Supplementary Online Content

Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referable and vision-threatening diabetic retinopathy. *JAMA Netw Open*. 2021;4(11):e2134254. doi:10.1001/jamanetworkopen.2021.34254

eAppendix. FDA-Specified Analysis of the Subgroup of Participants Specified by the FDA to Support Clearance

eFigure 1. Participant Disposition and Cohorts Used for FDA-Specified Analyses to Support FDA Clearance

eTable 1. Baseline Characteristics From the FDA-Specified Population and Original Population

eTable 2. EyeArt Performance From the FDA-Specified Analysis of the Sequentially Enrolled Cohort

eTable 3. EyeArt Performance From the FDA-Specified Analysis of the Enrichment-Permitted Cohort

eTable 4. EyeArt Imageability From the FDA-Specified Analysis of Participants Enrolled at Primary Care Centers

eTable 5. EyeArt Imageability From the FDA-Specified Analysis of Participants Enrolled at Eye Care Centers

eTable 6. EyeArt Performance by Study Center Type in the Per-Protocol Population

eTable 7. Best- and Worst-Case Imputation Analyses in the Per-Protocol Population

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. FDA-Specified Analysis of the Subgroup of Participants Specified by the FDA to Support Clearance

Methods

A total of 942 participants were consented of which 915 participants met the per protocol study eligibility criteria. To better align the analysis population to establish substantial equivalence in support of FDA clearance, FDA-specified analyses were performed on a subset of participants specified by the FDA. After excluding those participants who did not meet these additional criteria (eg, participants 21 years or younger), 655 participants were included for analysis **(eFigure 1)**. These 655 participants were enrolled in the study at 11 US study sites, including both primary care centers and general eye care centers.

Overall methods and outcome measures in this FDA-specified analysis were the same as the per protocol analyses in comparing EyeArt performance in detecting more-than-mild diabetic retinopathy (mtmDR) and vision-threatening diabetic retinopathy (vtDR) from 2-field fundus images with the clinical reference standard of Early Treatment Diabetic Retinopathy Study (ETDRS) grading of 4-wide field stereoscopic dilated fundus photographs (equivalent to 7-field ETDRS photographs) by the Wisconsin Fundus Photograph Reading Center. In the FDA-specified analyses, participants were further divided into 2 cohorts for analysis based on type of enrollment (sequential vs enrichment). During the sequential enrollment period, 235 participants were enrolled, with 45 at primary care facilities and 190 at eye care sites. During the enrichment period, 420 participants were enrolled, with 335 from primary care sites and 85 from eye care sites. Baseline characteristics of these participants are available in eTable 1.

Results

In eTable 2A, the key performance measures are summarized for the sequentially enrolled cohorts at primary care and ophthalmology sites. All the false-negative eyes for the EyeArt mtmDR output for this cohort were ETDRS levels 35 only. In other words, all eyes with ETDRS level 43 or higher were correctly identified as mtmDR positive. In eTable 2B, the key performance measures are summarized for the enrichment-permitted cohorts at primary care and eye care sites. All the false-negative eyes for the EyeArt mtmDR output for this cohort were also ETDRS levels 35 only. In other words, all eyes with ETDRS level 43 or higher were correctly identified as mtmDR positive. The 2 eyes at primary care sites that were false negative for vtDR had a positive EyeArt mtmDR result and would have been identified for referral, nonetheless.

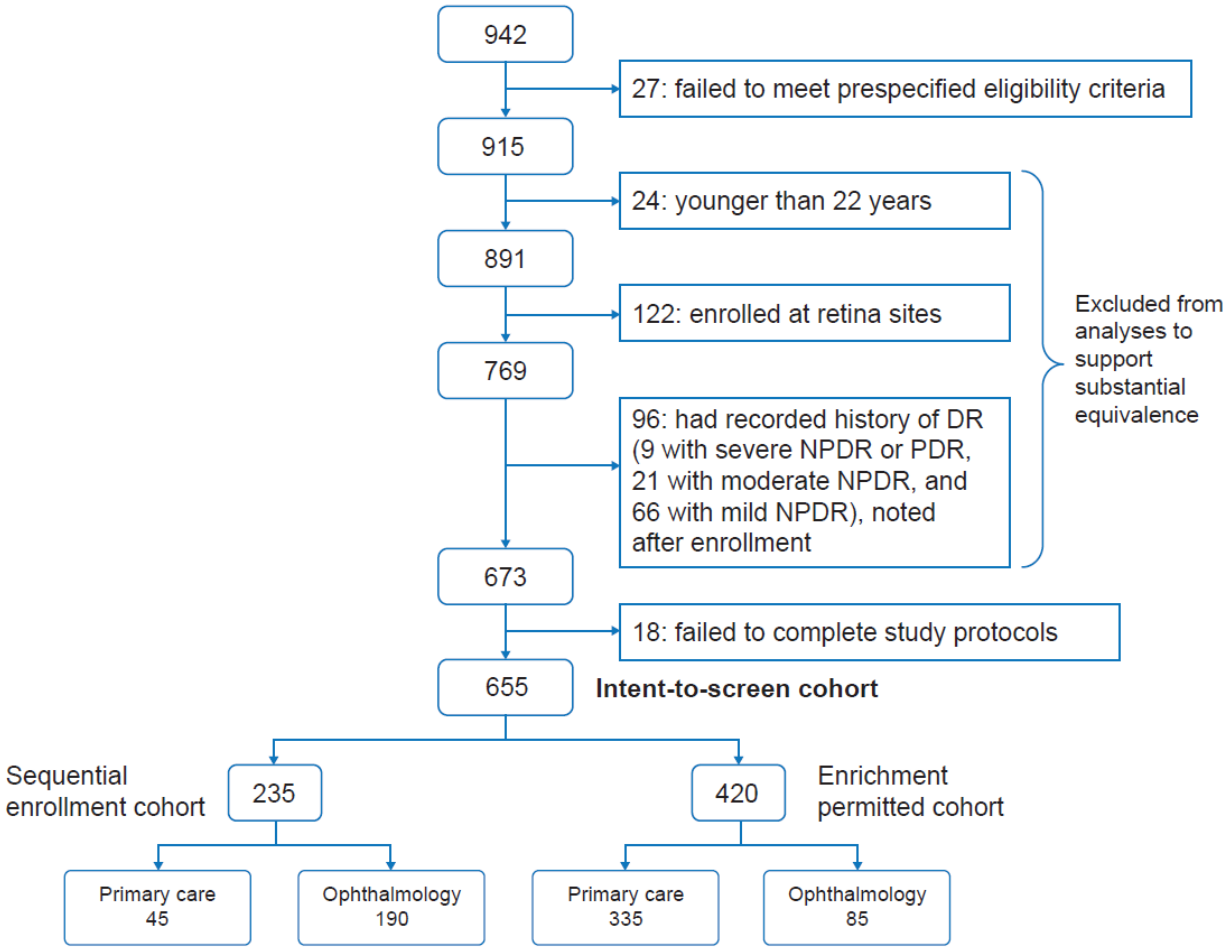
Of the participant eyes that received a completed FPRC grading, a large portion of participant eyes obtained an EyeArt disease result: 96.8% at primary care sites (eTable 3A) where most operators did not have any prior retinal imaging experience and 98.0% at ophthalmology sites (eTable 3B). As seen in these tables, imageability for the dilate-if-needed protocol is 96.5% or greater (ie, a large portion of participant eyes that received a completed FPRC grading obtained EyeArt disease results). Moreover, without dilation, the imageability is 89% or greater for primary care centers and 81.7% or greater for ophthalmology centers, demonstrating that only a small fraction of participants required dilation to obtain EyeArt disease results with the dilate-if-needed imaging protocol. The gradability (defined as the portion of participant eyes that obtained EyeArt disease results) without dilation is also high.

Discussion

The EyeArt system's sensitivity and specificity (ie, diagnostic accuracy) for detecting mtmDR and vtDR are high and comparable to the results in the per protocol analyses. Slight differences in performance are noted between sequential and enrichment-permitted enrollment periods, which is expected because enrichment can cause a shift in disease prevalence of the enrolled population. Sequential enrollment provides the most unbiased estimate of the device performance. We observe that in the enrichment-permitted period there are a larger number of participants who are in the borderline disease levels (having mild or moderate level of nonproliferative DR [NPDR]). The increased number of borderline cases in the enrichment-permitted cohort may be the reason for this slight performance difference due to spectrum effect. For example, for mtmDR determination mild NPDR or moderate NPDR cases could be considered borderline.

In conclusion, the FDA-specified analyses demonstrated that the EyeArt system achieved high performance when compared to the reference standard as determined by standardized, independent reading center grading of dilated 4-wide field stereo images for the autonomous detection of mtmDR and vtDR in people with diabetes but no history of mtmDR.

eFigure 1. Participant Disposition and Cohorts Used for FDA-Specified Analyses to Support FDA Clearance. Numbers indicate participant eyes. DR, diabetic retinopathy; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.



eTable 1. Baseline Characteristics From the FDA-Specified Population and Original Population

Characteristics		Sequential enrollment cohort		Enrichment permitted cohort		Overall cohort (FDA specified)	Overall cohort (per protocol)
		Primary care (n = 45 participants)	Eye care (n = 190 participants)	Primary care (n = 335 participants)	Eye care (n = 85 participants)	(N = 655 participants)	(N = 893 participants)
Age, years	Mean (SD)	51.9 (10.0)	60.5 (11.0)	51.5 (16.1)	60.0 (10.3)	55.2 (14.4)	53.9 (15.2)
	Median	52	61.5	54	60	57	56
HbA1c	Mean (SD)	9.2 (2.2)	7.0 (1.5)	7.8 (1.7)	7.5 (1.6)	7.7 (1.8)	7.9 (1.8)
	Median	9.1	6.8	7.5	7.2	7.3	7.4
Diabetes duration, years	Mean (SD)	7.9 (7.6)	11.3 (10.4)	14.4 (10.9)	15.7 (8.9)	13.2 (10.5)	13.5 (10.5)
	Median	5	8	13	13	11	11
Female, count (%)		20 (44.4%)	117 (61.6%)	149 (44.5%)	60 (70.6%)	346 (52.8%)	444 (49.7%)
Race, count (%)	American Indian or Alaska Native	0 (0.0%)	0 (0.0%)	1 (0.3%)	2 (2.4%)	3 (0.5%)	3 (0.3%)
	Asian	4 (8.9%)	5 (2.6%)	4 (1.2%)	0 (0.0%)	13 (2.0%)	22 (2.5%)
	Black or African American	2 (4.4%)	38 (20.0%)	50 (14.9%)	12 (14.1%)	102 (15.6%)	159 (17.8%)
	Native Hawaiian or Pacific Islander	0 (0.0%)	1 (0.5%)	0 (0.0%)	0 (0.0%)	1 (0.2%)	4 (0.4%)
	Other	0 (0.0%)	12 (6.3%)	5 (1.5%)	7 (8.2%)	24 (3.7%)	50 (5.6%)
	White	39 (86.7%)	134 (70.5%)	275 (82.1%)	64 (75.3%)	512 (78.2%)	655 (73.3%)
Ethnicity, count (%)	Non-Hispanic/Latino	8 (17.8%)	152 (80.0%)	290 (86.6%)	67 (78.8%)	517 (78.9%)	695 (77.8%)
	Hispanic/Latino	37 (82.2%)	38 (20.0%)	45 (13.4%)	18 (21.2%)	138 (21.1%)	198 (22.2%)
	Type 1	6 (13.3%)	7 (3.7%)	123 (36.7%)	3 (3.5%)	139 (21.2%)	206 (23.1%)

Diabetes type, count (%)	Type 2	39 (86.7%)	183 (96.3%)	212 (63.3%)	82 (96.5%)	516 (78.8%)	687 (76.9%)

eTable 2. EyeArt Performance From the FDA-Specified Analysis of the Sequentially Enrolled Cohort

	mtmDR			vtDR [95% CI]		
	Primary care (n = 90 eyes)	Eye care (n = 380 eyes)	Combined (N = 470 eyes)	Primary care (n = 90 eyes)	Eye care (n = 380 eyes)	Combined (N = 470 eyes)
Sensitivity, [95% CI]	100% [74.1%-100%][a]	94.9% [86.4%-100%]	96.0% [89.4%-100%]	100% [51.0%-100%][a]	88.9% NA	92.3% [70.0%-100%]
Specificity, [95% CI]	92.0% [85.1%-97.5%]	86.7% [82.1%-90.7%]	87.7% [83.9%-91.2%]	97.5% [93.4%-100%]	93.8% [90.4%-96.6%]	94.4% [91.7%-97.0%]
PPV, [95% CI]	64.7% [40.0%-85.7%]	46.2% [32.2%-59.0%]	49.5% [36.5%-61.0%]	66.7% NA	26.7% [11.1%-44.4%]	33.3% [15.4%-52.0%]
NPV, [95% CI]	100% [94.7%-100%][a]	99.3% [98.2%-100%]	99.4% [98.6%-100%]	100% [95.2%-100%][a]	99.7% [99.1%-100%]	99.8% [99.2%-100%]
Disease prevalence, [95% CI]	12.2% [4.4%-20.0%]	10.5% [6.6%-15.0%]	10.9% [7.2%-14.5%]	4.4% [0.0%-11.1%]	2.4% [1.0%-4.2%]	2.8% [1.1%-4.7%]

mtmDR, more-than-mild diabetic retinopathy; NA, not available; NPV, negative predictive value; PPV, positive predictive value; vtDR, vision-threatening diabetic retinopathy.

All 95% CIs are computed using the clustered bootstrap method that take into consideration the correlation between eyes of the same participant. “NA” indicates instances when this CI method fails due to small sample sizes.

[a] For cases with proportion of 100%, the 95% CIs using clustered bootstrap are [100%-100%], hence the Wilson method is used (however, this is not designed to consider eye correlation).

eTable 3. EyeArt Performance From the FDA-Specified Analysis of the Enrichment-Permitted Cohort

	mtmDR [95% CI]			vtDR [95% CI]		
	Primary care (n = 670 eyes)	Eye care (n = 170 eyes)	Combined (N = 840 eyes)	Primary care (n = 670 eyes)	Eye care (n = 170 eyes)	Combined (N = 840 eyes)
Sensitivity, [95% CI]	92.9% [87.1%-97.5%]	96.6% [87.5%-100%]	93.8% [89.2%-97.7%]	91.7% [80.0%-100%]	100% [51.0%-100%][a]	92.9% [82.6%-100%]
Specificity, [95% CI]	85.6% [82.2% – 89.1%]	85.2% [78.1%-91.5%]	85.5% [82.2%-88.5%]	92.2% [89.6%-94.6%]	89.8% [83.9% – 95.4%]	91.7% [89.3%-93.9%]
PPV, [95% CI]	54.4% [45.3%-63.6%]	58.3% [40.3%-74.5%]	55.3% [47.5% – 63.2%]	31.9% [19.7% – 44.4%]	20.0% [0.0% – 42.1%]	29.2% [18.5% – 40.2%]
NPV, [95% CI]	98.5% [97.3%-99.5%]	99.1% [97.2%-100%]	98.6% [97.6%-99.5%]	99.6% [99.1%-100%]	100% [97.3%-100%][a]	99.7% [99.3%-100%]
Disease prevalence, [95% CI]	15.5% [12.1%-19.3%]	19.4% [11.8%-27.6%]	16.3% [13.1%-19.6%]	4.2% [2.4%-6.3%]	2.4% [0.0%-5.9%]	3.8% [2.3%-5.5%]

mtmDR, more-than-mild diabetic retinopathy; NPV, negative predictive value; PPV, positive predictive value; vtDR, vision-threatening diabetic retinopathy.

All 95% CIs are computed using the clustered bootstrap method that take into consideration the correlation between eyes of the same participant.

[a] For cases with proportion of 100%, the 95% CIs using clustered bootstrap are [100%-100%], hence the Wilson method is used (however, this is not designed to consider eye correlation).

eTable 4. EyeArt Imageability From the FDA-Specified Analysis of Participants Enrolled at Primary Care Centers

	Primary care centers					
	mtmDR			vtDR		
	Sequentially enrolled (n = 90 eyes)	Enrichment permitted (n = 670 eyes)	Combined (N = 760 eyes)	Sequentially enrolled (n = 90 eyes)	Enrichment permitted (n = 670 eyes)	Combined (N = 760 eyes)
Imageability (dilate-if-needed) [95% CI]	96.6% [90.9%-100%]	96.8% [94.8%-98.5%]	96.8% [94.9%-98.4%]	96.5% [90.6%-100%]	96.7% [94.8%-98.5%]	96.7% [94.8%-98.4%]
Imageability (undilated) [95% CI]	94.4% [86.7%-100%]	89.0% [85.5%-92.3%]	89.6% [86.6%-92.6%]	94.2% [86.4%-100%]	89.3% [85.9%-92.5%]	89.9% [86.8%-92.7%]

mtmDR, more-than-mild diabetic retinopathy; vtDR, vision-threatening diabetic retinopathy.

All 95% CIs are computed using the clustered bootstrap method that take into consideration the correlation between eyes of the same participant.

eTable 5. EyeArt Imageability From the FDA-Specified Analysis of Participants Enrolled at Eye Care Centers

	Eye care centers					
	mtmDR			vtDR		
	Sequentially enrolled (n = 380 eyes)	Enrichment permitted (n = 170 eyes)	Combined (N = 550 eyes)	Sequentially enrolled (n = 380 eyes)	Enrichment permitted (n = 170 eyes)	Combined (N = 550 eyes)
Imageability (dilate-if-needed), [95% CI]	98.6% [97.0%-99.7%]	96.5% [91.8%-100%]	98.0% [96.3%-99.4%]	98.6% [97.0%-99.7%]	97.0% [92.9%-100%]	98.1% [96.6%-99.4%]
Imageability (undilated), [95% CI]	81.8% [76.0%-87.0%]	83.5% [75.3%-90.6%]	82.3% [77.9%-86.6%]	81.7% [75.9% – 87.0%]	84.9% [77.4%-92.2%]	82.7% [78.3%-86.9%]

mtmDR, more-than-mild diabetic retinopathy; vtDR, vision-threatening diabetic retinopathy.

All 95% CIs are computed using the clustered bootstrap method that take into consideration the correlation between eyes of the same participant.

eTable 6. EyeArt Performance by Study Center Type in the Per-Protocol Population

	mtmDR		vtDR	
	Primary care	Eye care	Primary care	Eye care
Sensitivity, [95% CI]	94.8% [91.0%-98.7%]	96.6% [92.1%-100%]	95.3% [90.7%-100%]	95.0% [83.7%-100%]
Specificity, [95% CI]	84.2% [80.9%-87.5%]	86.4% [83.4%-89.4%]	88.8% [86.1%-91.4%]	90.3% [87.5%-93.0%]
Imageability, [95% CI]	97.2% [95.8%-98.7%]	97.6% [96.2%-99.1%]	97.2% [95.7%-98.7%]	97.7% [96.3%-99.2%]

mtmDR, more-than-mild diabetic retinopathy; vtDR, vision-threatening diabetic retinopathy.

eTable 7. Best- and Worst-Case Imputation Analyses in the Per-Protocol Population

	mtmDR		vtDR	
	Worst-case imputation	Best-case imputation	Worst-case imputation	Best-case imputation
Sensitivity, [95% CI]	90.8% [86.7%-94.9%]	95.7% [92.9%-98.5%]	84.5% [74.7%-94.3%]	95.8% [91.4%-100%]
Specificity, [95% CI]	83.5% [81.2%-85.9%]	85.6% [83.4%-87.8%]	87.5% [85.5%-89.6%]	89.7% [87.8%-91.6%]

mtmDR, more-than-mild diabetic retinopathy; vtDR, vision-threatening diabetic retinopathy.