetic Retinopathy scale and macular edema by 3 masked independent retinal specialists and regraded with adjudication until consensus. The IDP analyzed the same images at a predetermined and fixed set point. We defined RDR as more than mild nonproliferative retinopathy and/or macular edema.

**Participants:** A total of 874 people with diabetes at risk for DR.

**Main Outcome Measures:** Sensitivity and specificity of the IDP to detect RDR, area under the receiver operating characteristic curve, sensitivity and specificity of the retinal specialists’ readings, and mean interobserver difference (κ).

**Results:** The RDR prevalence was 21.7% (95% CI, 19.0%-24.5%). The IDP sensitivity was 96.8% (95% CI, 94.4%-99.3%) and specificity was 59.4% (95% CI, 55.7%-63.0%), corresponding to 6 of 874 false-negative results (none met treatment criteria). The area under the receiver operating characteristic curve was 0.937 (95% CI, 0.916-0.959). Before adjudication and consensus, the sensitivity/specificity of the retinal specialists were 0.80/0.98, 0.71/1.00, and 0.91/0.95, and the mean intergrader κ was 0.822.

**Conclusions:** The IDP has high sensitivity and specificity to detect RDR. Computer analysis of retinal photographs for DR and automated detection of RDR can be implemented safely into the DR screening pipeline, potentially improving access to screening and health care productivity and reducing visual loss through early treatment.


Increasing health care productivity is a prerequisite to improve health care affordability. Automation has improved productivity in many sectors of the economy, whereas in health care, productivity has remained stagnant in the last 20 years. Regular eye examinations are necessary to diagnose diabetic retinopathy (DR) at an early stage, when it can be treated with the best prognosis and visual loss delayed or deferred. In 2010, US eye care practitioners examined less than 60% of the estimated 23 million people with diabetes, leaving millions of people at risk for potentially preventable visual loss and blindness. The hope of computer analysis of retinal images taken by ancillary staff is that it will increase DR screening accessibility and adherence and reduce cost.

Computer detection of DR analyzes retinal color images obtained by fundus cameras and triages those who have DR and require referral to an ophthalmologist from those who can be screened again in 1 year. The diagnostic accuracy of computer detection programs has been reported to be comparable to that of specialists and expert readers, but none of the semiautomated or fully automated computer detection programs have been validated in an independent cohort using an internationally recognized DR standard.

The International Clinical Diabetic Retinopathy (ICDR) severity scale was formulated by a consensus of international experts to standardize and simplify DR...
classification (Table 1) to improve communication and coordination of care among physicians caring for patients with diabetes. The ICDR classification simplified the Early Treatment Diabetic Retinopathy Study (ETDRS) classification for nonproliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) because the latter classification had proved unwieldy in clinical care. In the present study, 3 fellowship-trained retinal experts (D.P.H., J.D.W., D.F.W.) independently graded retinal images of each eye from people with diabetes using the ICDR severity level scales and a modified definition of macular edema (ME), namely, any retinal thickening, exudate, or microaneurysm within 1 disc diameter of the fovea.

The objective of the present study is to determine the sensitivity and specificity of the Iowa Detection Program (IDP) to detect referable diabetic retinopathy (RDR), which we defined as more than mild NPDR as defined by the ICDR and/or ME. We defined nonreferable DR as no NPDR or mild NPDR and no apparent ME. The IDP was compared against the consensus rating of the 3 retinal specialists reading the same images.

### METHODS

**PARTICIPANTS AND IMAGES**

Deidentified digital fundus color images of 1748 eyes in 874 people with diabetes were used. The images are publicly available for noncommercial use. A total of 186 individuals were photographed from January 1, 2005, through December 31, 2010, at Hôpital Lariboisière, Paris, France; 489 at Brest University Hospital, Brest, France; and 199 at Saint-Etienne University Hospital, St Etienne, France. Each department included consecutive people with diabetes, diagnosed according to the World Health Organization criteria in use at that time. Approval for the publication of the deidentified images and use of the aggregate demographic data was obtained from each hospital’s institutional review board, according to the tenets of the Declaration of Helsinki. Demographics for each participant were obtained retrospectively by medical record review at the 3 centers. To ensure deidentification of the images, only the aggregate mean (SD) of age and sex distribution was available to the authors. Participants underwent pharmacologic dilation at 2 centers (Paris and Saint-Etienne) and did not undergo dilation at the third center. They were then imaged using a color video 3CCD camera (Canon Europe BV) on a Topcon TRC NW6 nonmydriatic fundus camera (Topcon USA, Inc) with a 45° field of view centered on the fovea. The images were captured at 1440 × 960, 2240 × 1488, or 2304 × 1536 pixels and saved in tiff or jpeg format. All participants were imaged successfully.

### ADJUDICATED ICDR REFERENCE STANDARD

Three internationally recognized, fellowship-trained retinal specialists graded 1 image of each eye. The experts were masked to each other and to the IDP. Each expert assigned an ICDR retinopathy level (scale of 0 for best to 4 for worst) and an ME level (scale of 0 for no ME to 1 for ME) for each image. Experts used the presence of exudates, retinal thickening (if visible), or microaneurysms, all within 1 disc diameter of the fovea, as a sign of ME. We added the criterion of 1 or more microaneurysms because we were concerned that ME would be incorrectly missed on nonstereo photographs by the experts, and the isolated presence of 1 or more microaneurysms (s) can be the only sign of ME visible on nonstereo photographs. Any disagreements were adjudicated by rerating the images until consensus was reached by all 3 experts. The consensus ICDR and ME severity levels for each participant were dichotomized into a single adjudicated rating: either RDR, meaning either moderate or severe NPDR, or PDR, ME, or both. A person was deemed to have RDR if either or both eyes had these findings. A person was deemed to have nonreferable DR if there was no NPDR or mild NPDR and no ME in both eyes. Sensitivity and specificity determined by the 3 experts were estimated by comparing their individual RDR and nonreferable DR gradings before consensus against those of the other 2 experts.

### DR DETECTION PROGRAM

The IDP consists of previously published components for image quality assessment,20 microaneurysm and hemorrhage detection,21 detection of exudates and cotton wool spots,22 and a new component for detection of irregular lesions, including large hemorrhages and neovascularization.23 Generally, IDP examines each pixel in each image, analyzes it and its surrounding pixels, and combines the analysis of multiple neighboring pixels into multiple lesions or retinal structures, with their likelihood, size, shape, location, type, and other properties, as well as the quality of each image. The algorithms have all been published previously. A separate fusion algorithm, also previously published,23 combines these analyses of individual lesions and structures, as well as the image quality. The final output of the IDP is the DR index, a dimensionless number be-

<table>
<thead>
<tr>
<th>Measure</th>
<th>Score</th>
<th>Observable Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICDR severity level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No apparent retinopathy</td>
<td>0</td>
<td>No abnormalities</td>
</tr>
<tr>
<td>Mild nonproliferative diabetic retinopathy</td>
<td>1</td>
<td>Microaneurysms only</td>
</tr>
<tr>
<td>Moderate nonproliferative diabetic retinopathy</td>
<td>2</td>
<td>More than just microaneurysms but less than severe nonproliferative diabetic retinopathy</td>
</tr>
<tr>
<td>Severe nonproliferative diabetic retinopathy</td>
<td>3</td>
<td>Any of the following: &gt;20 intraretinal hemorrhages in each of 4 quadrants, definite venous beading in ≥2 quadrants, prominent intraretinal microvascular abnormalities in ≥1 quadrant, or no signs of proliferative retinopathy</td>
</tr>
<tr>
<td>Proliferative diabetic retinopathy</td>
<td>4</td>
<td>One or more of the following: neovascularization and/or vitreous or preretinal hemorrhage</td>
</tr>
<tr>
<td>Macular edema severity level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macular edema</td>
<td>0</td>
<td>Exudates or apparent thickening within 1 disc diameter from the fovea</td>
</tr>
<tr>
<td>No macular edema</td>
<td>1</td>
<td>No exudates and no apparent thickening within 1 disc diameter from the fovea</td>
</tr>
</tbody>
</table>

Abbreviations: ETDRS, Early Treatment Diabetic Retinopathy Study; ICDR, International Clinical Diabetic Retinopathy.
The study included 874 participants and 1748 images. The mean (SD) age of the patients was 57.6 (15.9) years, and 57.6% were male. No adverse events occurred. The 3 retinal experts graded all images from all participants. Of the 874 participants, 190 had RDR in the adjudicated reference standard, so the prevalence of RDR was 21.7% (95% CI, 19.0%-24.5%). The exact distribution of ICDR severity levels is given in Table 2.

### Table 2. Distribution of ICDR and Macular Edema Severity Levels in the Studied Population for Both Retinopathy and Diabetic Macular Edema

<table>
<thead>
<tr>
<th>Measure</th>
<th>Participants, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Retinal Specialist 1</td>
</tr>
<tr>
<td>ICDR severity level</td>
<td></td>
</tr>
<tr>
<td>No apparent retinopathy</td>
<td>78</td>
</tr>
<tr>
<td>Mild nonproliferative diabetic retinopathy</td>
<td>3</td>
</tr>
<tr>
<td>Moderate nonproliferative diabetic retinopathy</td>
<td>15</td>
</tr>
<tr>
<td>Severe nonproliferative diabetic retinopathy</td>
<td>3</td>
</tr>
<tr>
<td>Proliferative diabetic retinopathy</td>
<td>1</td>
</tr>
<tr>
<td>Macular edema severity level</td>
<td></td>
</tr>
<tr>
<td>No macular edema</td>
<td>89</td>
</tr>
<tr>
<td>Macular edema</td>
<td>11</td>
</tr>
</tbody>
</table>

Abbreviation: ICDR, International Clinical Diabetic Retinopathy.

Between 0 and 1, the DR index expresses the likelihood that the patient’s images will show RDR. The IDP calculates the DR index for a single individual (2 images) in less than 25 seconds on a computer equipped with a 2-core Intel i3 processor (Intel Corporation).

### DETECTION PROGRAM SET POINT

A previous version of the IDP was evaluated in a primary care setting, with 2 images per eye.24-25 We chose the set point for the present study as equal to the IDP set point in a previous study that resulted in a sensitivity for detecting RDR in that population of 91.6% before the IDP was tested on any image of the current study population.

### STATISTICAL ANALYSIS OF DETECTION PROGRAM PERFORMANCE

Analyses were conducted using SAS statistical software, version 9.2 (SAS Institute, Inc.). The IDP sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and CIs were calculated at the prefixed set point. Interobserver variability of the 3 experts was calculated with the $k$ statistic. Prevalence of RDR in the data set was determined from the adjudicated ICDR reference standard.

We calculated the receiver operating characteristic curve for all possible set points between 0 and 1. The area under the receiver operating characteristic curve (AUC) was determined using logistic regression (PROC LOGISTIC) and modeling the adjudicated reference standard as a function of the detection program or the theoretical maximum AUC given the characteristics of the 3 readers (captured by the average $k$) and the prevalence of disease in the population.7

### PRIMARY OUTCOME MEASURE

The primary outcome measure was IDP sensitivity and specificity for detecting RDR as measured against the adjudicated reference standard. The set point of the IDP was fixed at 0.101.

### RESULTS

The study included 874 participants and 1748 images. The mean (SD) age of the patients was 57.6 (15.9) years, and 57.6% were male. No adverse events occurred. The 3 retinal experts graded all images from all participants. Of the 874 participants, 190 had RDR in the adjudicated reference standard, so the prevalence of RDR was 21.7% (95% CI, 19.0%-24.5%). The exact distribution of ICDR severity levels is given in Table 2. The $k$ values were 0.85 (95% CI, 0.81-0.89) for experts 1 vs 2, 0.82 (95% CI, 0.78-0.87) for experts 1 vs 3, and 0.79 (95% CI, 0.75-0.84) for experts 2 vs 3, and the mean $k$ value was 0.822.

Sensitivity of the IDP to detect RDR was 96.8% (95% CI, 94.4%-99.3%), and specificity was 59.4% (95% CI, 55.7%-63.0%) at the preselected set point (Figure 1).
The PPV was 39.8% (95% CI, 35.2%-44.3%), whereas the NPV was 98.5% (95% CI, 97.4%-99.7%). Of the 874 participants, 184 had true-positive, 6 had false-negative, 406 had true-negative, and 278 had false-positive results. The corresponding false-negative rate was 0.00687. No participant with a false-negative result was assigned an ICDR severity level of severe NPDR or PDR (grade 3 or 4) or apparent diabetic macular edema (DME) by any retinal expert. Figure 2 shows the images of all 6 participants who had false-negative results (ie, were estimated by the IDP to not have RDR, whereas the consensus of the experts was that they had RDR).

The AUC (C statistic) against the consensus reference standard was 0.937 (95% CI, 0.916-0.959) (Figure 3). The AUC against the voted referenced standard was 0.935 (95% CI, 0.913-0.957), a nonsignificant difference. The estimated sensitivity and specificity of the 3 experts are given in Table 3; their sensitivity ranged from 71.4% to 91.0%. The theoretical maximum AUC measurable was 0.956, which was within the 95% CI of the AUC (95% CI, 0.913-0.957) that was measured.

The results reveal that the IDP has high sensitivity and specificity to differentiate patients with RDR from those without RDR compared with a consensus of 3 retinal specialists. On a practical level, the goal of a DR screening program is to identify those who need referral to an ophthalmologist for possible treatment. Individuals with no NPDR or mild NPDR and no ME have insufficient disease to require treatment and a low risk of advancing to treatment criteria within 1 year.28 The authors of the ICDR classification also stated that “the risk of significant progression over several years is very low in both groups,”13(p1679) referring to those who had no NPDR or mild NPDR without ME. There is excellent evidence, therefore, that people with these levels of DR should be screened again in 1 year. Ordinarily, a specificity of 50% would not be considered acceptable; however, in this situation, where all diabetic patients are urged to receive annual ophthalmologic screening, primarily for detection of vision-threatening retinopathy, this is a significant advance. Because the sensitivity is as high as for ophthalm-
mic screening, the number of patients requiring (and referred for) the expensive time of an ophthalmologist is reduced by more than half, with no diminution in the number of cases requiring attention the ophthalmologist would be required to see. Indeed, if all who screened by this system actually went to an ophthalmologist, more patients requiring intervention would be seen than the mean of 67% who now present for routine screening (only a fraction of whom actually need to be seen).

A major concern of automated computer detection programs is their potential to delay diagnosis of a treatable condition. Most of those with moderate NPDR do not progress to high-risk PDR within 1 year.29 The IDP missed 6 people with RDR at the set point of 0.101; all had moderate NPDR with no ME (Figure 2). Despite these false-negative results, IDP sensitivity exceeded the estimated sensitivity of any individual retinal expert: each of the retinal experts had a comparable or larger number of false-negative results. As explained in the Methods section, the IDP set point was determined before evaluating the first photograph with the expectation that it would result in a high sensitivity for the present study. Most of the disagreements among the experts that required adjudication, before consensus was reached, were around mild and moderate NPDR. This finding is easy to understand because, for instance, an image with only a few microaneurysms is rated as mild NPDR according to the ICDR, but if an expert thought that one of these microaneurysms was actually a small hemorrhage, he or she would grade that as moderate NPDR, again following the ICDR.

Authors have argued that the sensitivity for RDR detection in a screening program is not cost-effective over 80% or even 60%.28 There is some rationale to such a low sensitivity because the median sensitivity of the 3 retinal specialists in the present study was found to be 81%, whereas their median specificity was high at 98%. Most of the IDP false-positive results were in people with moderate NPDR, which corresponds to ETDRS stages 35 through 47. All of the IDP false-negative results had stage 35 NPDR, which has only a 4.2% chance of progressing to PDR in 1 year and a 1.2% chance of progressing to PDR with high-risk characteristics, indicating the need for immediate treatment.29 People with the most severe moderate NPDR have stage 47 disease, which has a 18.2% risk of developing any PDR at 1 year and only a 8.5% risk of PDR with high-risk characteristics.29 A case could be made to also rescreen patients with NPDR in 1 year, warning them to report promptly if they have vision loss or floaters, possibly indicating the development of ME or vitreous hemorrhage from proliferative disease.

The IDP set point, unlike a human expert, can be set at any value between 0 and 1. At the set point of 0.151, the IDP sensitivity is 94%, still higher than the median estimated sensitivity of the 3 retinal specialists, and the IDP specificity is then 74%. At this set point, 12 people with RDR would be missed, and only 176 people without RR would be referred unnecessarily. A review of the images of the 12 people with missed RDR revealed that all of them had only moderate NPDR without ME. An IDP set point higher than the predetermined 0.101 decreases the sensitivity at detecting RDR but increases specificity, which is also clear from the receiver operating characteristic analysis and Figure 1. A higher specificity reduces the number of false-positive results, meaning that fewer people with diabetes will be referred who do not need to be referred. In any screening program, high sensitivity is a patient safety issue, whereas high specificity is an efficiency issue that has the potential to increase productivity. A higher set point of 0.151 would not result in missing any person who was likely to need treatment within 1 year. Patients or institutions, including insurance companies or governments, that pay for health care may opt for this higher set point, knowing that there is a low risk of missing patients who need immediate treatment and a cost savings by reducing the number of people who are referred unnecessarily.

[Table 3. Sensitivity and Specificity (95% CIs) of Experts to Detect Referable Retinopathy Based on ICDR Severity Level Ratings by the 2 Remaining Experts]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts 1-3</td>
<td>0.803 (0.751–0.855)</td>
<td>0.983 (0.973–0.993)</td>
</tr>
<tr>
<td>Experts 2-3</td>
<td>0.714 (0.657–0.772)</td>
<td>0.995 (0.990–1.000)</td>
</tr>
<tr>
<td>Experts 1-2</td>
<td>0.910 (0.870–0.950)</td>
<td>0.950 (0.933–0.966)</td>
</tr>
</tbody>
</table>

Abbreviations: ICDR, International Clinical Diabetic Retinopathy; ME, macular edema.
Retinal fundus imaging with a nonmydriatic camera typically takes 10 minutes. The IDP outputs the DR index in less than 1 minute on standard hardware. This quick result can be provided at the point of care and, for most people, will eliminate the need for a separate visit to an ophthalmologist. The goal is to increase the number of people who are screened yearly, increase patient satisfaction, and reduce cost.

This study has potential limitations. The ICDR classification does not define an imaging protocol but refers to the ETDRS, which was based on 7 stereo photographic fields rather than the single field that was used in the present study. The genesis of the ICDR was because the process of taking and then grading images from 7 stereo fields was found to be too cumbersome to be widely used in clinical practice. Previous studies have found that reading a single field of the central retina is as good at detecting DR as reading multiple fields or a dilated retinal examination.

Another limitation is that the retinal images were not stereo photographs. The IDP detects microaneurysms, hemorrhages, and exudates in the macula, which are the usual components of ME. It cannot detect clear fluid that could be the cause of DME. Human expert detection of DME from exudates only in single images, however, was shown to be as sensitive (94%) as detection from clinical stereo biomicroscopic analysis of retinal thickening. The likelihood of vision-threatening ME, without the presence of IDP detectable microaneurysms, hemorrhages, or exudates, thus seems to be exceedingly rare. A further, but perhaps unnecessary, safeguard to avoid missing ME is to not screen anyone with visual loss, which is standard protocol in most screening programs.

In conclusion, the IDP performed well at differentiating between people with and without RDR. Although there were 6 false-negative results among the 874 participants, none of them had sight-threatening disease, and it seemed doubtful that any of them would need treatment within 1 year. Increasing the set point resulted in higher specificity but not in missing any person who needed immediate treatment or likely to need treatment within 1 year. Early detection with IDP seems to be at least as good at detecting RDR as a single retinal expert reader if our results are extrapolated to larger populations. These results add to the accumulating evidence that automated detection is ready for clinical use with the hope that it will increase the number of people screened, reduce vision loss caused by delayed treatment, and decrease the cost per person screened.

Submitted for Publication: September 20, 2012; final revision received October 23, 2012; accepted October 23, 2012. Correspondence: Michael D. Abràmoff, MD, PhD, Department of Ophthalmology and Visual Sciences, 11205 PFP, University of Iowa Hospital and Clinics, 200 Hawkins Dr, Iowa City, IA 52242 (michael-abramoff@uiowa.edu).

Conflict of Interest Disclosures: The corresponding author has the right to grant on behalf of all authors and does grant on behalf of all authors an exclusive license (or nonexclusive for government employees) on a worldwide basis. All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare the following: Drs Abr àmoff, Folk, Han, Walker, Williams, Russell, Massin, Cochener, Gain, Tang, Lamard, Moga, and Quellec received no support from any company for the submitted work. Drs Han, Walker, Williams, Massin, Cochener, Gain, Tang, Lamard, and Quellec have no relationships with any company that might have an interest in the submitted work in the previous 3 years. None of their spouses, partners, or children have financial relationships that may be relevant to the submitted work. Drs Abràmoff, Folk, Han, Walker, Williams, Russell, Massin, Cochener, Gain, Tang, Lamard, Moga, and Quellec have no nonfinancial interests that may be relevant to the submitted work. All authors were independent of the funders and sponsors of this study. Drs Abràmoff and Niemeijer are listed as inventors on patents and patent applications related to the study subject. Drs Abràmoff, Russell, Folk, and Niemeijer are shareholders in IDx, the company that has licensed the inventions from the University of Iowa on which the IDP was based.

Funding/Support: This study was supported by grants NEI EY017066, R01 EY018853, and R01 EY019112 from the Research to Prevent Blindness, New York, New York (University of Iowa and Medical College of Wisconsin) and grant UL1RR024979 from the National Center for Research Resources, National Institutes of Health. Dr Russell is the Dina J. Schrage Professor for Macular Degeneration Research. Dr Folk is the Judith Gardner and Donald H. Beisner, MD, Professor of Vitreoretinal Diseases and Surgery. Dr Han is the Jack A. and Elaine D. Klieger Professor of Ophthalmology.

Role of the Sponsor: No other sponsor had any role in the study design, conduct, collection, management, analysis, or interpretation of the data or preparation, review, or approval of the manuscript.

Disclaimer: Contents are solely the responsibility of the authors and do not necessarily represent the official views of the Clinical and Translational Science Award Consortium or National Institutes of Health.

Additional Contributions: Jean-Claude Klein, PhD, organized the Messidor study.

Additional Information: The images are publicly available for noncommercial use at http://latim.univ-brest.fr/index.php?option=com_content&view=article&id=61&Itemid=10034&lang=en (last accessed December 26, 2011) and from the corresponding author at michael-abramoff@uiowa.edu.

REFERENCES

19. Niemeier M, Abramoff MD, van Ginneken B. Image structure clustering for im-