Artificial Intelligence Improves Patient Follow-Up in a Diabetic Retinopathy Screening Program

Eliot R Dow, Karen M Chen, Cindy S Zhao, Austen N Knapp, Anuradha Phadke, Kirsti Weng, Diana V Do, Vinit B Mahajan, Prithvi Mruthyunjaya, Theodore Leng, David Myung

1Department of Ophthalmology, Byers Eye Institute at Stanford, Palo Alto, CA, USA; 2Department of Ophthalmology, Duke Eye Center, Durham, NC, USA; 3Department of Internal Medicine, Stanford Health Care, Palo Alto, CA, USA

Correspondence: David Myung, Byers Eye Institute at Stanford, Palo Alto, CA, 94303, USA, Tel +1 650 723-6995, Email djmyung@stanford.edu

Purpose: We examine the rate of and reasons for follow-up in an Artificial Intelligence (AI)-based workflow for diabetic retinopathy (DR) screening relative to two human-based workflows.

Patients and Methods: A DR screening program initiated September 2019 between one institution and its affiliated primary care and endocrinology clinics screened 2243 adult patients with type 1 or 2 diabetes without a diagnosis of DR in the previous year in the San Francisco Bay Area. For patients who screened positive for more-than-mild-DR (MTMDR), rates of follow-up were calculated under a store-and-forward human-based DR workflow (“Human Workflow”), an AI-based workflow involving IDx-DR (“AI Workflow”), and a two-step hybrid workflow (“AI–Human Hybrid Workflow”). The AI Workflow provided results within 48 hours, whereas the other workflows took up to 7 days. Patients were surveyed by phone about follow-up decisions.

Results: Under the AI Workflow, 279 patients screened positive for MTMDR. Of these, 69.2% followed up with an ophthalmologist within 90 days. Altogether 70.5% (N=48) of patients who followed up chose their location based on primary care referral. Among the subset of patients that were seen in person at the university eye institute under the Human Workflow and AI–Human Hybrid Workflow, 12.0% (N=14/117) and 11.7% (N=12/103) of patients with a referrable screening result followed up compared to 35.5% of patients under the AI Workflow (N=99/279; χ² df=2 = 36.70, p < 0.00000001).

Conclusion: Ophthalmology follow-up after a positive DR screening result is approximately three-fold higher under the AI Workflow than either the Human Workflow or AI–Human Hybrid Workflow. Improved follow-up behavior may be due to the decreased time to screening result.

Keywords: telemedicine, teleophthalmology, fundus photography, deep learning, machine learning, referral

Introduction

Early detection of diabetic retinopathy (DR) through regular retinal screening is critical to preventing vision loss. However, only around 50% to 60% of all diabetic patients in the United States receive the exam recommended by DR screening guidelines, and fewer than 40% of patients at high risk of vision loss receive appropriate treatment. The prevalence of DR challenges Wilson and Jungner’s principle that facilities for diagnosis and treatment should be available for screening and follow-up. During the past two decades, progress in teleophthalmology-based DR screening, consisting of remote, asynchronous interpretation of fundus photographs by experienced human readers, has offered an avenue to overcome some of these barriers. More recently, artificial intelligence (AI)-based technologies for DR screening have the potential to further address this challenge by offering expert-level diagnostic ability at a low marginal cost.

The ultimate public health impact of a screening program, though, is determined by whether patients identified as having referable disease during screening subsequently receive appropriate follow-up care. A variety of barriers may prevent follow-up of a positive DR screening result including a lack of awareness of the result, a lack of understanding about the appropriate next step, and a lack of access to an ophthalmologist. Whereas human-based store and forward teleophthalmology systems involve some degree of lag between photography and interpretation, AI-based technology can...
return a result within minutes. This proximity of the result with the timing of the screening encounter may alleviate some of the barriers to follow-up. Indeed, previous studies have suggested that AI-based screening for DR may increase the rate of adherence to recommended referral services.\textsuperscript{14,15}

To expand access to DR screening, the Stanford Teleophthalmology Automated Testing and Universal Screening (STATUS) program was developed as a collaboration between Byers Eye Institute of Stanford and affiliated primary care clinics throughout the San Francisco Bay Area. Five sites grew to seven total sites during the study period within the program. The first 18 months of the STATUS program involved a store-and-forward teleophthalmology program in which fundus photographs were interpreted asynchronously by remote retina specialists. Subsequently, the FDA-approved autonomous AI screening device for DR, IDx-DR (recently renamed LumineticsCore\textsuperscript{TM}), was integrated into the screening program.\textsuperscript{10} As detailed in our previous work, an AI–Human Hybrid Workflow was established in which fundus images that were ungradable by the AI system were evaluated in the human-based teleophthalmology workflow.\textsuperscript{16}

In this paper, we present results on patient follow-up across three years of a program for DR screening involving human teleophthalmology and AI. The goals of this study were three-fold. First, we wished to determine the proportion of patients who followed up with an ophthalmologist after a referrable result from their eye screening encounter. Second, we sought to understand the reasons that patients followed up or failed to do so after a referrable screening result. Finally, we wanted to determine if the proportion of patients who followed up was positively or negatively impacted by the AI screening system compared to a human-based teleophthalmology system. We hypothesized that the AI system would increase the rate of patient follow-up after a positive screening result.

\textbf{Methods}

\textbf{Program Overview}

The Stanford Teleophthalmology Automated Testing and Universal Screening (STATUS) program is a DR screening program that was initiated between the Byers Eye Institute at Stanford and Stanford-affiliated primary care and endocrinology clinics in the San Francisco Bay Area starting in September 2019. For the first 18 months of the program, screening was performed through store-and-forward human-based teleophthalmology workflow (“Human Workflow”). Subsequently, patient screening was performed with the IDx-DR (Digital Diagnostics Inc., Coralville, IA; recently renamed LumineticsCore\textsuperscript{TM}), an autonomous AI-based medical device approved for the evaluation of DR in primary care offices (“AI Workflow”). During this phase of the program, fundus images that received an ungradable result from the AI system were evaluated by the human-based teleophthalmology workflow (“AI–Human Hybrid Workflow”).

Regardless of workflow, all patients received their results either via a health system-based web portal, where patients can view test results, access their medical records, and conduct other health-related activities or through a phone call or letter. Patients were alerted to the new result by email. Aside from AI- or human-based DR screening and an automated notification of results sent via the patient’s web-based health portal, the STATUS program did not otherwise alter patient notification and referral placement processes already established at participating sites. In general, activities at the sites included clinic staff members notifying patients of their results by phone or email and notifying them of a referral to ophthalmology, if indicated. Specifically, any patient with more-than-mild-DR (MTMDR) or an ungradable AI–Human Hybrid Workflow screening result was called regarding their result, and in these cases the primary care physician (PCP) placed a referral to ophthalmology for in-person examination. Patients were not assumed to have an established ophthalmologist and had the option to choose any qualified ophthalmologist depending on insurance coverage, scheduling availability, location, and other matters of patient preference. After the referral was placed, patients either contacted their ophthalmologist of choice or were contacted by the accepting ophthalmologists’ office to schedule an in-person appointment. Consequently, the only prescribed differences between the Human Workflow, the AI Workflow, and the AI–Human Hybrid Workflow were the differences in the process and time to notification of the screening result due to varying image interpretation methods. In the Human Workflow and the AI–Human Hybrid Workflow, results were released within 7 days after the screening encounter when the retinal specialist had completed the remote image evaluation; the referral process described above began thereafter. In the AI Workflow, the patient received the health
portal result within 48 hours. The human work flow involved notification via CCed chart whereas the AI workflow results were sent to PCPs results inbox.

**Patient Evaluation and Image Interpretation**

Patients 18 years or older with type 1 or type 2 diabetes mellitus without a prior DR diagnosis or a DR exam in the past 12 months being seen at one of the affiliated STATUS program clinics (see Program Overview) were offered the opportunity to have DR screening performed during their primary care visit.

During the initial phase of the program under the Human Workflow, one 45-degree macula-centered image and one optic disc-centered image for each eye were obtained with a non-mydriatic fundus camera (TopCon NW400, Welch Allyn Inc., Skaneateles Falls, NY) in the absence of pharmacologic pupil dilation and saved to Picture Archiving and Communication Software (PACS). A rotating pool of fellowship-trained retina specialists at the Stanford Ophthalmic Reading Center (STARC) performed single-reader evaluation using the International Clinical Diabetic Retinopathy Severity Scale to determine the stage of DR (no DR, mild non-proliferative diabetic retinopathy (NPDR), moderate NPDR, severe NPDR, proliferative DR) and the presence or absence of diabetic macular edema (DME). Readers could also label the fundus images as ungradable if the quality was not sufficient to diagnose DR or DME. Image interpretation was performed on the native PACS software that allowed readers to manipulate image brightness and contrast (Zeiss Forum, Carl Zeiss Meditec, Dublin, CA). In this human-based teleophthalmology workflow, the time of screening to release of results took up to 7 days. During the AI Workflow, image acquisition was identical through the same TopCon NW400 camera system; the only difference was that image interpretation was performed by the IDx-DR through a secure cloud-based algorithm. Within five minutes, the device’s software displayed a screening result of MTMDR-positive, MTMDR-negative, or ungradable. Patients were informed within 48 hours of their test result as described above. Results of the AI testing were handled in the same way as lab results, with automated batch release of the results to the patients at 48 hours.

**Data Collection and Analysis**

For all patients screened in the STATUS program who received a MTMDR-positive or ungradable result during the Human Workflow, AI Workflow, or AI–Human Hybrid Workflow, the electronic medical record at Stanford Health Care was searched for subsequent encounters with an ophthalmologist. The data were cleaned and scraped using Pandas. The patient was considered to have successful follow-up if the in-person encounter with dilated fundus examination took place within 90 days of the screening encounter.

To test the hypothesis that sooner release of screening result under the AI Workflow would improve follow-up rates, we compared the rates of follow-up between the human-teleophthalmology phase (“Human Workflow”) and the AI screening phase (“AI Workflow”). Additionally, to account for any cohort differences between the two phases, as well as other effects introduced by the AI system apart from a more rapid result, we also looked at the rate of follow-up for patients who received an initial inconclusive screening result from the AI system that was then read as MTMDR in the human-teleophthalmology workflow up to 7 days later as described above (“AI–Human Hybrid Workflow”). Comparison was performed between rates of follow-up at the university health system since documentation of follow-up in the community was not available across all conditions. The chi-square test of independence was used to compare rates between these groups.

Patients who received a diagnosis of MTMDR-positive in the AI Workflow were randomly sampled and contacted by phone for a survey regarding follow-up behavior (Supplementary Material 1). Among 152 patients who were sampled and called, 93 (63.4%) were reached by phone, and all 93 provided consent to participate in the survey. Patient follow-up outside Stanford Health Care was calculated based on the self-reported behavior of the sample of 93 survey respondents. Of 93 patients who provided consent for the phone survey, 68 had followed up with an ophthalmologist, whereas 25 had not.

Approval for this study was obtained from the Stanford University Institutional Review Board in Protocol #57104. All guidelines outlined in the Declaration of Helsinki were followed, and informed consent was obtained from study participants.
Results

Under the AI Workflow, the program screened 2243 patients during the study period and identified 279 of them as positive for MTMDR. Among these 279 patients who received a screening result of MTMDR from the AI system, 69.2% followed up with an ophthalmologist within 90 days of the screening encounter, including 35.5% of patients who presented to the university eye center and 33.7% who presented to a community eye doctor (Figure 1; Supplementary Table 1). Among the patients who followed up at the university eye institute, 9.1% received treatment for diabetic eye disease at the first follow-up encounter (N=8 intravitreal anti-VEGF injection; N=1 pan-retinal photocoagulation). The screening program served racially and ethnically diverse populations (Asian 33.1%, White Non-Hispanic/Latino 25.0%, Hispanic/Latino 21.2%, Black 7.4%, Native American/Pacific Islander 2.5%, Other/Unknown 10.8%), with four self-reported demographic racial or ethnic groups representing a plurality at different clinic sites (Figure 2).

To further understand patients’ follow-up behavior, we conducted a survey by phone of patients with a MTMDR-positive screening result from the AI system. Of 93 patients who provided consent for the phone survey, 68 had followed up with an ophthalmologist, whereas 25 had not. For patients who followed up with an ophthalmologist after a referable screening result, the most frequent reason was that they were referred there by their PCP, which was given 70.5% of the time (N=48/68). Other reasons included that the location had been recommended as reputable and that it was in their insurance network (Figure 3A). Patients who did not follow-up after a MTMDR diagnosis most commonly conveyed a knowledge gap about the recommended follow-up (despite being contacted by their care team about the referral for an in-person appointment with an ophthalmologist), which was stated 52% of the time (N=13/25). Other reasons for failure to follow up included a language barrier or difficulty with translation, issues with scheduling, feeling that follow-up was not a priority in their lives at that time, and having fear of an unwanted diagnosis (Figure 3B). For patients who did not follow up after their encounter, 76.0% (N=19/25) of respondents said that they did not have an established eye doctor. A variety of other demographic and health factors showed no statistically significant difference between the patients who followed up or did not follow up including gender, age, glucose levels, body-mass index, or smoking status (Supplementary Tables 2 and 3). We also investigated differences in socioeconomic status proxies such as zip code and home value estimates and found no statistically significant difference between groups.

Among the patients that followed up in-person at the university eye institute, we found significant differences in follow-up patterns between the three workflows. Under the Human Workflow, 12.0% of patients with a screening result warranting referral to an ophthalmologist followed up for an eye exam (Referrable, N=14/117; MTMDR-positive result, N=8/40; ungradable result, N=6/77). In comparison, after a MTMDR-positive result in the AI Workflow 35.5% of

![Figure 1](https://doi.org/10.2147/OPTH.S422513) A flowchart representing patients involved in the study. “Human Workflow” indicates a store-and-forward human-based teleophthalmology system in which fundus images were interpreted by a retina specialist one to five days after acquisition. “AI Workflow” indicates assessment by the AI system without human teleophthalmology involvement. “AI-Human Hybrid Workflow” indicates fundus images that received an ungradable result from the AI system and were subsequently evaluated by the human-based teleophthalmology system, producing a result in one to five days.
patients followed up for an eye exam at the university eye institute (N=99/279). Finally, for patients under the AI–Human Hybrid Workflow, 11.7% of patients followed up at the university eye institute (Referrable, N=12/103; MTMDR-positive result, N=1/20; ungradable result, N=11/83) (Figure 1). This represented a three-fold increase in follow-up with an ophthalmologist under the AI Workflow compared to either the Human Workflow or AI–Human Hybrid Workflow (AI vs Human Workflow $\chi^2_{df=1} = 21.22, p < 0.000004$; AI vs AI–Human Hybrid $\chi^2_{df=1} = 19.59, p < 0.000001$; Human vs AI–Human Hybrid $\chi^2_{df=1} = 0, p = 1$).

**Discussion**
As rates of diabetes mellitus steadily increase both in the US and around the world, a public health crisis of DR looms. AI-based diabetic eye screening may help to meet the rising volume of patients. However, the number of patients screened is only an initial target in mitigating DR; the ultimate goal of DR screening is for patients identified as positive...
for MTMDR to receive the necessary subspeciality-specific monitoring and treatment. To this end, although teleophthalmology programs have expanded access to screening, many patients who screen positive do not follow up for eye care after referral.\textsuperscript{18} Whereas human-based teleophthalmology workflows typically have a delay between screening encounter and receipt of screening result, AI algorithms can return a result rapidly, and has the potential to do so at the point of care. This study assessed differences in follow-up rate between AI and Human Workflows in our system. The only prescribed differences between AI and Human Workflow were the method of image analysis, which altered time to result (within 48 hours hours for AI Workflow, within 7 days for Human Workflow) but was unknown to the patient as well as the process of referring provider notification. There were no baseline differences between study populations. Hence, we inferred that the differences in follow-up rate may have been driven by differences in time and process by which the result was released to the referring provider and, in turn, to the patient. Further comparison with the AI–Human Hybrid Workflow, which also took up to 7 days and would isolate other differences such as the Hawthorne Effect of the AI Workflow, found a similar follow-up rate between the AI–Human Hybrid and the Human Workflow. Overall, DR screening under the AI Workflow with shorter time to result had a three-fold increase in follow-up with an ophthalmologist compared to the workflows involving human intervention. Of note, the reason that the current AI Workflow takes up to 48 hours is that the results are handled in the same way that lab results are in our system, which are scheduled for automated release to the patient after 48 hours.

This work has several strengths that complement previous studies. First, the study takes place in an ethnically and racially diverse United States population. Whereas previous studies have investigated referral behavior in a Rwandan population or a predominantly African American population, our cohort is composed of large numbers of self-reported Asian, Hispanic and Latino, Black, White non-Hispanic/Latino, and Native American or Pacific Islander patients.\textsuperscript{14,15} We further investigated differences between demographics and socioeconomic status proxies such as zip code and home value estimates and found no difference to suggest that differences in follow-up rates were driven by demographic or socioeconomic status (Supplementary Tables 2 and 3). Second, our study was conducted with patients screened with IDx-DR, the first approved AI-based medical device for DR evaluation. We compared rates of follow-up in the AI Workflow, which screened with IDx-DR, to two comparison groups: a human-based teleophthalmology workflow (“Human Workflow”) and an AI–Human Hybrid Workflow. The comparison to the AI–Human Hybrid Workflow is important to minimize confounding effects that could be present by only comparing to a historical cohort. Moreover, our analysis of patients who were initially tested by AI and then screened by a human provider served as a control for changes attributed to patient behavior factors other than those related to the time interval to an image interpretation. The study reports both relative rate of follow-up as well as an overall estimate of follow-up after AI screening. Finally, our analysis of follow-up rates is complemented by survey results by the patients themselves that support the principle of why an AI system may enhance referral compliance.

Patients who failed to follow up with an ophthalmologist most often reported a knowledge gap about the next step in management, despite being informed by their care team about the result and that a referral was placed. Although clinic staff ensured that each patient was contacted with their screening result and had a referral to ophthalmology placed, a variety of communication, educational, and/or technological barriers may have stood in the way. Patients also reported language barriers to scheduling follow-up, that addressing the diagnosis was a low priority, and being fearful of an unwanted diagnosis of DR. These barriers could potentially be addressed through clinic workflow changes. For language barriers specifically, our health system lists patients’ preferred language, and schedulers are asked to contact patients in their preferred language using phone-based interpreter services. However, it is unknown whether these services are available in all community ophthalmology clinics which served patients in the study. Opportunities for improvement in follow-up include an automated process for generating referrals in the patient’s native language. Similarly, the top reason for selecting the location of follow-up among patients who did follow up after a positive screening result was that they were referred by their primary care doctor, highlighting the importance of the primary care encounter in directing health behaviors.

The study has several limitations. First, the descriptive and retrospective analyses leave open the potential for bias not addressed by randomization. Nevertheless, the results establish a foundation for future randomized controlled trials, particularly in a US population to complement previous international studies.\textsuperscript{14} A second limitation is that the overall rate of follow-up in this study was calculated from a sample of MTMDR patients. Patients who were not included in the sample or who did not answer the phone may have differed from those from whom the survey results were collected.
Eliciting responses from a greater proportion of patients would strengthen the conclusions drawn from the data. Nevertheless, the survey results would not affect the relative rates of follow-up calculated between the AI Workflow and other workflows. Finally, while surveyed patients did not directly report that the time to screening result impacted their follow-up behavior, they may have been anchored by the provided list of response options or may not have consciously recognized the impact of a delayed result on their own health care decision-making. To adhere to best practices for conducting surveys, phone calls were performed with a standardized flow of questions (Supplementary Material 1). The survey in this study was conducted over a year after patients in the Human Workflow had been screened, so patient recall of behavior and reasoning was likely not sufficient for explicit qualitative investigation of relationship between time to result and choice of follow-up. However, such questions could be explored in future studies, and there is reason to believe that the qualitative responses provided in the surveys here still give insight into the general factors that influenced patients’ health care decision-making.

In conclusion, this study reports that more than two-thirds of patients who received a positive result for MTMDR by the AI Workflow followed up with an ophthalmologist within 90 days. The rate of follow-up at the university eye clinic was three times greater in the AI Workflow compared to patients who received their screening result up to 1 week later under a human-based teleophthalmology workflow or AI–Human Hybrid Workflow. Survey results suggested several factors influencing patient follow-up behavior that may be addressed through modifications in post-visit messaging at or soon after the point of care between the patient and their provider, including clear communication of results, emphasis on the importance of ophthalmology follow-up, availability of results and further instructions in multiple languages, and counseling that treatment options may be available if follow-up is completed in a timely manner. The research was conducted in a diverse United States population across multiple primary care clinics using an FDA-approved medical device. In addition to expanding access to DR screening and lowering costs, AI-based devices may also improve eye care by synchronizing testing results with the patient-provider encounter.

**Funding**


**Disclosure**

This research was funded in part by departmental core grants from Research to Prevent Blindness and the National Institutes of Health (NIH P30 EY026877) as well as the Stanford Diabetes Research Center (SDRC). Authors with financial interests or relationships to disclose are listed after the references. Dr Diana V Do reports grants, personal fees from Regeneron, personal fees from Apellis, personal fees from Kodiak Sciences, personal fees from Iveric Bio, during the conduct of the study. Dr Vinit B Mahajan reports other from Digital Diagnostics, LLC, during the conduct of the study; Dr Theodore Leng reports grants, personal fees from Astellas, personal fees from Roche/Genentech, personal fees from Protagonist, personal fees from Nanoscope, personal fees from Alcon, personal fees from Apellis, personal fees from Regeneron, personal fees from Graybug, personal fees from Kanaph, personal fees from Boehringer Ingelheim, outside the submitted work. None of the remaining authors have any conflicts of interest to disclose.

**References**


Clinical Ophthalmology

**Publish your work in this journal**

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on PubMed Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/clinical-ophthalmology-journal

3330  Facebook  Twitter  LinkedIn  DovePress

Clinical Ophthalmology 2023:17