The ROYAL COLLEGE *of*
OPHTHALMOLOGISTS

REVIEW ARTICLE

Prognostic prediction models for diabetic retinopathy progression: a systematic review

Sajjad Haider[1] · Salman Naveed Sadiq[2] · David Moore[1] · Malcolm James Price[1] · Krishnarajah Nirantharakumar[1]

Abstract

With the increasing incidence of diabetic retinopathy and its improved detection, there is increased demand for diabetic retinopathy treatment services. Prognostic prediction models have been used to optimise services but these were intended for early detection of sight-threatening retinopathy and are mostly used in diabetic retinopathy screening services. We wanted to look into the predictive ability and applicability of the existing models for the higher-risk patients referred into hospitals. We searched MEDLINE, EMBASE, COCHRANE CENTRAL, conference abstracts and reference lists of included publications for studies of any design using search terms related to diabetes, diabetic retinopathy and prognostic models. Search results were screened for relevance to the review question. Included studies had data extracted on model characteristics, predictive ability and validation. They were assessed for quality using criteria specified by PROBAST and CHARMS checklists, independently by two reviewers. Twenty-two articles reporting on 14 prognostic models (including four updates) met the selection criteria. Eleven models had internal validation, eight had external validation and one had neither. Discriminative ability with c-statistics ranged from 0.57 to 0.91. Studies ranged from low to high risk of bias, mostly due to the need for external validation or missing data. Participants, outcomes, predictors handling and modelling methods varied. Most models focussed on lower-risk patients, the majority had high risk of bias and doubtful applicability, but three models had some applicability for higher-risk patients. However, these models will also need updating and external validation in multiple hospital settings before being implemented into clinical practice.

Introduction

There has been a global increase in the number of people with diabetes, rising from 108 million in 1980 to 422 million in 2014 [1]. The detection of retinopathy has also increased through better population screening [2]. While

These authors are Joint Senior Authors: David Moore, Malcolm James Price, Krishnarajah Nirantharakumar

✉ Sajjad Haider
drsajjadhaider@yahoo.co.uk

✉ Malcolm James Price
M.Price.2@bham.ac.uk

[1] Institute of Applied Health Research, University of Birmingham, Birmingham, UK

[2] York Teaching Hospital NHS Foundation Trust, York, UK

services may be organised differently from country to country, the care pathways are likely to be similar with patients at higher risk being provided closer monitoring and care. In the United Kingdom, diabetic retinopathy (DR) services are organised into diabetic eye screening programmes (DESP) and hospital eye services. DESP provides annual diabetic retinopathy screening to all patients with diabetes above 12 years of age. Screening uptake in the year 2015/16 was 82.5% [3]. If the screening findings indicate low risk (retinopathy stage R0, R1 and M0, see Fig. 1 and Table A1) they are retained within the DESP and reviewed yearly. When the disease progresses to sight-threatening diabetic retinopathy (STR) stage—(R2, R3 or M1), they are referred to the hospital eye services for closer observation and treatment (Fig. 1). In the United States, yearly screening is recommended to all type 2 diabetes mellitus (T2DM) patients at diagnosis and afterwards, type 1 diabetes mellitus (T1DM) patients are recommended to have screening on an annual basis, commencing 5 years after diagnosis [4, 5].

In the United Kingdom, ~50% of referrals with STR do not need intervention and are observed in the hospital eye

Fig. 1 Patient flow diagram[a].
[a]See Table A1 for more details on classification of diabetic retinopathy. R0, no retinopathy; R1, background retinopathy; R2, pre-proliferative retinopathy; R3, proliferative retinopathy; M0, no maculopathy; M1, maculopathy present; DESP, diabetic eye screening programme; HES, hospital eye services; STR, sight-threatening retinopathy; IVT, intra-vitreal therapy; VEGF, vascular endothelial growth factor; VR, vitreoretinal
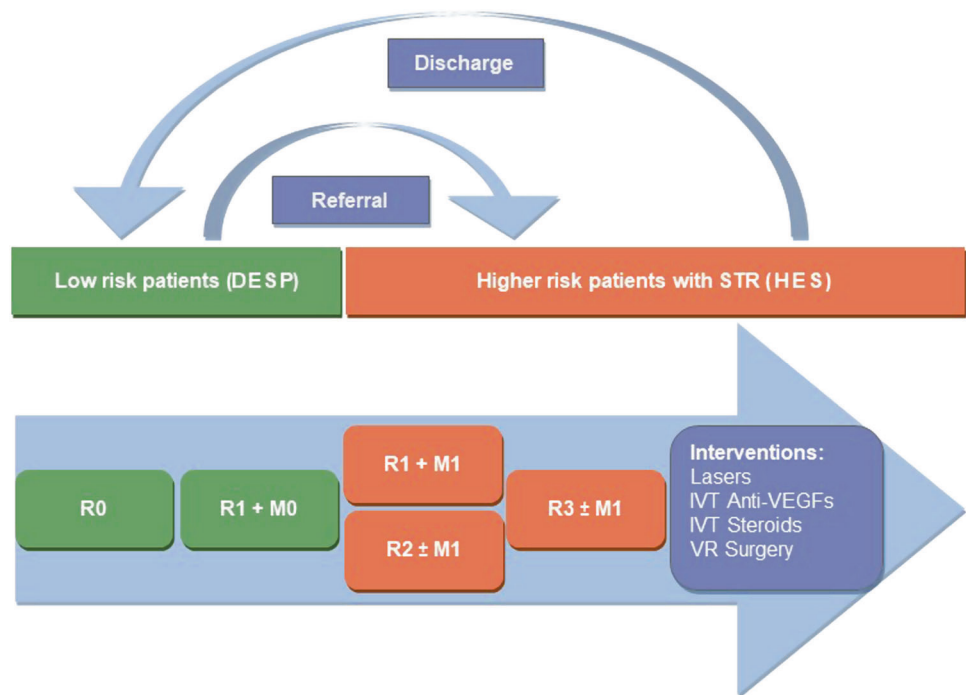
service for a variable period of time [6], placing extra burden on these services. Various modifications to improve the service have been proposed, like digital surveillance using optical coherence tomography (OCT), and virtual clinics within DESP [6].

Prognostic modelling/nomograms can aid decision-making [7]. There have been successful attempts at opti-mising diabetic screening services through stratification of patients by risk of progression of DR using a prognostic prediction model [8, 9]. A similar approach will also help optimise hospital eye services. Such a model combines multiple prognostic factors to predict which patients are at higher risk of progression to visual loss and thus need closer observation or treatment to prevent loss of vision [10–12]. The aim of this systematic review is to summarise the characteristics and performance of existing models in pre-dicting progression of retinopathy and their applicability for higher-risk DR patients under hospital care to predict need for treatment or loss of vision.

Methods/design

Patient group, selection criteria/study design/ inclusion criteria

Studies were included in the review of any design that developed, updated, validated, compared or evaluated a prognostic prediction model/tool, using multiple prognostic

factors to predict the risk of progression of diabetic reti-nopathy and/or vision loss. The searches covered studies reporting the development of a model, validation of a model and impact of a model in practice.

There was no restriction on the age of participants or type of diabetes. The primary outcomes of our review were predictive accuracy and applicability of the prog-nostic prediction models/tools in relation to progression of diabetic retinopathy from stages that required hospital referral (R2, R3 or M1) to treatment requiring stage or vision loss.

Search strategy and selection criteria

As prognostic model studies can be difficult to identify, several approaches were used. MEDLINE, EMBASE and COCHRANE CENTRAL (up to March 5, 2017) were searched using index and free text terms for diabetes, reti-nopathy and prognostic models. A sample search strategy for MEDLINE is shown in Table A2. We did not apply any restrictions on language or dates of publication. We also searched abstracts from the following national and inter-national conferences from 2014 to March 5, 2017.

- The Royal College of Ophthalmologists, American Academy of Ophthalmology, European Society of Retina Specialists (EURETINA), European Society of Ophthalmology (SOE) and Association for Research in Vision and Ophthalmology (ARVO)

- Diabetes conferences. American Diabetic Association (ADA), Diabetes UK and International Diabetes Federation

Reference lists of included studies were screened for additional studies and authors of relevant conference abstracts were identified and their publication lists checked for additional relevant studies using Pubmed, Google Scholar and Scopus.

Search results were recorded in Endnote (version x7.4Clarivate Analytics) and duplicate entries removed. Titles (and abstracts where available) were screened for relevance using predefined screening criteria. Full texts of all potentially relevant articles were obtained and assessed against the selection criteria. Reviewer decisions, including the reason why studies were excluded from the review were recorded.

Data extraction and quality assessment

The information extracted from each study included study characteristics, source of data, study design, participants characteristics, candidate predictors, their handling, outcomes assessed, sample size, missing data and its handling, modelling methods, methods for selection of final predictors, model performance measures (discrimination, calibration and classification measures), model validation and presentation of the final prediction models. Authors were contacted where necessary (mostly for reporting deficiencies). Some models have dealt with multiple outcomes other than of interest to our review. We only considered the ocular outcomes.

Critical appraisal was carried out using a risk assessment form by combining PROBAST [10] and CHARMS [13] checklists. Risk of bias and applicability was assessed mainly by using PROBAST tool [10], however, since the tool was being piloted, CHARM checklist [13] was also used to further refine assessment of the studies.

Study selection, data extraction and risk of bias assessment were carried out by two reviewers independently with disagreements resolved by mutual discussion or with a third reviewer where required.

Analysis

As some models were the subject of more than one study, they were grouped by specific models and organised by whether describing model development, internal validation or external validation. Details on each model are presented in evidence tables and narrative summaries are given on key model features (population, samples, predictors and performance).

This review was registered prospectively with PROSPERO (registration number CRD42017057767) [14] and is reported here in accordance with the PRISMA guidelines [15].

Results

Volume of the research literature available

The searches yielded 12,118 records of which 4893 were duplicates. After screening titles and abstracts, 62 relevant articles were identified of which 22 met all selection criteria and are included in the review (Fig. 2).

Overview

This systematic review analysed 14 prognostic prediction modelling studies, including four updates. Six models had both internal and external validation, five models performed only internal validation and two were only validated in external datasets. One model lacked both internal and external validation. No studies assessing the impact of a model were identified. All studies were published within the last two decades. Figure 3 shows how the studies fit into the evolution of the modelling process. The characteristics of individual studies are summarised in Table 1.

Population

The majority of studies have been conducted post year 2000 ($n = 19$). The latest data used were from 2009 to 2014 in ISDR study [16], and the oldest data were part of the UKPDS OM1 study (1977–1997) [17], though this model has been validated in recent data [18, 19]. Two studies were in T1DM [20, 21], eight in T2DM [17–19, 22–26] and 11 in mixed populations [8, 16, 27–34], with one unspecified [35]. Population studied ranged from newly diagnosed patients [17, 22, 35] to those with a relatively severe form of the disease [19]. The included studies have used large routinely available databases such as The Health Improvement Network (THIN) [19], Q Research and Clinical Practice Research Datalink (CPRD) [28], US claims databases [35], hospital databases [21, 26], diabetic screening data and clinical trials research data [17, 20, 22, 24].

Sample size, events per variable (EPV) and follow-up

Sample size ranged from 1441 [27] to 454,575 [28] in primary development studies and from 200 [21] to 206,050 [28] in validation studies. EPVs ranged from as low as 0.86 [20] to 424.27 [28]. External validation samples were generally small and yielded low EPV. Lack of reporting

Fig. 2 PRISMA flow diagram showing the search process from the volume of the research literature available
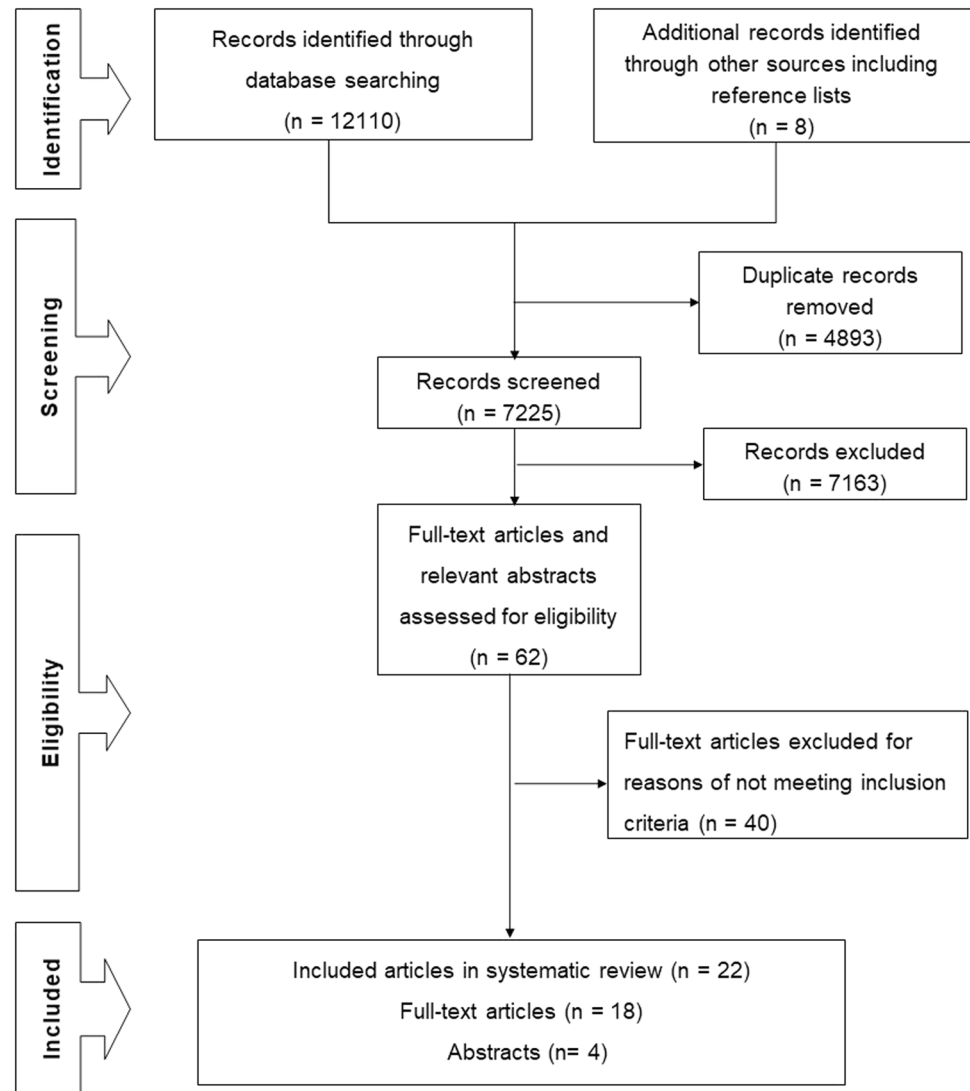
affected assessment of EPV in two models [20, 23]. Median duration of follow-up time was 5.1 years, ranging from 1.1 [30] to 17.6 years [22].

Predictors

The models contained 78 different candidate predictors, which can be grouped into nine broad groups (Table A3 in supplementary material). The number of candidate predictors ranged from 3 [32, 33] to 51 [27] in any one model and their selection was mostly based on literature reviews or clinical intuition. Forty nine different predictors appeared in the final models. The median number of final predictors used in a given model was 5 with a range of 2 [17] to 14 [35].

Standardised definitions and measurement methods were generally used for predictors. Predictor values were recorded or measured at baseline cohort entry, or soon after.

Categorisation of the continuous predictors was mostly avoided, but was not always reported [20, 24, 27, 34]. The method of selection of the final predictors was reported in 9 out of 14 primary development studies and was typically performed using backward elimination [16, 20, 22, 24, 28]. Four models [29, 31, 32, 35] used the full model approach by using all candidate predictors.

Biochemical predictors were the most common (Table A3 in supplementary material). HbA1c was the most popular predictor appearing in all studies except one [32]. It was followed by duration of diabetes ($n = 7$). Age was another important predictor used in various forms—as age ($n = 4$), age at diabetes diagnosis ($n = 4$) and age at DR diagnosis ($n = 1$). Half of the models used local predictors/ocular signs. One model [32] only used ocular signs as the sole set of predictors. They categorised baseline DR into three predictors (R0 in eyes, R1 one eye or R1 both eyes). This trend continued in subsequent related studies [34, 36],
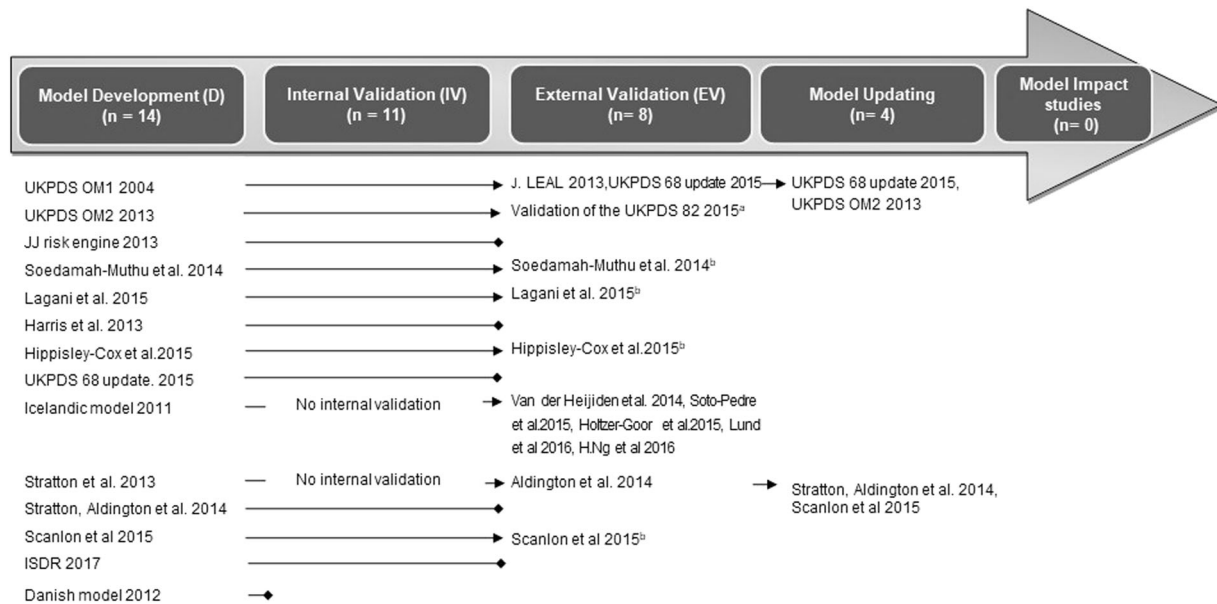
Fig. 3 Flow diagram showing the modelling process and studies in each part of this process (the arrow indicates continuation of the process; the box implies cessation of further progress). [a]Performed Internal Validation on UKPDS OM2 2013. [b]External Validation reported in same study

except R0 was later omitted and biochemical predictors were added.

Outcomes

Outcome definition varied from blindness only [17–20, 22, 23, 28], to STR [8, 25, 26, 29, 30, 32–34, 36], retinopathy progression [8, 24, 27, 35], need for treatment [31] and referable DR [16]. The definition of blindness was mostly defined as best corrected vision of less than 6/60. In some models, STR was used interchangeably with treatment [29, 31]. Validation studies defined the outcomes in a similar way to their respective model development studies.

Statistical model

The final models were based on Cox proportional hazards model in most studies [22, 24, 27, 28, 32, 34–36]. Weibull proportional hazards model was used in four studies [17, 19, 20, 29]. Logistic regression [31] and continuous-time Markov mathematical models [16] were used by one study each.

Model performance/evaluation

Performance measures were poorly reported in almost half of the studies. The remainder reported either Harrell's concordance index [20, 24, 28], area under the curve or both in the case of the ISDR model [16]. Model calibration where reported was mainly in the form of calibration plots. Five primary modelling studies [17, 22, 31, 32, 35] and one

validation study [26] failed to report any performance measures. For discriminatory power, one model [16] reported a c-statistic of > 0.80 (0.91) and the remainder reported moderate discrimination of 0.614 to 0.79, except for one poor c-statistic value of 0.57 [19].

Missing data

The amount of missing data varied in most studies [16, 18–20, 24, 27–29, 36], with some reporting more than 50% of participants missing at least one predictor value [19, 28]. Proportions of missing data have not always been reported [16, 18, 24, 27]. The mechanisms used to handle missing data included the last observation carried forward (average or mode) [18, 27, 29, 36], complete case analysis [36] and multiple imputations [16, 19, 20, 24, 28].

Summary of bias

Four modelling studies had moderate [16, 19, 28, 36] and one had low risk of bias [29] (Table 2). Risk of bias was unclear in one modelling study [34] taken from conference abstracts (lack of reporting). The remaining studies were at high risk of bias. High risk of bias was mostly due to low number of outcomes per variable (EPV < 10), lack of information on missing data, absence of external validation and lack of reporting of relevant performance measures. All validation studies have shown good discrimination ability, but were mostly at high risk of bias due to small numbers (<100) of outcomes [13] or lack of reporting [33].

Table 1 Characteristics of model development and validation studies included in the systematic review

Study ID	Study dates	Participants DM-type source	Predictors (n) candidate final	Outcome follow-up (years)	Sample size (n)	Events (n)	Events per variable	Model performance /evaluation	Presentation
Models development studies with external validation									
UKPDS OM1 2004 [17] (D, IV)	1977–1997	T2DM UKPDS trial	14 / 2	One eye blindness median 10.3y	3642	104	7	Not reported	Equation
Leal et al., 2013 [18] (EV of OM1)	1997–2007	T2DM PTM data	Same		4031	101		C-statistics (95% CI) 0.60 (0.55–0.65)	NA
UKPDS 68 update 2015 [19] (update, IV and EV of OM1)	2000–2009	T2DM (two risk groups) THIN database	14 / 12	Blindness low risk (LR) 5.1y Intermediate risk (IR) 6.4y	LR 54,169 IR 68,990	LR 176 IR 607	56	C-statistics LR: IV 0.59 and EV 0.57 IR: IV 0.59 and EV 0.58 Calibration plots	Risk groups (low and intermediate)
UKPDS OM2 2013 [22] (D, IV)	1997–2007	T2DM PTM data	24 / 7	One eye blindness median 17.6y	5102	271	11	Predicted curves in IV	Hazard ratios
UKPDS 82 2015 [23] (EV of OM2)	Not reported	T2DM	Same		Not reported	Not reported		MAPE (%) = 20.31 and $R^2 = 0.96$ Calibration plots	NA
Soedamah-Muthu et al., 2014 [20] (D, IV) EV in the same study three cohorts	1989–1999 1998–1999 1994–2009 2000–2002	T1DM EURODIAB trial data EDC FinnDiane CACTI	15 / 5 Same	Blindness median 7.4y Median 8.1y Median 7.5y Median 7.3y	1973 554 2999 580	12 29 Not reported 5	<1 (0.8)	Harrell's C-index: IV 0.74 C-statistics 0.79 0.82 0.73 Calibration plots	Equation score charts Kaplan–Meier plots for high, intermediate and low risk score groups
Lagani et al.,[a] 2015 [27] (D, IV) EV in the same study	1983–1993 2004–2014	T1DM DCCT, EDIC T1DM and T2DM Chorleywood, UK	51 / 5 Same	Retinopathy event[b] Mean 6.5y	1441(IV) T1DM 36 T2DM 294	969(IV) T1DM 17 T2DM 70	Not applicable	Concordance index: 0.66 ($p < 0.001$) T1DM 0.72 ($p = 0.002$) T2DM 0.55 ($p = 0.119$)	Equation recalibration NA
Hippisley-Cox et al., 2015 [28] (D, IV) EV in the same study	1998–2014	T1, T2DM Q Research CPRD, UK	18 / 9 Same	Blindness in one or both eyes Total 15y	454,575 (D) 142,419 (IV) 206,050 (EV)	8063 2651 2845	448	Harrell's C-index Women 0.73 (0.71–0.74) and men 0.75 (0.73–0.77) (IV) 0.73 (0.72–0.75) in both gender (EV) Calibration plots	Web calculator NA
Icelandic model 2011 [29] (D, EV)	Not reported	T1, T2DM Icelandic eye screening /Aarhus diabetic database	7 / 7	STR f/u not reported	5199	149	21	C-statistics (95% CI) 0.76 (0.74–0.78) and Calibration graph	Equation nomogram mobile app
Five separate EV studies for Icelandic model*									
*van der Heijden 2014 [25]	1998–2010	T2DM, Dutch	Same	STR Mean 4.4y	3319	76		0.83 (0.74–0.92)	NA
*Soto-Pedre et al., 2015 [30]	Not reported	T1, T2DM, Spain	Same	STR Median 1.1y	508	16		0.74 (0.62–0.85)	NA
*Holtzer-Goor, 2015 [26]	Not reported	T2DM, Dutch Hospital	Same	STR Mean 3.3y	888	47		Not reported	NA
*Lund et al., 2016 [8]	2010–2012	T1, T2DM, DESP, UK	Same	STR Total 2y	9687	531		T1DM 0.70 (0.67–0.73) and T2DM 0.80 (0.78–0.81)	NA

Table 1 (continued)

Study ID	Study dates	Participants DM-type source	Predictors (n) candidate final	Outcome follow-up (years)	Sample size (n)	Events (n)	Events per variable	Model performance /evaluation	Presentation
Models development studies with external validation									
*Ng et al., 2016 [21]	Not reported	T1DM, Dutch Hospital	Same	STR Mean 6.9y	200	22	22	0.75 (0.65–0.86)	NA
Stratton et al., 2013 [32] (D)	2005–2010	T1, T2DM DESP, UK	3 / 3	STR median 2.8y	14,554	803	268	Not reported	Kaplan–Meier plot
Aldington et al., 2014 [33] (EV)	Not reported	Two screening programmes, UK	Same	Total 4y	24,509 32,987	2328 3991		C-statistics 0.76 0.80	NA
Scanlon et al., 2015 [36] (D, IV) EV in the same study three programmes	2005–2011	T1, T2DM DESP, Gloucestershire (D)(IV) East Anglia South London Nottinghamshire	13 4 Same	STR total 5y Median 2.7y 3.8y 4.2y	7012(D) 5778(IV) 17,634 1223 1083	606(D) 490(IV) 845 94 81	47	C-statistics (95% CI) IV 0.79 (0.76–0.81) 0.78 (0.76–0.80) 0.84 (0.78–0.88) 0.82 (0.71–0.90)	Risk score NA
Models development studies with internal validation but no external validation									
JJ risk engine 2013 [24] (D, IV)	Not reported	T2DM, Japan Trial data JDCS, and J-EDIT	16 5	Progression of retinopathy Mean 7.2y	1748	178	11	Harrell's C-index (95% CI) 0.61 (0.52–0.70) Hosmer–Lemeshow $p = 0.13$	Equation web app
Harris et al., 2013 [35] (D, IV)	2001–2009	DM type not specified Claims database, US	14 14	Progression to PDR Median 1.7y	4617	307	22	Not reported	Equation risk score
Stratton et al., 2014 [34] (D, IV)	Not reported	T1, T2DM DESP, Gloucestershire (D)(IV)	6 5	STR Total 4y	6449(D) 5460 (IV)	555(D) 496(IV)	93	C-statistics (95% CI) 0.79 (0.78–0.81)	Not reported
ISDR 2017 [16] (D, IV)	2009–2014	T1, T2DM DESP, Liverpool and primary care data were combined	19 5	Referable DR Total 5y	11,806	388	20	Harrell's concordance index: 0.69 at 24 months AUC (95% CI): 0.91 (0.87–0.94) Calibration curve	Equation
Models development studies with no clear validation (performance testing only)									
Danish model 2012 [31] (D) Unclear if validated	1994–2007	T1, T2DM Århus database, Denmark	7 7	Treatment requiring DR f/u not reported	5311 1372	559 208	80	Not reported	Equation

PTM post-trial monitoring data, *DESP* diabetes eye screening programme, *T1DM* type 1 diabetes mellitus, *T2DM* type 2 diabetes mellitus, *D* development, *IV* internal validation, *EV* external validation, *U* model update, *STR* sight-threatening retinopathy, *PDR* proliferative diabetic retinopathy, *MAPE* mean absolute percentage error, R^2 coefficient of determination/validation, *THIN* The Health Improvement Network, *CPRD* Clinical Practice Research Datalink, *DCS* Diabetes Care System West-Friesland, *Sn* sensitivity, *Sp* specificity, *LR* low risk, *IR* intermediate risk

[a] Figures from DCCT trial and in electronic supplementary material (not given in the publication)

[b] Any worsening in the retina condition that lasted at least 6 months

Low risk of bias: The Icelandic model [29] used good sample size ($n = 5199$) and EPV of 21·29 (>10). Moreover, it is the only model to have multiple external validation studies ($n = 5$). Development of the model was a bit unusual, based on hazard ratios from UKPDS and Wisconsin studies and prevalence figures from Icelandic diabetic eye screening programme. They 'empirically' tested this model in Aarhus diabetic database, which could be interpreted as an external validation.

Moderate risk of bias: Hippisley-Cox et al. [28] used a large cohort with a long follow-up and good reporting. However, there was a lack of formal adjudication of outcomes and a high number of missing variable values (up to 80.9%), though dealt by multiple imputation. The study by Scanlon et al. [36] has sound methodology, logical process of evolution and pragmatic decisions about predictors. However, there is lack of multiple imputations for missing data. The recently published ISDR model [16] has no external validation yet (which according to PROBAST is high risk of bias). We assigned it a moderate risk of bias because of a large sample ($n = 11,806$) and a robust internal validation. UKPDS 2015 [19] lacks external validation but was judged to have moderate risk of bias as it had a large sample size and used multiple imputation.

Applicability of the models

The target population and the outcomes of interest in our review were progression of diabetic retinopathy reaching the treatment requiring stage or visual loss (blindness or low vision) in patients under close observation in a hospital setting (Table 2). In the models identified, the context was broadly the early, low-risk part of the disease pathway. The population was largely a mixed diabetic population of all ages and any duration of diabetes but mainly early-onset diabetic retinopathy. None of the models were fully applicable to our review question with regard to population characteristics or outcomes used.

Even though five out of 14 models used blindness as the outcome (defined as corrected vision of 6/60 or less) [17, 19, 20, 22, 28], these models were mostly designed for multiple outcomes. Therefore, their predictor sets were also less specific for the outcomes needed for DESP or hospital eye service population. Despite having blindness as their outcome, these limitations make them high risk for applicability. Lagani et al. [27] model also focussed on multiple outcomes and defined outcome as any diabetic retinopathy event. Therefore, this was also classified as high risk. JJ risk Engine [24] model used predictors and outcomes more relevant to DESP or hospital eye service population but their participants were only patients with type 2 diabetes. We therefore assigned it as medium risk for applicability.

There was good applicability to low-risk diabetic screening patients and partial applicability for higher-risk hospital patient population in the remaining seven models [16, 29, 31, 32, 34–36]. Mehlsen et al. [31] used the outcome of DR progression to treatment-requiring stage and another study [35] used the outcome of DR progression. The remainder used STR/referable DR [16, 29, 32, 34, 36]. Only three models had a moderate to low risk of bias and also have low risk for applicability [16, 29, 36].

There were 11 different types of final predictors in these three models. Duration of diabetes and HBA1c were common among all three, and systolic blood pressure was used by two models [16, 29]. Other predictors included in these three models were presence, grade of diabetic retinopathy [29], presence of background diabetic retinopathy in one or both eyes [36], gender [29], type of diabetes [29], age at diagnosis [16] and total serum cholesterol [16].

Because of the heterogeneity introduced due to differences in populations, outcome measurements and the context in which they were studied, a meta-analysis could not be carried out.

Discussion

This systematic review summarises the details of 14 predictive model development studies, including four updates. Most of the studies dealt with the question of diabetic retinopathy progression up to the level of referral to the hospital for treatment or closer observation (lower-risk part of the disease pathway). The perspective has been largely individual patients' risk stratification in diabetic screening services using diabetic screening databases (12 out of 22 studies). Only very limited evidence was fully applicable to high-risk patients. Five modelling studies had moderate to low risk of bias and out of them, only three studies also had potential for applicability [16, 29, 36].

To our knowledge, no systematic review looking at the predictive accuracy and applicability of predictive models for patients with DR beyond referable sight-threatening diabetic retinopathy has been published. Our review considered all levels of the modelling process from model development through to validation studies. We have included all diabetic populations (both type 1 and type 2 diabetic patients). Thorough electronic and manual searches were conducted. Following up authors of relevant conference abstracts identified two studies that were not found by other means, one of which was published prior to the searches of bibliographic databases. Models from Europe, United States and Japan are included giving a global picture.

Some studies failed to report important information such as sample size and EPV [23], predictor handling, mathematical algorithm/equation (22, 28), follow-up period

Table 2 Summary of bias assessment and applicability as per PROBAST tool mainly and adding CHARMS input where necessary

Study ID	Participants	Predictors	Outcome	Sample size and participant flow	Model development and evaluation	External validation	Overall bias High Moderate Low	Applicability concern High Moderate concern Low concern	Comments for bias
UKPDS OM1 [17] 2004	L	L	L	H	H	Yes	H	H	Low EPV; Did not report relevant performance measures[a]
J. Leal et al. [18] 2013	L	L	L	H	L	NA	H	H	Low EPV
UKPDS 68 update [19] 2015	L	L	L	L	L	No	M	H	High number of missing values (up to 57%) but dealt with multiple imputations. No EV after update
UKPDS OM2 [22] 2013	L	L	L	H	H	Yes	H	H	Missing data dealt with LOCF, no multiple imputations
Validation of the UKPDS 82 [23] 2015	L	L	L	H	L	NA	H	H	No. of outcomes and sample size not reported
JJ risk engine [24] 2013	L	H	L	L	L	No	H	M	Categorised continuous predictor (BMI) lacks external validation
Soedamah-Muthu et al. [20] 2014	L	L	L	H	L	Yes	H	H	Very low EPV and fewer number of outcomes
Lagani et al. [27] 2015	L	L	L	H	L	Yes	H	H	Low EPV
Harris et al. [35] 2013	L	H	L	L	H	No	H	L	Did not report relevant performance measures[a]; Lacks external validation
Hippisley-Cox et al. [28] 2015	L	L	L	L	L	Yes	M[b]	H	High number of missing values, but dealt with multiple imputations
Icelandic model [29] 2011	L	L	L	L	L	Yes	L[b]	L	
Van der Heijden [25] 2014	L	L	L	H	L	NA	H	L	Low number of outcomes (<100)
Soto-Pedre et al. [30] 2015	L	L	L	H	L	NA	H	L	Low number of outcomes (<100)
Holtzer-Goor [26] 2015	L	L	L	H	L	NA	H	L	Low number of outcomes (<100)
Lund et al. [8] 2016	L	L	L	L	L	NA	L[b]	L	
H. Ng Keunen et al. [21] 2016	L	L	L	H	L	NA	H	L	Low number of outcomes (<100)
Danish model [31] 2012	L	L	L	L	H	No	H	L	Did not report relevant performance measures[a]; No validation

Table 2 (continued)

Study ID	Participants	Predictors	Outcome	Sample size and participant flow	Model development and evaluation	External validation	Overall bias High Moderate Low	Applicability High concern Moderate concern Low concern	Comments for bias
Stratton et al. [32] 2013	L	L	L	L	H	Yes	H	L	Did not report relevant performance measures[a]
Aldington et al. [33] 2014	L	L	L	L	L	NA	Unclear[b]	L	Lack of information as meeting abstract only
Stratton Aldington et al. [34] 2014	L	L	L	L	L	No	Unclear[b]	L	Lack of information as meeting abstract only Lacks external validation
Scanlon et al. [36] 2015	L	L	L	H	L	Yes	M[b]	L	High amount of missing information and did CCA
ISDR [16] 2017	L	L	L	L	L	No	M[b]	L	Lacks external validation

EPV events per candidate variables, *EV* external validation, *LOCF* last observation carried forward, *NA* not applicable, *CCA* complete case analysis, *L* low, *M* moderate, *H* high

[a]Calibration, discrimination and/or classification measures

[b]See text

[29, 31] and model performance measures [17, 22, 26, 31, 32, 35]. Out of 14 primary modelling studies, three had no internal validation, and six lacked an external validation. On the whole, external validation studies had smaller samples. The studies are heterogeneous principally due to differing population characteristics, disease classification, outcomes, predictors, their handling/numbers and type of statistical models.

There have been three systematic reviews on topics related to diabetic retinopathy. Lagani et al. [37] examined the probability of complications developing in diabetic patients which included the incidence of diabetic retinopathy. However, they did not consider models for progression of diabetic retinopathy or vision loss (our precise review question). Van der Heijden's systematic review [38] is a conference abstract, so insufficient information was available to make comparison. The population of interest was only T2DM (our review includes T1DM as well). The context was screening and detection of DR (earlier low-risk part of the disease pathway), rather than progression to vision loss and treatment (higher-risk patients, the population of interest in our review). Taylor-Phillips et al. [39] investigated annual against longer screening intervals and concluded that there is insufficient evidence to support extending screening of diabetic patients for STR beyond 1 year. They based their conclusions on the lack of quality and use of different definitions of the low-risk group of patients. Their question was different from ours.

The three models identified as moderate to low risk of bias and low risk for applicability have already shown some impact in diabetic screening in lower-risk patients. Recent work by Scanlon et al. [36], Aspelund et al. [29] and ISDR [16] has clearly shown that individual patient's risk assessment and prediction can be safely and effectively achieved through the use of routine data in pre-STR patients. The evidence from Scanlon et al. [36] is also expected to have an impact on DESP, to risk-stratify patients into those suitable for 2-yearly screening and the relatively higher-risk group for yearly screening. Aspelund et al. [29] have the largest number of external validation studies and the model is being used in practice in Aarhus University. ISDR model is already the subject of an impact study [40]. One of these models could also be updated and tested on a higher-risk hospital patient population as well.

In conclusion, in countries with developed DR screening, for patients who have been referred for treatment or closer observation, a model is needed to determine their individual risk of progression to treatment stage/loss of vision, to direct the resources appropriately and further optimise the services especially for higher-risk patients. This review highlights some of the useful models available for the said purpose. Scanlon et al. [36], Aspelund et al. [29] and ISDR model [16] seem to be appropriate in terms of contemporary

participant data, accessible predictors and sound methodology, though they do not directly address the outcome of our interest. They need further external validation in diverse high-risk settings before being implemented into clinical practice. In addition to these three models, we have listed the predictors and performance of all other models. This means anyone, dependent on their own particular datasets, could use one that suits their needs such as to define screening intervals, or to target screening in poorly resourced countries.

More primary modelling studies particularly for use in hospital eye services will be useful as long as they are well-structured with good reporting. Studies should also ensure to present the final model in a simplified way, to make it easier for clinicians or policy makers to implement. The model's integration into electronic medical records can help decision-making and needs to be the goal in future models.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. NCD Risk Factor Collaboration (NCD-RisC)). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet. 2016;387:1513–30.
2. Mathur R, Bhaskaran K, Edwards E, Lee H, Chaturvedi N, Smeeth L, et al. Population trends in the 10-year incidence and prevalence of diabetic retinopathy in the UK: a cohort study in the Clinical Practice Research Datalink 2004–2014. BMJ Open. 2017;7:e014444.
3. NHS Diabetic Eye Screening Programme Summary statistics for England. In: England PH, editor. Gov.UK; 2016.
4. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. Arch Ophthalmol. 1984;102:520–6.
5. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. Arch Ophthalmol. 1984;102:527–32.
6. RCOPHTH. The Way Forward Executive Summary 2017 20 Jan. 2018; (300117), p. 6–7. https://www.rcophth.ac.uk/wp-content/uploads/2015/10/RCOphth-The-Way-Forward-Executive-Summary-300117.pdf.
7. Ross PL, Gerigk C, Gonen M, Yossepowitch O, Cagiannos I, Sogani PC, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. Semin Urol Oncol. 2002;20:82–8.
8. Lund SH, Aspelund T, Kirby P, Russell G, Einarsson S, Palsson O, et al. Individualised risk assessment for diabetic retinopathy and optimisation of screening intervals: a scientific approach to reducing healthcare costs. Br J Ophthalmol. 2016;100:683–7.
9. Echouffo-Tcheugui JB, Ali MK, Roglic G, Hayward RA, Narayan KM. Screening intervals for diabetic retinopathy and incidence of visual loss: a systematic review. Diabet Med: a J Br Diabet Assoc. 2013;30:1272–92.
10. Wolff R, Whiting P, Mallett S, editors. PROBAST: a risk of bias tool for prediction modelling studies. Cochrane Colloquium Vienna; 2015.
11. Moons KA, Wolff R, Whiting P. Prediction modelling: Where are we now and where do we need to go? 2018. http://www.meduniwien.ac.at/wbs/Moons170126.pdf.
12. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10:e1001381.
13. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11:e1001744.
14. PROSPERO. Prognostic prediction models for the progression of diabetic retinopathy (DR) and vision loss in patients with sight-threatening diabetic retinopathy (STDR): protocol for a systematic review; 2017. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017057767.
15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6:e1000100.
16. Eleuteri A, Fisher AC, Broadbent DM, Garcia-Finana M, Cheyne CP, Wang A, et al. Individualised variable-interval risk-based screening for sight-threatening diabetic retinopathy: the Liverpool Risk Calculation Engine. Diabetologia; 2017;60:2174–2182.
17. Clarke PM, Gray AM, Briggs A, Farmer AJ, Fenn P, Stevens RJ, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). Diabetologia. 2004;47:1747–59.
18. Leal J, Hayes AJ, Gray AM, Holman RR, Clarke PM. Temporal validation of the UKPDS outcomes model using 10-year posttrial monitoring data. Diabetes Care. 2013;36:1541–6.
19. McEwan P, Bennett H, Ward T, Bergenheim K. Refitting of the UKPDS 68 risk equations to contemporary routine clinical practice data in the UK. Pharmacoeconomics. 2015;33:149–61.
20. Soedamah-Muthu SS, Vergouwe Y, Costacou T, Miller RG, Zgibor J, Chaturvedi N, et al. Predicting major outcomes in type 1 diabetes: a model development and validation study. Diabetologia. 2014;57:2304–14.
21. Ng H, Keunen J, Tack C, Nijpels G, Van Der Heijden AAW. Validation of a prediction model to optimise retinopathy screening in type 1 diabetes. Diabetologia. 2016;59(1Supplement 1):S476–7.
22. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. Diabetologia. 2013;56:1925–33.
23. McEwan P, Ward T, Bennett H, Bergenheim K. Validation of the UKPDS 82 risk equations within the Cardiff Diabetes Model. Cost Eff Resour Alloc. 2015;13:12.
24. Tanaka S, Tanaka S, Iimuro S, Yamashita H, Katayama S, Akanuma Y, et al. Predicting macro- and microvascular complications in type 2 diabetes: the Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. Diabetes Care. 2013;36:1193–9.
25. van der Heijden AA, Walraven I, van 't Riet E, Aspelund T, Lund SH, Elders P, et al. Validation of a model to estimate personalised

screening frequency to monitor diabetic retinopathy. Diabetologia. 2014;57:1332–8.

26. Holtzer-Goor KM, Van Der Heijden AA, Jonker M, Stolk E, Nijpels G. Validation of an algorithm to predict the risk of sight threatening retinopathy in a multi-ethnic patient group treated in a Dutch hospital. Diabetologia. 2015;1:S525–6.

27. Lagani V, Chiarugi F, Thomson S, Fursse J, Lakasing E, Jones RW, et al. Development and validation of risk assessment models for diabetes-related complications based on the DCCT/EDIC data. J Diabetes its Complicat. 2015;29:479–87.

28. Hippisley-Cox J, Coupland C. Development and validation of risk prediction equations to estimate future risk of blindness and lower limb amputation in patients with diabetes: Cohort study. BMJ (Online). 2015;351:h5441.

29. Aspelund T, Porisdottir O, Olafsdottir E, Gudmundsdottir A, Einarsdottir AB, Mehlsen J, et al. Individual risk assessment and information technology to optimise screening frequency for diabetic retinopathy. Diabetologia 2011;54:2525–32.

30. Soto-Pedre E, Pinies JA, Hernaez-Ortega MC. External validation of a risk assessment model to adjust the frequency of eye-screening visits in patients with diabetes mellitus. J Diabetes its Complicat. 2015;29:508–11.

31. Mehlsen J, Erlandsen M, Poulsen PL, Bek T. Individualized optimization of the screening interval for diabetic retinopathy: a new model. Acta Opthalmologica. 2012;90:109–14.

32. Stratton IM, Aldington SJ, Taylor DJ, Adler AI, Scanlon PH. A simple risk stratification for time to development of sight-threatening diabetic retinopathy. Diabetes care. 2013;36:580–5.

33. Aldington SJ, Stratton IM, Scanlon PH. Validation of a simple stratification algorithm for progression to sight threatening diabetic retinopathy. Diabetologia. 2014;1:S474–5.

34. Stratton IM, Aldington SJ, Farmer AJ, Scanlon PH. Personalised risk estimation for progression to sight-threatening diabetic retinopathy: How much does clinical information add to screening data? Diabet Med. 2014;31:23–4.

35. Harris Nwanyanwu K, Talwar N, Gardner TW, Wrobel JS, Herman WH, Stein JD. Predicting development of proliferative diabetic retinopathy. Diabetes care. 2013;36:1562–8.

36. Scanlon PH, Aldington SJ, Leal J, Luengo-Fernandez R, Oke J, Sivaprasad S, et al. Development of a cost-effectiveness model for optimisation of the screening interval in diabetic retinopathy screening. Health Technol Assess. 2015;19:1–116.

37. Lagani V, Koumakis L, Chiarugi F, Lakasing E, Tsamardinos I. A systematic review of predictive risk models for diabetes complications based on large scale clinical studies. J Diabetes its Complicat. 2013;27:407–13.

38. Van der Heijden AA, Badloe F, Nijpels G, Beulens JW, editors. Prediction models for the risk of Retinopathy in people with type 2 Diabetes. A Systematic Review. 27th European Association for the Study of Diabetes Eye Complications Study Group; May 2017.

39. Taylor-Phillips S, Mistry H, Leslie R, Todkill D, Tsertsvadze A, Connock M, et al. Extending the diabetic retinopathy screening interval beyond 1 year: Systematic review. Br J Ophthalmol. 2016;100:105–14.

40. ISDR. http://www.isdrprojectcouk/rcthtml; 2018.