ORIGINAL ARTICLE

ICROP Third-Edition Phenotypes of Retinopathy of Prematurity and Their Association with Gestational Age in Indonesian Preterm Infants: A Retrospective Study

I Wayan Eka Sutyawan[1*], I Wayan Ardy Paribrajaka[2], Ni Made Ayu Surasmiati[1], Willy Yahya[2]

[1] Pediatric Ophthalmology and Strabismus Division, Department of Ophthalmology, Faculty of Medicine, Universitas Udayana, Denpasar, Indonesia

[2] Ophthalmology Residency Program, Faculty of Medicine, Universitas Udayana, Denpasar, Indonesia

ARTICLE INFO

ARTICLE TYPE
Original Article

ARTICLE HISTORY
Received: June 30, 2026
Revised: July 20, 2026
Accepted: August 18, 2026
Available online: August 28, 2026
Published: August 31, 2026

KEYWORDS
Gestational age; Infant, premature; Neonatal screening; Retinopathy of prematurity; Vision disorders

CORRESPONDING AUTHOR
I Wayan Eka Sutyawan
E-mail: eka_sutyawan@unud.ac.id

ORCID
0000-0002-0649-3438

DOI
https://doi.org/10.59345/sjped.v4i1.333

ABSTRACT

Background: Retinopathy of prematurity (ROP) is a leading cause of preventable childhood blindness, and middle-income countries face a third epidemic as neonatal survival outpaces screening capacity. Indonesian data using the third ICROP edition are scarce.

Objective: To describe the ICROP third-edition phenotype distribution among Indonesian infants with ROP and evaluate the associations of severity with gestational age and birth weight using exact inference and partial identification.

Methods: This retrospective STROBE-compliant study characterised the ICROP phenotype of 39 infants with ROP at a tertiary referral hospital in Denpasar between January 2020 and December 2022, and examined severity against gestational age and birth weight. Proportions carry 95% Wilson intervals; the gestational-age-by-stage cross-classification was reconstructed under the published marginals, tested by exact methods, and bounded across every reconstruction.

Results: Mean gestational age was 28.7 ± 2.7 weeks and mean birth weight was 1091 ± 320 g; 94.9% were of very low birth weight. Stage was documented for 38 infants: stage 1, 68.4%; stage 2, 28.9%; and stage 3, 2.6%. Pre-plus and plus disease each affected 7.7%, and aggressive ROP affected 2.6%. Stage ≥2 occurred in 70.0% of infants born before 28 weeks versus 17.9% of those born later (odds ratio 9.16, 95% CI 1.90–44.10; exact P = 0.005; Cramér's V = 0.370), with a monotone trend across strata (exact P = 0.004). Across all 30 admissible reconstructions, the exact P ranged from 0.004 to 0.115 and the odds ratio from 0.03 to 9.16. Birth weight carried no usable information because 94.9% of the cohort occupied a single stratum.

Conclusion: Gestational age, not birth weight, should anchor ROP surveillance intensity in Indonesian tertiary units, while a broad birth-weight entry threshold is retained.

1. Introduction

Retinopathy of prematurity (ROP) is a vasoproliferative disorder of the developing retina and one of the leading causes of avoidable blindness in children worldwide.[1,2] Preterm birth, its substrate, is extraordinarily common: an estimated 13.4 million babies were born preterm in 2020, and southern Asia together with sub-Saharan Africa accounted for roughly two-thirds of that total.[3,4] Indonesia contributes one of the largest absolute numbers of preterm and low-birth-weight births in the region, with national survey data placing low birth weight at a level that has changed little in two decades, and complications of preterm birth remain among the principal causes of death before the age of five.[5,6] As neonatal intensive care expands, more extremely preterm infants survive and the pool of children at risk of ROP grows accordingly; where screening and treatment capacity does not expand in step, the result is the pattern described as the third epidemic of ROP, in which middle-income countries carry a disproportionate share of ROP-related childhood blindness.[2,7]

The pathogenesis of ROP is specific to the immature infant and has no adult counterpart. Retinal vascularisation begins at approximately 16 weeks of gestation and advances centrifugally from the optic disc towards the ora serrata, reaching the periphery only near term. Preterm delivery interrupts this process and exposes an incompletely vascularised retina to an extrauterine environment for which it is not prepared. In the first phase, relative hyperoxia and the abrupt loss of placentally derived insulin-like growth factor 1 (IGF-1) suppress vascular endothelial growth factor (VEGF) signalling and arrest vessel growth; pooled clinical data confirm that low postnatal serum IGF-1 predicts subsequent ROP, and postnatal weight gain, which tracks IGF-1, performs well as a surrogate in prediction algorithms.[8,9,10] In the second phase the avascular periphery becomes hypoxic, hypoxia-inducible factor 1α drives a VEGF surge, and disorganised extraretinal neovascularisation follows; experimental blockade of the miR-17–5p/HIF-1 alpha/VEGF axis suppresses exactly this neovascular response in the oxygen-induced retinopathy model.[11] The

intensity of that second phase is modulated by postnatal exposures that are, in principle, controllable: oxygen saturation targeting, cumulative hyperoxia, neonatal sepsis and red-cell transfusion have each been linked to ROP incidence or severity in contemporary cohorts.[12,13,14,15] Because the depth of the first phase is fixed by how much retina remains avascular at birth, gestational age is a direct index of the biological substrate of the disease, whereas birth weight is additionally shaped by intrauterine growth and is therefore an indirect surrogate, as work on small-for-gestational-age and growth-restricted infants has shown.[16,17]

The International Classification of Retinopathy of Prematurity (ICROP) provides the common language in which this biology is recorded, describing disease by zone, stage and vascular activity. The third edition, published in 2021, refined the definitions of pre-plus and plus disease, replaced aggressive posterior ROP with the broader concept of aggressive ROP, and introduced explicit guidance on regression and reactivation.[18] These refinements matter because vascular activity, rather than stage alone, drives the decision to treat, and because the plus-disease threshold has demonstrably drifted between editions: a quantitative comparison of the published reference images shows a measurable shift in the vascular characteristics that define the standard.[19] The ICROP third-edition committee has since introduced a reference-image grading scale for the pre-plus to plus continuum, and artificial-intelligence vascular severity scores have been shown to raise both accuracy and inter-observer agreement among expert graders, with automated plus-disease detection now externally validated across multiple centres.[20,22] Every one of these tools consumes structured phenotype data as its input, so incomplete documentation of zone and vascular activity is not a clerical problem but a barrier to adopting them.

The relationship between ROP severity and the two variables recorded for every newborn, gestational age and birth weight, has been examined repeatedly, and both show an inverse association with disease severity. The relative weight of the two is not constant. A large claims-based analysis that stratified risk factors separately by birth weight and by gestational age, and a cohort study that stratified ROP risk by gestational age alone, both found gestational age to be the more discriminating axis, and a recent meta-analysis restricted to very-low-birth-weight infants reached the same conclusion within that narrow weight band.[23,24,25] In low- and middle-income settings, larger and more mature infants continue to develop treatment-requiring disease. Contemporary Asian series illustrate the range: a ten-year single-unit trend among extremely preterm infants in China, a 70-unit Chinese multicentre cohort of severe ROP, prediction models built on Thai infants, a rural Thai tertiary cohort, and Indian series from urban level III units and from rural Karnataka and North India all report ROP profiles that differ from high-income norms.[26,27,28,29,30,31,32] Screening rules calibrated in one environment therefore transfer imperfectly to another, and locally generated phenotype data are needed before national criteria can be set.

Indonesian evidence on ROP has grown but remains thin and Java-centred. A national multicentre survey documented the scale of the problem and the gaps in screening coverage; a risk-scoring model for ROP progression was subsequently developed for Indonesian practice and then externally validated; and a single-centre study has examined the determinants of an immature retina at first screening.[33,34,35,36] None of this work reports the ICROP third-edition phenotype of an Indonesian cohort in full, and data from tertiary referral centres outside Java, which receive the most complex neonatal cases and act as regional hubs for screening and treatment, are almost absent. Ngoerah General Hospital in Denpasar is the tertiary referral centre for Bali and the surrounding provinces and provides the paediatric ophthalmology service for the region. This study therefore aimed to: (1) describe the ICROP third-edition phenotype distribution, including stage and vascular activity, among infants diagnosed with ROP at this centre between 2020 and 2022; (2) quantify the association between ICROP stage and gestational age and birth weight using exact methods appropriate to a sparse contingency table; and (3) establish, by partial identification, how much of that association is determined by the data and how much remains open, so that the strength of the evidence is stated at its true resolution rather than overstated.

2. Methods

Study design and reporting standard

This was a single-centre, retrospective, cross-sectional analysis of clinical records. It is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for cross-sectional studies.[37] Conduct and reporting followed the recommendations of the Committee on Publication Ethics (COPE), the International Committee of Medical Journal Editors (ICMJE) and the World Association of Medical

Editors (WAME). The flow of participants from first examination to analysis, including the point at which the staging and vascular-activity denominators diverge, is set out in Figure 1.

Ethical considerations

Ethics approval and the consent waiver are reported in the Declarations.

Setting and participants

The study was conducted at Ngoerah General Hospital, Denpasar, Bali, Indonesia, a tertiary referral hospital serving Bali and neighbouring provinces. Records of infants examined by the paediatric ophthalmology service between 1 January 2020 and 31 December 2022 were reviewed. Infants were eligible if they had a documented diagnosis of ROP established on binocular indirect ophthalmoscopy and classified according to ICROP criteria.[18] Infants whose ophthalmological documentation was incomplete to the point that ROP status could not be confirmed were excluded, as recorded in Figure 1. The infant, not the eye, was the unit of analysis, and ocular findings were classified according to the more severely affected eye. Each infant contributed one record.

Ophthalmological examination and phenotyping

Examinations were performed by paediatric ophthalmologists of the Pediatric Ophthalmology and Strabismus Division after pupillary dilatation, using binocular indirect ophthalmoscopy with a 28-dioptre lens and scleral indentation where tolerated. Disease was recorded as stage 1 to 5, together with the presence of pre-plus disease, plus disease and aggressive posterior ROP, following the ICROP third-edition definitions.[18] The clinical records use the term aggressive posterior ROP; the third edition replaced it with the broader term aggressive ROP, because the phenotype is not confined to the posterior pole. The original label is retained here for fidelity to the record and the two are treated as equivalent throughout. Grading of the pre-plus to plus continuum followed the reference images of the third edition, the limitations of which are now well documented.[19,20]

For the purpose of analysis, an ordered severity variable was defined as stage 1, stage 2 and stage 3 or higher, and a dichotomous variable was defined as stage 1 versus stage 2 or higher. Vascular activity was analysed separately from stage, because pre-plus and plus disease modify the treatment decision independently of stage. Treatment-warranted, or type 1, disease is defined by ICROP and current treatment guidance as zone I disease of any stage with plus disease, zone I stage 3 disease without plus disease, or zone II stage 2 or 3 disease with plus disease.[18] Because retinal zone was not recorded in these notes, type 1 status cannot be assigned to an individual infant. The number of infants meeting it was therefore bounded rather than counted, taking aggressive ROP as treatment-requiring by definition and the union of plus disease with stage 3 disease as the upper limit; the resulting bound is reported together with the phenotype distribution.

Perinatal variables and definitions

Gestational age at birth in completed weeks and birth weight in grams were extracted from the neonatal record. Gestational age was grouped as less than 28 weeks, 28 to 33 weeks and 34 to 37 weeks; birth weight was grouped as less than 1500 g, 1500 to 2000 g and more than 2000 g, following the strata used in Indonesian ROP screening practice.[34] Very low birth weight was defined as a birth weight below 1500 g, following the World Health Organization definition. The remaining strata are those of the source screening protocol and are not the World Health Organization categories, which divide preterm birth at 28, 32 and 37 completed weeks and birth weight at 1000, 1500 and 2500 g; the divergence is stated because it limits direct comparison with series that report the World Health Organization strata. Sex and district of referral were also recorded; the complete variable list, with the distribution of every one of them, is presented in Table 1. Postnatal exposures relevant to ROP, including duration and mode of supplemental oxygen, culture-proven sepsis, red-cell transfusion, bronchopulmonary dysplasia and postnatal weight gain, were not consistently documented in the retrospective record and could not be analysed. Retinal zone and extent in clock hours were likewise not recorded systematically and are therefore absent from this report.

Sample size, precision and detectable effect

The study was a complete enumeration of eligible cases over a fixed three-year window rather than a sample drawn to a pre-specified power calculation; no a priori sample size calculation was therefore performed, and we report instead the precision the achieved sample delivers. At n=39, a 95% Wilson score interval for a proportion near 50% has a half-width of 15.0 percentage points, and for a proportion near 10% a half-width of 9.8 percentage points. Estimating a 10% phenotype prevalence to within ±5 percentage points would require 139 infants, and a 50% prevalence 385 infants. For

the comparison of stage ≥2 between gestational-age strata, the minimum detectable effect was obtained by Fisher exact simulation with 8,000 replicates at α = 0.05 (two-sided) and a fixed seed, holding the observed stratum sizes fixed. All of these precision quantities are reported together with the birth-weight analysis.

Statistical analysis: descriptive estimation

Analyses were performed in Python 3.11 with NumPy 2.4 and SciPy 1.17; a fixed random seed of 20260818 was used for every simulation. Categorical variables are summarised as counts with percentages, and continuous variables as mean ± standard deviation. Every proportion is accompanied by a 95% Wilson score interval, computed as in Equation 1, which is preferred to the Wald interval at these sample sizes because it does not collapse when the observed count is zero or equal to the denominator. Clopper–Pearson intervals are reported alongside as a conservative check. Denominators are stated explicitly for every proportion, because stage was documented for 38 of 39 infants whereas vascular activity was documented for all 39. Bounds on the prevalence of treatment-warranted disease were derived from the Fréchet inequalities for the union of the component phenotypes: because retinal zone was not recorded, plus disease alone does not establish type 1 disease, so the lower bound rests only on the single infant with aggressive ROP, while aggressive ROP carries plus disease by definition and is contained within it, so the upper bound is the union of plus disease with stage 3 disease.

$$CI_{95\%} = [\, \hat{p} + z^2/(2n) \pm z\sqrt{(\, \hat{p}(1-\hat{p})/n + z^2/(4n^2)\,)}\,] / [\, 1 + z^2/n\,]$$

Equation 1. Wilson score interval for a binomial proportion, with $\hat{p}$ the observed proportion, n the denominator and z = 1.96.

Statistical analysis: association, exact inference and partial identification

Variables were abstracted from each clinical record onto a standard form and tabulated. The dataset available for the present analysis consisted of those tabulated marginal distributions; the infant-level linkage between gestational age and ICROP stage was not preserved in the transfer and had to be recovered. Two consequences follow, and both were addressed explicitly rather than ignored. First, the Pearson χ^2 test is not valid at these margins: with a single stage 3 case and a single infant born at 34 to 37 weeks, most expected cell counts fall far below 5. All association testing was therefore performed with exact methods, namely the Freeman–Halton generalisation of Fisher's exact test for the full cross-classification, Fisher's exact test for the dichotomised comparison, and an exact permutation version of the Cochran–Armitage test for trend across ordered gestational-age strata, whose statistic is given in Equation 3. Effect sizes are reported as Cramér's V (Equation 2), Kendall's τb, the odds ratio with a Haldane–Anscombe correction and its 95% confidence interval, the risk difference in percentage points and the risk ratio.

$$V = \sqrt{(\, \chi^2 / [\, n\,(k-1)\,]\,)}, \quad k = \min(r, c)$$

Equation 2. Cramér's V for an r × c table. Because V is a deterministic function of χ^2 and n, conditioning on a reported χ^2 P value fixes V by arithmetic rather than by evidence.

$$z = \Sigma_i\, s_i\,(x_i - n_i\,\bar{p}) \,/\, \sqrt{(\,\bar{p}\,(1-\bar{p})\,[\,\Sigma_i\, n_i\, s_i^2 - (\Sigma_i\, n_i\, s_i)^2 / N\,]\,)}$$

Equation 3. Cochran–Armitage statistic for trend across ordered gestational-age strata, with s_i = 1, 2, 3 the stratum scores, x_i the number with stage ≥ 2, n_i the stratum size, N the total and $\bar{p}$ the overall proportion. The reference distribution used here is the exact permutation distribution over all tables with the observed margins, not the normal approximation.

Second, the cross-classification underlying the reported association was recovered rather than assumed. Every integer-valued contingency table consistent with the tabulated marginal distributions and reproducing the reported χ^2 P value to the three decimal places at which it was recorded, that is to within 0.0005, was enumerated exhaustively; the sensitivity of the resulting bounds to a ten-fold wider tolerance was examined as well. The resulting admissible set defines exactly what the published summaries determine. One reconstruction is reported as the reference reconstruction and used for the descriptive cross-tabulation presented in the Results. It was selected by a rule fixed in advance: among admissible configurations take the smallest number of infants born at 34 to 37 weeks that admits any table at all, since a cohort in which 94.9% of infants weighed less than 1500 g can contain very few near-term infants. That rule yields exactly one table, and it is also the only admissible table whose gestational-age marginals match those reported in Table 1. Every marginal total was recomputed and checked before it was used. The rule embodies an assumption about a quantity the source does not report, and the consequences of relaxing it are given in full in the Results.

Sharp bounds on the exact P value, on Kendall's τb, on Cramér's V and on the dichotomised odds ratio were then obtained by evaluating each quantity over the entire admissible set, so that the reader can see which conclusions are identified by the data and which are not. The attainable range of the χ^2 P value was computed in the same way for birth weight, and the actual size of the Pearson χ^2 test at the observed margins was estimated by Monte Carlo simulation under independence with 200,000 replicates, excluding replicates in which a zero marginal total leaves the statistic undefined.

Analyses that were not performed

Multivariable logistic regression, adjusted odds ratios, Nagelkerke R^2, the Hosmer–Lemeshow goodness-of-fit test, receiver operating characteristic analysis and time-to-event methods were all considered and are all reported here as not performed. Multivariable modelling requires infant-level linkage between exposures and outcome, which the available aggregate tables do not supply; fitting such a model would mean inventing the joint distribution. With 12 infants reaching stage ≥2 and a rule of at least 10 events per candidate covariate, the cohort could not support a stable model even if linkage were available. Diagnostic accuracy analysis was not performed because no index test was applied against an independent reference standard. Survival analysis was not performed because progression dates were not recorded. These omissions are limitations of the source data, not analytical choices made for convenience.

3. Results

Cohort and perinatal characteristics

Between 1 January 2020 and 31 December 2022, 39 infants with a documented diagnosis of ROP met the eligibility criteria and formed the analytic cohort; the full participant flow is set out in Figure 1. Male infants accounted for 24 of 39 (61.5%, 95% CI 45.9–75.1). Referrals were concentrated in the two districts closest to the hospital, with Denpasar and Badung each contributing 12 infants (30.8%, 95% CI 18.6–46.4) and the remaining 15 infants (38.5%, 95% CI 24.9–54.1) referred from other districts of Bali and from neighbouring provinces, a distribution consistent with the hospital's function as a regional referral hub, as detailed in Table 1.

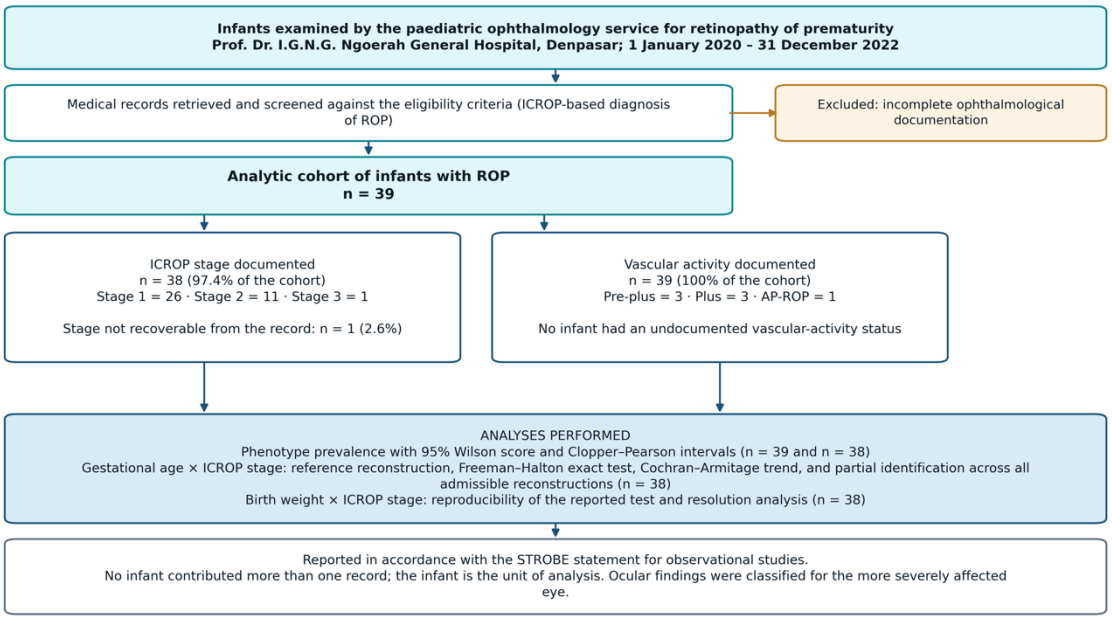

Figure 1. Participant flow for the retrospective review of infants with retinopathy of prematurity at Ngoerah General Hospital, Denpasar, 2020–2022, reported in accordance with the STROBE statement. ICROP stage was recoverable for 38 of the 39 infants in the analytic cohort, whereas vascular activity was documented for all 39.

Mean gestational age at birth was 28.7 ± 2.7 weeks. Ten infants (25.6%, 95% CI 14.6–41.1) were born before 28 weeks, 28 (71.8%, 95% CI 56.2–83.5) between 28 and 33 weeks and 1 (2.6%, 95% CI 0.5–13.2) between 34 and 37 weeks. Mean birth weight was 1091 ± 320 g, and 37 infants (94.9%, 95% CI 83.1–98.6) were of very low birth weight. Only 2 infants (5.1%, 95% CI 1.4–16.9) weighed 1500 g or more at birth. Every figure quoted in this paragraph, with its denominator and its interval, appears in Table 1, and the concentration of the cohort in a single birth-weight stratum shown there is the reason the birth-weight analysis carries so little information.

ICROP third-edition phenotype distribution

ICROP stage was recoverable for 38 of 39 infants (97.4%), whereas vascular activity was recorded for all 39, a divergence made explicit in Figure 1 and carried through every denominator in Table 2. Stage 1 was the commonest presentation, seen in 26 of 38 infants (68.4%, 95% CI 52.5–80.9), followed by stage 2 in 11 (28.9%, 95% CI 17.0–44.8) and stage 3 in a single infant (2.6%, 95% CI 0.5–13.5). No infant presented with stage 4 or stage 5 disease. Taken together, 12 infants (31.6%, 95% CI 19.1–47.5) had disease of stage 2 or greater.

Table 1. Demographic, perinatal and clinical characteristics of the cohort of infants with retinopathy of prematurity, Ngoerah General Hospital, Denpasar, 2020–2022

Characteristic	Category	n / N	%	95% CI (Wilson)
Sex	Male	24 / 39	61.5	45.9–75.1
	Female	15 / 39	38.5	24.9–54.1
District of referral	Denpasar	12 / 39	30.8	18.6–46.4
	Badung	12 / 39	30.8	18.6–46.4
	Other districts / provinces	15 / 39	38.5	24.9–54.1
Gestational age at birth	Mean ± SD, weeks	28.7 ± 2.7	—	—
	< 28 weeks	10 / 39	25.6	14.6–41.1
	28–33 weeks	28 / 39	71.8	56.2–83.5
	34–37 weeks	1 / 39	2.6	0.5–13.2
Birth weight	Mean ± SD, g	1091 ± 320	—	—
	< 1500 g (VLBW)	37 / 39	94.9	83.1–98.6
	≥ 1500 g	2 / 39	5.1	1.4–16.9
ICROP stage documented	Yes	38 / 39	97.4	—
Vascular activity documented	Yes	39 / 39	100.0	—

Values are n (%) with the 95% Wilson score interval unless otherwise stated. Denominators are stated in full because ICROP stage was recoverable for 38 of 39 infants whereas vascular activity was documented for all 39. VLBW, very low birth weight; SD, standard deviation. The gestational-age and birth-weight strata are those of the source screening protocol; only the definition of very low birth weight (< 1500 g) coincides with the World Health Organization categories, which otherwise divide preterm birth at 28, 32 and 37 completed weeks and birth weight at 1000, 1500 and 2500 g.

Vascular activity was infrequent. Pre-plus disease and plus disease were each documented in 3 infants (7.7%, 95% CI 2.7–20.3), and aggressive posterior ROP in one infant (2.6%, 95% CI 0.5–13.2), while 33 infants (84.6%, 95% CI 70.3–92.8) had neither pre-plus nor plus disease. Because retinal zone was not recorded, the number of infants meeting a type 1 treatment threshold cannot be counted: plus disease confined to zone III would not warrant treatment, whereas aggressive ROP does so by definition. The Fréchet bounds place the number between 1 and 4 infants, that is between 2.6% and 10.3% of the cohort, with 95% intervals around those bounds running from 0.5% to 23.6%. The complete phenotype distribution, including both interval methods, is presented in Table 2, and the same proportions are displayed against their Wilson intervals in Figure 2, where the width of every interval makes the limits of a 39-infant series immediately visible.

Internal consistency of the source summaries

Recomputing every published percentage against the integer numerators and denominators that could have produced it identified four inconsistencies that have a direct bearing on interpretation, and these are itemised line by line in Table 3. First, the three stage percentages resolve to 26, 11 and 1 of 38, not of 39: stage was not recoverable for one infant, a fact that was not stated in the source summaries. Second, the stage 3 percentage was printed as 2.7%, which corresponds to a denominator of 37; on the staging cohort of 38 the correct value is 2.6%. Third, the proportion of infants born at 28 to 33 weeks was given as 71.8% in one place and 59.0% in another, corresponding to 28 and 23 infants respectively, a discrepancy of five infants. Fourth, the proportion of male infants was rounded to 62% in one place and reported as 61.5% in another; both derive from 24 of 39.

Table 2. ICROP third-edition phenotype distribution, with 95% Wilson score and Clopper–Pearson intervals

Phenotype	n / N	%	95% CI (Wilson)	95% CI (Clopper–Pearson)
Stage 1	26 / 38	68.4	52.5–80.9	51.3–82.5
Stage 2	11 / 38	28.9	17.0–44.8	15.4–45.9
Stage 3	1 / 38	2.6	0.5–13.5	0.1–13.8
Stage 4 or 5	0 / 38	0.0	0.0–9.2	0.0–9.3
Stage ≥ 2 (composite)	12 / 38	31.6	19.1–47.5	17.5–48.7
Pre-plus disease	3 / 39	7.7	2.7–20.3	1.6–20.9
Plus disease	3 / 39	7.7	2.7–20.3	1.6–20.9
AP-ROP	1 / 39	2.6	0.5–13.2	0.1–13.5
No pre-plus or plus disease	33 / 39	84.6	70.3–92.8	69.5–94.1
Type 1 (treatment-warranted) disease, bounded	1–4 / 39	2.6–10.3	0.5–23.6 (on the bounds)	Fréchet bound; not point-identified

ICROP, International Classification of Retinopathy of Prematurity (third edition). Stage proportions use the staging cohort of 38 infants; vascular-activity proportions use the full cohort of 39. AP-ROP, aggressive posterior retinopathy of prematurity. Type 1 disease is bounded rather than counted because retinal zone was not recorded: plus disease confined to zone III is not treatment-warranted, so plus disease alone does not establish type 1 status, while aggressive ROP does so by definition. The lower bound therefore rests on the single infant with aggressive ROP and the upper bound on the union of plus disease with stage 3 disease. The clinical records use the pre-third-edition term AP-ROP; the third edition designates this phenotype aggressive ROP.

The resolution shown in Table 3 is not unique in every case: 2.6% is consistent with both 1 of 38 and 1 of 39, and it was assigned to the vascular-activity block on the basis of the surrounding text rather than of arithmetic alone. The larger gestational-age figure, 71.8%, is adopted throughout this report, and the consequences of adopting the alternative are carried explicitly through the partial-identification analysis below.

Gestational age and ICROP stage

The cross-classification of gestational age with ICROP stage was recovered under the published marginals. Under the clinically most plausible marginal configuration, in which one infant was born at 34 to 37 weeks and 10 before 28 weeks, exactly one contingency table is admissible. In that reference reconstruction, which is reproduced in full in Table 4 and displayed in panel A of Figure 3, 3 infants born before 28 weeks had stage 1 disease, 6 had stage 2 and 1 had stage 3, whereas 22 of the infants born at 28 to 33 weeks had stage 1 and 5 had stage 2, and the single infant born at 34 to 37 weeks had stage 1. Every marginal total of the reconstruction reproduces the published totals exactly, and that validation is recorded as its own block within Table 4.

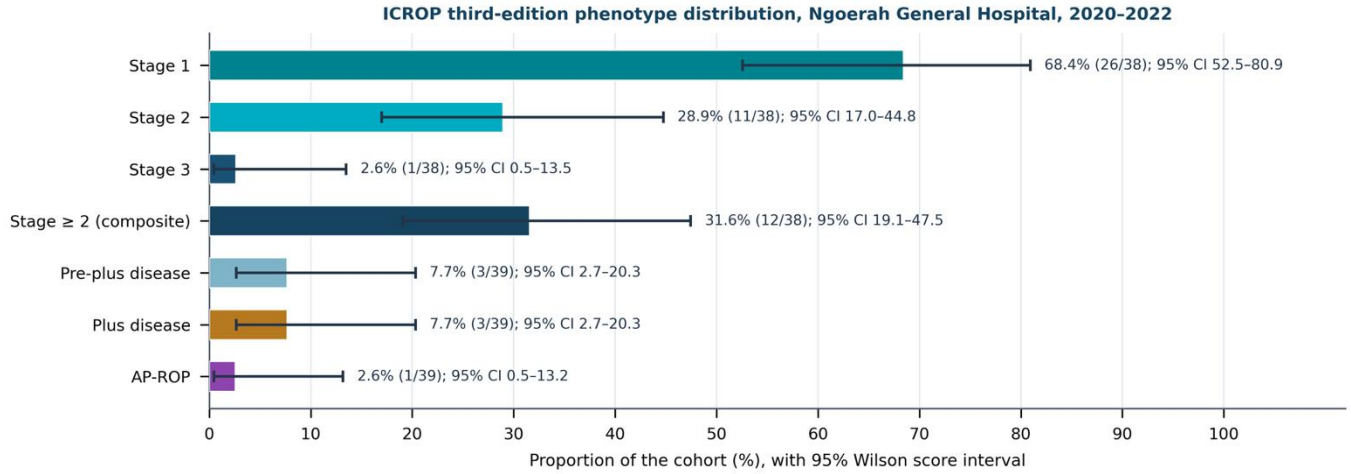

Figure 2. ICROP third-edition phenotype distribution with 95% Wilson score intervals. Stage proportions are expressed on the staging cohort of 38 infants and vascular-activity proportions on the full cohort of 39. The width of every interval reflects the small sample: at n = 39 a proportion near 50% carries a half-width of 15.0 percentage points.

On this reconstruction the association is substantial. Disease of stage 2 or greater affected 7 of 10 infants born before 28 weeks (70.0%) compared with 5 of 28 infants born at 28 weeks or later (17.9%), an odds ratio of 9.16 (95% CI 1.90–44.10; Fisher exact P = 0.005), a risk difference of 52.1 percentage points and a risk ratio of 3.92. Across the full three-by-three classification the Freeman–Halton exact test gave P = 0.006, appreciably smaller than the asymptotic χ^2 value of P = 0.034 reported in the source summaries, with Cramér's V 0.370 and Kendall's τb -0.495. The exact Cochran–Armitage test confirmed a monotone gradient across the ordered gestational-age strata (z = -3.00, exact permutation P = 0.004); that gradient, with the Wilson interval attached to each stratum, is plotted in panel B of Figure 3, and the corresponding test statistics are listed in Table 4. Simulation confirmed that the asymptotic test used in the source analysis is unreliable at these margins. In 200,000 samples generated under independence, only a fraction 0.407 yielded a defined χ^2 statistic at all, because a zero marginal total is common when a category holds a single infant; among the samples in which the statistic existed, the test rejected at 0.063 against a nominal level of 0.05, that is it was mildly liberal. The dichotomised test was correctly sized at 0.050.

What the published summaries do and do not determine

Exhaustive enumeration identified 30 contingency tables compatible with the published marginals and with the reported χ^2 P value, spanning both candidate gestational-age splits identified in the internal-consistency audit above. Evaluating each quantity of interest over that admissible set separates what the data fix from what they leave open, and the result, set out row by row in Table 5, is asymmetric. Cramér's V lies between 0.370 and 0.371 across every admissible reconstruction, a moderate to large effect by conventional benchmarks, but that apparent precision is arithmetic rather than empirical: as Equation 2 makes plain, V is a deterministic function of the χ^2 statistic and the sample size, so conditioning on a P value reported to three decimals necessarily pins it, and we do not offer it as evidence. The bias-corrected form of the statistic, which is preferable in sparse tables, is 0.296 for the reference reconstruction. The Freeman–Halton exact P ranges from 0.004 to 0.115 (median 0.041), and only 60.0% of admissible reconstructions reach

α = 0.05. Kendall's τb ranges from -0.495 to +0.495, and exactly 50.0% of reconstructions run in the direction in which severity increases as gestational age falls. For the dichotomised comparison, the odds ratio for stage ≥2 comparing infants born before 28 weeks with those born later is bounded only between 0.03 and 9.16, and 16.7% of reconstructions reach significance. The reference reconstruction sits at the very top of that range: its odds ratio of 9.16 is the largest value the admissible set contains, so it is the most favourable member rather than a typical one, and this must be weighed against it.

Table 3. Internal-consistency audit of the published summary statistics against their implied denominators

Item	Printed	Implied n / N	Exact %	Δ (pp)	Interpretation
Male sex (Abstract)	62%	24 / 39	61.5	+0.5	Rounding only; the Results value of 61.5% is exact
Gestational age 28–33 weeks (Abstract)	71.8%	28 / 39	71.8	0.0	Exact on 28 of 39
Gestational age 28–33 weeks (Results)	59.0%	23 / 39	59.0	0.0	Exact on 23 of 39; irreconcilable with the Abstract, a five-infant discrepancy
ICROP stage 1	68.4%	26 / 38	68.4	0.0	Denominator is 38, not the stated cohort of 39
ICROP stage 2	28.9%	11 / 38	28.9	0.0	Denominator is 38
ICROP stage 3	2.7%	1 / 38	2.6	+0.1	Printed value implies a denominator of 37; the correct value on the staging cohort is 2.6%
Pre-plus, plus disease	7.7%	3 / 39	7.7	0.0	Denominator is 39, so the staging and vascular-activity blocks use different denominators
AP-ROP	2.6%	1 / 39	2.6	0.0	Exact on 39
VLBW < 1500 g	94.9%	37 / 39	94.9	0.0	Exact on 39

Every printed percentage was resolved to the integer numerator and denominator that reproduce it to within 0.06 percentage points for denominators between 30 and 40. The resolution is unique for every item except 2.6%, which is consistent with both 1/38 and 1/39 and was assigned from the surrounding text. Δ is the printed value minus the exact value on the implied denominator. The audit is reported as a result rather than a footnote because two of the four items change the denominator on which the phenotype distribution rests.

Three conclusions follow, and all three are recorded in Table 5 and visualised in Figure 4. The reference reconstruction supports a strong, clinically coherent inverse gradient, and it is the configuration most consistent with the cohort's own birth-weight distribution. But an omnibus χ^2 statistic is direction-free, and the admissible set demonstrates this literally: all 30 of the 30 tables have their row-reversed mirror image also admissible, which is exactly why the bounds on Kendall's τb are symmetric and why precisely 50.0% run each way, as panel A of Figure 4 shows. The directional claim cannot be read off the reported P value at all; it requires the cross-classification, which is why the reconstruction was performed. Restricting the set to configurations with at most three infants born at 34 to 37 weeks, as the birth-weight distribution demands, leaves 3 tables with exact P between 0.004 and 0.010, of which 2 run in the direction reported here, and exactly 1 table reproduces the gestational-age marginals of Table 1. In addition, every one of the 30 admissible reconstructions violates Cochran's condition, with between 6 and 7 of the nine expected cell counts below 5 and a smallest expected count between 0.026 and 0.158. The asymptotic test was therefore inappropriate irrespective of which reconstruction is the true one. Finally, the bounds depend on how precisely the reported P value is taken: widening the tolerance from 0.0005 to 0.005 admits 254 tables and raises the upper bound on the exact P from 0.115 to 0.167, so the interval quoted here is the narrowest defensible one rather than a conservative one.

Birth weight: reproducibility, resolution and screening yield

The reported association between birth weight and ICROP stage was examined in the same way, with two findings, both tabulated in Table 6. First, the reported χ^2 P value of 0.917 could not be reproduced. Across 156 enumerated layouts, spanning three-level and dichotomised categorisations of both variables at 38 and at 39 infants, no table yields a P value equal to 0.917 to three decimal places. Among the layouts that respect the published stage distribution the largest attainable value is 0.9137, and relaxing the stage marginals raises the maximum only to 0.9192, as panel C of Figure 4 shows. We report the value as unreproduced rather than as impossible, since a still wider search cannot be ruled out; the most economical explanation is a transcription or rounding artefact, and the substantive conclusion is unaffected either way.

Table 4. Reference reconstruction of the gestational age × ICROP stage cross-classification, its validation, and exact tests of association (n = 38)

Quantity	Stage 1	Stage 2	Stage 3	Row total
Gestational age < 28 weeks	3	6	1	10
Gestational age 28–33 weeks	22	5	0	27
Gestational age 34–37 weeks	1	0	0	1
Column total	26	11	1	38
Validation of the reconstruction				
Stage marginals reproduced	26 / 11 / 1	exact	—	✓
Gestational-age marginals reproduced	10 / 28 / 1	exact	—	✓
Association, exact inference				
Freeman–Halton exact test (3 × 3)	P = 0.006	—	—	—
Pearson χ^2 reported in the source	χ^2 = 10.41, df = 4	P = 0.034	—	not valid at these margins
Cramér's V	0.370	—	—	moderate to large
Kendall's τb	−0.495	—	—	severity rises as gestational age falls
Cochran–Armitage trend (exact permutation)	z = −3.00	P = 0.004	—	monotone gradient
Dichotomised comparison, stage ≥ 2				
Born < 28 weeks	7 / 10	70.0%	—	—
Born ≥ 28 weeks	5 / 28	17.9%	—	—
Odds ratio (95% CI)	9.16	1.90–44.10	P = 0.005	Fisher exact
Risk difference	52.1 pp	—	—	—
Risk ratio	3.92	—	—	—

The reference reconstruction is the unique contingency table admissible under the clinically most plausible marginal configuration (fewest infants of 34–37 weeks, consistent with 94.9% very low birth weight) and reproducing the reported Pearson χ^2 P value. It is one of 30 admissible tables and it carries the largest odds ratio of any of them; see Table 5. Haldane–Anscombe correction applied to the odds ratio. The single unstaged infant is returned to the 28–33 week stratum when the full-cohort marginals are restored. pp, percentage points.

Second, and more importantly, that conclusion is uninformative rather than negative. With only 2 infants outside the very-low-birth-weight stratum, the comparison has almost no resolution: Fisher exact simulation showed a power of 0.017 if half of the non-very-low-birth-weight infants had stage ≥2 disease, 0.058 if 90.0% did, and 0.068 even in the extreme case in which every one of them did. No finite effect size is detectable at conventional power. By contrast, the gestational-age comparison retained usable resolution: the same simulation gave a power of 0.84 for a stage ≥2 prevalence of 70.0% before 28 weeks against the observed 17.9% thereafter, so the minimum detectable prevalence was approximately 68.0%, corresponding to an odds ratio of about 9.7. Considered as a screening rule rather than as a risk factor, birth weight nevertheless performs well: a birth-weight arm set at 1500 g would have captured 94.9% of the ROP cases detected here (95% CI 83.1–98.6), and an arm set at 2000 g would have captured between 94.9% and 100.0% of them, the exact figure being undetermined because the birth weights of the two infants weighing 1500 g or more are not reported. These screening-yield figures, the power curves and the precision of the design are collected together in Table 6.

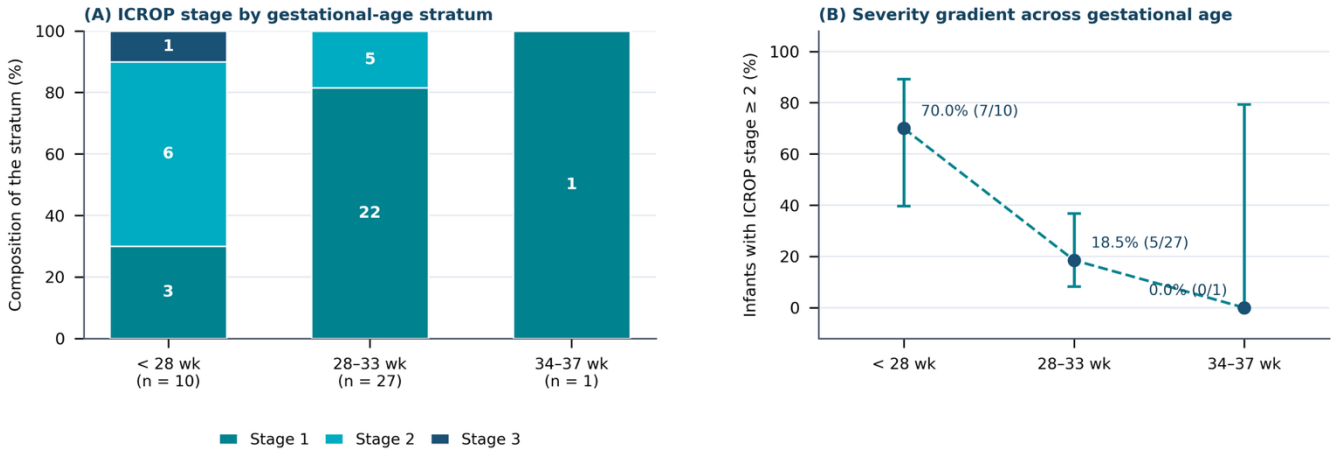

Figure 3. Reference reconstruction of the gestational age × ICROP stage cross-classification (n = 38). (A) Composition of each gestational-age stratum by ICROP stage, with cell counts printed inside the bars. (B) Proportion of infants with stage ≥ 2 disease by stratum, with 95% Wilson score intervals, together with the exact tests of association and trend. The gradient is monotone across the ordered strata.

Table 5. Partial identification: sharp bounds on every quantity of interest across all 30 contingency tables compatible with the published marginals and the reported χ^2 P value

Quantity	Lower bound	Upper bound	Status	Consequence
Freeman–Halton exact P	0.004	0.115	not identified	Median 0.041; only 60.0% of reconstructions reach α = 0.05
Odds ratio, stage ≥ 2, < 28 vs ≥ 28 weeks	0.03	9.16	not identified	16.7% significant; the reference reconstruction sits at the upper bound
Kendall's τb	−0.495	+0.495	not identified	Direction unconstrained; 50.0% run in the direction reported
Cramér's V	0.370	0.371	arithmetic	Deterministic in χ^2 and n; not evidence. Bias-corrected value 0.296
Structure of the admissible set				
Tables whose mirror image is also admissible	30	of 30	—	The omnibus χ^2 carries no directional information whatsoever
Tables matching the gestational-age marginals of Table 1	1	of 30	—	That table is the reference reconstruction
Tables with ≤ 3 infants born at 34–37 weeks	3	of 30	—	Exact P 0.004–0.010; 2 of 3 run in the direction reported
Sensitivity of the bounds to the P-value tolerance				
Tolerance ± 0.0005 (primary)	0.004	0.115	30 tables	The narrowest defensible interval
Tolerance ± 0.005	0.004	0.167	254 tables	Upper bound rises; the lower bound is unchanged
Validity of the asymptotic test at these margins				
Smallest expected cell count	0.026	0.158	—	Cochran's condition violated in every reconstruction
Expected cells below 5, of 9	6	7	—	Asymptotic χ^2 inappropriate irrespective of reconstruction
Replicates yielding a defined χ^2 (3 × 3)	0.407	0.407	—	A zero marginal leaves the statistic undefined in most null samples
Actual size given a defined statistic (3 × 3)	0.063	0.063	—	Mildly liberal against a nominal level of 0.05
Actual size, dichotomised 2 × 2	0.050	0.050	—	Correctly sized; the dichotomised analysis is the trustworthy one

The admissible set was obtained by exhaustive integer enumeration over both candidate gestational-age splits identified in Table 3, retaining every table whose Pearson χ^2 P value reproduces the reported 0.034 to within 0.0005. A quantity whose bounds coincide to three decimals is fixed by the published summaries; one whose bounds straddle a decision threshold is not. Cramér's V is fixed by arithmetic rather than by evidence, being a deterministic function of χ^2 and n. Monte Carlo estimates of the actual test size used 200,000 replicates under independence at the observed margins with seed 20260818, excluding replicates in which a zero marginal leaves the statistic undefined.

Table 6. Birth weight and ICROP stage: attainability of the reported test, resolution of the design, and screening-criterion yield

Analysis	Quantity	Value	95% CI or note
Reproducibility of the reported birth-weight test			
Reported Pearson χ^2 P	—	0.917	from the source summaries
Layouts enumerated	3-level and dichotomised, n = 38 and n = 39	156	—
Layouts reproducing 0.917 to three decimals	—	0	value not reproduced
Largest P consistent with the published stage distribution	—	0.9137	3 × 3 layout, χ^2 = 0.97, df = 4
Largest P with the stage marginals relaxed	—	0.9192	reported as unreproduced, not as impossible
Resolution of the birth-weight comparison			
Power if 50.0% of infants ≥ 1500 g had stage ≥ 2	baseline 17.9%	0.017	n = 2 in the comparison stratum
Power if 90.0% had stage ≥ 2	—	0.058	—
Power if 100.0% had stage ≥ 2	complete separation	0.068	no finite effect detectable at 80% power
Resolution of the gestational-age comparison			
Power at 60.0% stage ≥ 2 before 28 weeks	baseline 17.9%	0.66	n = 10 versus n = 28
Power at 70.0% stage ≥ 2 before 28 weeks	—	0.84	minimum detectable prevalence ≈ 68.0%
Power at 80.0% stage ≥ 2 before 28 weeks	—	0.95	corresponding odds ratio ≈ 9.7
Screening-criterion yield in this cohort			
Birth-weight arm at < 1500 g	captured / detected	94.9%	83.1–98.6
Birth-weight arm at < 2000 g	captured / detected	94.9–100.0%	bounded; the weights of the two infants ≥ 1500 g are not reported
Precision of the design			
95% CI half-width at p ≈ 0.50, n = 39	—	15.0 pp	n = 385 needed for ± 5 pp
95% CI half-width at p ≈ 0.10, n = 39	—	9.8 pp	n = 139 needed for ± 5 pp

Attainability was assessed by enumerating every contingency table consistent with 37 of 39 infants below 1500 g, across three-level and dichotomised layouts for both variables. Power figures are Fisher exact simulations with 6,000–8,000 replicates at α = 0.05 two-sided, seed 20260818, holding the observed stratum sizes fixed. Screening yield is the proportion of the ROP cases detected in this cohort that the birth-weight arm of each criterion would have captured; the gestational-age arm of each criterion would capture additional infants and the figures are therefore lower bounds on total yield.

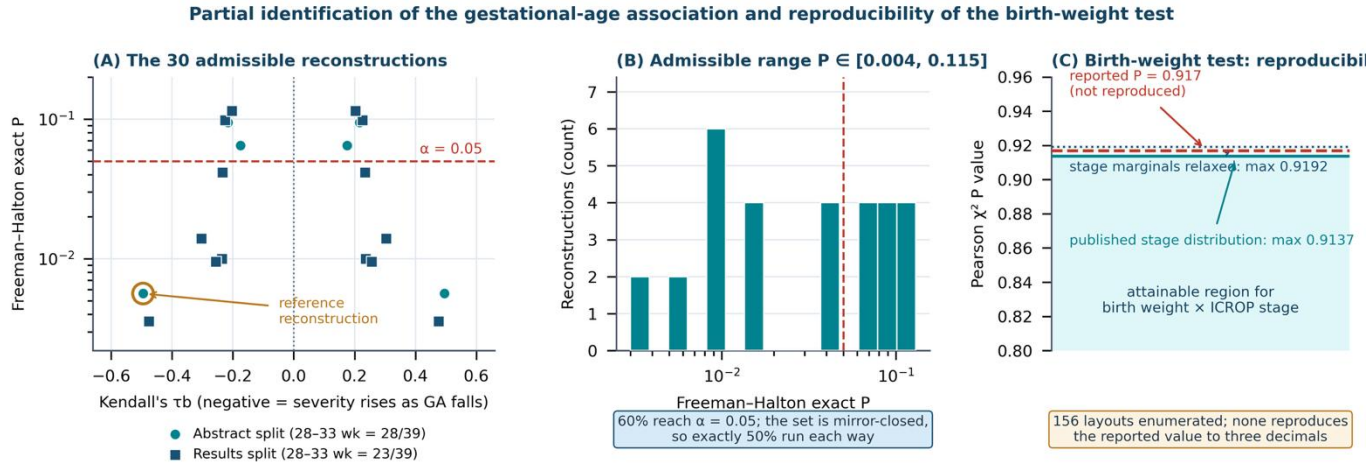

Figure 4. Partial identification of the gestational-age association and reproducibility of the birth-weight test. (A) Freeman–Halton exact P against Kendall's τb for each of the 30 admissible reconstructions, with the reference reconstruction circled; the set is closed under row reversal, so every point has a mirror image and the omnibus test carries no directional information. (B) Distribution of the exact P across the admissible set. (C) No enumerated layout reproduces the reported birth-weight P value of 0.917; the largest value consistent with the published stage distribution is 0.9137 and the largest with those marginals relaxed is 0.9192.

4. Discussion

This study characterised the ICROP third-edition phenotype of 39 infants with ROP managed at a tertiary referral centre in Bali and asked how far the association between disease severity and perinatal size is actually determined by the data. Three findings stand out. The phenotype distribution, set out in Table 2 and Figure 2, was dominated by early disease, with 68.4% of staged infants at stage 1 and treatment-warranted disease confined to between 2.6% and 10.3% of the cohort. Gestational age showed a strong inverse gradient with stage under the reference reconstruction, with stage ≥2 in 70.0% of infants born before 28 weeks against 17.9% thereafter (odds ratio 9.16, 95% CI 1.90–44.10); that figure is the largest the admissible set permits, and the honest range is 0.03 to 9.16. Birth weight carried no usable information about severity, not because it is unrelated to ROP but because 94.9% of this cohort occupied a single birth-weight stratum, as the perinatal profile in Table 1 makes plain.

The predominance of early-stage disease is consistent with contemporary series from middle-income tertiary centres, where severe stages have become uncommon wherever screening is systematic. A rural Thai tertiary cohort, an urban Indian level III unit and series from rural Karnataka and North India all report distributions weighted towards the milder stages, and a ten-year Chinese single-unit trend documents the same shift over time.[26,29,30,31,32] The distribution summarised in Table 2 and plotted in Figure 2 places our centre firmly in that group. Two interpretations are possible and cannot be distinguished with these data. The optimistic reading is that referral and screening pathways in Bali are identifying disease before it becomes sight-threatening. The cautious reading is that severe disease is under-represented because infants who developed it were not referred, died before referral, or were treated elsewhere; a 70-unit Chinese multicentre study of severe ROP shows how much case-mix can vary between units within a single health system, and a retrospective single-centre design cannot exclude the same effect here.[27]

The gestational-age gradient we recovered is consistent in direction and magnitude with the wider literature. A cohort study that stratified ROP risk by gestational age and a large claims-based analysis that stratified risk factors separately by gestational age and by birth weight both identified gestational age as the more discriminating axis, and a Thai risk-factor modelling study reached the same conclusion in a South-East Asian population.[23,24,28] Our reconstructed odds ratio of 9.16 for stage ≥2 below 28 weeks, shown in Table 4 and Figure 3, is larger than most published estimates, which is expected: our denominator is a cohort of infants who already have ROP, so we are describing severity conditional on disease rather than incidence, and conditioning on the presence of disease concentrates the gradient. The comparison is therefore with severity-conditional series rather than with incidence cohorts, a distinction that ROP papers frequently blur.

The failure of birth weight to show an association deserves a more careful reading than the source summaries gave it. Published evidence links small-for-gestational-age status and intrauterine growth restriction to ROP risk, with a direct comparison of small-for-gestational-age against appropriate-for-gestational-age neonates and a large multicentre cohort of very preterm infants both reporting an

effect, and a meta-analysis confined to very-low-birth-weight infants quantifying the residual gradient within that band.[16,17,25] What our data show is not the absence of that relationship but the absence of the resolution needed to see it, a distinction the two-phase architecture of Figure 5 helps to frame. With 2 infants outside the very-low-birth-weight stratum, the power to detect even a complete separation of outcomes was 0.068, as the power block of Table 6 records, and the attainability ceiling plotted in panel C of Figure 4 shows how little room the published marginals leave for any test at all. Reporting such a result as evidence that birth weight does not matter would be a textbook misreading of a null; quantifying the detectable effect converts an uninformative test into an honest statement about what the design could and could not see.

The biology, set out in Figure 5, explains why gestational age retains information where birth weight does not. The depth of the first, vaso-obliterative phase is set by how much retina remains unvascularised at delivery, which is a direct function of gestational age, while the intensity of the second, proliferative phase is modulated by postnatal exposures. Serum IGF-1 and the postnatal weight gain that tracks it predict subsequent disease, and duration of parenteral nutrition adds further prognostic information, so nutritional and growth trajectories are part of the causal picture rather than noise.[8-10,38] Oxygen saturation targeting, cumulative hyperoxia, neonatal sepsis and red-cell transfusion have all been associated with ROP in recent cohorts, and experimental blockade of the HIF-1 alpha/VEGF axis abolishes the neovascular response in the oxygen-induced retinopathy model.[11-15] These are precisely the variables our record review could not recover, so residual confounding by them cannot be excluded and is the most plausible explanation for any variation in severity that gestational age does not account for, and they act on the second phase of Figure 5 rather than the first.

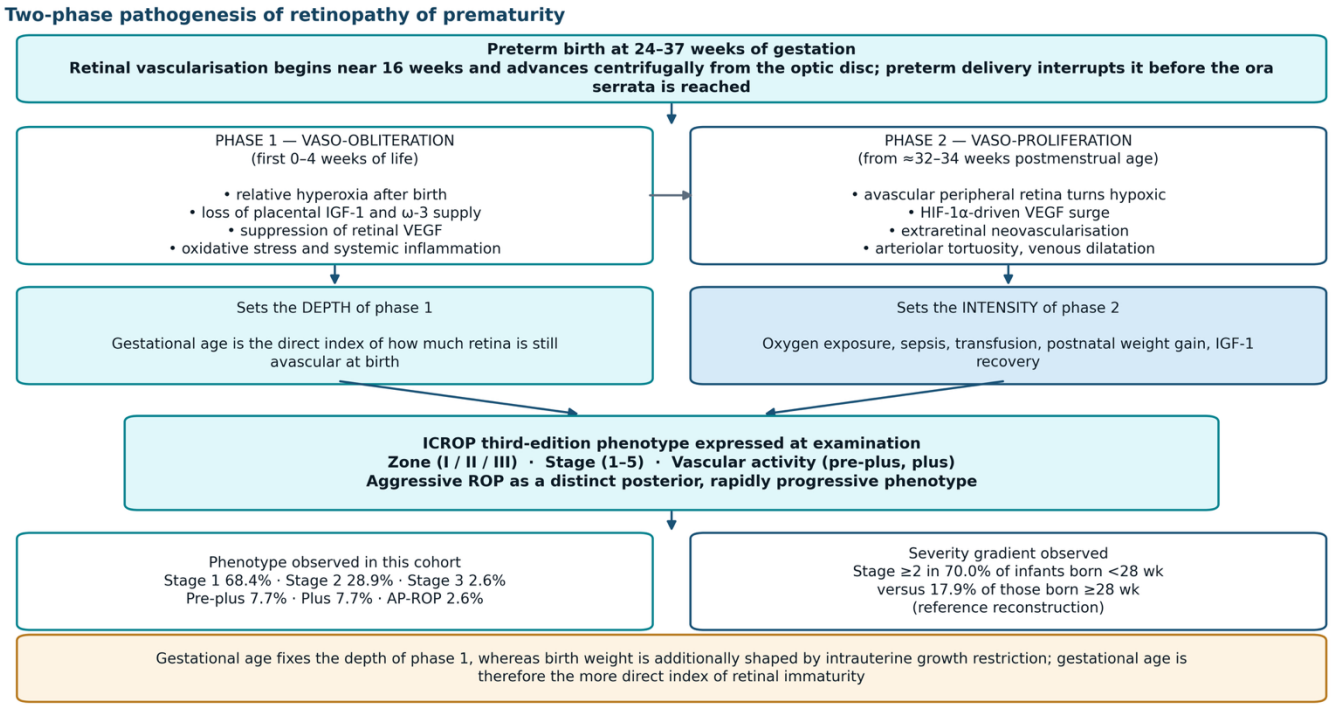

Figure 5. Two-phase pathogenesis of retinopathy of prematurity, showing why gestational age indexes the depth of the vaso-obliterative phase while postnatal exposures modulate the vaso-proliferative phase. The ICROP third-edition phenotype and the severity gradient observed in the present cohort are shown at the foot of the diagram.

A methodological point emerges that is not specific to ROP. The source analysis applied a Pearson χ^2 test to a table in which most expected cell counts were below 5, and then read a direction off the resulting P value. Neither step is defensible. Our simulation showed that at these margins the statistic is undefined in most samples generated under the null, and that among the samples in which it can be computed it rejects at 0.063 rather than at 0.05, so the asymptotic P value is not trustworthy in either direction. The deeper problem is interpretative, because an omnibus statistic carries no directional information at all, a point the admissible set makes concrete: every one of its 30 tables has an equally admissible mirror image, as panel A of Figure 4 shows. Partial identification made the boundary explicit: the exact P value is bounded only to 0.004 to 0.115, the odds ratio only to 0.03 to 9.16, and the direction is unconstrained, with 50.0% of admissible reconstructions running each way. Cramér's V appears pinned between 0.370 and 0.371, but

Equation 2 shows why that is an arithmetic consequence of conditioning on a reported P value and not an empirical finding. The bounds are set out in Table 5, and the arithmetic discrepancies that motivated them in Table 3. Reporting the bounds alongside the point estimate is not a counsel of despair; it states the finding at the resolution the data actually deliver.

For clinicians at Ngoerah General Hospital and at comparable Indonesian centres, the practical implication is that gestational age should govern the intensity of surveillance once an infant has entered the screening programme, while birth weight remains valuable as an entry criterion. The two roles are different and are often conflated. As an entry rule, a birth-weight threshold of 1500 g would have captured 94.9% of the ROP cases detected in this cohort and a threshold of 2000 g at least 94.9%, as the screening-yield block of Table 6 records. The wider evidence supports keeping the Indonesian threshold broad: validation of the postnatal growth criteria in a Pakistani tertiary hospital, and in Thai, Slovenian and Turkish cohorts, shows that criteria derived in one population do not transfer unchanged, and the DIGIROP decision-support models required revalidation even within Sweden.[39-43] For Indonesia specifically, a national risk-scoring model for ROP progression now exists and has been externally validated, and it should be paired with, not replaced by, a gestational-age-driven surveillance schedule of the kind set out in Figure 6.[34,35]

Two structural recommendations follow. The first is documentation. One infant in this cohort could not be staged at all, retinal zone was not recorded systematically, and neither postnatal oxygen exposure nor sepsis could be retrieved. A structured ICROP third-edition proforma recording zone, stage, vascular activity and extent in clock hours at every examination, of the kind embedded in the pathway of Figure 6, would close most of that gap at negligible cost. It would also make the unit's data usable by the tools that are now reaching clinical maturity: a reference-image grading scale for the pre-plus to plus continuum, artificial-intelligence vascular severity scores that raise expert accuracy, externally validated automated plus-disease detection, and smartphone-based telescreening that has been shown to work in an Indian LMIC setting.[20,21,22,44] The second is the referral letter. Where severity is communicated as ROP present rather than as a phenotype, receiving centres cannot triage and regional surveillance becomes impossible; the drift in the plus-disease standard between ICROP editions makes explicit phenotype recording more, not less, important.[18,19]

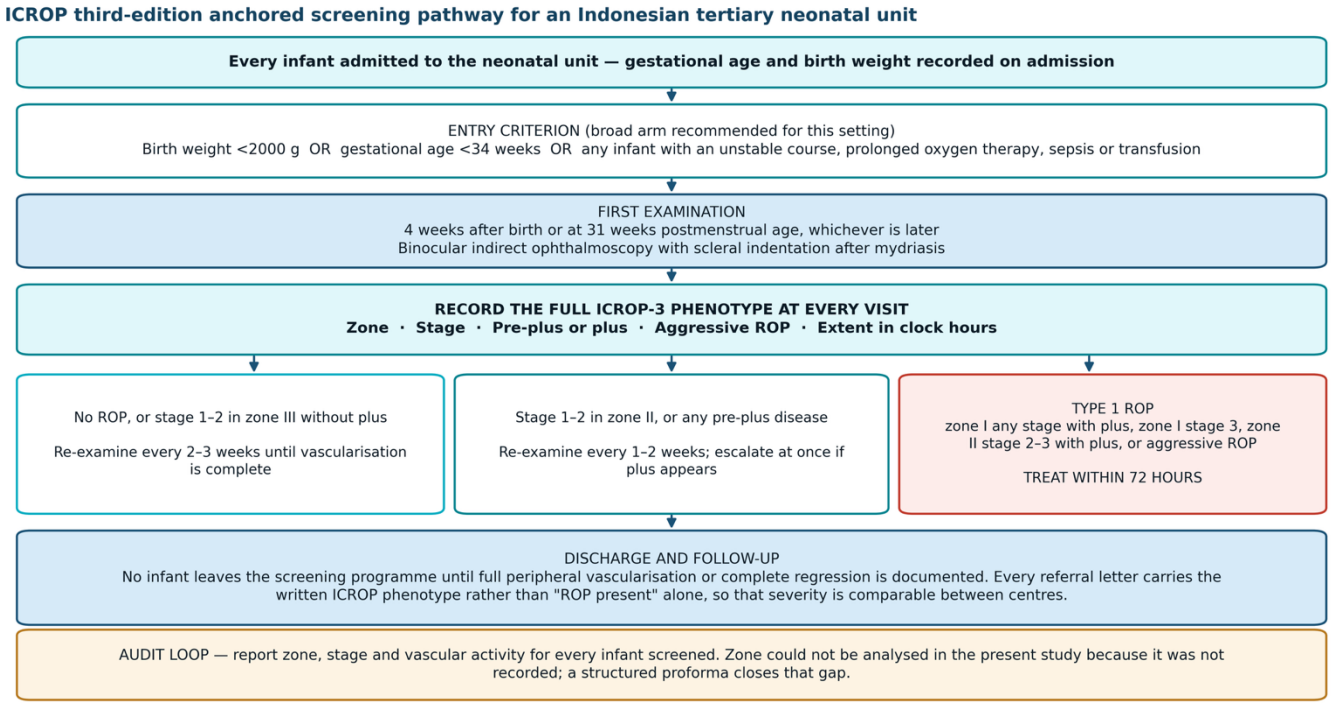

Figure 6. An ICROP third-edition anchored screening pathway for an Indonesian tertiary neonatal unit, with a broad birth-weight entry criterion, gestational-age-driven surveillance intensity, a structured phenotype proforma at every examination, and an audit loop that closes the documentation gap identified in this study.

Accurate phenotyping matters because it determines treatment, and treatment has changed. Anti-vascular endothelial growth factor agents are now an established alternative to laser for type 1 disease: the FIREFLEYE trial

compared intravitreal aflibercept with laser and its three-year extension has since reported ocular and developmental safety, while the RAINBOW extension and five-year follow-up provide the longest randomised comparison of ranibizumab against laser.[45-48] A network meta-analysis has quantified retreatment and reactivation risk across agents, a meta-analysis of neurodevelopmental outcomes after bevacizumab found no consistent excess of harm while stressing limited certainty, refractive outcomes appear more favourable after anti-VEGF than after laser at one year, and outcomes differ measurably between aggressive ROP and conventional type 1 disease.[49-52] None of these decisions can be taken from a record that states only that ROP is present, which is why the phenotype discipline argued for here is a clinical rather than merely an administrative requirement. The present study reports phenotype at diagnosis only, and the absence of treatment and visual-outcome data is a substantive limitation.

The strengths of this study are the application of the ICROP third edition, including vascular activity, which earlier Indonesian reports largely omitted; the use of exact rather than asymptotic inference at margins where the asymptotic approximation fails; the explicit audit of internal consistency in Table 3, which identified four discrepancies in the source summaries; and the partial-identification analysis in Table 5, which states which conclusions the data determine and which they do not. Equally, the analyses that the data could not support, including multivariable modelling, diagnostic accuracy and survival analysis, are reported as not performed rather than fabricated.

The limitations are substantial and constrain interpretation. The design is retrospective and cross-sectional, so no causal inference is possible and the temporal ordering of exposure and disease progression cannot be established. The study is confined to a single tertiary referral centre, and a referral cohort over-represents complex neonatal cases while under-representing infants managed in district facilities; generalisation to the wider Indonesian preterm population is therefore not warranted. The sample of 39 infants gives wide intervals for every proportion, with a half-width of 15.0 percentage points near 50%, as Figure 2 makes visible. Because the analysis proceeded from marginal tables rather than infant-level records, the cross-classification is a reconstruction: 30 tables remain compatible with the published summaries, the reference reconstruction is the one that maximises the odds ratio among them, and the rule that selects it rests on an assumption about the number of near-term infants that the source does not report. The reconstructed odds ratio should be read as the upper limit of a range rather than as an estimate. Retinal zone was not recorded, so type 1 disease could only be bounded and no infant could be classified as treatment-requiring with certainty. Postnatal oxygen exposure, sepsis, transfusion and postnatal weight gain were not retrievable, so residual confounding cannot be excluded and no adjusted estimate is offered. Finally, treatment and visual outcomes were not analysed, so the study speaks to phenotype at diagnosis and not to prognosis.

5. Conclusion

Among 39 preterm infants diagnosed with ROP at an Indonesian tertiary referral centre, disease was predominantly early stage, with 68.4% at stage 1 and treatment-warranted disease in between 2.6% and 10.3% of the cohort. Gestational age was strongly associated with severity, with stage ≥2 in 70.0% of infants born before 28 weeks against 17.9% of those born later (odds ratio 9.16, 95% CI 1.90–44.10, the upper limit of what the published summaries permit), whereas birth weight carried no usable information because almost the whole cohort occupied one stratum. Indonesian units should therefore let gestational age govern the intensity of surveillance while retaining a broad birth-weight entry criterion, and should adopt the structured ICROP third-edition proforma and surveillance pathway summarised in Figure 6, so that zone, stage and vascular activity are recorded for every infant. Prospective multicentre work with infant-level data on postnatal oxygen exposure, sepsis and weight gain is needed before national screening thresholds are revised.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Ethics Committee under approval number 2588/UN14.2.2.VII.14/LT/2023 and was conducted in accordance with the principles of the Declaration of Helsinki. Because the study used anonymised paediatric records collected during routine clinical care, individual written informed consent from parents or guardians was waived by the ethics committee under its protocol for retrospective analyses of anonymised data. Child assent was not applicable: every participant was an infant undergoing ROP screening, so no participant had reached the age of seven years at which assent is sought under Indonesian and

international standards for research involving children. No identifying information appears in this report or in any accompanying image, and the dataset analysed contained no direct identifiers.

Consent for publication

Separate consent for publication was not required because the manuscript contains no identifiable personal information or patient images.

Data availability

The data are not publicly available because they were derived from confidential clinical records. Requests for access to de-identified data may be directed to the corresponding author and will be considered in accordance with applicable institutional and ethical requirements.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no financial or non-financial conflicts of interest relevant to this work.

Author contributions

All authors contributed to the conception and design of the study, data acquisition, analysis and interpretation, drafting or critical revision of the manuscript, and approval of the final version. All authors agree to be accountable for the integrity and accuracy of the work.

Acknowledgements

The authors have no acknowledgements to declare.

Use of artificial intelligence

No artificial intelligence (AI) tools were used in the preparation of this manuscript.

References

1. Steinmetz JD, Bourne RRA, Briant PS, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health. 2021;9(2):e144-60. doi:10.1016/S2214-109X(20)30489-7
2. Wang S, Liu J, Zhang X, et al. Global, regional and national burden of retinopathy of prematurity in childhood and adolescence: a spatiotemporal analysis based on the Global Burden of Disease Study 2019. BMJ Paediatr Open. 2024;8(1):e002267. doi:10.1136/bmjpo-2023-002267
3. Ohuma EO, Moller AB, Bradley E, et al. National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. Lancet. 2023;402(10409):1261-71. doi:10.1016/S0140-6736(23)00878-4
4. Okwaraji YB, Krasevec J, Bradley E, et al. National, regional, and global estimates of low birthweight in 2020, with trends from 2000: a systematic analysis. Lancet. 2024;403(10431):1071-80. doi:10.1016/S0140-6736(23)01198-4
5. Wulandari F, Mahmudiono T, Rifqi MA, et al. Maternal characteristics and socio-economic factors as determinants of low birth weight in Indonesia: analysis of 2017 Indonesian Demographic and Health Survey (IDHS). Int J Environ Res Public Health. 2022;19(21):13892. doi:10.3390/ijerph192113892
6. Perin J, Mulick A, Yeung D, et al. Global, regional, and national causes of under-5 mortality in 2000-19: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet Child Adolesc Health. 2022;6(2):106-15. doi:10.1016/S2352-4642(21)00311-4
7. Penzias RE, Ogero MO, Tillya R, et al. Small and sick newborn care: changes in service readiness scoring between baseline and 2023 for 65 neonatal units implementing with NEST360 in Kenya, Malawi, Nigeria, and Tanzania. PLOS Glob Public Health. 2025;5(6):e0004367. doi:10.1371/journal.pgph.0004367
8. Fu Y, Lei C, Qibo R, et al. Insulin-like growth factor-1 and retinopathy of prematurity: a systemic review and meta-analysis. Surv Ophthalmol. 2023;68(6):1153-65. doi:10.1016/j.survophthal.2023.06.010
9. Athikarisamy S, Desai S, Patole S, et al. The use of postnatal weight gain algorithms to predict severe or type 1 retinopathy of prematurity. JAMA Netw Open. 2021;4(11):e2135879. doi:10.1001/jamanetworkopen.2021.35879
10. Yildirim M, Coban A, Bulut O, et al. Postnatal weight gain and retinopathy of prematurity in preterm infants: a population-based retrospective cohort study. J Matern Fetal Neonatal Med. 2024;37(1):2337720. doi:10.1080/14767058.2024.2337720
11. Zhao K, Jiang Y, Zhang J, et al. Celastrol inhibits pathologic neovascularization in oxygen-induced retinopathy by targeting the miR-17-5p/HIF-1alpha/VEGF pathway. Cell Cycle. 2022;21(19):2091-108. doi:10.1080/15384101.2022.2087277
12. Srivatsa B, Malcolm K, Clark RH, et al. Effect of a novel oxygen saturation targeting strategy on mortality, retinopathy of prematurity, and bronchopulmonary dysplasia in neonates born extremely preterm. J Pediatr. 2021;234:33-37.e3. doi:10.1016/j.jpeds.2021.03.007
13. Aggarwal V, Bhatia R, Tan K. Oxygen saturation levels and retinopathy of prematurity in extremely preterm infants: a case control study. BMC Pediatr. 2023;23(1):449. doi:10.1186/s12887-023-04278-6
14. Boskabadi H, Abrishami M, Shoeibi N, et al. Incidence of retinopathy of prematurity in neonates with neonatal sepsis. Arch Iran Med. 2023;26(5):248-51. doi:10.34172/aim.2023.38
15. Wang X, Rao R, Li H, et al. Red blood cell transfusion for incidence of retinopathy of prematurity: prospective multicenter cohort study. JMIR Pediatr Parent. 2024;7:e60330. doi:10.2196/60330
16. Arora P, Mahapatra M, Arora K. Retinopathy of prematurity: incidence and risk factor analysis in small for gestational age neonates compared to appropriate for gestational age. J Pharm Bioallied Sci. 2023;15(Suppl 2):S1266-9. doi:10.4103/jpbs.jpbs_130_23
17. Du J, Gu X, Jiang S, et al. Association of intrauterine growth with retinopathy of prematurity risk in very preterm infants: a multicenter cohort study. Neonatology. 2026;123(1):46-56. doi:10.1159/000548929
18. Chiang MF, Quinn GE, Fielder AR, et al. International classification of retinopathy of prematurity, third edition. Ophthalmology. 2021;128(10):e51-68. doi:10.1016/j.ophtha.2021.05.031
19. Wang W, Freedman SF, Wallace DK, et al. Quantitative comparison of vascular characteristics of plus disease images published over time: International Classification of Retinopathy of Prematurity. J AAPOS. 2025;29(1):104101. doi:10.1016/j.jaapos.2024.104101
20. Binenbaum G, Stahl A, Coyner AS, et al. P score: a reference image-based clinical grading scale for vascular change in retinopathy of prematurity. Ophthalmology. 2024;131(11):1297-303. doi:10.1016/j.ophtha.2024.05.019
21. Coyner AS, Young BK, Ostmo SR, et al. Use of an artificial intelligence-generated vascular severity score improved plus

disease diagnosis in retinopathy of prematurity. Ophthalmology. 2024;131(11):1290-6. doi:10.1016/j.ophtha.2024.06.006

22. Bai A, Dai S, Hung J, et al. Multicenter validation of deep learning algorithm ROP.AI for the automated diagnosis of plus disease in ROP. Transl Vis Sci Technol. 2023;12(8):13. doi:10.1167/tvst.12.8.13
23. Moir JT, Hyman MJ, Skondra D, et al. Risk factors for severe retinopathy of prematurity stratified by birth weight and gestational age in privately insured infants. J AAPOS. 2024;28(6):104049. doi:10.1016/j.jaapos.2024.104049
24. Wu T, Rao R, Gu H, et al. Retinopathy of prematurity: risk stratification by gestational age. J Perinatol. 2023;43(6):694-701. doi:10.1038/s41372-023-01604-9
25. Chen W, Liu T, Xiang D. Risk factors and prevalence of retinopathy of prematurity in very low birth weight preterm infants: a systematic review and meta-analysis. Transl Pediatr. 2026;15(2):29. doi:10.21037/tp-2025-aw-785
26. Dai Y, Zhu L, Zhou Y, et al. Ten-year trend of retinopathy of prematurity among extremely preterm infants in one neonatal intensive care unit in China. Front Pediatr. 2021;9:717090. doi:10.3389/fped.2021.717090
27. Chen X, Xu Y, Wang Y, et al. Incidence, risk factors and evolving treatment of severe retinopathy of prematurity in China: a retrospective multicenter cohort study in 70 neonatal intensive care units. Front Med (Lausanne). 2025;12:1652727. doi:10.3389/fmed.2025.1652727
28. Najmuangchan N, Ngerncham S, Piampradad S, et al. Risk factor-based models to predict severe retinopathy of prematurity in preterm Thai infants. Indian J Ophthalmol. 2024;72(Suppl 3):S514-20. doi:10.4103/ijo.ijo_1640_23
29. Ittarat M, Chansaengpetch S, Chansangpetch S. Incidence and risk factors for retinopathy of prematurity at a rural tertiary hospital in Thailand. J Ophthalmic Vis Res. 2023;18(1):81-7. doi:10.18502/jovr.v18i1.12728
30. Sanghi G, Sawhney JS, Kaur S, et al. Evaluation of clinical profile and screening guidelines of retinopathy of prematurity in an urban level III neonatal intensive care unit. Indian J Ophthalmol. 2022;70(7):2476-9. doi:10.4103/ijo.ijo_1925_21
31. Shaik R. Prevalence, risk factors and severity of retinopathy of prematurity in preterm infants in a tertiary care hospital in rural Karnataka. Indian J Clin Exp Ophthalmol. 2023;9(2):232-40. doi:10.18231/j.ijceo.2023.044
32. Parihar K, Gupta PK, Singh V, et al. Retinopathy of prematurity: incidence, risk factors and outcome in North Indian rural and semi-urban population. J Nepal Paediatr Soc. 2022;42(1):119-23. doi:10.3126/jnps.v42i1.39034
33. Siswanto JE, Bos AF, Dijk PH, et al. Multicentre survey of retinopathy of prematurity in Indonesia. BMJ Paediatr Open. 2021;5(1):e000761. doi:10.1136/bmjpo-2020-000761
34. Siswanto JE, Adisasmita AC, Ronoatmodjo S, et al. A risk scoring model to predict progression of retinopathy of prematurity for Indonesia. PLoS One. 2023;18(2):e0281284. doi:10.1371/journal.pone.0281284
35. Siswanto JE, Adisasmita AC, Ronoatmodjo S, et al. Validation of a screening score model to predict the development of retinopathy of prematurity. Sci Rep. 2025;15(1):40571. doi:10.1038/s41598-025-24303-1
36. Kartamihardja WI, Mutiara G, Harun YZ, et al. Risk factors of immature retina on the first screening for retinopathy of prematurity. Paediatr Indones. 2023;63(3):189-94. doi:10.14238/pi63.3.2023.189-94
37. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7. doi:10.1016/S0140-6736(07)61602-X
38. Pivodic A, Holmstrom G, Smith LEH, et al. Prognostic value of parenteral nutrition duration on risk of retinopathy of prematurity. JAMA Ophthalmol. 2023;141(8):716-24. doi:10.1001/jamaophthalmol.2023.2336
39. Tayyab H, Mirza A, Choudhary RA, et al. Validation of postnatal growth and retinopathy of prematurity (G-ROP) screening guidelines in a tertiary care hospital of Pakistan: a report from low-middle income country. PLoS One. 2024;19(5):e0302534. doi:10.1371/journal.pone.0302534
40. Bunyavee C, Hemarat K. Accuracy of the postnatal growth and retinopathy of prematurity screening criteria in predicting prethreshold retinopathy of prematurity in the tertiary hospital, Bangkok, Thailand. Taiwan J Ophthalmol. 2023;13(4):535-9. doi:10.4103/tjo.tjo-d-23-00027
41. Zupan M, Binenbaum G, Tekavcic Pompe M. Validation of the postnatal growth and retinopathy of prematurity (G-ROP) criteria in a Slovenian cohort. J AAPOS. 2025;29(1):104113. doi:10.1016/j.jaapos.2025.104113
42. Ekinci DY, Bayraktar H, Sero OL, et al. Retrospective validation of the postnatal growth and retinopathy of prematurity (G-ROP) and Colorado retinopathy of prematurity (CO-ROP) models in a Turkish cohort. Turk J Pediatr. 2023;65(3):398-405. doi:10.24953/turkjped.2022.843
43. Pivodic A, Smith LEH, Hard AL, et al. Validation of DIGIROP models and decision support tool for prediction of treatment for retinopathy of prematurity on a contemporary Swedish cohort. Br J Ophthalmol. 2022;107(8):1132-8. doi:10.1136/bjophthalmol-2021-320738
44. Young BK, Cole ED, Shah PK, et al. Efficacy of smartphone-based telescreening for retinopathy of prematurity with and without artificial intelligence in India. JAMA Ophthalmol. 2023;141(6):582-8. doi:10.1001/jamaophthalmol.2023.1466
45. Stahl A, Sukgen EA, Wu WC, et al. Effect of intravitreal aflibercept vs laser photocoagulation on treatment success of retinopathy of prematurity. JAMA. 2022;328(4):348-59. doi:10.1001/jama.2022.10564
46. Stahl A, Nakanishi H, Lepore D, et al. Three-year outcomes of intravitreal aflibercept versus laser therapy for retinopathy of prematurity: interim analysis of the efficacy and safety outcomes in the FIREFLEYE next trial. Neonatology. 2026;123(3):276-85. doi:10.1159/000549717
47. Marlow N, Stahl A, Lepore D, et al. 2-year outcomes of ranibizumab versus laser therapy for the treatment of very low birthweight infants with retinopathy of prematurity (RAINBOW extension study): prospective follow-up of an open label, randomised controlled trial. Lancet Child Adolesc Health. 2021;5(10):698-707. doi:10.1016/S2352-4642(21)00195-4
48. Marlow N, Reynolds JD, Lepore D, et al. Ranibizumab versus laser therapy for the treatment of very low birthweight infants with retinopathy of prematurity (RAINBOW): five-year outcomes of a randomised trial. eClinicalMedicine. 2024;71:102567. doi:10.1016/j.eclinm.2024.102567
49. Chang E, Josan AS, Purohit R, et al. A network meta-analysis of retreatment rates following bevacizumab, ranibizumab, aflibercept, and laser for retinopathy of prematurity. Ophthalmology. 2022;129(12):1389-401. doi:10.1016/j.ophtha.2022.06.042
50. Baiad AA, Kherani IZ, Popovic MM, et al. A meta-analysis of neurodevelopmental outcomes following intravitreal bevacizumab for the treatment of retinopathy of prematurity. Neonatology. 2023;120(5):577-88. doi:10.1159/000531541
51. Kiran Yenice E, Petricli IS, Kara C. One-year refractive outcomes after intravitreal bevacizumab versus laser photocoagulation for retinopathy of prematurity. Int Ophthalmol. 2023;43(7):2197-202. doi:10.1007/s10792-022-02615-9
52. Comez A, Karakucuk Y, Ozmen MC, et al. The results of intravitreal bevacizumab monotherapy for treating aggressive posterior retinopathy of prematurity and type 1 retinopathy of prematurity. Eye (Lond). 2021;35(12):3302-10. doi:10.1038/s41433-021-01413-4