Retinopathy of Prematurity: Improving vision outcomes by development of a clinical decision-making model using digital analysis of retinal images.

Retinal Image Digital Analysis (RIDA) study.

Study Protocol. B W Fleck. June 2012.

Summary

The aim of the programme is to develop a clinical decision-making model using digital analysis of retinal images, in order to achieve better vision in children who develop ROP following premature birth.

Retinopathy of Prematurity (ROP) is a blinding condition affecting premature babies. Retinal laser treatment can improve visual outcome. ROP is one of the few eye conditions affecting children in which disability and blindness should now be preventable by timely treatment. However, visual impairment due to ROP remains a major health problem in the UK, and worldwide. Eye examinations to detect ROP require a high level of expertise not available in many countries. Each year in the UK, 8,000 babies are screened for ROP and about 500 require laser treatment. The challenge is greater because the decision to treat ROP rests on clinical judgements about the calibre and tortuosity of retinal vessels (Plus disease).

The BOOST (Benefits Of Oxygen Saturation Targeting) II study is an MRC-funded multicentre study comparing the outcomes from two levels of supplemental oxygen for babies. Very detailed neonatal and ophthalmic data has been collected from 973 babies, including ~ 5,000 digital retinal photographs. Recruitment closed on 24th December 2010. Vision and eye health outcome measurements are currently being obtained at age 2 years (corrected for gestation).

Using BOOST-II clinical and digital image trial data we will study neonatal and ophthalmic factors associated with sight-threatening disease. We will use image analysis software (RetVas) to automatically quantify the clinically most important feature of sight threatening ROP – "plus" disease (dilatation and tortuosity of the posterior retinal blood vessels). We will also study the factors which determine outcome following treatment. We aim to

develop and test a clinical decision-making model for retinopathy of prematurity (ROP). We anticipate that this will substantially improve clinical decision making, so that we can reduce the numbers of children with visual impairment due to ROP. The work will also lead to the development of screening protocols by nurses, in which the expertise of ophthalmologists may be concentrated on validating the presence of sight-threatening ROP, and treating it. This will reduce the ophthalmic workload in the UK and promote access to ophthalmic services for babies elsewhere.

Introduction

Retinopathy of Prematurity (ROP) is a disorder of retinal vascular development in premature infants. It remains a major cause of childhood blindness worldwide. ROP is one of the few eye conditions affecting children in which disability and blindness should now be preventable by timely treatment. However, visual impairment due to ROP remains a major health problem in the UK, and worldwide(Gilbert 2008). Eye examinations to detect ROP require a high level of expertise. Each year in the UK, 8,000 babies are screened for ROP and about 500 require laser treatment (Fielder, Haines et al. 2002; Haines, Fielder et al. 2002; Haines, Fielder et al. 2005).

Most very premature babies require supplemental oxygen for several weeks after birth, but high arterial oxygen tensions are known to be toxic to the developing retina, causing ROP. A global collaboration of multicentre randomized controlled trials of reduced oxygen saturation therapy has recently completed recruitment. Infants were randomized to oxygen saturation range of 85 – 89% or 91 – 95%, using oxygen saturation monitors that were offset in a masked way to produce one of these two ranges, while appearing to show the "normal" range of 89 – 92% to those undertaking clinical care of the infants. One of these trials is the MRC-funded Benefits of Oxygen Saturation Targeting (BOOST-II UK) trial. Recruitment closed on 24th December 2010. Vision and eye health outcome measurements are currently being obtained at age 2 years (corrected for gestation). Very detailed neonatal and ophthalmic data is available from 973 babies. Preliminary data from the trial has recently been published (Stenson B, Brocklehurst P, Tarnow-Mordi W; Increased 36-week survival with high oxygen saturation target in extremely preterm infants. U.K. BOOST II trial; Australian

BOOST II trial; New Zealand BOOST II trial. N Engl J Med. 2011 Apr 28;364(17):1680-). The composite primary outcome for the trial is mortality and major disability at age 2 years (corrected for prematurity). Secondary outcomes include retinopathy of prematurity as well as duration of oxygen therapy, chronic lung disease, growth and health service utilisation. The early ophthalmic outcome measures of the trial are blindness or retinal detachment. The 2 year outcomes will include an assessment of visual function and of retinal appearance.

The UK trial addresses some issues not tackled by any of the other trials. These include a detailed description of the structural and functional retinal damage that occurs in infants within the trial. The BOOST-II UK trial is the first ROP trial in which a significant proportion of trial ophthalmologists use retinal digital imaging as their usual method of retinal screening examination (RetCam). Thus, a very large digital retinal images dataset from recruited babies has been collected for analysis. The BOOST-II UK Retinal Image Digital Analysis Study (BOOST-II UK RIDA Study) is a nested study within the BOOST-II UK Trial, and is fully endorsed by the Trial Steering Group.

Retinal blood vessel changes seen in ROP

The international classification of ROP (ICROP) (International Committee for the Classification of Retinopathy of 2005) has provided the basis of all multicentre studies in ROP over the past 2 decades as it permits a standardised approach that allows robust comparison between clinicians and between centres.

The retinal blood vessels grow centrifugally out from the optic disc, commencing around 14 weeks gestation and by 40 weeks the retinal is fully vascularised and the risk of developing ROP has passed.

ROP has 4 descriptors:

Severity by stage. Stages 1-5 and aggressive posterior ROP;

Extent. How much of the retinal circumference in clock hours is affected.

Location. There are 3 zones which describe how far the retinal blood vessels have grown out from the optic disc. When the blood vessels are confined to Zone 1, retinal blood vessel development is very immature and the baby has a high risk of developing severe and sight-threatening ROP while when vascularisation has reached zone 3 the risk of vision impairment is close to zero.

A key feature of the classification is "Plus" disease. "Plus" disease refers to dilatation and tortuosity of the retinal blood vessels close to the optic disc. A recent randomised trial has amended the indications for treatment so that the presence of plus disease is now the main driver for intervention (Early Treatment For Retinopathy Of Prematurity Cooperative 2003). Over the past decade digital imaging has been developed that is capable of retinal imaging in premature infants. We have performed a RCT to compare the performance of Retinal Digital Imaging as an alternative to binocular indirect ophthalmoscope examination screening(Dhaliwal, Wright et al. 2009). Retinal Digital Imaging performed well. Digital imaging allows for the first time quantitative image analysis of retinal parameters, rather than reliance on the subjective interpretation of expert readers, which is not robust. Inter-observer agreement of classification of "plus" disease is low (Chiang, Gelman et al. 2007; Wallace, Quinn et al. 2008; Shah, Wilson et al. 2011; Slidsborg, Forman et al. 2012), and this may partly explain variation in reported rates of ROP treatments in different hospitals(Darlow, Hutchinson et al. 2005). The quantification of "plus" disease by image analysis is therefore an important goal of software development. A number of software programs have been developed to quantify blood vessel dilation and tortuosity, including a program developed by members of our group(Wilson, Cocker et al. 2008). The use of objective measures of "plus" disease is likely to substantially improve the diagnostic accuracy of ROP treatment decisions. This represents a very important factor in the management of premature infants.

A large library of digital retinal images from babies in the BOOST II UK Trial is now being built. The images are held on "RetCam" digital camera systems at each trial site. A research clinician is currently in the process of visiting each site in order to select and download groups of images from each RetCam machine, select the best quality images, and upload the images to a secure study server. Robust and Secure IT systems are in place. Software has been written in order to facilitate secure on line reading of the retinal images.

A secure, anonymised link to the clinical information of each baby in the BOOST II UK trial has been established, so that the images can be matched to the clinical condition of each baby. This data will provide a unique opportunity to study risk factors for the development of ROP; the natural history of ROP up to the point of treatment intervention; and the potential diagnostic role of quantified retinal blood vessel image analysis in ROP.

We have previously developed methods of rapid objective quantification of ROP using digital analysis of retinal photographs. Using BOOST-II clinical and digital image trial data we will study neonatal and ophthalmic factors associated with sight-threatening disease. We will also study the factors which determine outcome following treatment. We aim to develop and test a clinical decision-making model for retinopathy of prematurity (ROP). We anticipate that this will substantially improve clinical decision making, so that we can reduce the numbers of children with visual impairment due to ROP.

Research plan

Hypotheses

Hypothesis 1. Infants managed with lower oxygen saturation will have more advanced retinal vascularisation, and their ROP will be more peripherally located in the retina and be less severe. They will be less likely to need therapeutic intervention (the two arms of the BOOST II UK trial).

Hypothesis 2. Infants at risk of developing severe ROP may be identified at an early postnatal age and individual risk factors and their relative significance may be identified.

Hypothesis 3. Retinal structural and functional outcomes are influenced by: the severity of ROP at the time of treatment and by the completeness of treatment (area of avascular retina covered by laser burns).

Hypothesis 4. Digital retinal photographs may be used for remote "telemedicine" reading and diagnostic decisions.

Hypothesis 5. The severity of "plus" disease (posterior retinal blood vessel dilatation and tortuosity) may be quantified by computerised image analysis, and matched with clinical findings.

Hypothesis 6. A clinical model of ROP that uses quantitative retinal image analysis information will enable reliable and prompt treatment decisions to be made.

Experiments planned to test each hypothesis.

Hypothesis 1. Infants managed with lower oxygen saturation will have more advanced retinal vascularisation, and their ROP will be more peripherally located in the retina and be less severe. They will be less likely to need therapeutic intervention (the two arms of the BOOST II UK trial).

Experiment 1

"Infants managed with lower oxygen saturation will have more advanced retinal vascularisation, and their ROP will be more peripherally located in the retina."

We will measure the location of retinal vascularisation in two ways: measurements of RetCam images of retinal blood vessels, and clinical grading of zones (1,2,3) from data obtained during the BOOST II UK trial. The timing of measurements is key in this context. We will compare the two groups of the BOOST II UK trial at time points when any separation of the two groups is likely to be maximal, and when we have the maximum number of retinal examination data: 32, 34 and 36 weeks corrected gestational age.

Details. RetCam image measurements will be done by the Clinical Fellow, using the highest quality image available for analysis, at the specified gestational ages. The measurements will be of the optic disc margin to the edge of retinal vascularisation on the horizontal meridian, nasally and temporally, in pixels). Measurements will only be made if the extent of vascularisation is clearly defined – either a clearly defined extent of vascularised retina meeting avascular retina, or an ROP line or ridge. The results will be tabulated for oxygen

treatment arm and corrected gestational age. They will be stratified for gestational age at birth. The significance of any differences between the treatment arms will be measured.

Experiment 2

Severity of ROP in the two BOOST II UK groups.

This information will be derived from RetCam images and from clinical grading of disease stages from data obtained during the BOOST II UK trial. We will measure the maximum disease severity (ROP stage and plus disease) for each infant and report the proportion of infants in the two study groups who developed a maximum stage of 0, 1, 2, 3, (4); and the proportion who developed no plus disease, pre plus disease, and plus disease. Data from the RetCam images will be used to validate clinical reporting of the BOOST II UK trial Details. The RetCam images will be read by one of a team of 6 expert readers, and the maximum stage of ROP viewed during each examination will be recorded. The gestational age at the time of the examination, and the gestational age at birth will also be recorded. Inter-observer variation in reading decisions will be measured using 60 images of various disease severities shown to all 6 readers. The results will be tabulated for oxygen treatment arm and corrected gestational age. They will be stratified for gestational age at birth. The significance of any differences between the treatment arms will be measured.

Experiment 3

The number of infants who required ROP treatment intervention in the two arms of the BOOST II UK trial will be recorded, using BOOST II UK trial data. The significance of any difference between the treatment arms will be measured.

The information will also be stratified be gestational age at birth, and birth weight. The gestational age at the time of treatment will also be recorded. The number of treatments in each group performed using diode laser, other laser, or anti VEGF agents will be recorded.

Hypothesis 2. Infants at risk of developing severe ROP may be identified at an early postnatal age and individual risk factors and their relative significance may be identified.
Experiment 4

The effects of neonatal variables including: different amounts of oxygen therapy (the 2 arms of the BOOST trial); differing postnatal weight gain; IVH, systemic infection, NEC, blood transfusions, will be studied in relation to the development of ROP that required treatment, and the gestational age at which treatment was performed. BOOST II UK clinical data will be used.

Details: The two groups to be analysed will be: ROP treated and ROP not treated. Differences between the two groups in the following parameters will be tested: birth weight; gestational age at birth; oxygen group; postnatal weight gain (absolute, and relative, at 6 weeks post natal); presence or absence of IVH; presence or absence of systemic sepsis; presence or absence of NEC; presence or absence of blood transfusion. A regression analysis will be performed, to test for independent associations. Relative risks for each factor will be calculated. A regression risk model will be developed.

The gestational age at which treatment was performed will be studied using correlation with: gestational age at birth; birth weight; postnatal weight gain (absolute, and relative, at 6 weeks post natal). Interactions will be measured. The difference in the mean gestational age at which treatment was performed will be tested between the two oxygen groups.

.

Hypothesis 3. Retinal structural and functional outcomes are influenced by: the severity of ROP at the time of treatment and by the completeness of treatment (area of avascular retina covered by laser burns).

Experiment 5

The severity of ROP at the time of treatment. RetCam images taken from each eye immediately prior to treatment will be graded by image readers. For infants where RetCam images are not available, the clinical grading at the time of treatment will be used. The structural and anatomical outcome at 2 years will be compared to the severity of disease present immediately prior to treatment. The outcomes of pairs of eyes in patients with asymmetric disease will be compared.

Details: . Reference photographs (ICROP) will be used to inform expert reader grading decisions. Image reading will be done by 3 expert readers and the consensus (majority) grading will be used. Grading will be by maximum stage of ROP present; the presence of no plus disease, pre-plus disease or plus disease; the lowest zone of ROP present. These results

will be compared with the clinical grading recorded by the examining ophthalmologist at the time of treatment (or at the time when a decision to treat was made if a separate grading at the time of treatment was not recorded). Agreement will be measured for stage, plus and zone. These measures test the reliability of BOOST II UK clinical data for treatment decisions.

The severity of ROP at the time of treatment will be grouped into mild, moderate and severe. The groups will be based on zone, stage and plus disease. Results will be presented as both "per eye" (all eyes included) and "per infant" (determined by the eye with worst severity at the time of treatment). The outcomes at 2 years will be formed into composite groupings of mild, moderate and severe deficit. The composites will be made up from anatomical abnormality of the fundi; visual acuity; amblyopia; refraction; strabismus. The proportions of outcome category for each disease severity category will be measured, and differences tested. In the "per eye" analysis, the composite outcomes will be measures that are independent of ocular interaction (i.e. exclude strabismus and amblyopia): anatomy of the retina; refraction. In the "per infant" analysis, the composite outcomes will include the presence or absence of strabismus and the presence or absence of amblyopia (reduced visual acuity that is not explained by anatomical abnormality)

For pairs of eyes with asymmetrical disease at the time of treatment where one eye falls out with ETROP Type 1 treatment criteria, the outcome measures will be tabulated for the "requires treatment" eyes and the "fellow eyes" and differences will be tested.

Experiment 6
The completeness of treatment (area of avascular retina covered by laser burns). RetCam images taken from each eye following treatment will be graded by image readers. The grades will measure the proportion of avascular retina covered by laser burns in each treated eye. The grades will be 98% – 100%; 80% - 98%; 50% - 80% and less than 50%. A pilot study will be used to confirm or alter the appropriateness of these grades. The following measures will be correlated with the completeness of treatment: need for re-treatment; time interval following treatment until complete resolution of plus disease (measured by reading plus disease in images taken during follow up after treatment, and from BOOST II UK clinical returns); structure and function outcome measures at 2 years

(using composites as outlined in experiment 5). The effect of the severity of disease at the time of treatment (experiment 5) will be included in the analysis.

Hypothesis 4. Digital retinal photographs may be used for remote "telemedicine" reading and diagnostic decisions.

Experiment 7

Diagnostic decisions made by trial ophthalmologists at the time of examination will be compared with decisions made by a panel of experts who read the photographs remotely. A sample of images of ROP of various degrees of severity, including images taken immediately prior to laser treatment, will be read by expert readers and compared with the record made by the examining ophthalmologist at the time of the examination. The disease severity score variation between these groups will be measured. The degree of consistency of diagnostic scoring, and particularly treatment decision making, will be measured. We will use approximately 200 images for these experiments, using a panel of 6 UK and international ROP experts. This process will also quality assure BOOST-II UK ophthalmology data.

Details. Reference photographs (ICROP) will be used to inform expert reader grading decisions. Image reading will be done by 6 expert readers and the consensus (majority) grading will be used. Grading will be by maximum stage of ROP present; the presence of no plus disease, pre-plus disease or plus disease; the lowest zone of ROP present. These results will be compared with the clinical grading recorded by the examining ophthalmologist at the time of the clinical examination as per the BOOST II UK data return. Agreement will be measured for stage, plus and zone.

When pre plus or plus disease is thought to be present, additional responses will be requested: the degree of arteriolar and venular tortuosity and dilatation adjacent to the optic disc in each quadrant of each eye. Semi quantitative scores using reference photographs will be requested. This data will be used to calibrate image analysis software scores of plus disease.

Hypothesis 5. The severity of "plus" disease (posterior retinal blood vessel dilatation and tortuosity) may be quantified by computerised image analysis, and matched with clinical findings.

Experiment 8

The images and expert reading processes described in experiment 7 will be used. In addition, the Clinical Fellow will analyse each image for plus disease using RetVas (see Appendix 2). We will correlate these measurements with functional and anatomic outcomes at 2 years of both treated and untreated infants.

Details. For each measurement (arteriolar and venular tortuosity and dilatation in each quadrant of each eye), the median expert reader score will be obtained. For each expert score group, the confidence intervals of the RetVas value will be reported. The RetVas values that best correlate with expert readers decisions on plus disease, and pre-plus disease will be analysed. The RetVas values that best define plus disease will then be determined. E.g. Plus disease may be defined as arteriolar tortuosity score above "x", with confidence intervals.

Composite outcomes for eyes and for infants (see experiment 5) will be measured for groups "no plus", "pre-plus" and "plus". The composite outcomes will also be measured for each RetVas measurement (arteriolar and venular tortuosity and dilatation in each quadrant of each eye), and the measurement that best predicts outcome will be determined. An attempt will then be made to determine the RetVas value (of one type of RetVas measurement, or of a combination of measurements) that best correlates with a reduced composite outcome. The composite outcomes will also be compared for eyes that are measured as "plus" using RetVas software (following clinical calibration experiments described above) , against eyes that are measured as "pre-plus" and as "no plus".

Hypothesis 6. A clinical model of ROP that uses quantitative retinal image analysis information will enable reliable and prompt treatment decisions to be made.
Experiment 9

We will develop clinical models for the detection of ROP that requires treatment. These will use clinical data, and image analysis data.

Details. Results obtained from experiments 4 and 8 will be combined. Infants who were treated for ROP, and a matched group who were not treated for ROP will be compared. RetCam images obtained during the weeks prior to the point at which the treatment decision was made will be analysed. This will also be done for the control group. At each

gestational age in weeks, RetVas analysis will be done. RetVas values that predict the later need for treatment will be determined. This will be done by:

1. Plotting the change in RetVas values over time during the weeks preceeding the need for treatment. What is the pattern of change – linear / exponential / a "tipping point" inflexion?

2. Determining at what time prior to the point of need for treatment the RetVas becomes predictive of the need for treatment. The RetVas values will be included in the regression model obtained in experiment 4 to determine at what point the RetVas value becomes an independent predictor of the need for treatment. One possible outcome may be that RetVas values are not independent predictors of the need for treatment until the point at which treatment is needed has been reached. In this situation, RetVas would not be useful as a predictor of the future need for treatment, but would might remain useful as a more objective diagnostic measure of the need to treat than conventional clinical decision making.

3. When combined with data from experiment 4, a probability estimate of the later need for treatment might be calculated. E,g. if an infant born at 26 weeks gestation and 800g birth weight with a postnatal weight gain of X, when examined at 33 weeks postnatal has a RetVas plus score of Y, then the probability of later needing ROP treatment is 88 – 96%.

We cannot be clear as to whether the numbers of cases we will have available to study will be sufficient to reach a robust disease risk model as described above. However, our results may provide pilot data for future studies.

The team, and their roles.

Edinburgh

Clinical Fellow PhD student. An ophthalmology trainee. Employed by the University of Edinburgh and based in Department of Population Health Sciences (Professor Harry Campbell), with periods of image analysis work done in the Clinical Research Imaging Centre (CRIC), Edinburgh Royal Infirmary, where the fellow will join a group of other clinical PhD students who are working in image analysis projects, including retinal blood vessel image analysis. During the early part of the work the Fellow will visits London to learn RetVas

analysis techniques, and Oxford to become familiar with the BOOST II UK clinical dataset, and the working processes of NPEU.

Professor Brian Fleck. Department of Ophthalmology, Edinburgh. Principal Investigator; team leader; supervision of Clinical Fellow PhD student. Professor Fleck is a Clinical Ophthalmologist with laboratory and clinical interest in Retinopathy of Prematurity, (and co-author of one paper on the theory of retinal blood vessel mathematical analysis), and lead ophthalmologist for the BOOST II UK trial.

Professor Harry Campbell, Co-Director Centre for Population Health Sciences, University of Edinburgh, and colleagues in his department will facilitate analysis of the experiments, and act as the senior PhD supervisor.

Dr Tom MacGillivary, medical physicist, manager Image Analysis Core, CRIC. Day to day facilitating of image analysis activities of the Clinical Fellow PhD student.

Professor Ben Stenson. Department of Neonatology, Edinburgh. Lead neonatologist, and neonatology supervisor of the Clinical Fellow PhD student. Lead neonatologist of the BOOST II UK trial.

London

Dr Clare Wilson: Lecturer in Ophthalmology, University College London. Dr Wilson successfully undertook a PhD in retinal blood vessel image analysis in ROP, using RetVas and other software programs. She has personally performed a range of image analysis experiments using expert readers and software analysis of a similar nature to those that will be performed by the clinical fellow PhD student. She will train the clinical fellow PhD student in the use of RetVas, and provide guidance in the design and conduct of image analysis experiments.

Professor Alistair Fielder, Professor Emeritus of Ophthalmology, City University, London. Professor Fielder remains a leading international figure in the field of ROP, and in the role of

image analysis in the diagnosis of ROP. He supervised Dr Wilson's PhD, and remains part of a group in international leaders in ROP who undertake telemedicine image reading experiments from time to time. It is anticipated that a number of experts from this group will provide expert image reading input.

Mr Ken Cocker, programmer and data processor, has worked closely with Prof Fielder and Dr Wilson for a number of years, and has co-authored a number of papers on image analysis in ROP. Mr Cocker built the image database used for this study, and is the data processor for the image database. He has contributed to the development of the RetVas image analysis software program. He supported Dr Wilson in setting up on line image reading experiments for her PhD, and will also support the clinical fellow PhD student in all aspects of IT support for this project. His support will be primarily required in developing and maintaining the online database and Web-study interfaces.

Oxford

Mr Ed Jusczack, Senior Medical Statistician and Head of Trials, National Perinatal Epidemiology Unit, University of Oxford. Mr Jusczack is an experienced triallist, and has been fully involved in the BOOST II UK trial. He will facilitate understanding of certain aspects of the BOOST II UK trial, and working aspects of the BOOST II UK trial clinical database.

Working arrangements.

The clinical fellow PhD student will work in Edinburgh, with PhD supervision by Professor Fleck, Dr Stenson, and Professor Campbell. A number of departments within Edinburgh University will fully support the clinical fellow PhD student, primarily the Departments of Population Health Sciences, and CRIC.

The clinical fellow PhD student will visit London for training in RetVas techniques by Dr Wilson, and will visit Oxford for orientation with the BOOST II UK study clinical database.

There will be regular electronic collaboration communication with members of the team outside Edinburgh, and occasional collaboration meetings.

Analysis – overall approach

The work may be broadly broken down into analysis of clinical ROP data from the BOOST trial, and analysis of the retinal image database using expert reader experiments and RetVas software.

1. Analysis of clinical ROP data from the BOOST trial. This will be based in the Department of Population Health Sciences (Prof. Campbell).

2. Analysis of the retinal image database using RetVas software. This will be based in CRIC (Dr Tom MacGillvary).

3. Development of a decision-making clinical model, using the results of the clinical data analysis, and of RetVas experiments. This will be based in the Department of Population Health Sciences (Prof. Campbell).

Chiang, M. F., R. Gelman, et al. (2007). "Plus disease in retinopathy of prematurity: an analysis of diagnostic performance." <u>Trans Am Ophthalmol Soc</u> **105**: 73-84; discussion 84-75.

> PURPOSE: To measure agreement and accuracy of plus disease diagnosis among retinopathy of prematurity (ROP) experts; and to compare expert performance to that of a computer-based analysis system, Retinal Image multiScale Analysis. METHODS: Twenty-two recognized ROP experts independently interpreted a set of 34 wide-angle retinal photographs for presence of plus disease. Diagnostic agreement was analyzed. A reference standard was defined based on majority vote of experts. Images were analyzed using individual and linear combinations of computer-based system parameters for arterioles and venules: integrated curvature (IC), diameter, and tortuosity index (TI). Sensitivity, specificity, and receiver operating characteristic areas under the curve (AUC) for plus disease diagnosis were determined for each expert and for the computer-based system. RESULTS: Mean kappa statistic for each expert compared to all others was between 0 and 0.20 (slight agreement) in 1 expert (4.5%), 0.21 and 0.40 (fair agreement) in 3 experts (13.6%), 0.41 and 0.60 (moderate agreement) in 12 experts (54.5%), and 0.61 and 0.80 (substantial agreement) in 6 experts (27.3%). For the 22 experts, sensitivity compared to the reference standard ranged from 0.308 to 1.000, specificity from 0.571 to 1.000, and AUC from 0.784 to 1.000. Among individual computer system parameters compared to the reference standard, venular IC had highest AUC (0.853). Among linear combinations of parameters, the combination of

arteriolar IC, arteriolar TI, venular IC, venular diameter, and venular TI had highest AUC (0.967). CONCLUSION: Agreement and accuracy of plus disease diagnosis among ROP experts are imperfect. A computer-based system has potential to perform with comparable or better accuracy than human experts, but further validation is required.

Darlow, B. A., J. L. Hutchinson, et al. (2005). "Variation in rates of severe retinopathy of prematurity among neonatal intensive care units in the Australian and New Zealand Neonatal Network." Br J Ophthalmol **89**(12): 1592-1596.

AIM: To analyse variations in rates of severe retinopathy of prematurity (ROP) among neonatal intensive care units (NICUs) in the Australian and New Zealand Neonatal Network (ANZNN), adjusting for sampling variability and for case mix. METHODS: 25 NICUs were included in the study of 2105 infants born at less than 29 weeks in 1998 and 1999, who survived to 36 weeks post-menstrual age and were examined for ROP. The observed NICU rates of severe ROP were adjusted for case mix using logistic regression on gestation, weight for gestational age and sex, and for sampling variability using shrinkage estimates. The corrected rate in the best 20% of NICUs was identified and NICU variations in rates were compared with those in 2000-1. RESULTS: The overall (unadjusted) rate of severe ROP in the NICUs was 9.6% (interquartile range 5.4-12.8%). After adjusting for both case mix and sampling variability there remained significant variation among the NICUs. 20% of NICUs had a rate of severe ROP </=5.9%. Variation in rates among NICUs showed a similar pattern in both time periods. If the overall network rate was reduced to 5.9%, the 20th centile of the adjusted rates, there would be 79 fewer cases in a 2 year period, in contrast with 26 fewer if rates in the two units with excess rates improved to the average. CONCLUSIONS: Considerable variation in rates of severe ROP among NICUs remained after adjustment for case mix and sampling variability. These data will facilitate investigation of potentially better practices associated with a reduced risk of severe ROP.

Dhaliwal, C., E. Wright, et al. (2009). "Wide-field digital retinal imaging versus binocular indirect ophthalmoscopy for retinopathy of prematurity screening: a two-observer prospective, randomised comparison." Br J Ophthalmol **93**(3): 355-359.

AIM: To compare the diagnostic accuracy of wide-field digital retinal imaging (WFDRI) with the current "gold standard" of binocular indirect ophthalmoscopy (BIO) for retinopathy of prematurity (ROP) screening examinations. METHODS: A consecutive series of premature infants undergoing ROP screening at Edinburgh Royal Infirmary were eligible for recruitment into this prospective, randomised, comparative study. Infants were screened using both WFDRI (Retcam II with neonatal lens) and BIO by two paediatric ophthalmologists who were randomised to the examination technique. Both examiners documented their clinical findings and management plans in a masked fashion. WFDRI eye findings were compared with those of BIO. RESULTS: A total of 81 infants were recruited, and information from 245 eye examinations was analysed. The sensitivity of WFDRI in detecting any stage of ROP, stage 3 ROP and "plus" disease was 60%, 57% and 80%, respectively, and specificity 91%, 98% and 98%, respectively. The proportional agreement between WFDRI and BIO was 0.96 for detecting stage 3 disease and 0.97 for detecting "plus" disease. There was very good agreement on management decisions (kappa 0.85). CONCLUSION: When used in a routine ROP screening setting, a randomised comparison of WFDRI and BIO, WFDRI showed relatively poor sensitivity in detecting mild forms of ROP in the retinal periphery. This resulted in difficulty in making decisions to discharge infants from the screening programme.

Sensitivity was better for more severe forms of ROP, but at present WFDRI should be regarded as an adjunct to, rather than a replacement for, BIO in routine ROP screening.

Early Treatment For Retinopathy Of Prematurity Cooperative, G. (2003). "Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial." Arch Ophthalmol **121**(12): 1684-1694.

OBJECTIVE: To determine whether earlier treatment using ablation of the avascular retina in high-risk prethreshold retinopathy of prematurity (ROP) results in improved grating visual acuity and retinal structural outcomes compared with conventional treatment. METHODS: Infants with bilateral high-risk prethreshold ROP (n = 317) had one eye randomized to early treatment with the fellow eye managed conventionally (control eye). In asymmetric cases (n = 84), the eye with high-risk prethreshold ROP was randomized to early treatment or conventional management. High risk was determined using a model based on the Multicenter Trial of Cryotherapy for Retinopathy of Prematurity natural history cohort. At a corrected age of 9 months, visual acuity was assessed by masked testers using the Teller acuity card procedure. At corrected ages of 6 and 9 months, eyes were examined for structural outcome. Outcomes for the 2 treatment groups of eyes were compared using chi2 analysis, combining data for bilateral and asymmetric cases. RESULTS: Grating acuity results showed a reduction in unfavorable visual acuity outcomes with earlier treatment, from 19.5% to 14.5% (P =.01). Unfavorable structural outcomes were reduced from 15.6% to 9.1% (P<.001) at 9 months. Further analysis supported retinal ablative therapy for eyes with type 1 ROP, defined as zone I, any stage ROP with plus disease (a degree of dilation and tortuosity of the posterior retinal blood vessels meeting or exceeding that of a standard photograph); zone I, stage 3 ROP without plus disease; or zone II, stage 2 or 3 ROP with plus disease. The analysis supported a wait-and-watch approach to type 2 ROP, defined as zone I, stage 1 or 2 ROP without plus disease or zone II, stage 3 ROP without plus disease. These eyes should be considered for treatment only if they progress to type 1 or threshold ROP. CONCLUSIONS: Early treatment of high-risk prethreshold ROP significantly reduced unfavorable outcomes to a clinically important degree. Additional analyses led to modified recommendations for the use of peripheral retinal ablation in eyes with ROP. Long-term follow-up is being conducted to learn whether the benefits noted in the first year after birth will persist into childhood.

Fielder, A. R., L. Haines, et al. (2002). "Retinopathy of prematurity in the UK II: audit of national guidelines for screening and treatment." Eye (Lond) **16**(3): 285-291.

AIMS: To ascertain how closely in 1995, neonatologists and ophthalmologists were adhering to the national guidelines for the screening of retinopathy of prematurity (ROP, 1990) and those for screening and treatment (1995). METHODS: Questionnaires about the local arrangements for the screening and treatment of retinopathy of prematurity were sent to the entire consultant membership (n = 648) of the Royal College of Ophthalmologists (RCOphth) and to the clinical directors (n= 259) of neonatal units in the UK in 1995. RESULTS: One hundered and eighty-three ophthalmologists in the UK were identified as undertaking ROP screening and/or treatment, and ROP screening took place in 207 neonatal units. Seventy-seven per cent of the ophthalmologists either complied with or exceeded recommendations for determining which babies required screening, while 7% used criteria that would have resulted in substantially fewer babies being screened. Only 17% units and 12% ophthalmologists provided written information for parents, although 66% ophthalmologists talked to the parents of babies they screened. There was a lack of clarity about responsibilities for ensuring the continuation of screening on transfer to another

hospital or on discharge to home. There was a wide range of views on the ophthalmic criteria that determined when screening examinations could cease and on the indications for treatment. CONCLUSIONS: While ROP screening is almost universally adopted in the UK, there is a need for the process to be more efficient and effective. Despite the delay in reporting this survey several issues remain extant and future guidelines should clarify and refine the criteria for screening and treatment. There is a need for improved communication with parents, and particularly for written information.

Gilbert, C. (2008). "Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control." Early Hum Dev **84**(2): 77-82.

Globally at least 50,000 children are blind from retinopathy of prematurity (ROP) which is now a significant cause of blindness in many middle income countries in Latin American and Eastern Europe. Retinopathy of prematurity is also being reported from the emerging economies of India and China. The characteristics of babies developing severe disease varies, with babies in middle and low income countries having a much wider range of birth weights and gestational ages than is currently the case in industrialized countries. Rates of disease requiring treatment also tend to be higher in middle and low income countries suggesting that babies are being exposed to risk factors which are, to a large extent, being controlled in industrialised countries. The reasons for this "third epidemic" of ROP are discussed as well as strategies for control, including the need for locally relevant, evidence based criteria which ensure that all babies at risk are examined.

Haines, L., A. R. Fielder, et al. (2005). "UK population based study of severe retinopathy of prematurity: screening, treatment, and outcome." Arch Dis Child Fetal Neonatal Ed **90**(3): F240-244.

BACKGROUND: Retinopathy of prematurity (ROP) is one of the few causes of childhood blindness in which severe vision impairment is largely preventable. Ophthalmic screening for ROP is required to identify disease that requires treatment whereby the development of potentially blinding disease can be minimised. OBJECTIVES: To make the first UK population based estimate of the incidence of babies with severe ROP (stage 3 or more); to document their clinical characteristics and management and to evaluate the appropriateness of current ROP screening guidelines in the UK. PATIENTS: Cases were recruited through a national surveillance programme with 1 year ophthalmic follow up and data from clinician completed questionnaires. RESULTS: Between 1 December 1997 and 31 March 1999, 233 preterm babies with stage 3 ROP were identified. Severity (location, extent, and presence of plus disease) was associated with degree of prematurity, most severe in the most premature babies. Fifty nine percent were treated. The UK screening protocol was followed in two thirds of cases, but in the remainder it was begun too late or was too infrequent. Three quarters of the cases were followed up at 1 year, and 13% had a severe vision deficit as a result of ROP. CONCLUSIONS: Visual deficit as a result of ROP in premature babies continues to be a severe disability in some of the survivors of neonatal intensive care. Further efforts are needed to organise treatment regionally to improve outcome and standards of practice.

Haines, L., A. R. Fielder, et al. (2002). "Retinopathy of prematurity in the UK I: the organisation of services for screening and treatment." Eye (Lond) **16**(1): 33-38.

AIMS: To ascertain how closely services for the screening and treatment of retinopathy of prematurity (ROP) were organised on a national level in 1995. METHODS: Questionnaires about the local arrangements for the screening and treatment of retinopathy of prematurity

(ROP) were sent to the entire consultant membership (n = 648) of the Royal College of Ophthalmologists (RCOphth) and to the clinical directors (n = 259) of neonatal units and other units caring for preterm babies in the UK in 1995. RESULTS: 568/648 of UK consultants (88%) and 15 non-consultant ophthalmologists and 210/259 paediatricians (81%) and 19% paediatricians in non-neonatal units responded. Thirty-one per cent responding ophthalmologists were involved in the ROP service: of these 64% screened babies, 34% screened and treated babies, while 1% ophthalmologists treated ROP but did not screen. Ninety-six per cent units caring for preterm babies had their babies screened for ROP and for almost 95% of the screening took place in the neonatal unit. About 8200 babies were screened in 1994; 277 developed stage 3, of whom 54% received treatment. Nine per cent (n = 14) and 5% (n = 8) treated babies became blind in one and both eyes respectively. A sessional commitment was identified for 9% ophthalmologists, but for less than half this was included in the contracted work programme. Sixty-five ophthalmologists treated babies with ROP, but only 10 treated more than five babies in 1994. Training needs were identified by 71 respondents. CONCLUSIONS: Several aspects of ROP screening and treatment services require improvement. Hopefully, reducing the number of identified screeners would increase skills, confidence and the ability to recognise severe disease requiring treatment, and also facilitate incorporation of this work into consultant work plans.

International Committee for the Classification of Retinopathy of, P. (2005). "The International Classification of Retinopathy of Prematurity revisited." Arch Ophthalmol **123**(7): 991-999.

The International Classification of Retinopathy of Prematurity (ICROP) was published in 2 parts, the first in 1984 and later expanded in 1987. It was a consensus statement of an international group of retinopathy of prematurity experts. The original classification has facilitated the development of large multicenter clinical treatment trials and furthered our understanding of this potentially blinding disorder. With improved imaging techniques in the nursery, we are able to offer a more quantitative approach to some of the characteristics described in the ICROP. An international group of pediatric ophthalmologists and retinal specialists has developed a consensus document that revises some aspects of ICROP. Few modifications were felt to be needed. The aspects that differ from the original classification include introduction of (1) the concept of a more virulent form of retinopathy observed in the tiniest babies (aggressive, posterior ROP), (2) a description of an intermediate level of plus disease (pre-plus) between normal posterior pole vessels and frank plus disease, and (3) a practical clinical tool for estimating the extent of zone I.

Shah, D. N., C. M. Wilson, et al. (2011). "Comparison of expert graders to computer-assisted image analysis of the retina in retinopathy of prematurity." Br J Ophthalmol **95**(10): 1442-1445.

AIMS: To determine correlation of width and tortuosity between expert graders and computer-assisted image analysis of the retina in narrow-field images of eyes with retinopathy of prematurity. METHODS: 11 digital images were selected based on severity of retinopathy of prematurity (ROP). Narrow field images were analysed for width and tortuosity of vessels using computer-aided image analysis of the retina (CAIAR), an image analysis software, and by four ROP experts. Spearman correlation coefficients (rho) assessed the correlation of CAIAR grading with expert grading. Intra-class correlations assessed agreement among graders. Width and tortuosity were compared among severity of ROP and treatment status using analysis of variance and generalised estimating equations. RESULTS: Expert measurements correlated well with measures from CAIAR for venule width (rho=0.57-0.66) and arteriole tortuosity (rho=0.71-0.81). Measurements from four graders

agreed moderately well (intra-class correlations were 0.49 and 0.69 for venule width and arteriole tortuosity, respectively). Increased severity of ROP (no pre-plus/plus, pre-plus, plus) was associated with larger width (linear trend p=0.02 in two graders) and tortuosity (linear trend p<0.03 in all graders). Tortuosity measurements by CAIAR and graders were statistically different between treated and untreated eyes (p<0.002). CONCLUSIONS: We found moderate correlation between expert graders' assessment of vessel tortuosity and width and CAIAR using narrow-field images.

Slidsborg, C., J. L. Forman, et al. (2012). "Experts do not agree when to treat retinopathy of prematurity based on plus disease." Br J Ophthalmol **96**(4): 549-553.

OBJECTIVES: To investigate inter-reader agreement on five severity levels of central vascular changes (none, mild, moderate, severe pre-plus disease, plus disease) and aggressive posterior retinopathy of prematurity (ROP), and to see whether an unintended shift in indication for treatment occurred. METHODS: Four international ROP readers participated. Before the grading of the photographs, the readers were informed that a high proportion of advanced ROP cases were included. In total, 243 photographs/948 quadrants were available from 136 infants. As a standard series of photographs was available, grading was performed under optimised conditions. RESULTS: The four readers agreed on the quadrant scores of only 70 (7.38%) of the 948 quadrants--that is, on 1, 5, 15, 4 and 45 quadrants for scores 0, 1, 2, 3 and 4, respectively. The mean scores differed systematically between the readers (permutation test, p<0.0001). Agreement on presence of aggressive posterior ROP from all four readers was not obtained for any of the photographs. Readers scored plus disease in at least two quadrants in 95.5% of the eyes for which treatment was indicated. All four readers agreed on the scoring of indication for treatment for 195 eyes (80.2%); however, treatment was only recommended in 18 (7.4%) eyes. One reader was found to differ systematically from the others in indicating treatment (Rasch analysis; p=0.0001). Finally, a significant shift in indication for treatment occurred between birth period 2000-2002 and 2003-2006 (Mann-Whitney rank sum test, p<0.001). CONCLUSIONS: Inter-reader agreement on central vascular changes is poor, especially when based on more than two rating categories. The subjective nature of diagnosing such vascular changes possibly resulted in earlier treatment of preterm infants in Denmark over the entire study period (1997-2006). The recent increased incidence of treated infants in Denmark is, at least in part, explained by a significant shift in indication for treatment.

Wallace, D. K., G. E. Quinn, et al. (2008). "Agreement among pediatric ophthalmologists in diagnosing plus and pre-plus disease in retinopathy of prematurity." J AAPOS **12**(4): 352-356.

PURPOSE: Plus disease has become the major criterion for laser treatment in infants with retinopathy of prematurity (ROP), but its assessment is subjective. Our purpose was to compare quadrant-level and eye-level assessment of plus disease and pre-plus disease among 3 experienced ROP examiners and to report their rate of agreement. METHODS: One hundred eighty-one high-quality RetCam images from premature infants were graded by 3 of the authors. Dilation and tortuosity were judged separately using a scale of normal or sufficiently abnormal to meet criteria for pre-plus or plus disease. RESULTS: There was disagreement on the presence of plus disease for 18 images (10%), on tortuosity sufficient for plus disease (plus tortuosity) for 26 images (14%), and on dilation sufficient for plus disease (plus dilation) for 26 images (14%). Of 67 images judged to have pre-plus disease or worse, there was disagreement on the presence of plus disease for 18 images (27%), on plus tortuosity for 25 images (37%), and on plus dilation for 21 images (31%). For distinguishing

plus or pre-plus disease from normal, there was disagreement on pre-plus tortuosity for 38 of 181 images (21%) and on pre-plus dilation for 58 of 181 images (32%). CONCLUSIONS: Three experienced ROP examiners disagreed frequently on the diagnosis of plus or pre-plus disease when evaluating cropped clinical photographs of infants, many of which had borderline plus disease. Further study is required to determine the implications of these observations on clinical decision making.

Wilson, C. M., K. D. Cocker, et al. (2008). "Computerized analysis of retinal vessel width and tortuosity in premature infants." Invest Ophthalmol Vis Sci **49**(8): 3577-3585.

PURPOSE: To determine, with novel software, the feasibility of measuring the tortuosity and width of retinal veins and arteries from digital retinal images of infants at risk of retinopathy of prematurity (ROP). METHODS: The Computer-Aided Image Analysis of the Retina (CAIAR) program was developed to enable semiautomatic detection of retinal vasculature and measurement of vessel tortuosity and width from digital images. CAIAR was tested for accuracy and reproducibility of tortuosity and width measurements by using computer-generated vessel-like lines of known frequency, amplitude, and width. CAIAR was then tested by using clinical digital retinal images for correlation of vessel tortuosity and width readings compared with expert ophthalmologist grading. RESULTS: When applied to 16 computer-generated sinusoidal vessels, the tortuosity measured by CAIAR correlated very well with the known values. Width measures also increased as expected. When the CAIAR readings were compared with five expert ophthalmologists' grading of 75 vessels on 10 retinal images, moderate correlation was found in 10 of the 14 tortuosity output calculations (Spearman rho = 0.618-0.673). Width was less well correlated (rho = 0.415). CONCLUSIONS: The measures of tortuosity and width in CAIAR were validated using sequential model vessel analysis. On comparison of CAIAR output with assessments made by expert ophthalmologists, CAIAR correlates moderately with tortuosity grades, but less well with width grades. CAIAR offers the opportunity to develop an automated image analysis system for detecting the vascular changes at the posterior pole, which are becoming increasingly important in diagnosing treatable ROP.