Telemedicine for detecting diabetic retinopathy: a systematic review and meta-analysis

Lili Shi,[1,2] Huiqun Wu,[1] Jiancheng Dong,[1] Kui Jiang,[1] Xiting Lu,[3] Jian Shi[4]

ABSTRACT

Objective To determine the diagnostic accuracy of telemedicine in various clinical levels of diabetic retinopathy (DR) and diabetic macular oedema (DME).

Methods PubMed, EMBASE and Cochrane databases were searched for telemedicine and DR. The methodological quality of included studies was evaluated using the Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2). Measures of sensitivity, specificity and other variables were pooled using a random effects model. Summary receiver operating characteristic curves were used to estimate overall test performance. Meta-regression and subgroup analyses were used to identify sources of heterogeneity. Publication bias was evaluated using Stata V.12.0.

Results Twenty articles involving 1960 participants were included. Pooled sensitivity of telemedicine exceeded 80% in detecting the absence of DR, low- or high-risk proliferative diabetic retinopathy (PDR), it exceeded 70% in detecting mild or moderate non-proliferative diabetic retinopathy (NPDR), DME and clinically significant macular oedema (CSME) and was 53% (95% CI 45% to 62%) in detecting severe NPDR. Pooled specificity of telemedicine exceeded 90%, except in the detection of mild NPDR which reached 89% (95% CI 88% to 91%). Diagnostic accuracy was higher with digital images obtained through mydriasis than through non-mydriasis, and was highest when a wide angle (100–200°) was used compared with a narrower angle (45–60°, 30° or 35°) in detecting the absence of DR and the presence of mild NPDR. No potential publication bias was detected.

Conclusions The diagnostic accuracy of telemedicine using digital imaging in DR is overall high. It can be used widely for DR screening. Telemedicine based on the digital imaging technique that combines mydriasis with a wide angle field (100–200°) is the best choice in detecting the absence of DR and the presence of mild NPDR.

INTRODUCTION

In 2013, 382 million people across the globe had diabetes mellitus. If no action is taken, that number will rise to 592 million by 2035.[1] Diabetic retinopathy (DR) is the most frequently occurring complication of diabetes and one of the major causes of acquired blindness in the working-age population around the world.[2] The WHO has estimated that DR accounts for approximately 15–17% of all cases of total blindness in the USA and Europe, and 7% of all cases in China and Mongolia.[3] A pooled analysis of 35 studies showed that the overall prevalence of DR of any severity is 34.6%, and that the prevalence of the sight-threatening stages typified by proliferative diabetic retinopathy (PDR) and diabetic macular oedema (DME) is 6.96% and 6.81%, respectively.[4] The task of detecting and evaluating patients for the presence and severity of DR and DME is enormous and urgent, because early detection of DR and timely treatment can reduce the risk of visual impairment.

DR can be detected using various methods. Conventionally, it is found by an ophthalmologist examining the fundus using either an indirect ophthalmoscope or a slit-lamp biomicroscope following pupillary dilation. Alternatively, photographs of the fundus may be taken through a dilated pupil and subsequently examined by an ophthalmologist. The gold standard for DR detection is the Early Treatment Diabetic Retinopathy Study (ETDRS) using 30°, seven standard field stereoscopic 35 mm colour slides (7F-ETDRS) or fundus fluorescein angiography (FFA) taken by a skilled photographer and assessed by skilled readers.[5–7] The 7F-ETDRS is time consuming and expensive and leads to poor compliance,[8] which makes it impractical for routine clinical use; the potential adverse effects associated with FFA[9] make FFA impractical for use as a screening tool. Telemedicine based on digital photographs of the fundus that have been compressed (or not), stored and forwarded to a distant ophthalmologist electronically for analysis is being used with increasing frequency to detect DR, and especially for DR screening.

In a previous review, Vaziri et al[10] evaluated the validity and feasibility of telemedicine for the diagnosis of DR using a statistical agreement value (κ statistics) as an indicator. The aim of this review is to assess the diagnostic accuracy of telemedicine in the full range of DR and DME severity compared with the current gold standard.

METHODS

Search strategy and selection criteria

We searched PubMed, Embase and Cochrane databases for relevant citations in April 2013 and updated the search in February 2014. There was no time and language restriction. We used a combination of medical subject headings and text terms to generate two subsets of citations: one for DR or macular oedema, based on a search using such terms as 'diabetic retinopathy', 'macular edema' and 'diabetic maculopathy', and the other for telemedicine based on a search using such terms as 'telemedicine', 'remote consultation', 'telehealth', 'tele', 'ehealth', 'mobile health', 'teleophthalmology', 'telescreening', 'telecommunication', 'teleretinal', 'internet', 'web', 'online', 'digital imaging', 'retinal imaging', 'grading and ETDRS' and 'screening and ETDRS'. We combined the terms to generate a subset of citations relevant to our research

question. We also examined the reference lists in all known primary and review articles to identify additional studies that were not captured by the electronic searches.

We included studies if they met the following criteria: (1) recruited subjects with either type 1 or type 2 diabetes; (2) compared telemedicine with 7F-ETDRS for DR or DME detection; and (3) provided sufficient data to calculate the diagnostic accuracy of the technique used. Studies in which telemedicine was compared with other methods of DR detection were excluded, as were studies with no description of its subjects.

Two reviewers (LS and HW) examined the electronic searches independently and obtained a full report of each citation that was likely to meet the predefined selection criteria. Disagreements were resolved by consensus and after discussion with a third reviewer (JD).

Quality assessment and data extraction
The quality of the included studies was assessed using the Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2)[11] and RevMan 5.2. QUADAS is designed to evaluate the quality and validity of the studies by assessing their potential bias and applicability in patient selection, as an index test, the reference standard used and in terms of flow and timing. We considered a study to be of good quality if it was based on prospective consecutive recruitment; it included an adequate description of the study population, the index test and the reference standard; it provided full verification of the index test using a reference standard; the index test and the reference standard were evaluated and interpreted independently; and more than 90% of the subjects were included in the final analysis.

Two reviewers (LS and HW) extracted information about the study characteristics and assessed the quality of the included studies independently. Disagreements were resolved by consensus and after discussion with a third reviewer (JD).

Data analysis
Two reviewers (LS and HW) independently extracted the raw data regarding the true and false positives and true and false

negatives of each included study. Study authors were contacted if the raw data could not be extracted or calculated. The statistical analysis was performed using Meta-Disc v1.4 software.[12] We computed the measures of test accuracy for each included study (ie, sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and diagnostic OR (DOR) with corresponding 95% CI).

Heterogeneity induced by the threshold effect in the included studies was assessed by calculating the Spearman correlation coefficient and p value between the logit of sensitivity and logit of 1–specificity. Heterogeneity induced by non-threshold effects was assessed using a χ^2 test and quantified by I^2. In general, $I^2 > 50\%$ ($p < 0.05$) indicates that heterogeneity among the included studies could have an impact on outcomes.[13] If the heterogeneity was found to be significant, we pooled the measures of test accuracy using the random effects model. In addition, the estimate of the summary receiver operating characteristics (SROC) curve was used to describe the relationship between test sensitivity and specificity across the included studies. Q* values, defined by the point at which sensitivity equalled specificity, were calculated based on the SROC curve. The area under the SROC curve (AUC) was calculated to show the probability of the correctly ranked diagnostic test values for a random pair of subjects (one with disease and one without disease). Potential heterogeneity factors were analysed using meta-regression. Subgroup analyses were carried out to identify the test-related factors or other factors responsible for heterogeneity in test accuracy. Statistical significance was defined by a p value of <0.05.

Publication bias was determined using the Deeks test.[14] Stata V.12.0 software was used to explore the potential publication bias in this review.

RESULTS
The literature search yielded 720 references. Figure 1 presents a flowchart of the study selection process including reasons for study exclusion. Twenty studies with a total of 1960 participants were included, 19 in English and one in German.

Figure 1 Flow chart of study selection.
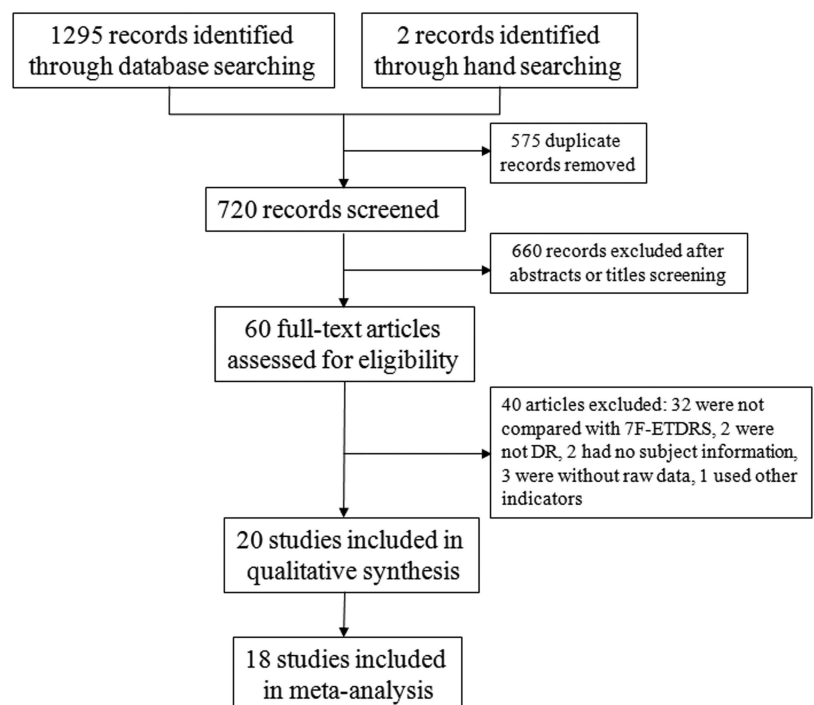
Summary characteristics of included studies

Both men and women were recruited for each study. With the exception of a study by Hubbard et al[15] which only recruited subjects with type 1 diabetes, the recruited subjects had either type 1 or type 2 diabetes. Some of the studies described how digital images were transferred. In one study they were transferred telemetrically,[16] in others via network,[17 18] and in another study they were uploaded to a secure web server.[19] The remaining studies did not include a description of the method of transmission.

In the study by Hanse et al,[20] digital imaging was carried out with mydriasis in 83 participants at the Department of Ophthalmology; an additional 59 subjects (who were demographically not statistically significantly different from the others) underwent digital imaging without mydriasis at an optician's shop. Li et al[21–25] conducted a series of studies to evaluate multiple digital photography versus 7F-ETDRS. One of these studies was designed specifically to evaluate the use of digital imaging in DME.[21] A description of the included studies is presented in table 1.

Quality of included studies

The methodological quality of the included studies is summarised in online supplementary table S1 and shown in figure 2.

The overall quality of the included studies was high. The subject selection method may have introduced high bias in two studies in which recruitment was weighted towards patients with less severe retinopathy,[20 26] and in a third study in which it was weighted towards patients with more severe retinopathy.[27] Inclusion and exclusion criteria were clearly defined in all studies and no inappropriate exclusions were noted. The method used to obtain digital images of the fundus to transmit for telemedicine were described well enough to allow the reviewer to answer the review question in all included studies. The number of subjects not included in the analysis and the reasons for their exclusion are listed in the notes in online supplementary table S1.

Heterogeneity and threshold effect analysis

The Spearman correlation coefficient and p value for various clinical levels of DR and DME (listed in online supplementary table S2) indicate that heterogeneity was not due to a threshold effect.

Accuracy of telemedicine in detecting DR

Among the 20 included studies, four studies (all conducted by Li and colleagues) used the same subjects.[22–24 28] For this reason, only one of these studies was included in the meta-analysis,[23] giving a total of 17. In seven studies[15–19 23 29]

Table 1 Characteristics of the included studies

Study ID	Sample size (patients/eyes)	Mean age/range (years)	Mean duration of diabetes/range (years)	Digital imaging technique	Image resolution (pixels)	Compression rate
Bursell et al[33]	54/108	48/20–75	17.7/3.0–42	3-field, 45°, stereoscopic, colour, non-mydriatic	640×480	10:1
Fransen et al[29]	290/549	ND/>18	Any duration	7-field, 30°, stereoscopic, colour, mydriatic	1152×1152	
Lin et al[16]	197/-	ND/>21	Any duration	1-field, 45°, monoscopic, monochromatic, non-mydriatic	640×480	
Massin et al[35]	74/147	52/25–74	8/0–23	5-field, 45°, non-stereoscopic, colour, non-mydriatic	800×600	
Boucher et al[27]	98/196	59.9/26–92	ND	2-field, 45°, stereoscopic, colour, non-mydriatic	600×800	
Hansen et al[20]	83/165	47/25–70	22/1–53	5-field, 45°, non-stereoscopic, colour, mydriatic	1450×1026	
	59/118	47/25–70	22/1–53	5-field, 45°, non-stereoscopic, colour, non-mydriatic	1450×1026	
Schiffman et al[30]	111/222	57/18–99	19/1–49	15-field, 55–60°, non-stereoscopic, monochromatic, mydriatic	ND	
Rudnisky et al[19]	102/204	ND/23–92	ND/1 months-35 years	7-field, 30°, stereoscopic (fields 1 and 2) and non-stereoscopic (fields 3 to fields 7), colour, mydriatic	3040×2008	16:1
Li et al[23]	85/152	59.4/33–83	ND	7-field, 35°, stereoscopic, colour, mydriatic	2392×2048	
Li et al[22]†	85/152	59.4/33–83	ND	7-field, 35°, monoscopic, colour, mydriatic	2400×2000	
Li et al[24]†	85/152	59.4/33–83	ND	7-field, 35°, stereoscopic, colour, mydriatic	2400×2000	37:1
Li et al[21]	85/152	59.4/33–83	ND	7-field, 35°, stereoscopic, colour, mydriatic	2400×2000	
				7-field, 35°, stereoscopic, colour, mydriatic	2400×2000	37:1
				7-field, 35°, monoscopic, colour, mydriatic	2400×2000	
				9-field, 45°, monoscopic, MosD (mosaic digital), colour, mydriatic	3000×2000	
Li et al[25]†	85/152	59.4/33–83	ND	9-field, 45°, monoscopic, MosD, colour, mydriatic	3000×2000	
Gangaputra et al[36]	96/157	62/37–86	19/ND	7-field, 30°, stereoscopic or 4-field 45–60° stereoscopic, colour, mydriatic*	ND	
Hubbard et al[15]	319/628	48/ND	27.2/ND	7-field, 30°, stereoscopic, colour, mydriatic	ND	20:1
Kernt et al[17]	34/66	62/ND	14.2/>3	1-field, 200°, non-stereoscopic, colour, non-mydriatic	ND	
Maker et al[26]	106/211	49/19–78	23.7/1–66	7-field, 30°, stereoscopic, colour, mydriatic	1280×1024	
Silva et al[31]	67/126	49/24–83	21.1/1–51	3 separate stereoscopic 45° fields, 2 30° stereoscopic fields, and an external image, colour, non-mydriatic	1000×1000	
Silva et al[32]	103/206	53.9/18–88	ND	1-field, 100° and 200°, stereoscopic, colour, non-mydriatic	1200×1600	
Kernt et al[18]	141/212	64/25–78	12/3–39	1-field, 180–200°, non-stereoscopic, colour, non-mydriatic	3900×3072	

*48 (31%) digital images were obtained at a 45–60° angle and 109 (69%) with the common 30° angle.
†There were three comparable groups in the three studies: two of these groups were identical. One used the reference standard 7F-ETDRS; the other involved 7-field 35° stereoscopic colour images with mydriasis. The third comparable group differed from those in the three studies and is presented in the table.
ETDRS, Early Treatment Diabetic Retinopathy Study; ND, no data available.

Figure 2 Results of the Quality Assessment for Diagnostic Accuracy Studies (QUADAS)-2 evaluation of each study.

the photographs were graded by ETDRS levels (levels 10, 14, 15, 20……). In 10 studies[20 26 27 30–36] the ETDRS levels were converted to the analogous clinical scale: absence of DR (or no DR): levels 10 and 14; mild non-proliferative diabetic retinopathy (NPDR): levels 15, 20 and 35; moderate NPDR: levels 43 and 47; severe NPDR: level 53; low-risk PDR: levels 60, 61 and 65; and high-risk PDR: levels 71, 75 and 80.[37] We used an analogous clinical scale for data analysis; however, not all of the included studies covered the full range of severity for DR and DME.

Figure 3 shows the sensitivity and specificity of each study and the pooled estimates when telemedicine was used to determine various clinical levels of DR and DME. The pooled sensitivity exceeded 70%, except in telemedicine detection of severe NPDR in which it was 53% (95% CI 45% to 62%). The pooled specificity exceeded 90%, except in telemedicine detection of

mild NPDR in which it was 89% (95% CI 88% to 91%). Most of the I^2 in sensitivity and specificity forest plots exceeded 50% with p<0.05, which suggests the existence of significant heterogeneity among the included studies.

The corresponding SROC curves (figure 4) showed AUC >0.9 in the detection of the full range of severity of DR and DME and pooled diagnostic accuracy (Q*) ≥0.9 or close to 0.9 except in telemedicine detecting mild NPDR (Q*=0.8389, SE=0.0313). These findings indicate an overall high degree of accuracy for telemedicine in the detection of DR.

Meta-regression analysis
We attempted to explore the sources of heterogeneity across the included studies using meta-regression analysis with several demographic covariates including mean age, duration of diabetes, subjects recruited (consecutive or otherwise), and digital imaging techniques including mydriasis versus non-mydriasis, observation angle, stereoscopy versus non-stereoscopy, number of fields, use of colour versus monochromatic techniques, image resolution and image compression versus non-compression. We found no evidence that any of these factors corresponded with heterogeneity.

Subgroup analyses
We carried out four subgroup analyses of the accuracy of digital image-based telemedicine in the detection of DR and DME based on the use of mydriasis versus non-mydriasis, the observation angle, stereoscopic versus non-stereoscopic techniques and compression versus non-compression. The results indicate that telemedicine using mydriasis had higher sensitivity and specificity, PLR, NLR and DOR than telemedicine without mydriasis in detecting the absence of DR and the presence of mild NPDR. We also found that an observation angle of 100–200° provided the greatest diagnostic accuracy, followed by an observation angle of 30° or 35°, 45–60° (tables 2 and 3).

We divided 10 of the included studies into four subgroups according to the digital imaging technique (see online supplementary table S3). In two studies[20 35] using non-mydriatic, 5-field, 45° angle, non-stereoscopic, colour, non-compressed techniques, heterogeneity in telemedicine detection of severe NPDR was greatly reduced. In three studies[17–18 32] using non-mydriatic, single-field, 100–200° angle, non-stereoscopic, colour, non-compressed techniques, heterogeneity in telemedicine detection of the absence of DR, mild NPDR and moderate NPDR was greatly reduced. The other two subgroup analyses, which evaluated the use of mydriatic, 7-field, 30°, stereoscopic, colour digital imaging technique in which compressed images were used in one subgroup[15 24] and non-compressed images in another,[23 26 29] revealed the existence of unexplained heterogeneity.

Publication bias
We used the raw data generated in studies of various clinical levels of DR and DME to detect publication bias. As shown in online supplementary figure S1, these studies did not demonstrate statistical significance for telemedicine in the use of DR detection (p>0.05). The results indicated no potential publication bias.

DISCUSSION
Telemedicine is the transfer of digital photographs of the retina to a reading centre where they can be evaluated by an ophthalmologist. It allows clinicians to detect DR in a non-ophthalmological setting and thus allows them to serve patients
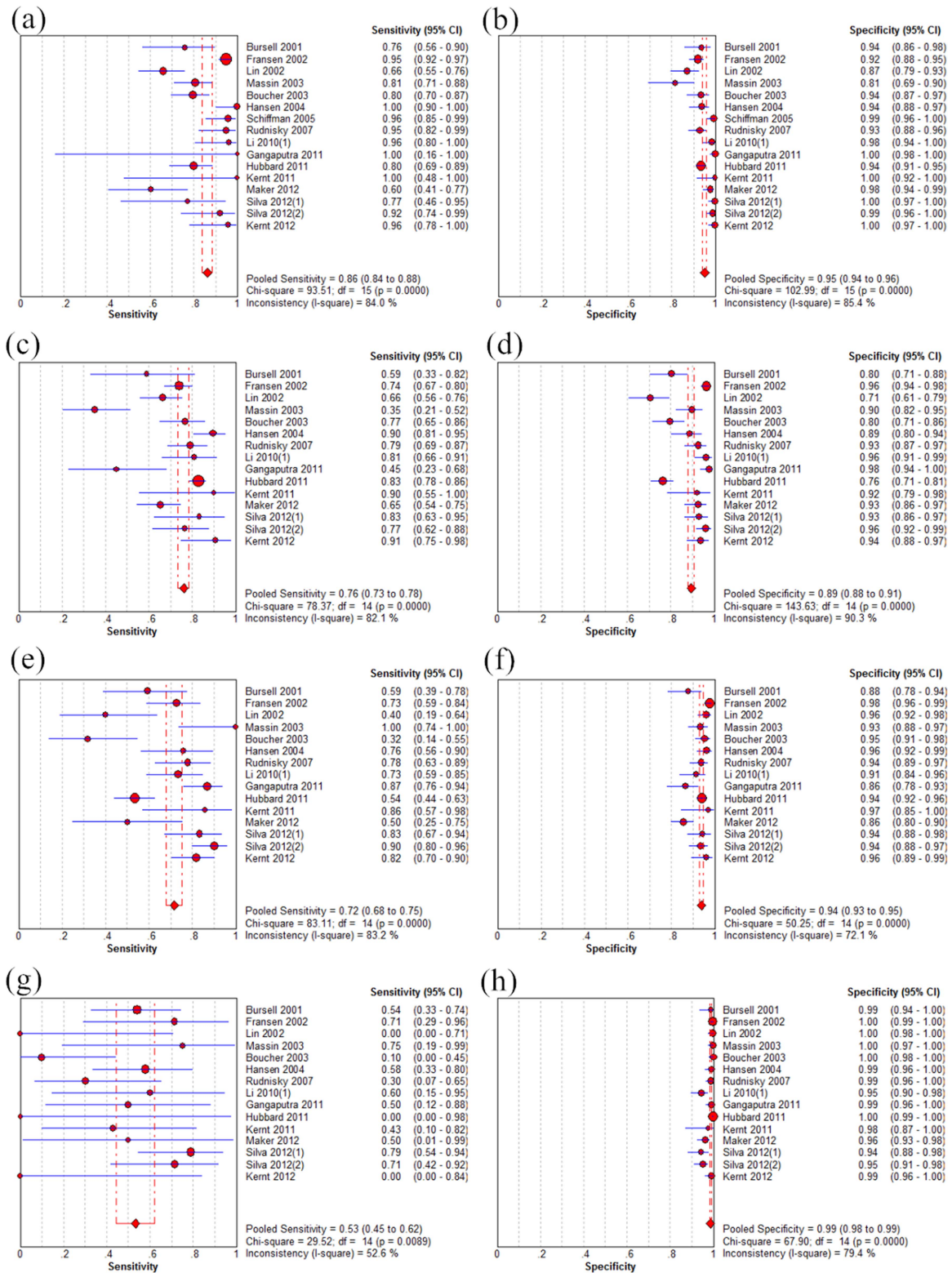
Figure 3 Forest plots of sensitivity and specificity in telemedicine detection of the absence of DR (A, B), mild NPDR (C, D), moderate NPDR (E, F), severe NPDR (G, H), low-risk PDR (I, J), high-risk PDR (K, L), DME (M, N) and CSME (O, P). DR, diabetic retinopathy; NPDR, non-proliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy; DME, diabetic macular oedema; CSME, clinically significant macular oedema.

in rural, remote and hard-to-reach locations. Previous literature reviews and studies have indicated that telemedicine is a cost-effective means of screening patients for DR and can alleviate some of the burden of DR.[38] Patients have expressed a high degree of satisfaction with the telemedicine examination.[39]

All of the included studies except the study by Lin et al[16] were based on the number of eyes rather than the number of patients. We believe that 'eye count' versus 'person count' had little relationship with the results, so we included both in the meta-analysis. The overall high accuracy of telemedicine in

(i)

Sensitivity (95% CI)

Study	Sensitivity	95% CI
Bursell 2001	0.80	(0.28 - 0.99)
Fransen 2002	0.67	(0.45 - 0.84)
Lin 2002	0.00	(0.00 - 0.84)
Hansen 2004	0.50	(0.07 - 0.93)
Rudnisky 2007	0.88	(0.64 - 0.99)
Li 2010(1)	0.90	(0.68 - 0.99)
Gangaputra 2011	0.90	(0.79 - 0.96)
Hubbard 2011	0.94	(0.86 - 0.98)
Kernt 2011	1.00	(0.63 - 1.00)
Maker 2012	0.57	(0.18 - 0.90)
Silva 2012(1)	0.69	(0.41 - 0.89)
Silva 2012(2)	0.72	(0.57 - 0.84)
Kernt 2012	1.00	(0.80 - 1.00)

Pooled Sensitivity = 0.84 (0.79 to 0.88)
Chi-square = 41.50; df = 12 (p = 0.0000)
Inconsistency (I-square) = 71.1 %

(j)

Specificity (95% CI)

Study	Specificity	95% CI
Bursell 2001	0.96	(0.90 - 0.99)
Fransen 2002	1.00	(0.99 - 1.00)
Lin 2002	0.99	(0.96 - 1.00)
Hansen 2004	0.99	(0.97 - 1.00)
Rudnisky 2007	0.98	(0.95 - 1.00)
Li 2010(1)	0.98	(0.95 - 1.00)
Gangaputra 2011	0.94	(0.87 - 0.98)
Hubbard 2011	0.99	(0.97 - 0.99)
Kernt 2011	1.00	(0.91 - 1.00)
Maker 2012	0.99	(0.97 - 1.00)
Silva 2012(1)	0.99	(0.95 - 1.00)
Silva 2012(2)	0.97	(0.93 - 0.99)
Kernt 2012	0.94	(0.89 - 0.98)

Pooled Specificity = 0.98 (0.98 to 0.99)
Chi-square = 33.32; df = 12 (p = 0.0009)
Inconsistency (I-square) = 64.0 %

(k)

Sensitivity (95% CI)

Study	Sensitivity	95% CI
Bursell 2001	0.75	(0.19 - 0.99)
Fransen 2002	0.67	(0.09 - 0.99)
Lin 2002	0.50	(0.01 - 0.99)
Rudnisky 2007	0.80	(0.44 - 0.97)
Li 2010(1)	1.00	(0.69 - 1.00)
Gangaputra 2011	1.00	(0.48 - 1.00)
Hubbard 2011	0.00	(0.00 - 0.84)
Maker 2012	1.00	(0.54 - 1.00)
Silva 2012(1)	1.00	(0.29 - 1.00)
Silva 2012(2)	0.40	(0.05 - 0.85)
Kernt 2012	1.00	(0.40 - 1.00)

Pooled Sensitivity = 0.81 (0.69 to 0.91)
Chi-square = 23.92; df = 10 (p = 0.0078)
Inconsistency (I-square) = 58.2 %

(l)

Specificity (95% CI)

Study	Specificity	95% CI
Bursell 2001	1.00	(0.97 - 1.00)
Fransen 2002	1.00	(0.99 - 1.00)
Lin 2002	1.00	(0.98 - 1.00)
Rudnisky 2007	0.99	(0.97 - 1.00)
Li 2010(1)	0.99	(0.96 - 1.00)
Gangaputra 2011	0.99	(0.96 - 1.00)
Hubbard 2011	1.00	(0.99 - 1.00)
Maker 2012	0.91	(0.84 - 0.96)
Silva 2012(1)	1.00	(0.97 - 1.00)
Silva 2012(2)	0.99	(0.97 - 1.00)
Kernt 2012	1.00	(0.97 - 1.00)

Pooled Specificity = 0.99 (0.99 to 1.00)
Chi-square = 41.80; df = 10 (p = 0.0000)
Inconsistency (I-square) = 76.1 %

(m)

Sensitivity (95% CI)

Study	Sensitivity	95% CI
Bursell 2001	0.30	(0.07 - 0.65)
Fransen 2002	0.80	(0.68 - 0.89)
Hansen 2004	0.73	(0.45 - 0.92)
Kernt 2011	0.88	(0.70 - 0.98)
Silva 2012(1)	0.85	(0.71 - 0.94)
Silva 2012(2)	0.52	(0.31 - 0.73)

Pooled Sensitivity = 0.76 (0.69 to 0.82)
Chi-square = 20.97; df = 5 (p = 0.0008)
Inconsistency (I-square) = 76.2 %

(n)

Specificity (95% CI)

Study	Specificity	95% CI
Bursell 2001	0.97	(0.91 - 0.99)
Fransen 2002	0.95	(0.92 - 0.96)
Hansen 2004	0.96	(0.91 - 0.99)
Kernt 2011	0.82	(0.57 - 0.96)
Silva 2012(1)	0.98	(0.92 - 1.00)
Silva 2012(2)	0.94	(0.89 - 0.97)

Pooled Specificity = 0.95 (0.93 to 0.96)
Chi-square = 6.90; df = 5 (p = 0.2283)
Inconsistency (I-square) = 27.5 %

(o)

Sensitivity (95% CI)

Study	Sensitivity	95% CI
Bursell 2001	0.38	(0.09 - 0.76)
Massin 2003	1.00	(0.72 - 1.00)
Li 2010(4)	0.83	(0.70 - 0.93)
Gangaputra 2011	0.86	(0.78 - 0.92)
Hubbard 2011	0.50	(0.28 - 0.72)
Kernt 2011	0.93	(0.68 - 1.00)
Maker 2012	0.47	(0.24 - 0.71)
Silva 2012(1)	0.77	(0.55 - 0.92)
Silva 2012(2)	0.57	(0.41 - 0.72)

Pooled Sensitivity = 0.75 (0.70 to 0.80)
Chi-square = 43.97; df = 8 (p = 0.0000)
Inconsistency (I-square) = 81.8 %

(p)

Specificity (95% CI)

Study	Specificity	95% CI
Bursell 2001	1.00	(0.96 - 1.00)
Massin 2003	0.99	(0.96 - 1.00)
Li 2010(4)	0.99	(0.95 - 1.00)
Gangaputra 2011	0.84	(0.71 - 0.92)
Hubbard 2011	0.98	(0.95 - 0.99)
Kernt 2011	0.97	(0.83 - 1.00)
Maker 2012	0.98	(0.95 - 0.99)
Silva 2012(1)	0.98	(0.93 - 1.00)
Silva 2012(2)	0.94	(0.88 - 0.97)

Pooled Specificity = 0.97 (0.96 to 0.98)
Chi-square = 33.46; df = 8 (p = 0.0001)
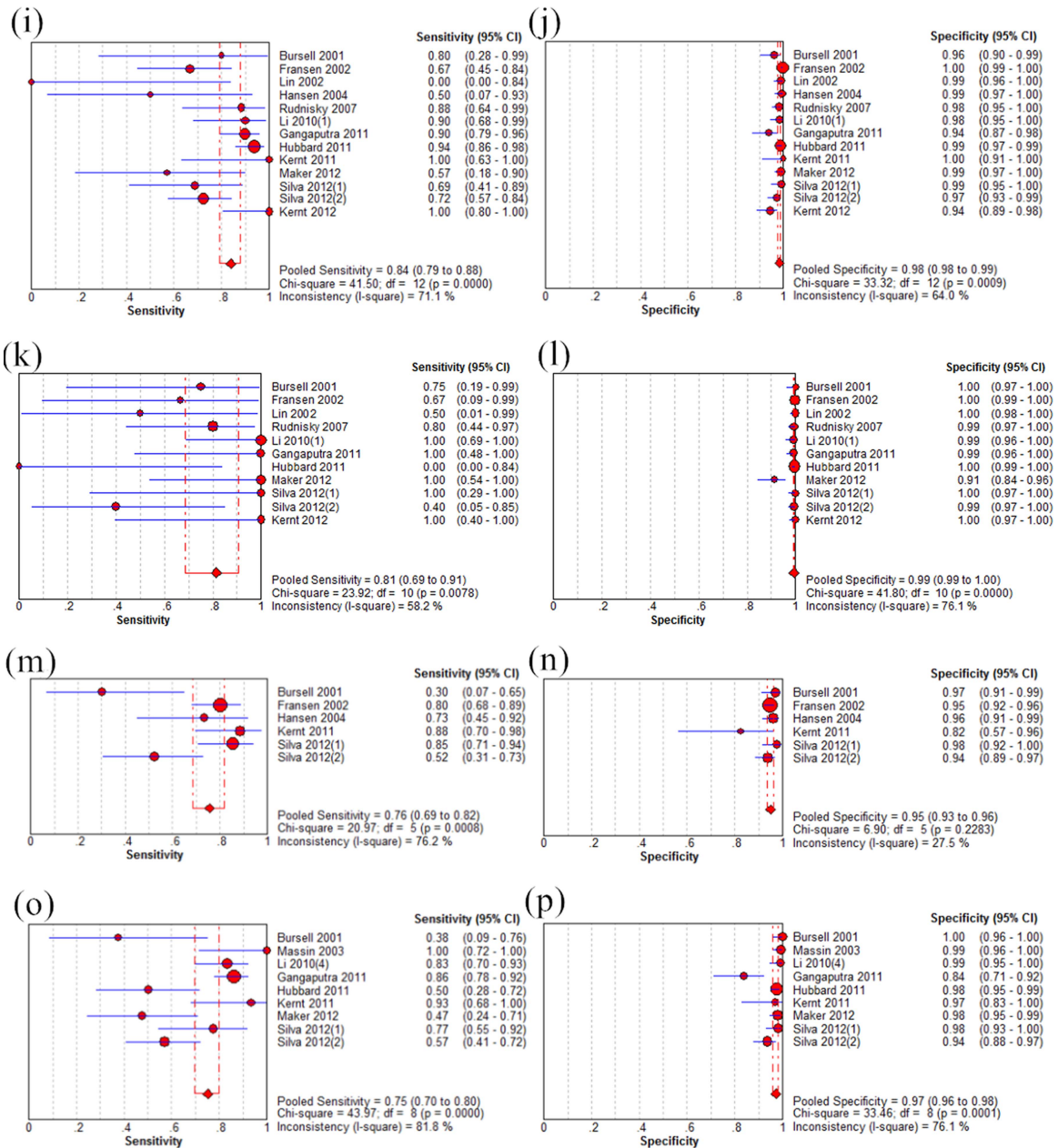Inconsistency (I-square) = 76.1 %

Figure 3 Continued.

detecting DR suggests that it is valuable for this purpose. When telemedicine was used to detect the absence of DR and mild NPDR, the diagnostic accuracy was higher when the images were acquired through mydriasis than through non-mydriasis and a wide angle (100–200°) of observation rather than a narrower angle (45–60° or the more commonly used 30°or 35°).

Sensitivity varied widely among primary studies, falling as low as zero (see figure 3). A potential explanation for this variability might be the small sample size and ungradable digital images. A previous meta-analysis was limited to the detection of the presence or absence of DR.[40] In this review we analysed the full range of severity of DR and DME. It would have been ideal to use the entire sample size. When the total population was assigned to different clinical levels of DR, however, the sample sizes became quite small. In the study by Hubbard et al,[15] of the total of 628 eyes, the 7F-ETDRS detected severe NPDR in one eye that was not detected by telemedicine; this resulted in a

sensitivity of zero. Compared with 7F-ETDRS, digital imaging was more likely to result in images that could not be graded. The ungradable slides were excluded, but the ungradable digital images were included in the analysis.[16 29] This reduced the sensitivity of telemedicine.

When telemedicine was used to detect the absence of DR, a positive number represented the absence of DR and a negative number represented any clinical level of DR within a 2×2 (true positive vs false positive and false negative vs true negative) array. We were also able to calculate sensitivity and specificity based on the distribution of results within the same 2×2 array. The sensitivity was just the specificity of DR absence detection and the specificity was just the sensitivity of DR absence detection. Thus, when any DR was detected, the pooled sensitivity was 95% (95% CI 94% to 96%) and the pooled specificity was 86% (95% CI 84% to 88%). Sensitivity is the effectiveness of telemedicine in detecting DR in patients who have DR, and
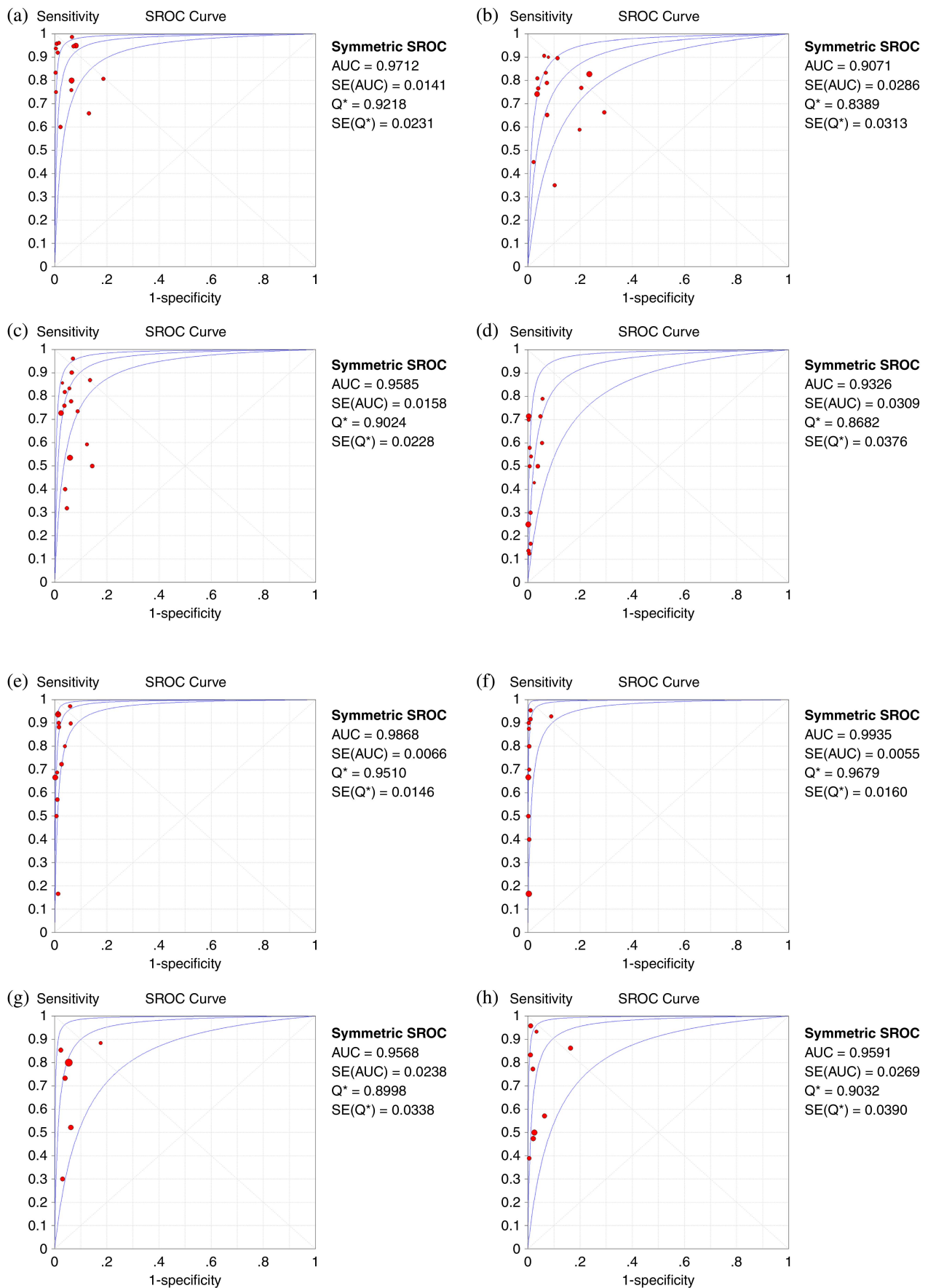
Figure 4 Summary receiver operating characteristic (SROC) curves of various clinical levels of DR and DME: (A) absence of DR; (B) mild NPDR; (C) moderate NPDR; (D) severe NPDR; (E) low-risk PDR; (F) high-risk PDR; (G) DME; and (H) CSME. AUC, area under the curve; DR, diabetic retinopathy; DME, diabetic macular oedema; PDR, proliferative diabetic retinopathy; NPDR, non-proliferative diabetic retinopathy; CSME, clinically significant macular oedema.

Table 2 Subgroup analysis of mydriasis versus non-mydriasis in telemedicine detection of absence of DR and mild NPDR

	Absence of DR		Mild NPDR	
	Non-mydriasis	Mydriasis	Non-mydriasis	Mydriasis
Se (95% CI)	80% (75% to 84%)	91% (89% to 94%)	70% (65% to 75%)	74% (67% to 80%)
Sp (95% CI)	95% (93% to 96%)	95% (94% to 96%)	87% (85% to 89%)	96% (94% to 98%)
PLR (95% CI)	15.96 (7.90 to 32.25)	17.70 (11.82 to 26.50)	5.91 (3.58 to 9.74)	20.94 (12.19 to 35.96)
NLR (95% CI)	0.21 (0.14 to 0.32)	0.09 (0.04 to 0.22)	0.32 (0.21 to 0.50)	0.27 (0.21 to 0.34)
DOR (95% CI)	110.11 (37.04 to 327.31)	254.95 (99.37 to 654.11)	19.70 (8.63 to 44.96)	42.52 (22.15 to 81.64)

DOR, diagnostic OR; DR, diabetic retinopathy; NLR, negative likelihood ratio; NPDR, non-proliferative diabetic retinopathy; PLR, positive likelihood ratio; Se, sensitivity; Sp, specificity.

Table 3 Subgroup analysis of observation angle in telemedicine detection of absence of DR and mild NPDR

	Absence of DR			Mild NPDR		
	30°, 35°	45–60°	100–200°	30°, 35°	45–60°	100–200°
Se (95% CI)	90% (87% to 93%)	80% (76% to 84%)	94% (84% to 99%)	77% (74% to 80%)	72% (67% to 76%)	83% (74% to 90%)
Sp (95% CI)	95% (94% to 96%)	94% (93% to 96%)	99% (98% to 100%)	91% (89% to 93%)	84% (82% to 87%)	95% (92% to 97%)
PLR (95% CI)	16.17 (10.67 to 24.51)	14.11 (7.20 to 27.67)	95.82 (30.97 to 296.40)	11.54 (5.08 to 26.24)	4.63 (2.98 to 7.21)	15.44 (9.52 to 25.04)
NLR (95% CI)	0.12 (0.05 to 0.31)	0.22 (0.14 to 0.34)	0.07 (0.03 to 0.19)	0.29 (0.22 to 0.39)	0.32 (0.19 to 0.54)	0.17 (0.09 to 0.34)
DOR (95% CI)	159.55 (66.27 to 384.13)	93.65 (31.39 to 279.41)	1343.43 (279.51 to 6456.95)	39.65 (19.40 to 81.04)	14.83 (6.46 to 34.06)	99.86 (44.93 to 221.92)

DOR, diagnostic OR; DR, diabetic retinopathy; NLR, negative likelihood ratio; NPDR, nonproliferative diabetic retinopathy; PLR, positive likelihood ratio; Se, sensitivity; Sp, specificity.

thus reduces the risk of missing DR. The analogous sensitivity and specificity were also displayed in the subgroup analysis using digital imaging. The high sensitivity in detecting any clinical level of DR suggests the good performance of telemedicine in ophthalmology and supports its common use in DR screening.[8] The pooled sensitivity for detecting severe NPDR was lower than for other levels of DR; this was also shown by the two subgroup analyses (see online supplementary table S3).

One limitation of this review was the presence of heterogeneity. The results of meta-regression analysis using covariates such as subject characteristics, recruitment methods and digital imaging techniques did not provide valuable information. Even after the subgroup analyses based on the digital imaging technique, heterogeneity was only partially ameliorated. In a previous study, the authors reported that the prevalence of DR of any clinical level was higher in people with type 1 diabetes than in those with type 2 diabetes.[4 41] Even when we excluded the study that recruited subjects with type 1 diabetes only,[15] heterogeneity was not ameliorated (data not shown). In the 20 studies that were included, the percentage of subjects with no DR varied from 1.27%[36] to 59.86%;[35] this may also have been a source of heterogeneity. Another limitation of our findings was that the raw data provided by three studies were unavailable and the data used in the meta-analysis were only from published papers. This, in turn, introduces the possibility of study selection bias. Additionally, the quality of telemedicine largely depends on the experience of readers who grade the digital images and the technician who takes the images. Thus, the use of telemedicine is limited.

In summary, the diagnostic accuracy of telemedicine using digital imaging in DR is overall high. The high sensitivity of its detection of any clinical level of DR indicates that telemedicine can be used widely for DR screening. Telemedicine based on digital imaging techniques that incorporate mydriasis and a wide (100–200°) field is the best choice for detecting the absence of DR and mild NPDR. Further research should be carried out to evaluate the diagnostic accuracy of telemedicine based on each digital imaging technique using a study design that may prevent heterogeneity across studies.

REFERENCES

1 International Diabetes Federation. *Diabetes atlas*. 6th edn. http://www.idf.org/diabetesatlas
2 Porta M, Bandello F. Diabetic retinopathy: a clinical update. *Diabetologia* 2002;45:1617–34.
3 Resnikoff S, Pascolini D, Etya'ale D, *et al*. Global data on visual impairment in the year 2002. *Bull WHO* 2004;82:844–51.
4 Yau JW, Rogers SL, Kawasaki R, *et al*. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 2012;35:556–64.
5 [No authors listed]. 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. *Ophthalmology* 1991;98(5 Suppl):786–806.
6 [No authors listed]. Classification of diabetic retinopathy from fluorescein angiograms. ETDRS report number 11. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology* 1991;98(5 Suppl):807–22.
7 Hazin R, Barazi MK, Summerfield M. Challenges to establishing nationwide diabetic retinopathy screening programs. *Curr Opin Ophthalmol* 2011;22:174–9.
8 Zimmer-Galler IE, Zeimer R. Telemedicine in diabetic retinopathy screening. *Int Ophthalmol Clin* 2009;49:75–86.
9 Lira RP, Oliveira CL, Marques MV, *et al*. Adverse reactions of fluorescein angiography: a prospective study. *Arq Bras Oftalmol* 2007;70:615–18.
10 Vaziri K, Moshfeghi DM, Moshfeghi AA. Feasibility of telemedicine in detecting diabetic retinopathy and age-related macular degeneration. *Semin Ophthalmol*. Published Online First: 30 Oct 2013.
11 Whiting PF, Rutjes AW, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
12 Zamora J, Abraira V, Muriel A, *et al*. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;6:31.

13 Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.

14 Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58:882–93.

15 Hubbard LD, Sun W, Cleary PA, et al. Comparison of digital and film grading of diabetic retinopathy severity in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Arch Ophthalmol* 2011;129:718–26.

16 Lin DY, Blumenkranz MS, Brothers RJ, et al. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol* 2002;134:204–13.

17 Kernt M, Pinter F, Hadi I, et al. Diabetic retinopathy: comparison of the diagnostic features of ultra-widefield scanning laser ophthalmoscopy Optomap with ETDRS 7-field fundus photography. *Ophthalmologe* 2011;108:117–23.

18 Kernt M, Hadi I, Pinter F, et al. Assessment of diabetic retinopathy using nonmydriatic ultra-widefield scanning laser ophthalmoscopy (Optomap) compared with ETDRS 7-field stereo photography. *Diabetes Care* 2012;35:2459–63.

19 Rudnisky CJ, Tennant MT, Weis E, et al. Web-based grading of compressed stereoscopic digital photography versus standard slide film photography for the diagnosis of diabetic retinopathy. *Ophthalmology* 2007;114:1748–54.

20 Hansen AB, Sander B, Larsen M, et al. Screening for diabetic retinopathy using a digital non-mydriatic camera compared with standard 35-mm stereo colour transparencies. *Acta Ophthalmol Scand* 2004;82:656–65.

21 Li HK, Hubbard LD, Danis RP, et al. Comparison of multiple stereoscopic and monoscopic digital image formats to film for diabetic macular edema evaluation. *Invest Ophthalmol Vis Sci* 2010;51:6753–61.

22 Li HK, Hubbard LD, Danis RP, et al. Monoscopic versus stereoscopic retinal photography for grading diabetic retinopathy severity. *Invest Ophthalmol Vis Sci* 2010;51:3184–92.

23 Li HK, Hubbard LD, Danis RP, et al. Digital versus film fundus photography for research grading of diabetic retinopathy severity. *Invest Ophthalmol Vis Sci* 2010;51:5846–52.

24 Li HK, Florez-Arango JF, Hubbard LD, et al. Grading diabetic retinopathy severity from compressed digital retinal images compared with uncompressed images and film. *Retina* 2010;30:1651–61.

25 Li HK, Esquivel A, Hubbard LD, et al. Mosaics versus Early Treatment Diabetic Retinopathy seven standard fields for evaluation of diabetic retinopathy severity. *Retina* 2011;31:1553–63.

26 Maker MP, Noble J, Silva PS, et al. Automated Retinal Imaging System (ARIS) compared with ETDRS protocol color stereoscopic retinal photography to assess level of diabetic retinopathy. *Diabetes Technol Ther* 2012;14:515–22.

27 Boucher MC, Gresset JA, Angioi K, et al. Effectiveness and safety of screening for diabetic retinopathy with two nonmydriatic digital images compared with the seven standard stereoscopic photographic fields. *Can J Ophthalmol* 2003;38:557–68.

28 Li HK, Danis RP, Hubbard LD, et al. Comparability of digital photography with the ETDRS film protocol for evaluation of diabetic retinopathy severity. *Invest Ophthalmol Vis Sci* 2011;52:4717–25.

29 Fransen SR, Leonard-Martin TC, Feuer WJ, et al. Clinical evaluation of patients with diabetic retinopathy: accuracy of the Inoveon diabetic retinopathy-3DT system. *Ophthalmology* 2002;109:595–601.

30 Schiffman RM, Jacobsen G, Nussbaum JJ, et al. Comparison of a digital retinal imaging system and seven-field stereo color fundus photography to detect diabetic retinopathy in the primary care environment. *Ophthalmic Surg Lasers Imaging* 2005;36:46–56.

31 Silva PS, Walia S, Cavallerano JD, et al. Comparison of low-light nonmydriatic digital imaging with 35-mm ETDRS seven-standard field stereo color fundus photographs and clinical examination. *Telemed J E Health* 2012;18:492–9.

32 Silva PS, Cavallerano JD, Sun JK, et al. Nonmydriatic ultrawide field retinal imaging compared with dilated standard 7-field 35-mm photography and retinal specialist examination for evaluation of diabetic retinopathy. *Am J Ophthalmol* 2012;154:549–59 e2.

33 Bursell SE, Cavallerano JD, Cavallerano AA, et al. Stereo nonmydriatic digital-video color retinal imaging compared with Early Treatment Diabetic Retinopathy Study seven standard field 35-mm stereo color photos for determining level of diabetic retinopathy. *Ophthalmology* 2001;108:572–85.

34 Vujosevic S, Benetti E, Massignan F, et al. Screening for diabetic retinopathy: 1 and 3 nonmydriatic 45-degree digital fundus photographs vs 7 standard early treatment diabetic retinopathy study fields. *Am J Ophthalmol* 2009;148:111–18.

35 Massin P, Erginay A, Ben Mehidi A, et al. Evaluation of a new non-mydriatic digital camera for detection of diabetic retinopathy. *Diabet Med* 2003;20:635–41.

36 Gangaputra S, Almukhtar T, Glassman AR, et al. Comparison of film and digital fundus photographs in eyes of individuals with diabetes mellitus. *Invest Ophthalmol Vis Sci* 2011;52:6168–73.

37 [No authors listed]. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology* 1991;98(5 Suppl):823–33.

38 Rein DB, Wittenborn JS, Zhang X, et al. The cost-effectiveness of three screening alternatives for people with diabetes with no or early diabetic retinopathy. *Health Serv Res* 2011;46:1534–61.

39 Kumari Rani P, Raman R, Manikandan M, et al. Patient satisfaction with tele-ophthalmology versus ophthalmologist-based screening in diabetic retinopathy. *J Telemed Telecare* 2006;12:159–60.

40 Bragge P, Gruen RL, Chau M, et al. Screening for presence or absence of diabetic retinopathy: a meta-analysis. *Arch Ophthalmol* 2011;129:435–44.

41 Kempen JH, O'Colmain BJ, Leske MC, et al. The prevalence of diabetic retinopathy among adults in the United States. *Arch Ophthalmol* 2004;122:552–63.