Development and validation of a new clinical decision support tool to optimize screening for retinopathy of prematurity

Aldina Pivodic ![ORCID],[1] Helena Johansson,[2,3] Lois E H Smith,[4] Anna-Lena Hård,[1] Chatarina Löfqvist,[1,5] Bradley A Yoder,[6] M Elizabeth Hartnett,[7] Carolyn Wu,[4] Marie-Christine Bründer,[8] Wolf A Lagrèze,[9] Andreas Stahl ![ORCID],[8] Abbas Al-Hawasi,[10] Eva Larsson ![ORCID],[11] Pia Lundgren ![ORCID],[1,12] Lotta Gränse,[13] Birgitta Sunnqvist,[14] Kristina Tornqvist,[13] Agneta Wallin,[15] Gerd Holmström ![ORCID],[11] Kerstin Albertsson-Wikland,[16] Staffan Nilsson,[17,18] Ann Hellström ![ORCID][1]

ABSTRACT

Background/Aims Prematurely born infants undergo costly, stressful eye examinations to uncover the small fraction with retinopathy of prematurity (ROP) that needs treatment to prevent blindness. The aim was to develop a prediction tool (DIGIROP-Screen) with 100% sensitivity and high specificity to safely reduce screening of those infants not needing treatment. DIGIROP-Screen was compared with four other ROP models based on longitudinal weights.

Methods Data, including infants born at 24–30 weeks of gestational age (GA), for DIGIROP-Screen development (DevGroup, N=6991) originate from the Swedish National Registry for ROP. Three international cohorts comprised the external validation groups (ValGroups, N=1241). Multivariable logistic regressions, over postnatal ages (PNAs) 6–14 weeks, were validated. Predictors were birth characteristics, status and age at first diagnosed ROP and essential interactions.

Results ROP treatment was required in 287 (4.1%)/6991 infants in DevGroup and 49 (3.9%)/1241 in ValGroups. To allow 100% sensitivity in DevGroup, specificity at birth was 53.1% and cumulatively 60.5% at PNA 8 weeks. Applying the same cut-offs in ValGroups, specificities were similar (46.3% and 53.5%). One infant with severe malformations in ValGroups was incorrectly classified as not needing screening. For all other infants, at PNA 6–14 weeks, sensitivity was 100%. In other published models, sensitivity ranged from 88.5% to 100% and specificity ranged from 9.6% to 45.2%.

Conclusions DIGIROP-Screen, a clinical decision support tool using readily available birth and ROP screening data for infants born GA 24–30 weeks, in the European and North American populations tested can safely identify infants not needing ROP screening. DIGIROP-Screen had equal or higher sensitivity and specificity compared with other models. DIGIROP-Screen should be tested in any new cohort for validation and if not validated it can be modified using the same statistical approaches applied to a specific clinical setting.

INTRODUCTION

Retinopathy of prematurity (ROP) is a sight-threatening disease occurring mainly in extremely preterm infants.[1] Screening for severe ROP, for which treatment can prevent blindness, comprises repeated eye examinations following national screening guidelines, mostly using birth parameters, gestational age (GA) and birth weight.[2] These examinations are stressful, costly and very inefficient.[3–6] In Sweden and in the USA, only ~6% of screened infants need treatment for ROP.[7 8] The number of ROP examinations and need for treatment are increasing over time with improved neonatal healthcare that increases the number of infants surviving extreme prematurity.[9 10] A prediction model including known risk factors at birth and postnatal parameters using statistical approaches enabling risks to vary over time could identify the time to safely end ROP screening as well as identify low-risk infants requiring fewer or no ROP examinations. Such a clinical decision support tool would be valuable both for infants, and health economics. Reducing the number of examinations would not only reduce the stress and pain, but also for example, avoid the transport of infants to the screening unit, change of daily routines and potential exposure to infections during transport and at the hospital. Even if stress is minimised during ROP screening, the examinations may still affect the infants systemically with such as increased tachycardia and apnoeic episodes. From a health economics perspective, such models would help optimise the use of healthcare personnel to focus on the babies who need careful monitoring.

Many models predicting ROP requiring treatment have been published during the past two decades, such as weight, insulin-like growth factor 1, neonatal, ROP (WINROP), Colorado-ROP (CO-ROP), Children's Hospital of Philadelphia-ROP (CHOP-ROP), postnatal growth and retinopathy of prematurity (G-ROP) and Omaha-ROP (OMA-ROP).[11–20] A systematic review of 23 studies, performed by the American Academy of Ophthalmology (AAO) in 2016, developing or validating prediction models for different ROP outcomes found no model development study, and only one model validation study judged as good quality.[21 22] The AAO concluded that prediction model development at the time was still in its early phase and

needed rigorous implementation of guidelines for generating prognoses, including larger sample sizes and assessment of generalisability.

Our research group has previously published the prediction model (WINROP) which was based on birth parameters with the addition of first longitudinal serum insulin-like growth factor 1 (IGF-1) levels (that were difficult to obtain), then based on postnatal growth reflecting postnatal IGF-1.[11 12 22 23] This model, used to identify low-risk and high-risk infants, did not always achieve 100% sensitivity and had variable specificity. Recently, we published a prediction model for ROP requiring treatment, DIGIROP-Birth, for infants born at GA 24–30 weeks, estimating individual risks at an early stage based on birth characteristics alone (GA, birth weight and sex), as weight measurements at specific postnatal periods are not always available to the screening ophthalmologist and/or neonatologist. We applied statistical methods enabling description of the actual development of risk for severe ROP postnatally for each individual infant.[24]

In the current study, we extended DIGIROP-Birth into DIGIROP-Screen to also include ROP progression data. Based on the estimated predictions we created a clinical decision support tool to reduce the burden of ROP screening sessions. As well as identifying infants who do not develop severe ROP in our cohort, we also sought to identify the time point when the longitudinal screening process could safely end in infants who had some risk of developing severe ROP during their postnatal course. To our knowledge this has not been studied previously. Internal and external validations, and comparisons (with respect to sensitivity and specificity of predicting severe ROP) to four other published models (WINROP, CO-ROP, CHOP-ROP, OMA-ROP), were performed.[12 16–18] The aim was to develop and validate models with 100% sensitivity to capture all infants requiring treatment and the highest specificity to reduce examinations in infants not developing severe ROP in our cohort and the validation cohorts using parameters that were easily available to ophthalmologists. The algorithm must be validated in any new cohort before being adopted to show that the same 100% sensitivity and high specificity apply. If 100% sensitivity and high specificity are not validated, using the same statistical

approaches used in DIGIROP-Screen development the prediction model can be modified for any new clinical setting.

MATERIALS AND METHODS

This study has followed the guidelines for Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis.[25]

Study population

The data, including infants' birth characteristics and the timing of progression of ROP through stages and treatment, originate from the Swedish National Registry for ROP (SWEDROP) that is part of the Swedish Neonatal Register, and was initiated in 2007.[7 26] The registry has high coverage, ~97%, and collects data about the number of eye exams, dates for first and last eye exam, presence of ROP at the first eye screening, ROP stage, zone, plus disease, treatment, type of ROP, maximum stage and most central zone for ROP left and right eye and the date for first observation of respective ROP stage. The incomplete and missing data were validated against medical records. A study flowchart describing *model development group* and *validation groups* is presented in figure 1.

Model development group (DevGroup)

Of 7031 infants born at GA 24–30 weeks, between 1 January 2007 and 24 October 2017, 6991 (99.4%) were eligible for inclusion in the *model development group*. Twenty-four (0.3%) infants were excluded due to missing birth characteristics data and 13 (0.2%) due to missing or inconsistent follow-up data. Additionally, three infants were identified as outliers during model development. They were treated despite not fulfilling treatment criteria for type 1 ROP (ROP stage 3 zone III, at the most one clock hour).

Validation groups (ValGroups)

Infants born at GA 24–30 weeks between 1 November 2017 and 7 August 2018 (n=318) and registered in SWEDROP were considered for inclusion in the *Swedish temporal validation*
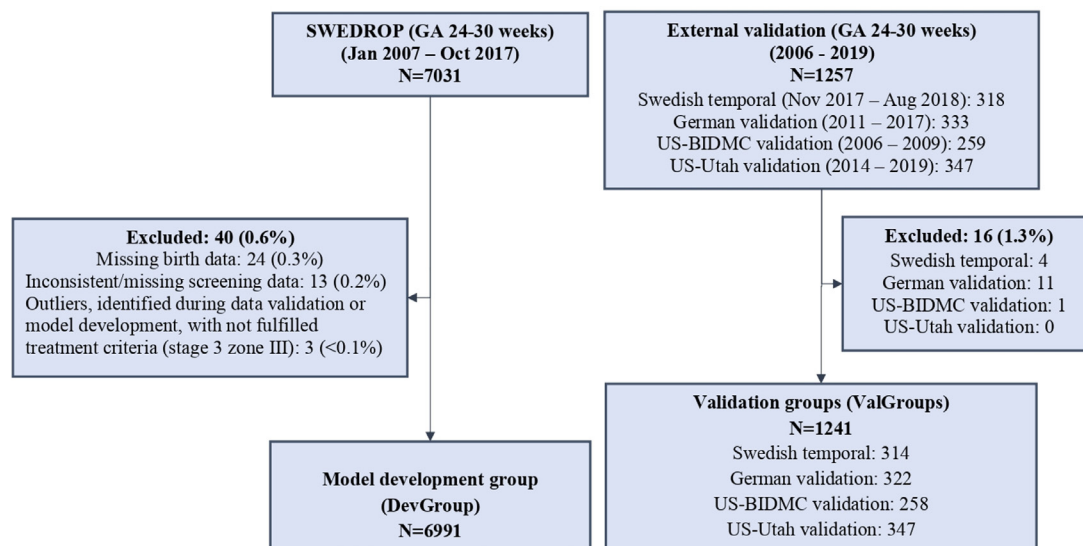

Figure 1 Study flowchart. BIDMC, Beth Israel Deaconess Medical Center; GA, gestational age; ROP, retinopathy of prematurity; SWEDROP, Swedish national Registry for ROP.

group. Four (1.3%) infants were excluded for missing data, leaving 314 (98.7%) eligible infants for validation.

Retrospectively collected data from 2011 to 2017 from a German site in Freiburg included 322 (96.7%) out of 333 infants born at GA 24–30 weeks and served as the *German validation group.*[27] Eleven (3.3%) infants were excluded due to either missing birth weight or GA (n=4), or unavailable ROP progression data (n=7).

The *US-BIDMC validation group* included 258 (99.6%) out of 259 infants born at GA 24–30 weeks between 2006 and 2009 from the US site Beth Israel Deaconess Medical Center (BIDMC), Boston, Massachusetts.[22] One infant was excluded due to unavailable ROP progression data. For this cohort, information about race and ethnicity was available in 240 (93.0%)/258 and used to test the model's predictive ability for a white (n=177) and a non-white (n=63) population.

The *US-Utah validation group* included 347 (100%)/347 infants born at GA 24–30 weeks between 2014 and 2019 from the US site John A. Moran Eye Center, Salt Lake City, Utah.

The two US cohort files contained infants' weekly weights and were used to compare four other ROP models (WINROP, CO-ROP, CHOP-ROP, OMA-ROP) using postnatal weight gain as input.[12 16–18]

In total, 1241 infants were included in the *validation groups* (ValGroups).

Study procedures

Fetal ultrasound was used to estimate GA in all cohorts. The postnatal age (PNA), postmenstrual age and GA are defined according to the American Academy of Pediatrics policy.[28] Birth weight standard deviation scores (BWSDS) were calculated based on birth weight, GA and sex using a Swedish reference for 800 000 healthy singletons (of ~1 million born) born at GA ≥24 weeks during 1990–1999.[29]

Study outcome and predictors

The study outcome is ROP treated following early treatment for ROP criteria or if judged required by the examining/treating ophthalmologist.[30] ROP stages were defined by the International Classification of ROP.[31] The infant's status (yes/no), age at the first sign of ROP and weeks since the first sign of ROP were potential predictors tested for inclusion in the DIGIROP-Screen model, besides the log-odds for the DIGIROP-Birth probabilities (log(probability/(1−probability))), GA, sex, BWSDS and important interactions. The final models included log-odds for the DIGIROP-Birth probabilities, infant's status and age at the first sign of ROP and interaction between them. Data were analysed on patient-level, including first occurrence of any ROP as predictor and first ROP treatment as outcome.

Statistical analysis

Continuous variables were presented by mean, SD, median and range and categorical variables by number and percentage. The difference between DevGroup and ValGroups was tested using Fisher's exact test for dichotomous variables, Mantel-Haenszel χ^2 trend test for ordered categorical variables and Mann-Whitney U test for continuous variables. The estimated risk predictions from DIGIROP-Birth were applied at birth. Multivariable logistic regression was used for PNAs 6–14 weeks. PNA week 6 is the earliest week when infants are starting their screening per guidelines. By PNA week 14 it was expected that majority of ROP treatments have occurred. GA-specific cut-offs based on estimated probabilities for 100% sensitivity were

retrieved from the models performed on DevGroup and used for implementation of the clinical decision support tool. Specificities and cumulative specificities that is a fraction of infants below the cut-off at the current time or earlier among the non-treated infants were obtained with 95% CI. Internal validation, examining the model's reproducibility in its cohort, was performed by 10-fold cross-validation. The Hosmer-Lemeshow test examined goodness-of-fit and calibration of observed versus the estimated number of events. External validation, analysing the model's generalisability/transportability in the cohorts from other healthcare settings, populations and periods were assessed on ValGroups by describing sensitivity, specificity and cumulative specificity with 95% CI based on cut-offs for 100% sensitivity obtained from DevGroup. In order to achieve the recommended lower 95% limit for 100% sensitivity of 99%, ~300 events (ROP treatment) were needed, that was fulfilled by the DevGroup sample. Sensitivity and cumulative specificity/specificity with 95% CI were presented for DIGIROP-Screen and the four ROP comparison models based on the two US external validation cohorts combined. Detailed descriptions of the statistical methods are available in online supplemental eappendix 1. Graphical workflow of DIGIROP-Screen and four comparison models are presented in online supplemental efigure 1.

All tests were two-tailed and p<0.05 was considered significant. Analyses were performed using SAS software V.9.4 (SAS Institute Inc, Cary, North Carolina, USA).

RESULTS
Study population

Birth characteristics and ROP progression for the DevGroup (N=6991) and the four cohorts included in the ValGroups (N=1241) are presented in table 1. In the DevGroup, 3158 (45.2%) were girls. Mean GA was 28.3 (SD 1.9) weeks, mean birth weight 1146 (SD 339, range 307–3245) grams and mean BWSDS −1.03 (SD 1.37). In DevGroup and ValGroups, respectively, 2026 (29.0%) and 502 (40.5%) were diagnosed with any ROP, and 287 (4.1%) and 49 (3.9%) were treated for ROP.

ValGroups included more girls, had lower average birth weight, differed with respect to the birth year and more infants experienced any ROP compared with DevGroup. Online supplemental etable 1 describes infant characteristics for the validation cohorts and online supplemental etable 2 for treated and not treated infants.

DIGIROP-Screen in model development group (DevGroup)

The multivariable logistic models for DIGIROP-Screen at birth and over PNAs 6–14 weeks are presented in online supplemental etable 3 and cut-offs based on estimated probabilities in online supplemental etable 4. Estimated probabilities for ROP treatment stratified by GA at birth (24–30 weeks) for different PNA are presented in online supplemental efigure 2 A–J. The area under the receiver operating characteristic curve (AUC) ranged between 0.91 and 0.93 (online supplemental etable 5, efigure 3). For selected cut-offs for 100% (95% CI: 98.7% to 100%) sensitivity in DevGroup, specificity at birth was 53.1% (95% CI: 51.9% to 54.3%), cumulatively at 8 weeks 60.5% (95% CI: 59.3% to 61.7%) and cumulatively at 12 weeks PNA 75.5% (95% CI: 74.5% to 76.5%) (table 2, online supplemental etable 6, efigure 4). The prediction models' contribution at 6, 7 and 14 weeks PNA to the increase of cumulative specificity was negligible. Among infants flagged as not needing ROP screening already at birth 3179 (89.2%) were diagnosed with no ROP, 202 (5.7%) with ROP stage 1, 137 (3.8%) with untreated stage 2 and

Table 1 Infants' characteristics at birth, first sign of ROP and ROP treatment

	Model development group (N=6991)	Validation groups (N=1241)	P value*
Girls	3158 (45.2%)	607 (48.9%)	0.016
GA at birth (weeks)	28.3 (1.9) 28.6 (24.0; 30.9)	28.0 (1.9) 28.3 (24.0; 30.9)	<0.0001
GA (full weeks)			0.0007
24	427 (6.1%)	94 (7.6%)	
25	597 (8.5%)	114 (9.2%)	
26	781 (11.2%)	129 (10.4%)	
27	914 (13.1%)	187 (15.1%)	
28	1141 (16.3%)	239 (19.3%)	
29	1419 (20.3%)	236 (19.0%)	
30	1712 (24.5%)	242 (19.5%)	
Birth weight (grams)	1146 (339) 1135 (307; 3245)	1068 (319) 1065 (335; 2450)	<0.0001
Birth weight SDS (Niklasson and Albertsson-Wikland 2008)	−1.03 (1.37) −0.77 (−8.56; 4.93)	−1.31 (1.57) −0.99 (−9.92; 2.75)	<0.0001
Birth year			<0.0001
2006–2007	543 (7.8%)	139 (11.2%)	
2008–2009	1331 (19.0%)	119 (9.6%)	
2010–2011	1303 (18.6%)	9 (0.7%)	
2012–2013	1369 (19.6%)	103 (8.3%)	
2014–2015	1445 (20.7%)	249 (20.1%)	
2016–2017	1000 (14.3%)	331 (26.7%)	
2018–2019	0 (0.0%)	291 (23.4%)	
Any ROP	2026 (29.0%)	502 (40.5%)	<0.0001
PNA at first diagnosed ROP (weeks)	8.35 (2.22) 8.14 (3.43; 18.71)	8.07 (2.56) 7.71 (3.14; 19.00)	
ROP treatment	287 (4.1%)	49 (3.9%)	0.87
PNA at ROP treatment (weeks)	12.8 (2.8) 12.4 (7.0; 21.9)	12.3 (2.4) 11.9 (7.1; 19.6)	

Model development group includes data from Swedish National Registry for ROP, born at GA 24–30 weeks (2007–2017).
Validation groups consist of four external validation cohorts, one from Sweden (later time period than in the model development group), one from Germany and two from USA.
For categorical variables, n (%) is presented. For continuous variables, mean (SD)/median (min; max) is given.
*P values should be interpreted with caution due to the large cohorts. Conclusions should be made based on the clinically relevant differences.
GA, gestational age; PNA, postnatal age; ROP, retinopathy of prematurity; ROP, retinopathy of prematurity; SDS, standard deviation score.

Table 2 Specificity with 95% CI for 100% sensitivity at birth and over postnatal weeks for model development group (N=6991), and external validation groups (N=1241)

	Model development group (DevGroup)* Specificity (95% CI) N=6991	External validation (ValGroups)† Specificity (95% CI) N=1241
At birth	53.1 (51.9 to 54.3)	46.3 (43.4 to 49.2)
Cumulatively at PNA 6 weeks	53.3 (52.0 to 54.5)	46.4 (43.5 to 49.3)
Cumulatively at PNA 7 weeks	54.2 (53.0 to 55.4)	47.2 (44.4 to 50.1)
Cumulatively at PNA 8 weeks	60.5 (59.3 to 61.7)	53.5 (50.6 to 56.4)
Cumulatively at PNA 9 weeks	67.6 (66.5 to 68.7)	61.2 (58.3 to 63.9)
Cumulatively at PNA 10 weeks	72.1 (71.0 to 73.2)	65.9 (63.1 to 68.5)
Cumulatively at PNA 11 weeks	75.3 (74.3 to 76.4)	69.3 (66.6 to 71.9)
Cumulatively at PNA 12 weeks	75.5 (74.5 to 76.5)	69.6 (66.9 to 72.2)
Cumulatively at PNA 13 weeks	80.6 (79.6 to 81.5)	75.2 (72.6 to 77.6)
Cumulatively at PNA 14 weeks	80.6 (79.7 to 81.6)	75.2 (72.6 to 77.6)

Model development group includes data from SWEDROP, born at GA of 24–30 weeks (2007–2017).
Validation groups consist of four external validation cohorts, one from Sweden (later time period than in model development group), one from Germany and two from USA.
Cumulative specificity at a certain PNA is calculated as a union of specificities up to and including that certain PNA.
*Cut-offs selected in model development group for sensitivity 100%.
†For validation groups, cut-offs obtained from model development group are applied.
Sensitivity 100% for all postnatal weeks except for one infant at birth, and PNA 6 and 7 weeks (sensitivity 48/49 at those time points), with severe comorbidity profile.
GA, gestational age; PNA, postnatal age; ROP, retinopathy of prematurity; SWEDROP, Swedish National Registry for ROP.

and cumulatively up to PNA 8 weeks was 60.0% (95% CI: 58.8% to 61.1%) for internal validation.

Hosmer-Lemeshow test was non-significant at all PNAs (online supplemental etable 3), indicating goodness-of-fit accepted as satisfactory, and showed a well-calibrated estimated versus the observed number of events.

External validation of DIGIROP-Screen in validation groups (ValGroups)
Individual risk predictions stratified by GA at birth (24–30 weeks) over PNA are presented in online supplemental efigure 5.

Applying the same cut-offs on ValGroups, as those obtained for DevGroup (for 100% sensitivity), the specificities were 46.3% (95% CI: 43.4% to 49.2%) at birth, 53.5% (95% CI: 50.6% to 56.4%) cumulatively at 8 weeks and 69.6% (95% CI 66.9% to 72.2%) cumulatively at 12 weeks PNA (table 2, online supplemental etable 6, efigure 6). In ValGroups, sensitivity was 100% (95% CI: 92.7% to 100%) for all models except for one infant at birth and PNAs 6 and 7 weeks. By inclusion criteria for current ROP screening, this infant should have been followed and screened because of the medical indication. At birth (GA 30 weeks) the infant had VACTERL association (vertebral defects, anal atresia, cardiac defects, tracheo-esophageal fistula, renal abnormalities, limb abnormalities) with severe intrauterine growth restriction.

Stratified by GA <28 and ≥28 weeks, specificity at birth was 11.3% and 69.7%, and cumulatively up to 12 weeks 35.4% and 92.6%, respectively.

Figure 2C and online supplemental efigure 7 illustrate the number of infants who could potentially be released from ROP screening cumulatively over PNAs according to last examination reported in SWEDROP, according to DIGIROP-Screen in DevGroup and ValGroups.
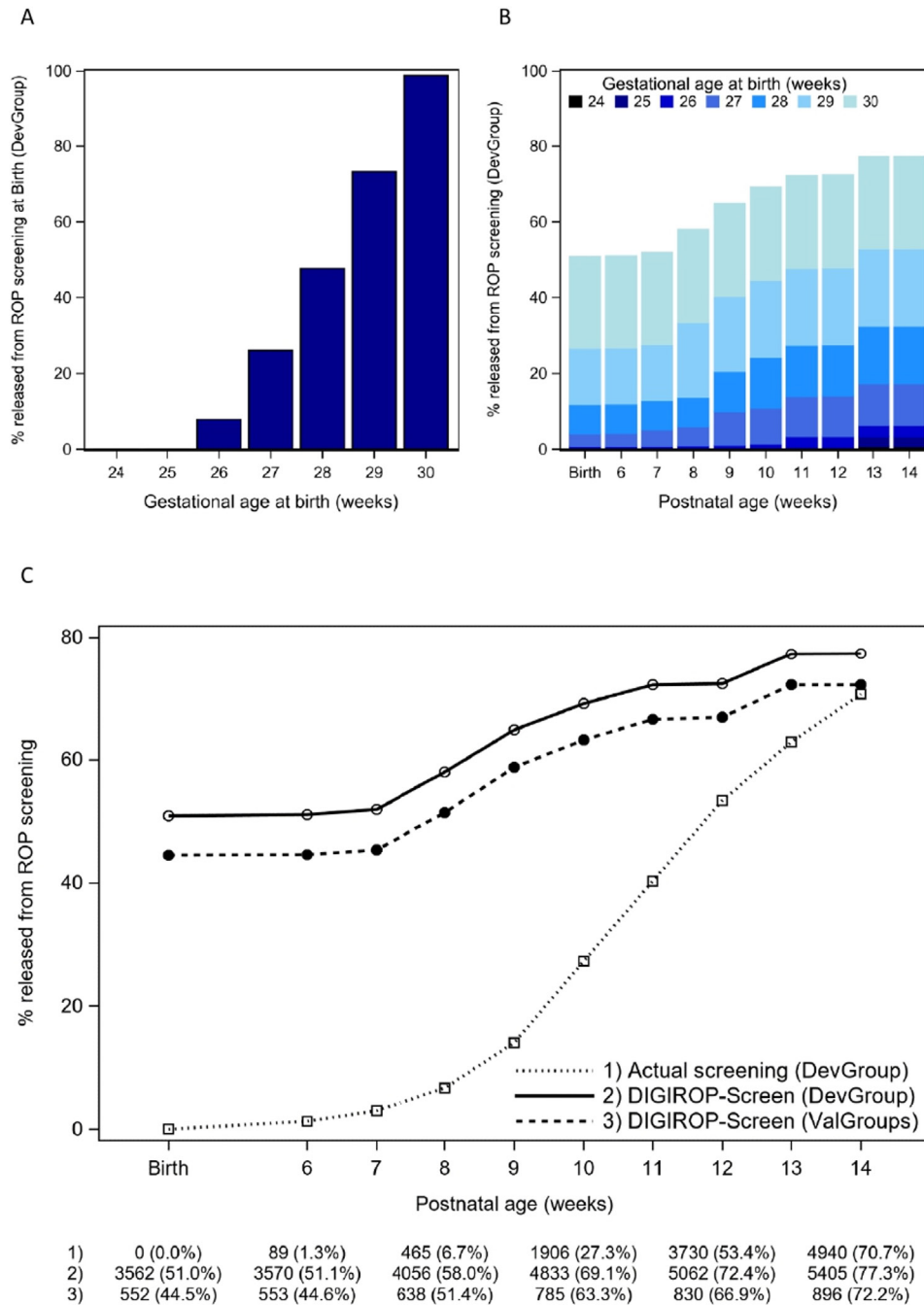
44 (1.2%) with untreated stage 3 (online supplemental etable 7). No infants born at GA 24 and 25 weeks could be released from ROP screenings at birth (figure 2A). Percentages of infants identified as possible to be released from ROP screenings over PNA stratified by GA at birth are presented in figure 2B.

Stratified by GA <28 and ≥28 weeks, specificity at birth was 11.9% and 76.8%, and cumulatively up to 12 weeks 40.6% and 95.5%, respectively (online supplemental etable 6). The corresponding specificities for GA <30 weeks were 37.4% at birth and 67.1% cumulatively up to 12 weeks PNA.

Internal validation of DIGIROP-Screen in model development group (DevGroup)
Specificity, cumulative specificity and AUC with 95% CI obtained from the 10-fold cross-validation were obtained from logistic regression models developed on DevGroup (online supplemental etable 5 and 6). The AUC ranged between 0.90 and 0.94 (online supplemental etable 5). The specificity at birth was 48.0% (95% CI: 46.8% to 49.2%),

Figure 2 Illustration of infants born 24–30 weeks of gestational age released from screening for ROP according to: (A) risk predictions from DIGIROP-Screen at birth by gestational age in model development group (DevGroup), (B) risk predictions from DIGIROP-Screen over postnatal ages by gestational age in DevGroup, (C) last examination date reported in SWEDROP, and risk predictions from DIGIROP-Screen in DevGroup and validation groups (ValGroups). In (C), n and % are presented for time points: birth, postnatal ages 6, 8, 10, 12 and 14 weeks. ROP, retinopathy of prematurity; SWEDROP, Swedish National Registry for ROP.

Information about race and ethnicity was available in the *US-BIDMC validation group*. Stratifying by infants reported as white (n=177, one required ROP treatment) and those reported as non-white (n=63, three required ROP treatment) infants, specificity at birth was 54.5% and 38.3%, and cumulatively up to 12 weeks 65.9% and 56.7%, respectively (online supplemental etable 6).

The AUC for the models at birth and over different PNAs ranged between 0.88 and 0.92 (online supplemental etable 5).

Comparison of DIGIROP-Screen to other ROP prediction models

DIGIROP-Screen was compared with four other published models using *US validation groups*, as comparison cohorts table 3.

With 100% sensitivity cut-off, DIGIROP-Screen versus CHOP-ROP[17] had better specificity (48.7% vs 27.5%) at 8 weeks and better specificity at 12 weeks PNA (63.6% vs 27.9%). DIGIROP-Screen versus OMA-ROP[18] had the same sensitivity

Table 3 Comparison of DIGIROP-Screen versus other existing ROP prediction models

Comparison model and time point	N/N Specificity (95% CI) (%)		N/N Sensitivity (95% CI) (%)	
	DIGIROP-Screen	Comparison ROP model	DIGIROP-Screen	Comparison ROP model
DIGIROP-Screen (up to PNA 8 w) vs CHOP-ROP[17] (up to PNA 8 w)	278/571 48.7 (44.5 to 52.9)	157/571 27.5 (23.9 to 31.4)	26/26 100.0 (86.8 to 100.0)	26/26 100.0 (86.8 to 100.0)
DIGIROP-Screen (up to PNA 12 w) vs CHOP-ROP[17] (up to PNA 12 w)	362/569 63.6 (59.5 to 67.6)	159/569 27.9 (24.3 to 31.8)	26/26 100.0 (86.8 to 100.0)	26/26 100.0 (86.8 to 100.0)
DIGIROP-Screen (up to PMA 36 w) vs OMA-ROP[18] (up to PMA 36 w)	250/541 46.2 (41.9 to 50.5)	206/541 38.1 (34.0 to 42.3)	24/25 96.0 (79.6 to 99.9)	24/25 96.0 (79.6 to 99.9)
DIGIROP-Screen (up to WINROP risk flag or last measurement) vs WINROP[12]	256/568 45.1 (40.9 to 49.3)	257/568 45.2 (41.1 to 49.4)	25/26 96.2 (80.4 to 99.9)	23/26 88.5 (69.8 to 97.6)
DIGIROP-Screen (at birth) vs CO-ROP[16] (at PNA 4 w)	231/564 41.0 (36.9 to 45.1)	54/564 9.6 (7.3 to 12.3)	25/26 96.2 (80.4 to 99.9)	25/26 96.2 (80.4 to 99.9)

CHOP-ROP, Children's Hospital of Philadelphia-ROP; CO-ROP, Colorado-ROP; OMA-ROP, Omaha ROP; PMA, postmenstrual age; PNA, postnatal age; ROP, retinopathy of prematurity; w, weeks; WINROP, weight, insulin-like growth factor 1, neonatal, ROP.

(96.0% vs 96.0%), but better specificity (46.2% vs 38.1 %). DIGIROP-Screen versus WINROP[12] had better sensitivity (96.2% vs 88.5%) and similar specificity (45.1% vs 45.2%). DIGIROP-Screen applied at birth versus CO-ROP[16] had similar sensitivity (96.2% vs 96.2%) and better specificity (41.0% vs 9.6%).

Clinical implications
The DIGIROP-Screen prediction tool comprising automatically calculated individual risk predictions for infants born at GA 24–30 weeks is available at www.digirop.com.[32] Additionally, evaluations of the risks based on defined cut-offs provide information whether any/further ROP examinations are required or not for 100% sensitivity (in these cohorts). Example illustrations following a specific infant over screening PNAs planned for availability in the application are presented in online supplemental efigure 8.

DISCUSSION
In this study, we developed an ROP clinical decision support tool, DIGIROP-Screen, for infants born at GA 24–30 weeks, suitable for longitudinal use with ROP screening. The tool is developed to identify the time point for safe release of an infant from the ROP screening. DIGIROP-Screen is based on the infants' birth characteristics (GA, birth weight and sex) and ROP data that are easily obtained at almost all medical facilities while performing routine ROP screening. Other models use longitudinal weights at specific intervals which are less readily available to ophthalmologists and less retrievable on a national level for all screened infants. The prediction tool applied to several cohorts of infants screened for ROP by current criteria in advanced neonatal intensive care unit (NICU) settings, identified early ~45% of infants as not needing any ROP screening using only neonatal characteristics and identified an additional 25% for whom screening may be terminated earlier than with today's screening practice, thus potentially substantially and safely reducing the number of screening examinations. The prediction tool is made available as an online application, www.digirop.com, to clinicians worldwide. This tool must be validated and assessed in each specific clinical setting, before being implemented for routine use. Using the same statistical approaches used in DIGIROP-Screen

development the prediction model can be modified for any new clinical setting.

Studying a low-incidence disease requiring 100% sensitivity, that is, correctly identified all high-risk infants requiring ROP treatment, implies the need for access to very large datasets. The lower 95% confidence limit for sensitivity would need to approach 99%, as previously discussed.[21 33] Our study, which included ~7000 infants and 287 endpoints (ie, ROP treatment), reaches this goal. Larger datasets imply larger individual variability and thus also increased risk for outlying data in the cohort. Having the diagnostic cut-offs in such large datasets based on the individually estimated risks (potentially including outliers) together with the requirement of 100% sensitivity, most often results in low specificity, that is, correctly identified all low-risk infants not needing treatment that might be released from the ROP screening. In the external validation, DIGIROP-Screen demonstrated specificity of 46% at birth (11% for GA <28 weeks, 70% for GA ≥28 weeks), and 70% for data used up to postnatal week 12 (35% for GA <28 weeks, 93% for GA ≥28 weeks), compared with 11% with the updated CHOP-ROP model using longitudinal weekly weights and 33% for the G-ROP algorithm that screens all infants <28 weeks of GA at birth.[17 34] Smaller datasets, on the other hand, including 191 to 560 infants in model development have resulted in higher specificities ranging from 62% to 85% for achieved 100% sensitivity.[12 18 20] Nonetheless, in our US validation cohorts of ~600 infants, DIGIROP-Screen appeared to be a more accurate prediction model than the four comparison models, for both sensitivity and specificity. Unfortunately, the weight measurements at 10, 19, 20, 29, 30 and 39 postnatal days were not available for DIGIROP-Screen precluding a full comparison to the G-ROP screening criteria.[34]

The high performance of DIGIROP-Screen even at birth, applying only DIGIROP-Birth risk estimations, is achieved due to the availability of a large model development dataset and the most prominent risk factors for ROP treatment, GA and birth weight. However, as well known, these are not the only important risk factors, which is why the obtained probabilities showed high variability between GA that resulted in the decision to apply GA-specific cut-offs as scores rather than probabilities in the prediction tool.

The infant with congenital VACTERL association was incorrectly flagged as not needing ROP screening. In current clinical practice, any very preterm baby with severe congenital malformations would have had continuous clinical and medical surveillance for ROP. The medical evaluation is of paramount importance, no matter how high predictive ability is achieved of any model. Likewise, babies born before 24 weeks of GA are all at a very high risk for developing ROP (89%) and all should be followed closely.[24] Optimisation of screening through general prediction models for these babies is inapplicable and inappropriate.

Our study's strength is the large national cohort, the validation datasets originating from two continents and the selected statistical methods. Development, validation and evaluation of the prediction tool followed the prognostic research guidelines.[25] Another strength is the wide availability of birth and ROP progression data and easy access to the model that might facilitate screening for ophthalmologists. Identification of infants as potentially requiring no further screening after a defined date may safely decrease the number of unnecessary examinations for low-risk infants after affirming that the model applies to the particular cohort under consideration.

Our study's limitation is its retrospective design and registry data, although intense efforts were made to validate incomplete data points. Ongoing research including photographic documentation and telemedicine will certainly decrease the variability in ROP diagnostics between ophthalmologists, and hence also improve sensitivity and specificity of prediction models.[35] A second limitation is the small subgroup of non-white infants used for validation. In many countries, screening of infants born <31 weeks of GA is mandatory, although in Sweden the current guidelines from 2020 recommend screening for GA <30 weeks.[7 36–38] Our tool was developed to study infants born at GA 24 weeks (+0 days) to 30 weeks (+6 days). However, infants born at 31 weeks of GA or later who require screening based on a medical indication should be monitored closely and carefully, as should infants born <24 weeks of GA, all of whom have a high risk for developing ROP needing treatment. This algorithm which is aimed at identifying the time point for ending ROP screening is thus of limited value to these babies as they need screening according to guidelines. Another limitation is that no validation has been performed on populations from low-income countries where more mature infants need treatment for ROP due to the risk associated with unmonitored oxygen exposure, but also from countries with high-level neonatal care but with limited facilities and personnel. Continued validation, performed on similar and different populations is needed. The model parameters or even the model selection, including other important variables, might need to be updated to match some specific healthcare settings. The future implementation of this tool at our or any other NICU should concomitantly initiate a clinical study monitoring its effectiveness (including stress reduction), impact on patient safety as well as on the clinical workload and health economics.

In conclusion, the DIGIROP-Screen, an internally and externally validated ROP prediction tool, is available to be applied to infants, born at GA 24–30 weeks, at birth and also applied during the routine ROP screening process. The tool may allow ophthalmologists to reduce the number of stressful examinations and optimise screening efficiency by potentially and safely releasing many infants from unnecessary eye examinations. DIGIROP-Screen appears to be one of the more robust models predicting severe ROP requiring treatment.

Author affiliations
[1]Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[2]Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Victoria, Australia
[3]Sahlgrenska Osteoporosis Centre, Institute of Medicine, University of Gothenburg, Gothenburg, Sweden
[4]Department of Ophthalmology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[5]Learning and Leadership for Health Care Professionals, Institute of Health Care Science, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[6]Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah, USA
[7]Department of Ophthalmology, John A Moran Eye Center, University of Utah, Salt Lake City, Utah, USA
[8]Department of Ophthalmology, University Medicine Greifswald, Greifswald, Germany
[9]Department of Ophthalmology, Eye Center, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[10]Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden
[11]Department of Neuroscience/Ophthalmology, Uppsala University, Uppsala, Sweden
[12]Department of Ophthalmology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden
[13]Department of Clinical Sciences, Ophthalmology, Skåne University Hospital, Lund University, Lund, Sweden
[14]Länssjukhuset Ryhov, Jönköping, Sweden
[15]St. Erik Eye Hospital, Stockholm, Sweden
[16]Department of Physiology/Endocrinology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[17]Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden
[18]Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

ORCID iDs
Aldina Pivodic http://orcid.org/0000-0002-2224-3236
Andreas Stahl http://orcid.org/0000-0002-8013-7597
Eva Larsson http://orcid.org/0000-0001-9674-0094
Pia Lundgren http://orcid.org/0000-0002-7731-1988
Gerd Holmström http://orcid.org/0000-0002-5600-7186
Ann Hellström http://orcid.org/0000-0002-9259-1244

REFERENCES

1 Lee J, Dammann O. Perinatal infection, inflammation, and retinopathy of prematurity. *Semin Fetal Neonatal Med* 2012;17:26–9.
2 Mora JS, Waite C, Gilbert CE, *et al*. A worldwide survey of retinopathy of prematurity screening. *Br J Ophthalmol* 2018;102:9–13.
3 Mitchell AJ, Green A, Jeffs DA, *et al*. Physiologic effects of retinopathy of prematurity screening examinations. *Adv Neonatal Care* 2011;11:291–7.
4 Szigiato A-A, Speckert M, Zielonka J, *et al*. Effect of eye masks on neonatal stress following dilated retinal examination: the MASK-ROP randomized clinical trial. *JAMA Ophthalmol* 2019. doi:10.1001/jamaophthalmol.2019.3379. [Epub ahead of print: 05 Sep 2019].
5 Moral-Pumarega MT, Caserío-Carbonero S, De-La-Cruz-Bértolo J, *et al*. Pain and stress assessment after retinopathy of prematurity screening examination: indirect ophthalmoscopy versus digital retinal imaging. *BMC Pediatr* 2012;12:132.
6 Zupancic JAF, Ying G-S, de Alba Campomanes A, *et al*. Evaluation of the economic impact of modified screening criteria for retinopathy of prematurity from the postnatal growth and ROP (G-ROP) study. *J Perinatol* 2020;40:1100–8.
7 Holmström G, Hellström A, Gränse L, *et al*. New modifications of Swedish ROP guidelines based on 10-year data from the SWEDROP register. *Br J Ophthalmol* 2020;104:943–9.
8 Quinn GE, Ying G-S, Bell EF, *et al*. Incidence and early course of retinopathy of prematurity: secondary analysis of the postnatal growth and retinopathy of prematurity (G-ROP) study. *JAMA Ophthalmol* 2018;136:1383–9.
9 Blencowe H, Lawn JE, Vazquez T, *et al*. Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010. *Pediatr Res* 2013;74 Suppl 1:35–49.
10 Kim SJ, Port AD, Swan R, *et al*. Retinopathy of prematurity: a review of risk factors and their clinical significance. *Surv Ophthalmol* 2018;63:618–37.
11 Löfqvist C, Andersson E, Sigurdsson J, *et al*. Longitudinal postnatal weight and insulin-like growth factor I measurements in the prediction of retinopathy of prematurity. *Arch Ophthalmol* 2006;124:1711–8.
12 Hellström A, Hård A-L, Engström E, *et al*. Early weight gain predicts retinopathy in preterm infants: new, simple, efficient approach to screening. *Pediatrics* 2009;123:e638–45.
13 Slidsborg C, Forman JL, Rasmussen S, *et al*. A new risk-based screening criterion for treatment-demanding retinopathy of prematurity in Denmark. *Pediatrics* 2011;127:e598–606.
14 Eckert GU, Fortes Filho JB, Maia M, *et al*. A predictive score for retinopathy of prematurity in very low birth weight preterm infants. *Eye* 2012;26:400–6.
15 Binenbaum G, Ying G-S, Quinn GE, *et al*. The CHOP postnatal weight gain, birth weight, and gestational age retinopathy of prematurity risk model. *Arch Ophthalmol* 2012;130:1560–5.

16 Cao JH, Wagner BD, Cerda A, *et al*. Colorado retinopathy of prematurity model: a multi-institutional validation study. *J Aapos* 2016;20:220–5.
17 Binenbaum G, Ying G-S, Tomlinson LA, *et al*. Validation of the children's Hospital of Philadelphia retinopathy of prematurity (CHOP ROP) model. *JAMA Ophthalmol* 2017;135:871–7.
18 McCauley K, Chundu A, Song H, *et al*. Implementation of a clinical prediction model using daily postnatal weight gain, birth weight, and gestational age to risk stratify ROP. *J Pediatr Ophthalmol Strabismus* 2018;55:326–34.
19 Binenbaum G, Bell EF, Donohue P, *et al*. Development of modified screening criteria for retinopathy of prematurity: primary results from the postnatal growth and retinopathy of prematurity study. *JAMA Ophthalmol* 2018;136:1034–40.
20 Ahmed IS, Badeeb AA. The Alexandria retinopathy of prematurity model (Alex-ROP): postnatal weight gain screening algorithm application in a developing country. *Int J Ophthalmol* 2019;12:296–301.
21 Hutchinson AK, Melia M, Yang MB, *et al*. Clinical models and algorithms for the prediction of retinopathy of prematurity: a report by the American Academy of ophthalmology. *Ophthalmology* 2016;123:804–16.
22 Wu C, Löfqvist C, Smith LEH, *et al*. Importance of early postnatal weight gain for normal retinal angiogenesis in very preterm infants: a multicenter study analyzing weight velocity deviations for the prediction of retinopathy of prematurity. *Arch Ophthalmol* 2012;130:992–9.
23 Löfqvist C, Hansen-Pupp I, Andersson E, *et al*. Validation of a new retinopathy of prematurity screening method monitoring longitudinal postnatal weight and insulinlike growth factor I. *Arch Ophthalmol* 2009;127:622–7.
24 Pivodic A, Hård A-L, Löfqvist C, *et al*. Individual risk prediction for sight-threatening retinopathy of prematurity using birth characteristics. *JAMA Ophthalmol* 2019:1–9.
25 Equator. Tripod guidelines. Available: https://www.equator-network.org/reporting-guidelines/tripod-statement/ [Accessed 22 Dec 2020].
26 Medscinet. SNQ register. Available: https://www.medscinet.com/pnq/default.aspx [Accessed 22 Dec 2020].
27 Larsen PP, Bründer M-C, Petrak M, *et al*. [Screening for retinopathy of prematurity: Trends over the past 5 years in two German university hospitals]. *Ophthalmologe* 2018;115:469–75.
28 Engle WA. American Academy of pediatrics Committee on F, newborn. age terminology during the perinatal period. *Pediatrics* 2004;114:1362–4.
29 Niklasson A, Albertsson-Wikland K. Continuous growth reference from 24th week of gestation to 24 months by gender. *BMC Pediatr* 2008;8:8.
30 Early Treatment For Retinopathy Of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. *Arch Ophthalmol* 2003;121:1684–94.
31 International Committee for the Classification of Retinopathy of Prematurity. The International classification of retinopathy of prematurity revisited. *Arch Ophthalmol* 2005;123:991–9.
32 DIGIROP. Welcome to DIGIROP. Available: https://www.digirop.com/ [Accessed 22 Dec 2020].
33 Binenbaum G. Algorithms for the prediction of retinopathy of prematurity based on postnatal weight gain. *Clin Perinatol* 2013;40:261–70.
34 Binenbaum G, Tomlinson LA, de Alba Campomanes AG. Validation of the postnatal growth and retinopathy of prematurity screening criteria. *JAMA Ophthalmol* 2019.
35 Valikodath N, Cole E, Chiang MF. Imaging in retinopathy of prematurity. *Asia Pac J Ophthalmol* 2019;8:178–86.
36 Wilkinson AR, Haines L, Head K, *et al*. Uk retinopathy of prematurity guideline. *Early Hum Dev* 2008;84:71–4.
37 American Academy of Pediatrics Section on Ophthalmology, American Academy of Ophthalmology, The American Association for Pediatric Ophthalmology and Strabismus. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics* 2018;142.
38 Augenärztliche Screening-Untersuchung bei Frühgeborenen. German ROP screening guideline (Augenärztliche Screening-Untersuchung bei Frühgeborenen). Available: https://www.awmf.org/uploads/tx_szleitlinien/024-010l_S2k_Augenaerztliche_ Screening-Untersuchung_Fr%C3%BChgeborene_2020-07.pdf [Accessed 22 Dec 2020].