QUANTIFYING OCULAR INFLAMMATION IN UVEITIS USING OPTICAL COHERENCE TOMOGRAPHY

by

XIAOXUAN LIU

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Academic Unit of Ophthalmology
Institute of Inflammation and Ageing
College of Medical and Dental Sciences
University of Birmingham
2021

Abstract

Xiaoxuan Liu, Academic Unit of Ophthalmology, Institute of Inflammation and Ageing, College of, Medical and Dental Sciences, University of Birmingham

Inflammation is the key underlying physiological process in uveitis. It drives the onset of acute flares, causes permanent structural damage and can result in sight-threatening complications. Being able to accurately detect and measure changes in inflammatory activity is crucial for managing uveitic flares and rationalising therapeutic decisions. Unfortunately, many of the current methods for quantifying inflammation are imperfect, due to the fact that they are based on subjective and unreliable clinician estimates.

In this thesis, I evaluated the potential for imaging-based technologies such as optical coherence tomography (OCT) to measure key markers of intraocular inflammation in uveitis. Whilst several key markers of inflammation are recognised, this thesis focuses on those with an existing clinical standard, which can be used as a comparator or reference test (anterior chamber cells, anterior chamber flare and vitreous haze).

I conducted a series of systematic reviews evaluating potential instrument-based techniques for measuring anterior chamber cells, anterior chamber flare and vitreous inflammation, respectively. These identified OCT and laser flare photometry as potential instruments for measuring anterior chamber cell and flare, and OCT and retinal photography for measuring vitreous inflammation. However, the interpretation of results in each review was limited by relatively few studies and the inclusion of highly heterogenous uveitic patient populations, varying severities of disease, and lack of a standardised image acquisition protocol.

Second, in the prospective study, OCTAVE (OCT-assisted vitreous evaluation), I found that our custom OCT-based vitreous analysis technique (EQUIP) demonstrated good repeatability in healthy and uveitic eyes, was able to detect vitreous inflammation and was associated with the current clinical vitreous haze grading. The EQUIP measurement was able to predict visual acuity whereas the current standard method (clinician grading

using the National Eye Institutevitreous haze scale) could not. Whilst these results were encouraging, there remains substantial overlap in the OCT measurement between NEI vitreous haze grades. It is not clear whether this is due to poor signal-to-noise ratio of the OCT technique, or a sign of poor reliability of the comparator (clinician-based grading using the NEI vitreous haze scale). Further investigation through longitudinal studies may be able to answer this question.

In summary, OCT has demonstrated potential for quantifying inflammation for multiple key measures in uveitis. However, a key limitation for the validation of all instrument-based measures has been the lack of a reliable reference test.

Acknowledgement

I am immensely grateful to my supervisors: Prof Alastair Denniston, who could not have given me a better start in the world of scientific research, and to whom I owe so much for his wisdom, patience and generosity; Dr David Moore, for his helpful guidance and thoughtful steer to help me grow my critical thinking; and Dr Pearse Keane, for his inspiration and for providing me with many exciting opportunities over the last few years.

I also want to thank Dr Giovanno Ometto, Dr Giovanni Montesano and Prof David Crabb, for their dedicated work on the OCT vitreous tool and for teaching me so patiently and imparting their wisdom. Thank you to Dr Alice Sitch for expertly advising on the statistical analysis of this study, as well as Dr Lavinia Ferrante Di Ruffano and Prof Jon Deeks for all their helpful advice on test evaluation. Thank you to Dr Lola Solebo, Dr Tasanee Braithwaite, Prof Phil Murray, Mr Sreekanth Sreekantam, for all their wisdom and advice and support at various points during this project.

My heartfelt thanks to the ophthalmology research team: Charlotte Radovanovic, Nick Capewell, Amy Price and Claire Arthur for help with study co-ordination, as well as their good humour and friendship on the days where things weren't going well. Thank you to the ophthalmology staff who helped run this study smoothly in clinic: Miss Rupal Morjaria, Dr Hina Khan, Miss Dipti Trivedi, Mr Jesse Panthagani and Dr Mathew Kurumthottical. Also, to the exceptionally talented students and junior (and some not-so-junior) doctors who have contributed to or given advice about this work: Aditya Kale, Ben Hui, Chris Way, Livia Faes and Siegfried Wagner. Thank you for Mr Geraint Williams who gave me the push I needed to pursue this PhD three years ago. I am also immensely grateful to the Wellcome Trust for funding this study.

I want to especially thank the patients and healthy volunteers who participated in the study, not only for their generosity in time, but also for their genuine curiosity which often led to the most enjoyable and rewarding conversations.

And finally, to my wonderful parents for their unconditional love and encouragement, and to my better half, Matt, who has been so exceptionally patient and supportive over the last three years.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

This thesis explores the use of instrument-based technologies, such as optical coherence tomography (OCT), to objectively measure intraocular inflammation in uveitis. Chapter 1 is an introduction to uveitis, describing inflammatory biomarkers and how they are used in clinical practice to inform therapeutic decision-making, how inflammation is currently quantified, how it could be measured using newer imaging modalities and the challenges of evaluating a new test for uveitis. Chapters 2, 3 and 4 are systematic reviews of instrument-based tools for measuring anterior chamber cells, anterior chamber flare and vitreous haze, respectively. Each systematic review identifies the range of instrument-based technologies currently available for quantifying inflammation and evaluates the level of association between the instrument-based measures and the current clinical standard, as well as the reliability of the instrument-based measurements. Chapters 5 and 6 describe our development of an OCT-based method for measuring vitreous inflammation. Chapter 5 summarises the evidence for the method prior to, and near the beginning of, my involvement with this project. It outlines prior validation of the technique, including the earliest proof-of-concept and subsequent studies seeking to refine the technique. Chapter 6 presents the OCTAVE study, a prospective observational study led by me, evaluating the OCT-based vitreous haze technique. The imaging protocol selected for the OCTAVE study was largely driven by prior evidence described in Chapter 5. The OCTAVE study explored the test reliability of our technique in healthy and uveitic eyes, as well as its association with other measures of uveitic inflammation and visual function. Chapter 7 summarises the findings of the Chapters 2-6 and discusses the unanswered questions and future work for the OCT-based vitreous haze technique, as well as general challenges of validating a new test for measuring uveitis inflammation.

1.1 Uveitis

1.1.1 Definition

Uveitis is defined as ocular inflammation of the uveal tract, which consists of the iris, ciliary body and choroid. The underlying pathological process is an inflammatory response, causing breakdown of the blood-ocular-barrier and disruption to the physiological environment in the eye. Any part of the uvea can become inflamed and this can be accompanied by, or extend to involvement of other structures such as the retina, sclera, cornea, vitreous and optic nerve. As such, the term uveitis is used to describe a range of diseases characterised by intraocular inflammation.

As an umbrella term, uveitis encompasses multiple disease entities with a range of underlying aetiologies, some of which are caused by an infectious agent (such as toxoplasmosis, tuberculosis, cytomegalovirus related uveitis) and others through autoimmune-mediated mechanisms. Many uveitic entities are associated with systemic autoimmune conditions, such as the HLA-B27 spondyloarthropathies, sarcoidosis, tuberculosis, Behçet's disease and others, but others are limited to the eye.(Rosenbaum and Dick, 2018)

1.1.2 Burden on vision

Uveitis is the fourth commonest cause of blindness in the working age population. (Darrell, Wagener and Kurland, 1962; Bodaghi *et al.*, 2001) It is responsible for approximately 10-15% of blindness in the developed world and approximately 25% in the developing world.(Rothova *et al.*, 1996; Wakefield and Chang, 2005; Rao, 2013) The estimated prevalence of uveitis is approximately 38 to 200 per 100,000 general population, with an annual incidence estimated to be 17.4 to 52.4 per 100,000.(Rothova *et al.*, 1996; Durrani *et al.*, 2004) Although any age group

may be affected, it tends to peak in the working-age group, accounting for a significant socio-economic impact.

Acute uveitic flare can cause a range of symptoms, the most concerning of which is loss of vision, which may be temporary for the duration of the flare or irreversible. Irreversible loss of vision may result from direct damage to the uveal structures, but more commonly is due to secondary tissue damage. Three of the most common sight-threatening complications arising secondary to uveitis are cataracts, glaucoma and macular oedema.(Jones, 2015) Other complications such as neovascularisation, scarring, atrophy and posterior synechiae less commonly cause visual loss, but can if they impact the central visual axis.

1.1.3 Pathophysiology of uveitis

The intraocular inflammatory response in uveitis occurs as a result of an insult, causing disruption of the ocular immune homeostasis.(Lee *et al.*, 2014) Much of our understanding of the mechanisms behind intraocular inflammation has come from animal models of experimental autoimmune uveoretinitis.(Caspi, 2011) Clinically, uveitis is categorised as infectious when in the presence of a clear pathogen, and non-infectious when no pathogenic agent is identified. In non-infectious cases, the underlying mechanism is thought to be driven primarily by a dysregulated immune response, and increasingly this is thought to be autoinflammatory (i.e. driven by the innate immune response) rather than autoimmune (i.e. driven by the adaptive immune system in response to self-antigen). This inflammation continues even after any postulated infectious trigger subsides, causing ongoing chronic or remitting episodes of inflammatory activity.(Forrester, Kuffova and Dick, 2018)

In all cases, the result is breakdown of the blood retinal barrier (BRB), or the blood aqueous barrier in the case of anterior uveitis.(Freddo, 2013) In healthy individuals, the BRB, composed

of tight junctions of the retinal vessels and the retinal pigmented epithelium (RPE), acts as a physical and immunological barrier to the eye. In uveitis, the integrity of the BRB is disrupted, leading to infiltration of infectious agents and/or damaging inflammatory cells.

Whilst it is not possible to directly observe the immune-mediated insult itself, or the invasion of a pathogen at the level of the ocular tissue, the downstream consequences of BRB breakdown can be observed. Typically there is leakage of inflammatory cellular infiltrates into various compartments of the eye which is visible on slit-lamp examination.(Muller, 2013; Forrester, Kuffova and Dick, 2018) Increased permeability of the iris and ciliary body vessels cause infiltration of leukocytes which can be visualised as floating particles in the aqueous humour, described as anterior chamber (AC) cells. Smaller protein exudates, such as albumin, leaking into the aqueous humour through gives rise to a hazy appearance of (a typically transparent) aqueous, described as AC flare;(Agrawal *et al.*, 2010) The aqueous appearance in AC flare is attributed to the Tyndall effect: the scattering of light in multiple directions giving rise to a turbidity which affects the clarity of the media. A similar finding to AC flare in the vitreous gel occurs as a result of exudates from the retinal vessels, described as vitreous haze.(Foster C and Vitale Albert, 2013)

As there are no readily available serological markers in uveitis, such as acute phase proteins or autoantibodies, the diagnosis and monitoring of inflammation is primarily informed by changes in these clinical biomarkers. These are typically observed by clinicians during biomicroscopic slit-lamp examination, and contribute to the overall clinical picture on presence and severity of inflammatory activity. In experimental settings, aqueous samples taken from the anterior chamber show increased levels of inflammatory infiltrates including T cells, B cells and macrophages.(Denniston *et al.*, 2011, 2012)

Depending on the cause of the inflammatory insult and the location of the infiltrates or anatomical changes, several classification systems have been proposed to categorise distinct aetiological and anatomical phenotypes. The different classification systems for uveitis are described in the next section.

1.1.4 Classification of uveitis

There are several approaches to classifying uveitis and the current approaches to classification were agreed upon in the 2005 Standardisation of Uveitis Nomenclature (SUN) workshop.(Jabs *et al.*, 2005) In practice, the most commonly used systems are anatomical and aetiological classification. Anatomical classification is based on the location of the foci of inflammation and is characterised as anterior, intermediate, posterior and panuveitis **(Table 1)**.(Jabs *et al.*, 2005)

Table 1. The SUN* Working Group Anatomic Classification of Uveitis

Type	Primary Site of Inflammation[†]	Includes
Anterior uveitis	Anterior chamber	Iritis Iridocyclitis
Intermediate uveitis	Vitreous	Pars planitis Posterior cyclitis Hyalitis
Posterior uveitis	Retina or choroid	Focal, multifocal or diffuse choroiditis Chorioreitinis Retinochoroiditis Retinitis Neuroretinitis
Panuveitis	Anterior chamber, vitreous and retina or choroid	

*SUN = Standardization of Uveitis Nomenclature; Standardization of Uveitis Nomenclature for Reporting Clinical Data. Results of the First International Workshop. Jabs DA, Nussenblatt RB, Rosenbaum JT. 2005.
[†]As determined clinically. Adapted from the International Uveitis Study Group anatomic classification (Deschenes et al., 2008)

Aetiological classification is broadly split into infectious (bacterial, viral, fungal and parasitic) and non-infectious causes (which may or may not be associated with systemic disease) and masquerade subtypes **(Table 2)**.(Jabs *et al.*, 2005) Other ways of describing uveitis entities relate to their clinical course and duration (acute, chronic, recurrent) and their pathological findings (granulomatous versus non-granulomatous; where glaucomatous uveitis are associated with the clinical features of large 'mutton fat' keratic precipitates (macrophages) and iris nodules).(Jabs *et al.*, 2005)

Diagnosis is based on a combination of: (1) clinical presentation (symptoms and signs); (2) medical history (known diagnoses and systemic review to identify potential undiagnosed autoimmune diseases) and (3) investigations (imaging and laboratory tests). However, defining aetiology in uveitis can be problematic. Tissue sampling is uncommon due to the risk of damage to the eye and most uveitic entities are not associated with clinical or investigative features that are of high specificity.

Table 2. International Uveitis Study Group (IUSG) Classification of Uveitis

(Deschenes *et al.*, 2008)*Deschenes* et al*., 2008)*

International Uveitis Study Group Classification of Uveitis	
Infectious	Bacterial Viral Fungal Parasitis Others
Noninfectious	Known systemic associations No known systemic associations
Masquerade	Neoplastic Nonneoplastic

For a large group of non-infectious uveitides there is still limited understanding as to what the underlying pathogenesis is – both for those with and without systemic association (presumed 'autoimmune uveitis'). Often diagnoses cannot be made on the first visit and earlier reviews of medical records have shown that only 17% of patients receive a definitive diagnosis on the first visit.(Rodriguez *et al.*, 1996)

1.1.5 Typical course of disease

The typical course of disease in uveitis is characterised by single or recurrent episodes of flares of inflammatory activity, which can cause acute manifestations in different parts of the eye. However, the course of uveitis can be widely variable. Some patients experience single episodes of uveitis with no recurrence, some experience recurrent episodes with periods of quiescence in between, and some experience chronic inflammation.

Inflammation, whether from recurrent acute episodes or chronic disease, can result in clinically significant consequences for the patient. These consequences of uveitis can be broadly thought of as disease activity (ongoing active inflammation) and disease damage (damage resulting from previous inflammation). Activity represents an ongoing inflammatory insult which can either self-resolve or may require therapeutic intervention to bring under control. Damage occurs as a result of disease activity and is permanent. Both disease activity and damage can exhibit symptoms and signs, including deteriorations in visual function, however the approach to treatment is radically different; to treat disease activity requires targeting the inflammatory process (almost always medically), whereas to treat disease damage requires targeting structural changes (which may include surgery). The ability to differentiate between disease activity and damage is therefore key for clinical decision-making and rationalising treatment. In this section, we model the relationship between activity and damage, as well as their

manifestations, using a number of theoretical thresholds. This model is described below using an illustrative example.

1.1.5.1 Threshold for symptoms and signs

Depending on the level of inflammatory activity, active flares can manifest with various symptoms (the range of typical symptoms in uveitis is described in **1.1.5.1 Symptoms**). The threshold at which inflammatory activity becomes symptomatic is variable and poorly defined. It is assumed that very low levels of inflammatory activity may fall below this threshold and therefore do not manifest as symptoms.

Figure 1 is a schematic representation of a time course of recurrent flares with varying severity. Whilst some episodes cross the threshold for symptom manifestation (black dotted line), others fall below it. Schematic time course representing recurrent flares of uveitic inflammatory activity of varying severity and return to periods of quiescence between episodes. Inflammatory episodes above a certain threshold manifest as symptoms whereas those below the threshold do not. For example, the first and fourth flares are below the threshold and would not manifest as symptoms.

In this time course, there are two episodes falling below the symptoms threshold. Whilst these episodes may have manifested with clinically detectable signs, they were unnoticed by the patient. It is therefore important to monitor patients, to ensure signs of asymptomatic disease, which can be both asymptomatic inflammatory activity or asymptomatic damage, are detected. The same concept can be applied to clinical signs. Only when inflammatory activity reaches beyond a certain threshold will clinical signs manifest and become detectable.

Figure 1. Schematic representation of the typical time course in recurrent uveitis flares, showing inflammatory activity and threshold for symptoms.
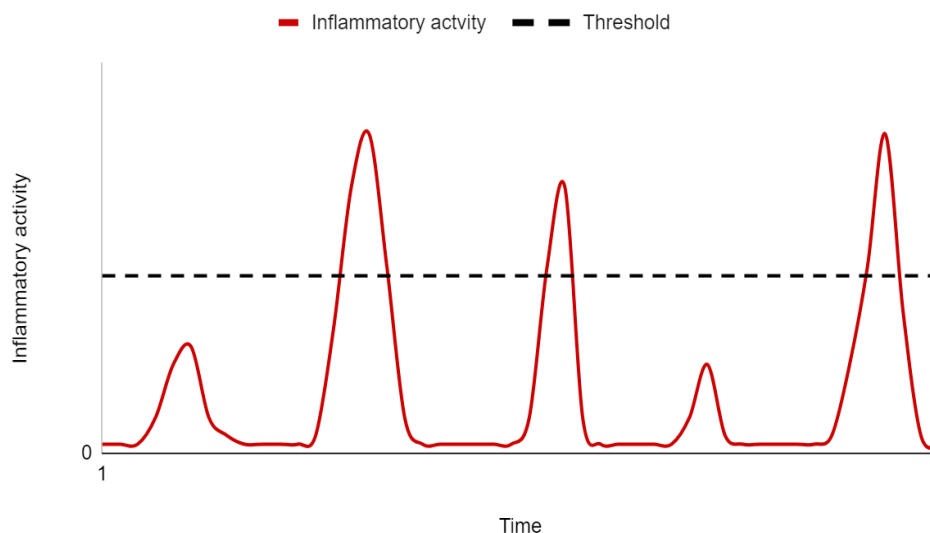
1.1.5.2 Threshold for damage

As previously mentioned, inflammatory flares may resolve spontaneously or with treatment, or it may leave behind permanent damage. For example, acute flares of posterior segment-involving uveitis may present with acute deteriorations in vision from inflammatory macular thickening. After the acute inflammatory activity has resolved, vision may be restored, or it may leave behind disruption of the retinal layers, scarring, atrophy or epiretinal membrane, which can result in permanent loss of visual function.(Karim *et al.*, 2013; Fardeau *et al.*, 2016) The threshold at which inflammatory activity causes permanent structural damage, as with the threshold for manifestation of symptoms, is also poorly defined and varies depending on the structures involved and the underlying cause of the flare. For example, central subfoveal thickness (as measured by OCT) in uveitic macular oedema and baseline anterior chamber flare (as measured using laser flare photometry) have been shown to be predictive of visual

loss.(Holland, 2007; Matas *et al.*, 2019) However, the additive effects of multiple inflammatory episodes over time is not well understood. **Figure 2** illustrates the accumulated damage (shaded red area) resulting from the same inflammatory flare events in **Figure 1**. This time, the threshold represents the minimum level of inflammation to cause permanent damage. Each episode above this threshold results in additional accumulated damage. The fourth episode falls below the threshold for symptoms, as shown in **Figure 1** and it also falls below the threshold for damage (therefore no increase in the red shaded area is seen). Under this assumption, it could be concluded that this episode is clinically insignificant. This is as opposed to the first episode, which falls below the symptom threshold in **Figure 1**, but reaches the threshold to cause damage in **Figure 2** and is therefore clinically significant.

Figure 2. Schematic representation of the typical time course in recurrent uveitis flares, showing inflammatory activity, threshold for irreversible damage and accumulated damage over time.
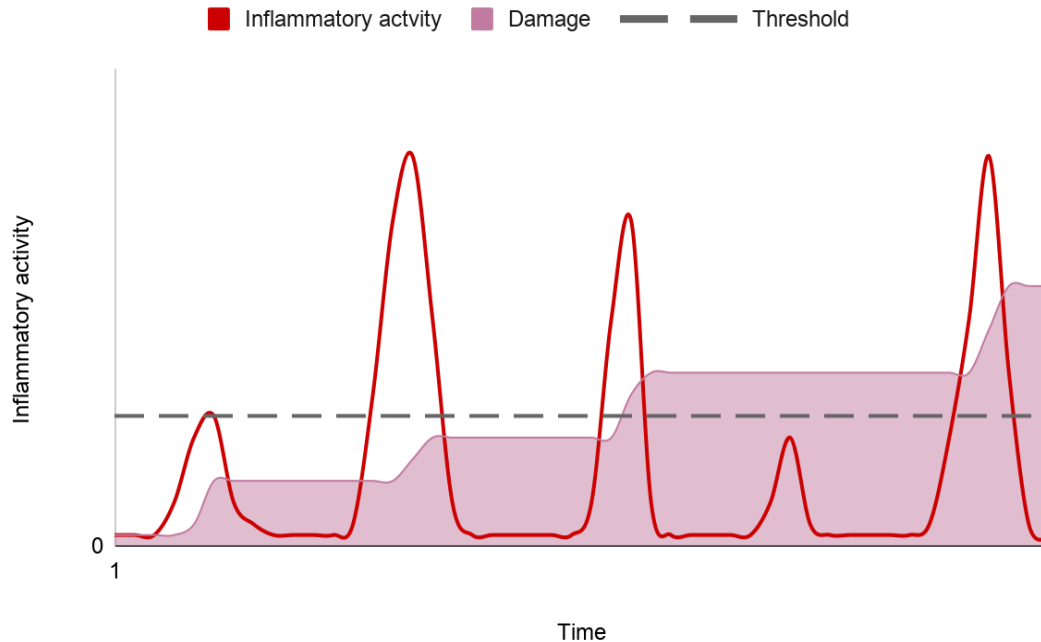
1.1.5.3 Relation to clinical decision-making

In order to apply these concepts in real-world clinical practice, two prerequisites are needed. First, there needs to be sufficiently reliable and sensitive markers of inflammatory activity to monitor patients over time. These markers need to be detectable and measurable with a monitoring test, so that the time course of individual patients as illustrated above can be tracked. Second, there needs to be knowledge of where the thresholds for symptoms and damage lie. Importantly, the size of the gap between the two thresholds needs to be understood (i.e. at what level disease activity can be detected and at what level damage occurs), as this is crucial for understanding how large or small the window of opportunity for prevention of irreversible damage is. These concepts will be revisited in **1.5.1 Biomarkers and surrogate endpoints.**

1.1.6 Manifestations of disease

 Among the many manifestations exhibited by uveitis, the nature and severity of them vary depending on many contributing factors, including the type and aetiology of the disease and the structures involved.

Denniston *et al* have previously described the outcome measures of uveitis according to the dimensions of disease: disease activity, disease damage, observed visual function, patient reported visual function and quality of life.(Denniston, Keane and Srivastava, 2017) This model is helpful as the sequential nature of each dimension of disease reflects the typical time course of disease, as well as the relationship between disease pathology to visual function to patient experience**. Figure 3** illustrates an adapted version of this model, which illustrates the relationship between each dimension and additionally accounts for the transient effects of

disease activity on downstream function and patient experience during active flares (dotted line).

Figure 3. Dimensions of disease. Solid black line denotes irreversible effects and dotted line denotes reversible effects caused by active disease.
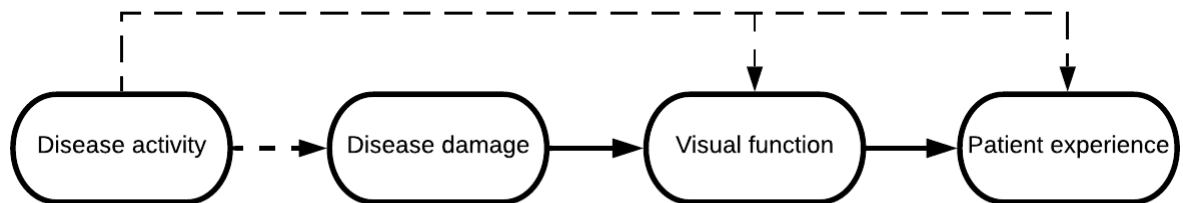
1.1.6.1 Symptoms

The symptoms experienced by patients with uveitis can be highly variable, but commonalities exist within uveitis subtypes. Uveitis of the anterior segment typically manifests with pain, photophobia and hyperaemia.(Foster C and Vitale Albert, 2013) Blurred vision is not a common symptom in acute presentations of anterior diseases; when it does occur it is usually a manifestation of inflammatory changes in the optical pathway through the anterior segment e.g. corneal oedema, cell deposition on the inner corneal surface (keratic precipitates), high levels of inflammatory cells or protein in the aqueous, or deposition of inflammatory cells on the anterior surface of the lens.(Gueudry and Muraine, 2018) Uveitis of the posterior segment usually results in a painless decrease or alteration in visual function, which can be generalised or focal. Based on symptoms alone, it is difficult to differentiate between inflammatory activity in the vitreous, retina and choroid. Intermediate uveitis presents most commonly with floaters and variable degrees of generalised blurred vision (either directly from vitritis or associated macular oedema); it often presents bilaterally.(Bonfioli *et al.*, 2005) Posterior uveitis can manifest with symptoms including generalised blurred vision (e.g. from inflammatory macular oedema or

diffuse inflammation of the choroid or retina), focal loss of vision (blind spots/scotomata arising from areas of focal inflammation), or distorted vision (arising from focal inflammation and its consequences); they may also report floaters and photopsia.(Tallouzi *et al.*, 2020)

As described in the previous section, uveitic inflammation can also occur without any symptoms. This silent asymptomatic inflammatory activity, which may only be detectable through clinical examination, can cause permanent structural damage with associated loss of vision. Certain uveitic diseases are more likely to be asymptomatic, for example juvenile idiopathic arthritis (JIA) associated uveitis is characterised by chronic inflammatory signs in the absence of symptoms.(Kotaniemi *et al.*, 2005) It is therefore important to routinely monitor even asymptomatic disease, to ensure asymptomatic inflammatory activity or any damage which has occurred, can be detected through clinical examination or monitoring tests.

1.1.6.2 Clinical signs

Inflammation affecting the anterior segment typically presents with leakage of exudates into the aqueous humour causing a light scattering effect, described as AC flare, and the egress of inflammatory cells into the anterior chamber (AC cells). Granulomatous anterior uveitis may present with mutton fat keratic precipitates and iris nodules. There may also be posterior synechiae, fibrin and spillover of cells into the anterior vitreous. Certain signs such as posterior synechiae, fibrin and cells stuck to the corneal endothelium (keratic precipitates) remain even after active inflammation has resolved.

Signs for posterior segment-involving uveitis tend to be more variable depending on the underlying aetiology. Intermediate uveitis is characterised by vitreous exudates causing floaters, vitreous haze and vitreous 'snow-banking' and 'snowballs' and may be accompanied by retinal signs such as retinal thickening and cystoid macular oedema (CMO). Posterior uveitis can

present with a wide range of signs including focal inflammatory lesions or 'spots' in the retina or choroid, vasculitic changes (such as non-perfusion or vessel sheathing) and CMO. There may be evidence of chronic or previous inflammation such as scarring (disruption or loss of the normal retinal structure in focal areas of previous inflammation) and generalised retinal atrophy.

1.1.7 Treatment

The approach in managing uveitis varies depending on the aetiology, the anatomical foci of inflammation, disease severity and other considerations such as previous response to treatment and management of complications arising from uveitis. For isolated anterior segment uveitis, topical corticosteroids is often sufficient. For disease affecting the posterior segment (intermediate, posterior or panuveitis), or where topical treatment has proved insufficient for anterior segment disease, then local or systemic therapy may be required. Local therapy is particularly suitable where there is unilateral or asymmetric disease, either on its own or as an adjunct to systemic therapy. Local therapy is primarily with long-acting corticosteroid implants, notably dexamethasone (Ozurdex; duration of action up to 6 months) or fluocinolone acetonide (Iluvien; duration of action up to 3 years).(Lowder *et al.*, 2011; Jaffe *et al.*, 2019; Thorne *et al.*, 2019) Systemic therapy, where needed, is staged in which corticosteroid is used in the initial phase to get rapid control of inflammation, but tapered with the aim of reducing to a safe maintenance dose (or to zero) once control is achieved. If there is recurrence whilst tapering, or there is other indication for long-term immunosuppression, a second-line agent is introduced. Typically these are disease modifying anti-rheumatic drugs (DMARDs) such as methotrexate, mycophenolate mofetil, azathioprine.(Kempen, Gewaily and Cw, 2016) Biologics are usually used third line after failure of at least one DMARD, with adalimumab now being licensed for the treatment of uveitis since the VISUAL trials reported. (Jaffe *et al.*, 2016; Nguyen *et al.*, 2016; Suhler *et al.*, 2018)

All immunosuppressive therapies used in uveitis carry the potential for adverse drug effects. Topical and local corticosteroids accelerate the onset of cataract formation and can increase the intraocular pressure, making the patient at risk of glaucoma.(Friedman *et al.*, 2013) Systemic corticosteroids can cause numerous side effects including weight gain, mood disturbance, sleep disturbance and bone mineral density loss.(Saag *et al.*, 1994) Many non-biologic DMARDs require regular blood tests to monitor for hepatic, renal or biochemical dysfunction and Adalimumab is associated with a risk of demyelination disorder (Gangaputra *et al.*, 2009; Pasadhika *et al.*, 2009; Daniel *et al.*, 2010; Ledingham *et al.*, 2017; Suhler *et al.*, 2020) Rationalising treatment is therefore always a balance between likelihood of reducing inflammatory activity and the risks of unwanted side effects. The ideal level of immunosuppression, which may or may not be possible, is just enough to suppress the inflammatory response, but not enough to cause complications of an over-suppressed immune system and other associated complications of the drug. In order to inform these decisions, there has to be a good measure of the level of inflammatory activity, so that the decision to commence treatment can be justified and treatment response can be monitored.

1.2 Current measures of inflammation in uveitis

1.2.1 How inflammatory activity is measured in uveitis

The assessment of uveitic inflammatory activity is complex, but crucially important for clinical decision-making. Assessment findings inform key treatment decisions which aim to prevent irreversible loss of vision and reduce the burden of disease for patients. Its complexity is in part because uveitis is a heterogeneous group of diseases with significant variations between different anatomical and aetiological subtypes, but also because many existing methods of measurement are based on surrogate markers which indirectly represent inflammatory activity.

For example, anterior chamber (AC) cells, the hallmark of anterior uveitis, increases as a consequence of blood-ocular-barrier breakdown and the leakage of cells into the aqueous. However, AC cells are not direct measures of the inflammatory activity itself. Similarly, CMO results from the escape of fluid into the intraretinal or subretinal spaces, following breakdown in the blood-retinal-barrier (BRB). Whilst a higher degree of BRB breakdown generally results in higher volumes of CMO (and therefore increased central retinal thickness), the CMO itself is not a direct measure of inflammatory activity. Because of this, there can be a lag between the onset of inflammation to the manifestation of inflammatory signs and vice versa, between the return of the eye to a quiescent state and the resolution of these signs.

Furthermore, the way in which these measures are assessed is inherently problematic. Many of the existing surrogate measures are detected through clinical examination by ophthalmologists using slit lamp biomicroscopy. This is naturally subjective and highly dependent on the skills and experience of the observer.(Kempen *et al.*, 2008) In an effort to standardise these techniques, a number of grading systems have been proposed for quantifying different measures of inflammation.(Jabs *et al.*, 2005) The limitations of the current approaches are discussed in detail in **1.2.2.1 Commonly used biomarkers for uveitic inflammation**.

Visual function, although often the most important outcome for patients, is recognised as a poor marker of inflammatory activity.(Denniston and Dick, 2013; Enoch *et al.*, 2019) This is because change in visual function is a downstream consequence of inflammatory activity, rather than a direct measure of it. The consequences of uveitic inflammation on visual function may be immediate, or may take months to years to develop. Furthermore, it is inherently subjective to the patient and can be confounded by many non-inflammatory variables. Use of visual function-specific tools such as NEI-VFQ-25 has been shown to be more useful for prediction disease activity than general quality of life metrics, however could only detect disease improvement and

not disease worsening.(Sugar *et al.*, 2011) Whilst visual function should always factor into clinical decision-making, it is not the most suitable outcome for determining the level of active inflammation.

1.2.2 Biomarkers for uveitic inflammation

1.2.2.1 Commonly used biomarkers for uveitic inflammation

As described above, a variety of common manifestations can present in uveitis. Some are considered more direct measures of inflammatory activity than others. At the 2005 SUN workshop, a number of important clinical features for grading inflammatory activity were recommended, including AC cells, AC flare, vitreous cells, vitreous haze, perivascular sheathing and vascular leakage.(Jabs *et al.*, 2005) Grading systems were standardised for AC cells, AC flare and vitreous haze at the SUN workshop. Before this, four systems existed for grading AC cells, three existed for AC flare, two existed for vitreous cells and three for vitreous haze. Whilst a standardised assessment method could be agreed upon for AC cells, AC flare and vitreous haze, no suitable grading system could be agreed upon for vitreous cells, perivascular sheathing or vascular leakage. Other manifestations commonly found in uveitis such as macular oedema, epiretinal membrane and subretinal neovascularisation were also discussed at the SUN workshop and were considered structural complications of inflammation rather than signs of activity. The workshop was a significant step forward in harmonising standardised grading schema for each measure. Additionally, to aid in the assessment of each domain, photographic aids were recommended by the SUN working group to assist grading and definitions of "improving" and "worsening" inflammation were agreed upon **(Table 3)**.

Table 3. The SUN* Working Group recommendation on Activity of Uveitis terminology

Term	Definition
Inactive	Grade 0 cells[†]
Worsening activity	Two step increase in level of inflammation (e.g. anterior chamber cells, vitreous haze) or increase from grade 3+ to 4+
Improved activity	Two step decrease in level of inflammation (e.g. anterior chamber cells, vitreous haze) or decrease to grade 0
Remission	Inactive disease for ≥3 months after discontinuing all treatments for eye disease

[†] Applies to anterior chamber inflammation.

* SUN = Standardization of Uveitis Nomenclature; Standardization of Uveitis Nomenclature for Reporting Clinical Data. Results of the First International Workshop. (Jabs DA, Nussenblatt RB, Rosenbaum JT. 2005).

In a clinical trial context, a wide range of outcomes measures have previously been adopted. Denniston *et al* reviewed the range of primary outcome measures used by 104 therapeutic clinical trials for posterior segment-involving uveitis up to October 2013.(Denniston *et al.*, 2015) This review found that 74% of trials (n = 94) included one or more measures of disease activity (listed in **Table 4**), with vitreous haze and macular oedema being the most commonly used measures. More recent trials have adopted composite outcome measures consisting of multiple measures of visual function and inflammatory activity.(Adalimumab vs. Conventional Immunosuppression for Uveitis Trial – ClinicalTrials.gov (NCT03828019); Jaffe et al., 2016; Nguyen et al., 2016)

Table 4. Primary outcome measures of activity used in registered trials for posterior segment involving uveitis (Denniston et al., 2015)

Primary outcomes in studies with single efficacy outcomes	Primary outcomes in studies with composite efficacy outcomes
Vitreous haze Macular oedema Treatment requirement	AC cells Vitreous haze Vitreous cells Snowballs Macular oedema Chorioretinal inflammatory lesions Retinovascular inflammation Treatment requirement

At a joint National Eye Institute (NEI)/ Food and Drugs Administration (FDA) Uveitis Endpoints Workshop held in 2015, Holland and associates further proposed a short-list of manifestations that should be regarded as the '*fundamental signs of inflammation*': AC cells, vitreous haze, retinal vascular leakage, retinal infiltrates and choroidal infiltrates.(Holland, 2015) These features were recommended as preferred endpoints with which to measure disease outcome in uveitis clinical trials. The following section describes each of these five key measures and the current recommended method for quantifying them, where available.

1.2.2.1.1 Anterior chamber cells

Anterior chamber cells can be observed due to the leakage of inflammatory cells into the aqueous humour as a result of blood-retinal barrier disruption. The current approach to assessing AC cells is using the SUN grading system for AC cells: a clinician's estimate of the number of cells seen on slit-lamp biomicroscopy within a 1 by 1 mm slit beam shone through the anterior chamber.(Jabs *et al.*, 2005) The clinician is required to focus the slit beam in the anterior chamber so that cells can be counted. As AC cells move, the clinician is required to mentally adjust for the natural current of the aqueous humour, which causes cells to drift in and out of the 1 by 1 mm slit beam area. The cell count is then placed into one of six grades **(Table 5).**

Table 5. The SUN* Working Group Grading Scheme for Anterior Chamber Cells

Grade	Cells in Field†
0	<1
0.5+	1–5
1+	6–15
2+	16–25
3+	26–50
4+	>50

† Field size is a 1 mm by 1 mm slit beam.
* SUN = Standardization of Uveitis Nomenclature; Standardization of Uveitis Nomenclature for Reporting Clinical Data. Results of the First International Workshop. Jabs DA, Nussenblatt RB, Rosenbaum JT. 2005.

1.2.2.1.2 Anterior chamber flare

AC flare describes the hazy appearance of the aqueous fluid between the cornea and the iris, caused by leakage of inflammatory material, such as proteins. The appearance is caused by an optical phenomenon called the Tyndall effect: the scattering of light by particles in a fine suspension, whereby the hazy appearance of the aqueous causes a decrease in clarity of the structures behind it. AC flare does not always present in conjunction with AC cells and is more prevalent in certain uveitic subgroups, such as juvenile idiopathic arthritis (JIA) related uveitis.(Holland, 2007) The accepted approach to grading AC flare is clinical examination by slit lamp biomicroscopy and classification of severity using the SUN grading system for AC flare **(Table 6)**. Each grade is estimated based on the clarity of iris detail behind the aqueous fluid and ranges from "none", to "faint", "moderate (iris and lens detail clear)", "marked (iris and lens details hazy)" and "intense (fibrin or plastic aqueous)". Prior to the SUN workshop, a number of alternative systems existed that quantified flare in a similar way.(Hogan, Kimura and Thygeson, 1959; Schlaegel, 1967; Foster and Vitale, 2002; Nussenblatt and Whitcup, 2004)

Table 6. The SUN* grading scheme for anterior chamber flare.

Grade	Flare	Description
0	None	
1+	Faint	Barely detectable
2+	Moderate	Iris and lens details clear
3+	Marked	Iris and lens details hazy
4+	Intense	Fixed coagulated aqueous with considerable fibrin

[*] SUN = Standardization of Uveitis Nomenclature; Standardization of Uveitis Nomenclature for Reporting Clinical Data. Results of the First International Workshop. Jabs DA, Nussenblatt RB, Rosenbaum JT. 2005.

1.2.2.1.3 Vitreous haze

Until recent years, the preferred clinical trial endpoint by regulatory authorities was the six-step National Eye Institute (NEI) Vitreous Haze (VH) scale, which was originally introduced in 1985 and subsequently adopted by the SUN workshop. Prior to the SUN workshop, three grading systems existed.(Hogan, Kimura and Thygeson, 1959; Nussenblatt *et al.*, 1985; Foster and Vitale, 2002) The NEI VH scale is a six-point grading system for estimating the vitreous clarity as seen through indirect ophthalmoscopy and is also referred to as the National Institute for Health (NIH) or Nussenblatt scale.(Nussenblatt *et al.*, 1985; Jabs *et al.*, 2005) The clinician's estimate is compared to a standardised set of photographs and given a score of 0, +0.5, +1, +2, +3 or +4 **(Table 7)**. In the presence of media opacity such as cataracts, it is suggested that the cataractous changes can be *"looked around fairly easily with an indirect ophthalmoscope"*, however whether this is practically possible, especially in diffuse global vitreous haze, is debatable.(Nussenblatt *et al.*, 1985)

Table 7. National Eye Institute/Nussenblatt/SUN* Scale for Grading of Vitreous Haze

Grade	Description
0	No evidence of vitreal haze
Trace/ 0.5+	Slight blurring of optic disc margin
1+	Obscured view but definition to optic nerve head and retinal vessels
2+	Obscured view but definition to retinal vessels
3+	Optic nerve head visualised but borders are very blurry
4+	Obscured fundal view

1.2.2.1.4 Choroidal and retinal inflammatory lesions

Certain uveitis subtypes are characterised by the occurrence of choroidal and retinal inflammatory lesions. Active inflammation may give rise to new chorioretinal lesions and/or cause enlargement or expansion of existing chorioretinal lesions. These lesions are traditionally detected during fundus examination on slit lamp biomicroscopy, with subtle signs such as colour and regularity suggestive of whether the lesion is actively inflamed. Unlike AC cells, AC flare and vitreous haze, there are no existing clinical systems for quantifying chorioretinal lesions and their use in clinical trials has been mainly reduced to presence/absence of new inflammatory lesions.(Denniston *et al.*, 2015; Kim *et al.*, 2015; Jaffe *et al.*, 2016; Nguyen *et al.*, 2016)

1.2.2.1.5 Retinal vascular leakage

Inflammatory breakdown of the BRB results in leakage of fluid and exudates from the retinal vessels, clinically termed retinal vasculitis.(Walton and Ashmore, 2003; El-Asrar *et al.*, 2010) Retinal vascular leakage is a physiological sign caused by breakdown of the BRB. As a marker of inflammation, it is somewhat different to the others described above. First, 'leakage' is a physiological process which describes the movement of exudates or fluid from the retinal

intravascular compartment to the surrounding structures. Although it is possible to observe

related signs using slit lamp examination (such as vessel sheathing or vessel filling defects),

dye-based tests such as fluorescein angiography (FA) are regarded as the gold standard for

observing actual vessel leakage in real time. FA demonstrates the integrity of the retinal

vasculature and allows visualisation of abnormalities such as dye leakage, staining or pooling.

The assessment of vascular leakage on FA is largely a qualitative assessment and is not

quantified routine practice.(Walton and Ashmore, 2003; Nussenblatt and Whitcup, 2004; El-

Asrar *et al.*, 2010)

1.2.2.2 Limitations of current methods

The SUN workshop was a significant step forward in harmonising standardised grading schema

for commonly recognised measures of inflammation. Despite this acceptance of a standardised

assessment nomenclature, there are still significant disadvantages with the current approach.

Assessment using the SUN grading systems are subjective, non-continuous and poorly

sensitive at lower levels of inflammation.(Kempen *et al.*, 2008; Hornbeak *et al.*, 2014) There is a

significant concern that these measures, which are the current standard in routine clinical care

and as clinical trial endpoints, are not fit-for-purpose and as a result can impact upon patient

care and hamper the development of potential new drug therapies. There is therefore an urgent

need for improved measures of inflammatory activity. Indeed, at the NEI/FDA Uveitis Endpoints

Workshop held in 2015, Director of the FDA's Division of Ophthalmic, Neurological and Ear,

Nose and Throat Devices, Dr Eydelman, said "*There is a clear need for objective assessments

of ocular inflammation, which are repeatable, reliable, and quantifiable*."

1.3 Imaging-based biomarkers of uveitis

Since the SUN working group meeting of 2005, widespread adoption of ophthalmic imaging, including high resolution modalities such as optical coherence tomography (OCT), has become routine in most ophthalmic clinics. New and improved imaging techniques provide new ways of quantifying retinal disease in a precise and objective way.(Pichi *et al.*, 2020) The best current example of this is the use of OCT to measure and monitor central macular thickness (CMT) to detect, quantify and track over time the development and resolution of macular oedema.(Browning *et al.*, 2004; Kempen *et al.*, 2013; Schmidt-Erfurth and Waldstein, 2016) Commercially available OCT devices are now able to utilise automated software to align sequential scans, to allow measurement of changes in the same anatomical area between visits. OCT as an imaging modality is discussed in more detail later **(1.4 Optical coherence tomography)**.

1.3.1 Advantages of imaging-based markers

Quantitative ophthalmic imaging has many characteristics which make it an ideal solution for measuring biomarkers. First, automated interpretation of images is more objective than clinician examination, as it eliminates variability between observers attributable to clinical examination skills.(Kempen *et al.*, 2008) Imaging-based approaches also overcome certain slit lamp-based issues met by the current clinical grading systems, such as the need to manually account for moving AC cells, as the image captures only one single snapshot in time. The second advantage of quantitative imaging is its potential for automation, which not only has the potential to bring further objectivity, but also improve time and cost efficiency by removing the need for a human grader. Third, most forms of ophthalmic imaging are non-invasive and light-based (photography, OCT and ultrasound) and therefore are free of radiation. This means repeated

exposures for the patient are safe and imaging is suitable as a monitoring tool. Fourth, most basic forms of ophthalmic imaging are fast and can be acquired in a matter of seconds. This means the image acquisition is not burdensome on patients and it would be feasible to fit them within clinical workflows. Lastly, clinical examination and application of existing grading systems typically requires an experienced ophthalmologist, whereas most ophthalmic imaging can be acquired by technicians. This opens the possibility of monitoring patients in settings outside the hospital, such as in high street optometrists and satellite clinics, as well as the possibility of remote monitoring and telemedicine.

In recognition of the potential utility of quantitative imaging, the follow-up meeting of the 2015 NEI/FDA Uveitis Endpoints Workshop took place in March 2019. This UCLA/AUS Workshop on 'Objective Measures of Intraocular Inflammation for Use in Clinical Trials' was strongly focused towards instrument and machine-based methods.(American Uveitis Society/UCLA Stein Eye Institute, no date) The aim of the workshop was to explore which instrument-based techniques may have the potential to replace current methods for the following key measures of inflammation: AC inflammation (including AC cells and AC flare), vitreous haze, retinal lesions, choroidal lesions and retinal vascular leakage.

1.3.2 Instrument-based markers for uveitic inflammation

Given that biomarkers for uveitic inflammation are currently detectable through slit-lamp examination, and the relative resolution of newer imaging devices far exceeds human-level optical resolution, it follows that such devices may be able to capture such biomarkers in a more objective way.(Leite *et al.*, 2011) The ability of OCT to detect and correlate with, histopathological proven inflammatory infiltrates (including vitreous cells, retinal vessel engorgement and peri-vascular exudates) has been previously shown in animal models.(Chen

et al., 2013; Chu *et al.*, 2016). Longitudinal imaging studies in animal models have been able to show the phenotypical sequalae detectable in the retina, as well as restoration of the retinal architecture after resolution of inflammatory activity.(Chen *et al.*, 2013) Imaging methods like OCT also offer the advantage of cross-sectional imaging, allowing visualisation along the axial length of the eye, and overcoming the problem of posterior ocular structures being obscured by anterior structures, as often is the case during slit-lamp biomicroscopy.

There is also known variation in the constituents of uveitic exudates depending on underlying aetiology. Granulomatous inflammation of the anterior chamber is known to cause mutton-fat keratic precipitates composed of epitheloid cells on the corneal endothelium, whereas non-granulomatous inflammation tends to result in smaller lymphocytic cells in the aqueous.(Harthan *et al.*, 2016) Whilst it is not possible to differentiate individual free-floating AC cell types on slit-lamp examination, differences in anterior segment OCT cell morphology (such as cell size and reflectance) has been shown to correlate with different cell types, including neutrophils, lymphocytes, monocytes and red blood cells.(Rose-Nussbaumer *et al.*, 2015; Qian *et al.*, 2019)

In the next section, each of the five measures of inflammation are discussed again, this time to highlight potential instruments and imaging techniques which have been proposed, or have shown potential in recent years, to provide objective measures of inflammation. These instruments and imaging techniques are briefly introduced here but are discussed in more detail throughout this thesis.

1.3.2.1 Anterior chamber cells

In recent years, it has become possible to visualise AC cells using anterior segment optical coherence tomography (AS-OCT).(Kumar *et al.*, 2015; Sharma *et al.*, 2015) This adaptation of OCT imaging uses an anterior segment lens to visualise the anterior segment of the eye and

allows acquisition of multiple cross-sectional B scans between the cornea and lens at different levels to capture a theoretical volume. AS-OCT has been shown to capture inflammatory cells as hyper-reflective spots in the aqueous humour and image analysis techniques have demonstrated the possibility of automating cell counts. The output is an absolute cell count (within the theoretical area captured within the OCT volume) or cell density.

1.3.2.2 Anterior chamber flare

As noted earlier, anterior chamber flare is the appearance given by a light scattering optical phenomenon called the Tyndall effect. Laser flare photometry, a device which measures this effect, has been available since the 1980s and has been validated for quantifying AC flare in uveitis. The device production has been led almost exclusively by Kowa (Kowa Company Ltd, Nagoya, Japan) and has gone through several generations of updating. Its output is a measurement in photons and studies have been conducted to correlate the photon count to clinician grading.(Tugal-Tutkun and Herbort, 2010; Konstantopoulou *et al.*, 2015) The only alternative technology proposed for AC flare measurement is AS-OCT. Invernizzi *et al* have demonstrated the ability to utilise pixel brightness of the aqueous on AS-OCT, as a surrogate for AC flare.(Invernizzi *et al.*, 2017)

1.3.2.3 Vitreous haze

Keane *et al* have previously demonstrated the repurposing of macular OCT scans to quantify vitreous haze.(Keane *et al.*, 2014) The technique, which has since been automated, uses signal intensity in the vitreous compartment captured on OCT as a surrogate for vitreous haze.(Keane *et al.*, 2015) This technique has been shown to generalise across different populations and OCT devices, and also has preliminary supporting evidence that it may detect treatment response. The output measurement is a continuous variable on an arbitrary scale which represents the signal intensity.

1.3.2.4 Chorioretinal lesions

Chorioretinal lesions can be captured sequentially over time on fundus cameras and ultra wide-field fundus photographs, such as using the Optos device (Optos Plc, Dunfermline, Scotland, UK).(Campbell *et al.*, 2012; Leder *et al.*, 2013) These devices allow a series of imaging records to be kept, allowing subjective comparisons of old and new lesions between visits. The ability to record this in an image is already an improvement upon clinician examination alone. However, as with the other biomarkers described above, the ideal scenario is if the lesions can be segmented and quantified. Madhusudhan *et al* have noted that the functions already in place for measuring retinal thickness could be employed to detect focal structural changes near inflammatory lesions.(Madhusudhan, Keane and Denniston, 2016) Attempts to automatically demarcate and measure 'area affected' have been attempted using multi-modal imaging (infrared and blue autofluoresence (AF) imaging using an OCT device), in combination with automated segmentation and summation of total area affected.(Ometto *et al.*, 2020) This provides the ability to follow the progression of individual lesions over time **(Figure 4)**.

Figure 4. Automated analysis of chorioretinal lesions using multimodal imaging and calculation of total affected area and expansion rate over time.(Ometto *et al.*, 2020)
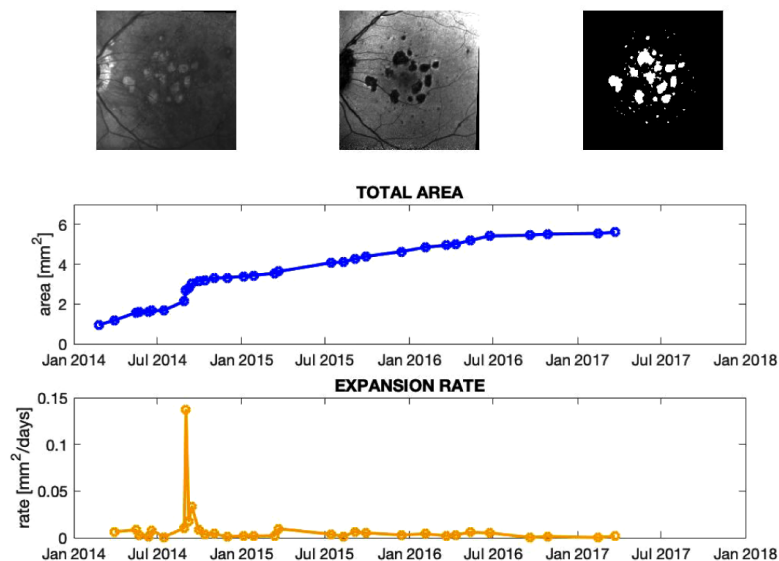
1.3.2.5 Retinal vascular leakage

Retinal vascular leakage, as described previously, is only truly demonstrable using dye-based tests such as fluorescein angiography (FA) and therefore is already reliant on imaging-based methods in routine practice. Retinal vascular inflammation is often found in the peripheral vessels and therefore recent advances in ultra-widefield fundus imaging have greatly improved FA image capture, by extending the field of view to up to 200-degrees (Optos, Optos PLC, Dunfermline, Scotland, UK) compared to the conventional 60-degree fundus cameras.(Mackenzie *et al.*, 2007) However, the interpretation of leakage on FA remains descriptive and only qualitatively assessed in routine clinical practice. The 'amount' of leakage seen on FA is often not quantified at all, or at best summarised in terms of location (central or peripheral) and a vague estimate of area involved (e.g. less or more than 50% of the retinal vasculature).(Pecen *et al.*, 2017) Both manual and automated systems for automated quantification have been proposed to provide various metrics (including total area of leakage, areas of peripheral versus central/macular leakage and areas of non-perfusion).(Karampelas *et al.*, 2015) Although the clinical relevance of each proposed metric is yet unclear, they at least provide a comparable measurement over time. Automated systems have been developed for detecting the retinal vasculature and vessel leakage of fluorescein dye on fundus imaging in other retinal diseases such as diabetic retinopathy.(Ehlers *et al.*, 2017; Jiang *et al.*, 2020)

1.3.2.6 Macular oedema

Macular oedema describes accumulation of intra-retinal fluid at the central macular region and is a common manifestation of posterior segment-involving uveitis. It was determined not to be a key measure of inflammation at the NEI/FDA workshop, as it was considered a complication of the inflammation rather than a sign of activity.(Holland, 2015) It is mentioned here for two reasons. First, macular oedema is often used in clinical practice to inform treatment decisions,

as its consequences are visually significant and therefore it is an important outcome for patients.(Sugar *et al.*, 2011) Second, of all the measurable dimensions of disease, OCT quantification of macular oedema is perhaps the most reliable and precise measurement available today.

Before OCT was available, macular oedema was diagnosed by the pathological elevated appearance of the macula, through slit lamp biomicroscopy. This method was only able to detect macular oedema in relatively severe cases and estimations of macular thickness was impossible to judge by eye, even by highly skilled clinicians.(Browning *et al.*, 2004) Fluorescein angiography also played a part in the assessment of macular oedema, characteristically appearing as diffuse leakage of fluorescein dye around the macular area and/or potentially pooling of dye in cystoid spaces.(Kempen *et al.*, 2013)

In the last decade, OCT has become widely accepted as the gold standard for imaging macular oedema, and its use is well established in retinal diseases beyond uveitis (such as diabetic retinopathy and wet age-related macular degeneration).(Browning *et al.*, 2004; Schmidt-Erfurth and Waldstein, 2016) It has enabled cross-sectional imaging of retinal volumes to examine structural changes. Features of the retina can be assessed and recorded quantitatively (thickness of the retina and individual retinal layers) and qualitatively (pattern of fluid accumulation, presence of cystoid spaces and changes in the architecture of retinal layers). Moreover, the addition of spatial tracking functions such as 'follow-up' of sequential scans allow precise longitudinal comparison of changes in measurements taken at the same anatomical point.(Giani *et al.*, 2012) The ability to track fluctuations in retinal thickness over time has therefore become highly reliable and precise. It is therefore no surprise that most recent clinical trials have adopted central retinal thickness as an outcome measure and most ophthalmic units

now carry out OCT monitoring scans as routine practice.(*Macular Edema Ranibizumab v. Intravitreal Anti-inflammatory Therapy Trial - ClinicalTrials.gov (NCT02623426)*, 2020)

1.4 Optical coherence tomography

1.4.1 Imaging modality

OCT is a non-invasive imaging method which was first introduced in the 1990s.(Huang *et al.*, 1991) It is based on the principles of measuring back-scattered light from different tissues (such as the retinal tissues within the eye) to construct cross-sectional depth signal profiles (A scans). Multiple A scans can be combined in a linear formation to form 2 dimensional images (B scans) and multiple B scans can be acquired along yet another axis to produce 3 dimensional volumes. Earlier generations of OCT used time-domain technology, which is based on the principles that tissue at different location introduces longitudinal time delays to the light signal.(Costa *et al.*, 2006) By comparing two beams of light (one reflected from the ocular tissue and another reference beam), the time-of-flight delay information can be used to indicate location of the reflective tissue. Significant advances have been achieved in the last decade with the introduction of spectral domain OCT (SD OCT), which is based on fourier domain detection and has significant improvements in speed, resolution and reliability.(Choma *et al.*, 2003; Leitgeb, Hitzenberger and Fercher, 2003) Current models of OCT achieve high axial resolution up to approximately 5 microns, almost the same size as a human erythrocyte (6.2-8.2 microns).(Mary and Turgeon, 2005; van Velthoven *et al.*, 2007; Leite *et al.*, 2011) The ability to view cross-sectional and 3D reconstructions of the retina, at a near-cellular level, is akin to *in vivo* histological analysis **(Figure 5)**.

Figure 5. OCT and histology image of the optic nerve from the reproduced from the first publication describing OCT (Huang et al 1991).

BV = blood vessel, RPE = retinal pigment epithelium, SRF = subretinal fluid.
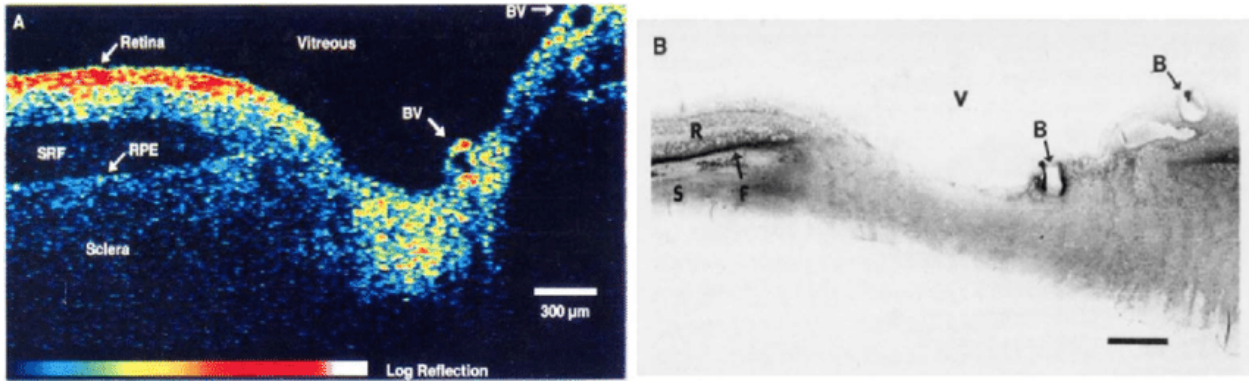
1.4.2 Common applications of OCT in routine practice

Most ophthalmic clinics now utilise OCT as a routine diagnostic tool, with the acquisition of images usually being carried out by trained imaging technicians. Given the speed and ease with which OCT images can be acquired, it is relatively convenient to incorporate OCT as part of the clinic flow.

The two most common applications for the use of OCT in ophthalmology are to image the central retina (macular scans) and the retinal nerve fibre layer (RNFL) surrounding the optic nerve. Macular scans provide detailed architectural information and repeatable measurements which can be used to detect pathological changes over time in a wide range of retinal diseases, including many of the leading causes of blindness including age related macular degeneration and diabetic macular oedema. OCT measurements such as retinal thickness have been demonstrated to be highly reliable in both healthy eyes(Tangelder *et al.*, 2008; Kakinoki *et al.*,

2009; Leung *et al.*, 2009; Liu, A. Kale, *et al.*, 2019) and eyes with pathology(Forooghian *et al.*, 2008; Garcia-Martin *et al.*, 2011; Krebs *et al.*, 2011).

For glaucoma, OCT provides RNFL thickness measurements for glaucoma monitoring. Automated segmentation of retinal layers can produce thickness measurements with each scan. This can be used to plot measurements at different time points for monitoring progression over time **(Figure 6).** Trends can be derived from historical data to assist determining the rate of deterioration.

Segmentation errors can occur and tend to be more common in relation to RNFL measurements at the peripapillary region (Mansberger *et al.*, 2017) and in the presence of pathology(Patel *et al.*, 2009; Aojula *et al.*, 2018), poor quality scans and movement artefact.(Sadda *et al.*, 2006) With widefield OCT, the area of capture can be increased to capture more peripheral retina (for example the Heidelberg SPECTRALIS device can be equipped with two different lenses offering 30 and 55 degrees of view). Another workaround for increasing the area captured is to create composite images of multiple OCTs images stitched together.(Mori *et al.*, 2012; Reznicek *et al.*, 2014; Uji and Yoshimura, 2015) Depth of imaging can be increased by shifting the focus posterior to the retina (i.e. 'enhanced depth imaging' (EDI)) to include better visualisation of the choroid.(Wong, Koizumi and Lai, 2011; Mrejen and Spaide, 2013) Indeed, the EDI technique has been used in the context of uveitis to measure the subfoveal choroidal thickness and characterise the vasculature.(Agrawal *et al.*, 2016; Kim *et al.*, 2016)

Anterior segment OCT (AS-OCT) became available more recently and can be used to image the anterior structures including the cornea, angle, iris and anterior ciliary body and lens. Unlike retinal and optic nerve head OCT, AS-OCT is less commonly used in current practice.

Figure 6. Example of Retinal Nerve Fibre Layer thickness from repeated visits and projected rate of decline (slope of RNFL thickness).
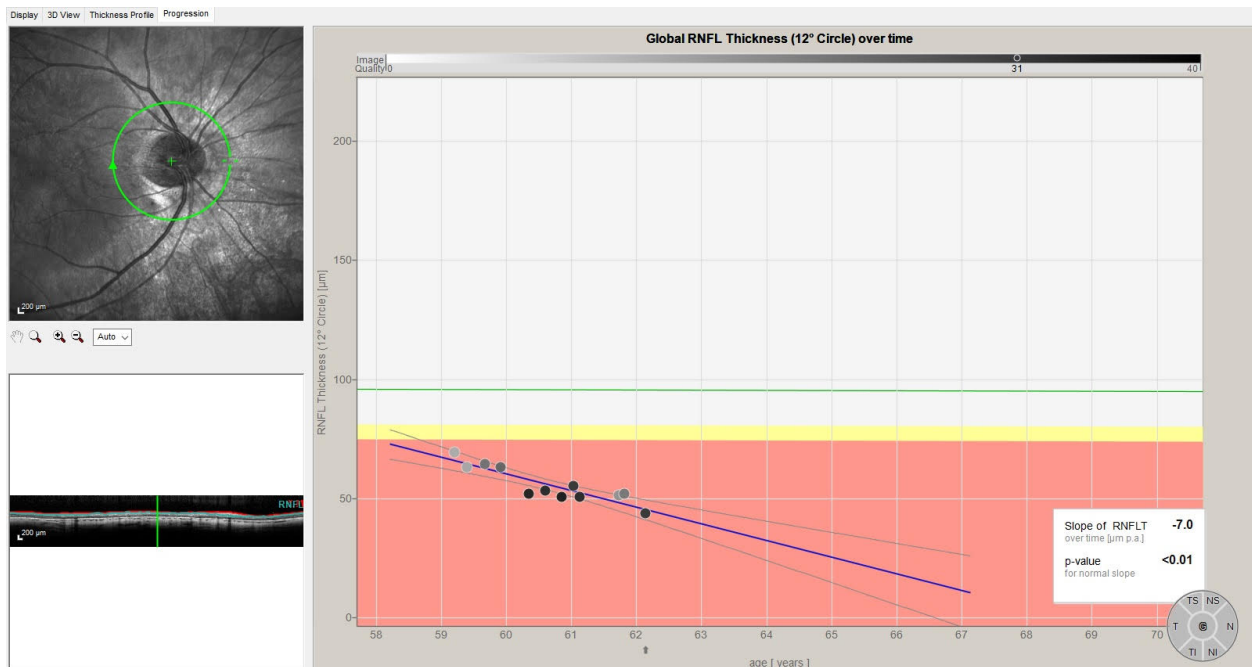
1.4.3 Practicality of OCT

There are many reasons why OCT has quickly found its place in routine ophthalmic practice. It satisfies many of the desired practicality attributes of a monitoring test: it is non-invasive, cheap to maintain and simple to perform.(Glasziou, Irwig and Aronson, 2008) Unlike many other forms of medical imaging (such as X-Rays, computed tomography and magnetic resonance imaging) OCT is fast and non-invasive, whilst still providing ultra high resolution visualisation of structures which cannot be seen through clinical examination.

Operation of most OCT devices is versatile and user-friendly, such that technicians can be easily trained to acquire routinely used scan protocols. OCT is now the most commonly used imaging platform in hospital eye services and in the last few years has been introduced into

high street optometrists and primary health care settings.(McCormick, 2020) Most OCT scans are low on burden for the patient, as the acquisition is fast and the only requirement for the patient is to be able to position adequately on the chin rest and maintain reasonable fixation. Patients with severe visual impairment may have difficulties with fixation, however a skilled operator can adapt the acquisition procedure to help the patient use their best seeing eye with the help of an external fixator. Acquisition in patients who are wheelchair bound or have severely limited cervical spinal flexibility may be difficult, but there are possible alternative setups for acquiring images in this group of patients.(Liu, A. U. Kale, *et al.*, 2019)

1.5 Evaluation of a new test

1.5.1 Biomarkers and surrogate endpoints

Biomarkers, clinical outcomes and surrogate endpoints are all different measures of a disease and its interactions with interventions. Their definitions are distinct and related to each other. The current definitions provided below are from the 2015 FDA-NIH Biomarker working group's BEST (Biomarkers, Endpoints, and other Tools) Resource.(FDA-NIH Biomarker Working Group, 2016)

- A **clinical outcome** is defined as '*a characteristic or variable that reflects how a patient feels, functions, or survives*'.
- A **biomarker** is defined as *'a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions.'*
- A **surrogate endpoint** is defined as *'An endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives.'*

Biomarkers are used for a variety of purposes in clinical contexts. For example, they can be used to predict individuals at risk of disease, to screen individuals for early stages of disease, to diagnose disease in symptomatic individuals, to stage severity of disease, to stratify likelihood of response to therapy, to determine treatment efficacy and to monitor disease stability and progression in the long term.(FDA-NIH Biomarker Working Group, 2016) Whether independently, or in combination with other markers, it can be used to diagnose, monitor and stage/quantify severity of disease and it can be used to determine treatment efficacy.

In uveitis, the clinical outcomes of interest are symptoms (such as pain and changes in visual function) and patient reported outcome measures such as quality of life.(Braithwaite *et al.*, 2019) The ideal biomarker is one that directly measures inflammatory activity, as this is the underlying pathology which drives all downstream consequences and it is also the physiological target for immunosuppressant therapies. The commonly used measures of inflammation (those described in **Section 1.2.2.1**) have not historically been considered *direct* measures of inflammation, but rather surrogate measures. They have also generally not been individually used as clinical trial endpoints for posterior segment-involving uveitis, as regulators have preferred treatment benefits to be demonstrated through clinical outcomes (reflecting how a patient feels, functions or survives) and therefore there has been a tendency for visual acuity to be the preferred outcome measure.(Denniston, Keane and Srivastava, 2017) Recent trials have adopted 'disease free status' or disease 'inactivity' as primary outcome measures, which incorporates eye examination findings and imaging based markers in the assessment.(Jaffe *et al.*, 2016; Nguyen *et al.*, 2016; *ADVISE*, 2020) This approach recognises that no one biomarker is adequate for all individuals with uveitis.

1.5.2 Choosing a suitable test

Once a suitable biomarker is identified, there needs to be a suitable test with which to detect it. To ensure a test is suitable, Mant describes in *Evidence-based Medical Monitoring: From Principles to Practice (*by Glasziou, Irwig and Aronson) four key criteria which should be met: validity, high signal-to-noise ratio, responsiveness and practicality.(Glasziou, Irwig and Aronson, 2008)

- **Validity**: the test should measure or predict a biomarker which indicates a clinically significant event. Earlier, the sequential relationship between disease activity, damage, visual function and patient experience has been described (Section 1.1.5). We have argued so far that an ideal test for uveitic inflammation should target the earliest part of this pathway to reflect the pathophysiological changes (rather than its consequences), but in order for it to be 'clinically significant', the test results should also also be related to downstream outcomes such as damage, visual function and patient experience.

- **High signal-to-noise ratio**: the test should be able to distinguish true changes in disease status compared to measurement variability. Measurement variability is the random variation observed around the true value. Two types of measurement variability are relevant to tests: technical variability (measurement errors) and biological variability.

  a. Technical variability describes the random changes in the variable caused by the measurement technique. For example, measurement variability from OCT could arise from various operator-dependent factors (such as inadequate focus) or machine-dependent factors (such as sensor-level variations).

  b. Biological variability describes real changes in the measured variable which fluctuate on a short-term basis, but which does not reflect a true change in disease state. For example, the diurnal variations in intraocular pressure are

considered normal biological variability rather than a true change in physiological or disease state.(David *et al.*, 1992)

Signal-to-noise ratio is often described in terms of reliability, which refers to the extent to which an instrument consistently measures its target. A test with good reliability should demonstrate consistency in measurements over time (test-retest reliability), but also within observers (within-observer reliability), between observers (inter-observer reliability) and between operators.

- **Responsiveness:** the test should be sensitive to changes in the clinical status. An ideal test for uveitis should be sensitive enough to detect clinically significant fluctuations in inflammatory activity, so that flares can be detected early. Similarly, if treatment is initiated, the test should be able to detect a response in repeated measurements. The biomarkers targeted by this test should therefore be rapidly responsive to the condition getting better and worse. If the test response time is less than the time taken for development of disease consequences (i.e. time to detect disease activity is shorter than time to disease damage, or visual function and patient experience), then a window of opportunity is available for early prevention of subsequent disease sequelae.

- **Practicality**: the test should be safe, easy to conduct and acceptable to the patient. Both factors are important in a test which is to be repeatedly performed. It should be feasible for repeated measurements and therefore ideally the test should also be cheap to perform. In the previous section, OCT was provided as an example of a highly practical test. This is as opposed to tests like fluorescein angiography, which have traditionally been used for providing highly useful diagnostic information in the uveitis clinic, but are not safe or practical enough to carry out repeatedly.

1.5.3 Test evaluation pathway

The test evaluation pathway can broadly be thought of as three phases: technical validity, clinical validity and clinical effectiveness. The evaluation is typically carried out in this order in an iterative fashion. **Figure 7** is an expanded version of this model which includes possible investigations at each stage, to provide supporting evidence of validity for imaging-based tests in uveitis (grey boxes).

Figure 7. Pathway to validation of a new test
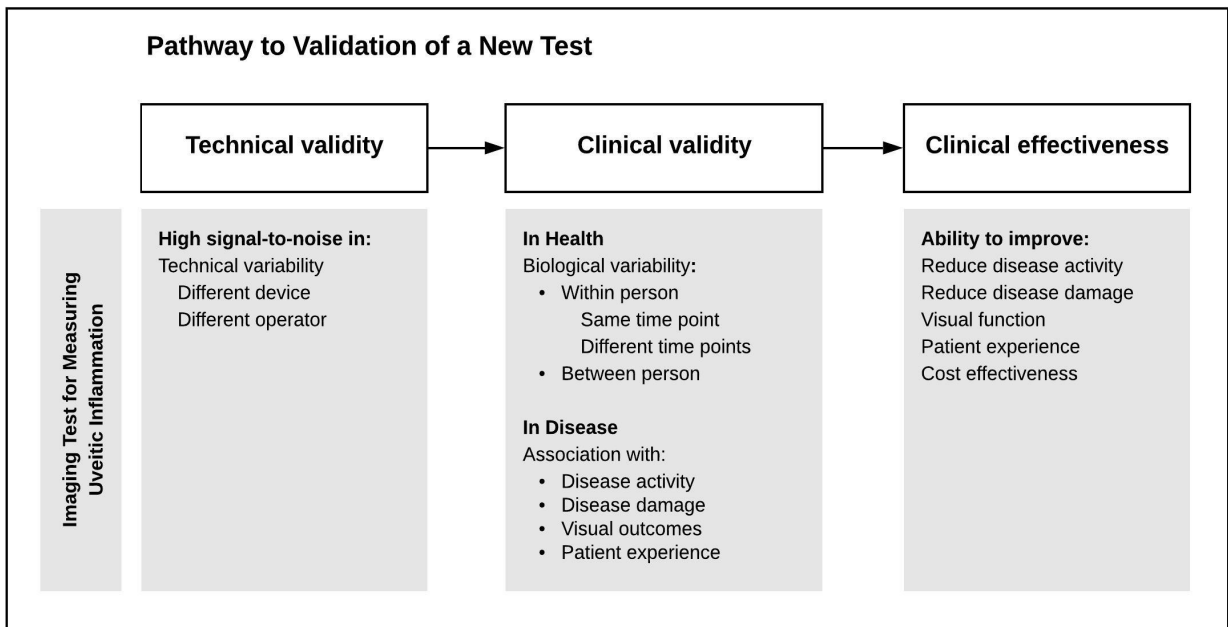
It is worth noting that in non-imaging tests for other use cases, additional parameters may need to be assessed (for example, blood tests may require additional technical investigations such as variability introduced by site of sampling, sample storage and handling, etc).

1.5.3.1 Technical validity

The test evaluation pathway typically begins with a technical evaluation phase, which begins as a proof-of-concept to investigate whether the test has internal validity. In this phase, the outcomes of interest include test repeatability, reproducibility and intra/inter-observer variability. Estimating measurement variability, or noise, requires investigation of repeated measures in test subjects where the disease state is presumed to be stable. Technical variability should be assessed using repeated measures taken on one occasion (same visit, same day) and can be measured at different levels (for example, if there is operator-dependency in performing the test, then technical variability within the same operator should be established first, before assessing variability introduced by different operators). For example, the SUN grading systems for AC cells, AC flare and vitreous haze have been shown to exhibit low to moderate interobserver agreement in uveitis eyes.(Kempen *et al.*, 2008; Hornbeak *et al.*, 2014)

Investigation of biological variability usually requires a longer time interval than repeated measures on the same occasion. This interval should be one where reasonable biological fluctuation could be expected, but where the true disease state is unlikely to have changed. Fluctuations in retinal thickness variability at different times of the day have been described in healthy eyes and in pathology(Frank *et al.*, 2004; Gupta *et al.*, 2009), suggesting that retinal thickness might decrease later in the day. It has also been suggested that posture and the gravitational effects of being upright, compared to supine, is associated with a decrease in retinal thickness.(Polito *et al.*, 2007; Almeida *et al.*, 2015) However, this is less of a concern as most conventional OCT devices acquire images in the same position (patient sitting upright).

1.5.3.2 Clinical validity

If the test demonstrates sufficient technical performance, it can progress to the clinical evaluation phase, which investigates whether the test can differentiate between healthy and

disease states and whether the test measurements are related to disease severity. In this phase, the outcome of interest is usually test accuracy for diagnosing or predicting disease and association with other meaningful clinical outcomes. For example, for a new test measuring vitreous haze, the outcome of interest may be diagnostic accuracy (ability to differentiate presence or absence of vitreous inflammation) or correlation with a meaningful clinical variable (such as clinician vitreous haze grading) or outcome (such as visual acuity). In this clinical evaluation phase, the test may need to be assessed in the context of a patient pathway, as its performance may be affected depending on its use. For example, diagnostic test accuracy is affected by pre-test probability and population disease prevalence, which varies at different points along a patient pathway.(Bossuyt *et al.*, 2006; Mallett *et al.*, 2012)

1.5.3.3 Clinical effectiveness

Ultimately, if a test can be shown to have good technical and clinical validity, the final stage of evaluation is to demonstrate whether the test can add clinical value by demonstrating a beneficial effect on patient outcomes. As tests usually form one part of a care pathway, their efficacy is closely tied to what happens before it (e.g. prior tests) and after it (e.g. the therapeutic decision following test results, the therapy that is subsequently implemented and the patient's adherence to it).(Bossuyt *et al.*, 2006) Therefore the test and its subsequent treatment strategy may need to be considered in tandem as an overall test-treatment intervention. Ferrante di Ruffano *et al* have proposed a framework for assessing the value of diagnostic tests, which not only considers treatment accuracy, but also other important aspects which may affect its clinical value (including acceptability, procedural harms and placebo effects, interpretability, timing and speed of diagnosis, user confidence and therapeutic yield).(Ferrante di Ruffano *et al.*, 2012)

1.6 Aims and Objectives of the Thesis

Being able to monitor uveitic inflammatory activity in a sensitive, objective and accurate way is essential for rationalising therapeutic decisions as part of the routine clinical care and for demonstrating drug efficacy in clinical trials. Current standard measures of assessment, based on clinical examination by a clinician, are inadequate. In recent years, with the widespread adoption of high resolution, non-invasive imaging, proposals for instrument-based technologies have emerged, which may be able to replace the current methods. These techniques offer objectivity and potentially higher sensitivity, therefore warrant further exploration.

The broad aims of this thesis are in two parts:
1. To identify potential technologies for measurement of key inflammatory variables which are currently quantified using SUN grading systems and assess their technical and clinical validity.
    a. This will be done through three systematic reviews for instrument-based tools measuring each of:
        i.   AC cells
        ii. AC flare
        iii. Vitreous haze.
    b. Each review will seek to evaluate the reliability of the proposed instrument and the strength of correlation between instrument measurements and clinical measures (clinician grading systems)
2. To prospectively evaluate the technical reliability and clinical validity of a novel OCT-based system for analysis of vitreous inflammation:

a. This is evaluated through the prospective OCT-Assisted Vitreous Evaluation (OCTAVE) study.

b. The first part of this study will evaluate the test-retest variability of the OCT-based system in healthy eyes.

c. The second part of the study evaluates the test-retest variability of the OCT-based system in uveitic eyes, as well as the association between the OCT-based measure with other markers of inflammation (including clinical vitreous haze grading, vision and macular thickness).

Chapter 2: Instrument-based tests for measuring anterior chamber cells in uveitis

This chapter explores potential instrument-based technologies for measuring AC cells. It consists of a protocol and systematic review which aimed to identify instruments used to measure AC cells in uveitis. This is the first systematic review in a series of three identifying potential technologies for measuring the key inflammatory variables, currently quantified using the SUN grading system. The systematic review evaluates the level of correlation between the instrument's measurements with clinician assessment (through the use of grading systems such as the SUN grading system) as well as the instrument's reliability.

The protocol for the systematic review was registered on PROSPERO (CRD42017084156) and is published at: *Liu, X, et al. (2019). Instrument-based tests for measuring anterior chamber cells in uveitis: a systematic review protocol. Syst Rev 8, 30.*

The report for the systematic review is published at: *Liu X, et al. (2019). Instrument-based tests for measuring anterior chamber cells in uveitis: a systematic review protocol. Syst Rev. 2019;8(1):30.*

PROTOCOL   **Open Access**

Instrument-based tests for measuring anterior chamber cells in uveitis: a systematic review protocol

Xiaoxuan Liu[1,2,4] , Ameenat L. Solebo[3], Pearse A. Keane[4], David J. Moore[5] and Alastair K. Denniston[1,2,4,6]*

Abstract

Background: Uveitis describes a group of inflammatory conditions affecting the eye. The ability to monitor inflammatory changes in anterior uveitis is crucial in clinical practice for making treatment decisions and in clinical trials for testing therapeutic agents. The current standard for quantifying anterior segment inflammation is clinical slit-lamp examination findings classified using the Standardisation of Uveitis Nomenclature (SUN) grading system. Such clinical grading systems rely on a subjective estimation using the slit lamp and are often non-linear and non-continuous scales, with large increases in cell count between each grade. Novel instrument-based technologies have emerged over the last few decades, which can provide objective and quantifiable measurements. This review will evaluate the reliability of such technologies and their level of agreement with anterior chamber (AC) cell count using clinical slit-lamp examination.

Methods: Standard systematic review methodology will be used to identify, select and extract data from studies that report the use of any instrument-based technology in the assessment of AC cells. Searches will be conducted through bibliographic databases (MEDLINE, EMBASE and Cochrane Library), clinical trial registries and the grey literature. No restrictions will be placed on language or year of publication. The outcomes of interest are the correlation of index test measurements of AC cells with clinical grading systems using slit-lamp examination and the reliability of each index test identified. Quality assessment will be undertaken using QUADAS2. Degree of correlation between the index and reference test measures will be pooled and meta-analysed if appropriate.

Discussion: A number of instrument-based tools are available for measuring AC cells. This review will evaluate the technologies available and measure the level of correlation of these alternative methods with clinical grading systems as well as their performance in reliability and repeatability. The findings of this review will identify those objective, instrument-based technologies which show good utility for measuring AC cells in a quantifiable way and which warrant further exploration for their sensitivity and reliability over the current standard.

Systematic review registration: PROSPERO CRD42017084156 (Liu X, Moore DJ, Denniston AK). Instrument-based tests for measuring anterior chamber (AC) cells in uveitis: a systematic review. 2017). Study screening stage is complete. Data extraction stage has not yet commenced.

Keywords: Systematic review, Uveitis, Anterior chamber cells, Aqueous humour, Monitoring test, Diagnostic test, Optical coherence tomography, Laser flare-cell photometry

* Correspondence: a.denniston@bham.ac.uk
[1]Ophthalmology Department, University Hospitals Birmingham NHS
Foundation Trust, Birmingham, UK
[2]Academic Unit of Ophthalmology, Institute of Inflammation and Ageing,
University of Birmingham, College of Medical and Dental Sciences,
Birmingham, UK
Full list of author information is available at the end of the article

Background

Uveitis is a group of inflammatory conditions affecting the eye. It is a major cause of blindness globally, with an estimated prevalence of 38 to 114.5 per 100,000 population [1–3]. Uveitis can occur in any age, but it has a tendency to affect the working-age group, thus having a high socio-economic impact [1].

The Standardisation of Uveitis Nomenclature (SUN) Working Group classifies uveitis based on the anatomical focus of inflammation: anterior (anterior chamber), intermediate (vitreous), posterior (retina and/or choroid) and panuveitis (all anatomical parts) [4]. This systematic review focuses on inflammation in the anterior chamber, where inflammation causes a disruption to the normal blood-aqueous barrier, resulting in leakage of cells into the aqueous humour. These cells can be observed as floating particles in the anterior chamber (AC). An increase or decrease in the number of AC cells can be indicative of improving or worsening disease and are critical in identifying active inflammation and rationalising treatment decisions [4].

Clinical examination of AC cells

The current standard measurement as defined by the SUN grading system is clinical examination by slit-lamp biomicroscopy, whereby a clinician aims a 1 × 1-mm slit beam through the anterior chamber and counts the number of cells visible in the lit area [4]. The cell count is placed into one of six grades in the SUN grading system (Table 1). Prior to the SUN grading system, a number of alternative systems existed which attempt to quantify cells in the same way; however, the SUN grading system is now the accepted standard for clinical practice and for clinical trials [5–9].

Optical coherence tomography

In the last few decades, novel ophthalmic imaging techniques such as optical coherence tomography (OCT) have become widely adopted and provide new ways of quantifying disease through instrument-based measurements. OCT is fast, non-invasive and provides a quantifiable measure of ocular structure that is more sensitive and reliable than clinical estimates. Anterior segment optical coherence tomography (AS-OCT) can provide cross-sectional imaging of the AC and is able to identify cells in the anterior chamber as hyper-reflective dots. Several studies have suggested that AC cell count on OCT correlates with clinical grading [10, 11]. Additionally, the number of dots/cells in a given volume of aqueous humour can be counted manually or with automated software [12].

Laser flare and cell photometry

The laser flare meter was introduced in 1988 for quantification of anterior chamber protein and cells [13]. It is a fast and non-invasive technique which measures the amount of light scatter from particles as a laser beam is projected into the anterior chamber. The amount of back-scattered light is proportional to the concentration and size of proteins and cells in the aqueous humour.

Laser flare-cell photometry is primarily used in the estimation of AC flare (the hazy appearance given to the aqueous fluid by inflammatory proteinaceous leakage); however, certain models can be used for counting AC cells. Counting of cells by laser photometry is reported to be less accurate than its use in measuring flare, particularly in the extreme ends of the spectrum [14].

Purpose

Although we are aware of two main instrument-based techniques for quantifying AC cells, it is possible that more technologies, newer generations of the same technologies or the same technologies accompanied by newer acquisition techniques and software automation are available. The aim of this systematic review is to investigate all instrument-based methods for quantifying AC cells and evaluate the correlation of these measures with clinical grading using slit-lamp examination in patients with uveitis. Where reported, we will also compare the level of reliability and repeatability for these methods to determine which yields the most reliable results. For the purposes of this review, we will refer to all instrument-based methods as 'index tests'.

Aim

To investigate which instrument-based technologies for measuring anterior chamber cells in uveitis are available and assess their level of validation.

The following questions are proposed:

- Which non-invasive instrument-based tests (index tests) have the potential to measure anterior chamber cells in uveitis?
- What is the level of agreement between each index test and grading by clinical examination?

Table 1 The Standardisation of Uveitis Nomenclature (SUN) working group grading scheme for anterior chamber cells

Grade	Cells in field
0	< 1
0.5 +	1–5
1 +	6–15
2 +	16–25
3 +	26–50
4 +	> 50

Standardisation of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. [4]

- What is the reliability and repeatability of each index test?

Methods
Protocol
This protocol is designed as per guidelines of Preferred Reporting Items for Systematic Review and Meta-Analysis Protocol (PRISMA-P) [15]. The systematic review will be reported in accordance to the PRISMA guidelines [16].

Systematic review registration
Our protocol has been registered on PROSPERO (CRD42017084156) [17].

Searches
Databases
The relevant data for this review is likely to come from studies assessing a variety of quantitative research questions in uveitis; therefore, our search strategy will reflect the pathological finding of interest 'anterior chamber cells' and the disease context 'uveitis'. No search terms will be applied for the 'technologies/tests' to maximise sensitivity of the search. For bibliographic databases, free text and index terms (where available) will be combined for each search element. A sample search strategy for MEDLINE can be found in the Appendix. We will search:

- Bibliographic databases of published studies
  - MEDLINE (Ovid), 1946–present
  - Embase (Ovid), 1947–present
  - The Cochrane library, database inception–present
  - Centre for Reviews and Dissemination database (Health Technology Assessments and the Database of Abstracts and Reviews of Effects), database inception - present
- Registers of clinical trials
  - Clinicaltrials.gov (www.clinicaltrials.gov)
  - WHO International Clinical Trials Registry Platform (ICTRP portal) (www.who.int/ictrp)
- Abstract and conference proceedings
  - British library Ethos
  - ProQuest (www.proquest.com)
- Dissertations and theses
  - British Library's ZETOC
  - Conference Proceedings Citation Index (Web of Science)
- Grey literature
  - OpenGrey (www.opengrey.eu)

No restrictions will be placed on year or language of publication. The literature search results will be entered onto EndNote × 8.2 (Clarivate Analytics, Philadelphia, PA) to facilitate removal of duplicates, study selection, recording decisions and references. References to other works will be considered for inclusion.

Selection criteria
Participants/population
The populations of interest are those with evidence of anterior chamber cells and/or a diagnosis of uveitis, irrespective of active or inactive inflammation. There will be no restrictions on age, gender, ethnicity, or underlying aetiology or anatomical subtype. Studies with only healthy participants and studies on animals will not be included.

Index test
Included studies should describe one or more instrument-based methods for counting the number of anterior chamber (AC) cells.

Comparator
Included studies should use a clinical grading system (such as the SUN grading system) as a comparator.

Primary outcome
The primary outcome is the correlation between index tests and the clinical grading system; the reported outcome should be a correlation coefficient for the two measures. Studies which do not report a correlation coefficient but report matched measurements for the index and reference tests can be included, and the correlation coefficient extrapolated during analysis.

Additional outcome
To assess the reliability of a test, inclusion of a comparator is not required. The outcome of interest is intra/inter-observer reliability and repeatability of an index test in the same population as above.

Type of study
There will be no restrictions on study design; however, evaluation of correlation between index test and clinical grading using slit-lamp examination requires both tests to be done in a cross-sectional manner. Only those studies where measurements are taken within a reasonable time point (within 24 h of each other) will be included.

Case reports involving only one subject, commentaries, opinion articles and pictorial articles will not be included.

Selection process
Titles and abstracts of records returned by the searches will be screened for relevance to the review, to remove obviously irrelevant studies Two independent reviewers

will carry out quality assessment and reach consensus by discussion or referral to a third reviewer.

Full text of potentially relevant articles will be retrieved and assessed for inclusion in the review against the full selection criteria.

Data extraction

Data will be extracted from the included studies using a standardised data extraction form in Microsoft Excel (Microsoft, Washington, UK). The extraction process will be carried out by two independent reviewers with referral to a third reviewer if necessary. Information to be extracted from all studies include the following:

- Study characteristics
  - Title, authors, publication year, journal and language
  - Sample size
  - Study design
  - Index test used
    1. Manufacturer and model (including resolution and scan speed)
    2. Image acquisition settings
    3. Area, volume and position scanned in the AC
    4. Software for image analysis and level of automation (manual, semi-automated or fully automated)
  - Clinical grading system measurements
    1. Observer (same observer or different observers)
    2. Description of grading system used, including the name of the system, number of grades and number of cells in each grade
    3. Slit lamp settings such as area and brightness of illumination
- Patients' characteristics
  - Age, gender and ethnicity
  - Underlying aetiologies (type of uveitis, anatomical subtype, aetiological classification)
  - Active or inactive disease
  - If the study involves a therapeutic intervention: treatment details (indication, drug, dosage, route, subject pre/post-treatment status and length of follow-up)
- Outcomes and findings
  - Data will be extracted in preparation for two separate analyses depending on the data reported:
    1. *Evaluation of correlation between index tests and a clinical grading system:* Studies which include AC cell measurement by both index test and a clinical grading system in the same population will be used to evaluate the level of correlation between the two tests. The correlation coefficient reported will be directly extracted. If no correlation coefficient

is reported, we will extract index and reference test measurements, providing they are matched, and calculate the correlation coefficient. If the two measurements are not matched, we will contact the authors for matched measurements.

    2. *Evaluation of reliability and repeatability of an index test:* Studies which report intra- and inter-observer reliability for a study will be analysed separately for assessment of repeatability. The reported kappa values will be extracted for intra-observer reliability and inter-observer reliability.

- Cross-sectional measurements may be nested within longitudinal studies whose primary aims are not to compare performance or correlation between two tests. In this situation, measurements at each time point should be extracted and analysed as individual cross-sectional comparisons
- Results of sub-analyses and sensitivity testing for uveitis subtype (aetiological and anatomical) and disease activity (active versus inactive disease)

Quality assessment

Quality assessment of all included studies will be based on elements from the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS2) [18]. Assessment will be carried out at the study level. Two independent reviewers will carry out quality assessment and refer to a third reviewer if needed.

It is anticipated that a small number of studies will meet the inclusion criteria, with many of which being early proof-of-concept studies with a high risk of bias. Therefore, the risk of bias assessment and quality of evidence will be reported but will not influence data synthesis.

For studies assessing the correlation between an index and reference test, the following four risk of bias domains will be rated as low, high or unclear [18]:

1. Selection of participants
   a. Selection of participants with different degrees of disease severity should be justified in the context of assessing the full spectrum of disease, as opposed to any prior knowledge of how disease severity may affect test performance. As severe inflammation (i.e. grade 4+) is less prevalent than mild inflammation (i.e. + 0.5), a random selection of participants would invariably result in more participants with milder disease than participants with severe disease.
   b. Exclusion of participants should be justified in the context of interference with index test measurement (i.e. corneal opacities may prevent visualisation of AC structures behind it, or the

presence of AC pigment may appear like AC cells but is caused by a different pathological process).

2. Index tests
   a. We will consider whether index test acquisition and analysis parameters were determined a priori and consistent for all participants in the study.
   b. If analysis of index tests were done manually (i.e. manual counting of AC cells in an image), we will consider whether the observer was blinded to the clinical AC cell grading, and vice versa.

3. Reference test
   a. We will consider whether assessment using the clinical grading system in each study was standardised, i.e. performed by the same clinician, consistent slit-lamp settings such as brightness of illumination and ambient settings such as room lighting.

4. Flow and timing
   a. Order of tests: The order of tests does not affect the interpretability of results if the observer of one test is blinded to the results of the other test.

For studies assessing the reliability of a single index test, the same assessment should be done for selection of participants and index tests as above. Additional considerations should be given for:

1. Intra-observer reliability studies: the conditions under which index test repeatability was performed should be reported and kept consistent (i.e. room lighting, slit lamp beam intensity).
2. Inter-observer reliability studies: as well as the points relating to intra-observer reliability, any differences between observers (such as seniority and experience) should be reported and accounted for.

Data synthesis

Data synthesis will be divided into analysis for the two outcomes: correlation with reference test and reliability of index test. For each outcome, a narrative synthesis of tabulated evidence will be conducted and supported with a meta-analysis where possible.

1. Evaluation of correlation between index and reference test
   a. These studies will be grouped by type of technology (i.e. OCT, laser flare cell photometry). For each type of technology, we are expecting small numbers of studies; therefore, we will analyse different platforms, generations or manufacturers of the same technology together.
   b. Studies will then be grouped by comparator (i.e. different clinical grading systems).
   c. If the data allows, correlation coefficients between each index test versus reference test will be compared and pooled for meta-analysis.

2. Evaluation of reliability and repeatability of an index test
   a. These studies will also be grouped by type of technology, as described above.
   b. Intra-observer and inter-observer reliability will be analysed separately for each technology. If the data allows, reliability measure (i.e. Kappa values) will be compared and pooled for meta-analysis.

Assessment of clinical and methodological homogeneity will determine whether studies are sufficiently similar to allow for appropriate data pooling by meta-analysis. Heterogeneity across studies in each meta-analysis will be quantified using the I^2 statistic. Irrespective of the ability to appropriately undertake any meta-analyses, data will be reported narratively across all studies for each grouping.

We will perform sensitivity analyses if there is significant heterogeneity between studies. If data permits, we will consider subgroup analyses for population groups (i.e. age, gender and ethnicity), anatomical subtype of uveitis and aetiological subtype. It is anticipated that only a small number of studies will be relevant to the inclusion criteria, and this may limit our ability to carry out meaningful subgroup analyses.

Discussion

The assessment of uveitis is complex: firstly, because uveitis describes a heterogeneous group of diseases with significant variations between different anatomical and disease-specific subtypes, and secondly, because many clinical measures in ophthalmology are based on visual function (namely visual acuity), which are subjective to the patient and do not always reflect active inflammation. Treatment for anterior uveitis, namely topical steroids, carry side effects such as increased risk of glaucoma and accelerated onset of cataract. The ability to reliably measure ocular inflammation in uveitis is important for rationalising treatment in clinical practice and assessing disease outcome in clinical trials.

Whilst the consensus Standardisation of Uveitis Nomenclature meeting of 2005 was a significant effort towards defining a systematic method in assessing uveitic inflammation, it is recognised that its reliance on clinical examination remains subjective, unreliable and poorly sensitive. A number of factors can affect the examiner's

ability to see AC cells including the slit lamp optics, the degree of illumination, and the observer's skills [19]. The problem exists that clinical grading through examination is error-prone and is therefore susceptible to intra-observer and inter-observer variability [19, 20]. Clinical grading systems also typically classify levels of severity into stepwise grades, in a non-continuous and non-linear fashion, with a wider range of cell counts in the higher grading groups (i.e. 26–50 cells in field for grade 3+ and > 50 cells in field for 4+) than lower grading groups [4]. According to the SUN criteria, "improved activity", or a decrease in inflammation, requires a two-step improvement or resolution to grade 0. However, a much larger reduction of AC cell count is required in higher grades of inflammation (i.e. 4+ to 2+ requires 50+ cells decreasing to 16–25 cells) than lower grades (i.e. 2+ to 0.5+ requires 16–25 cells decreasing to < 1 cells). This non-linear grading scale may not be able to detect small changes, especially within higher grades of inflammation, allowing potentially clinically meaningful changes to go undetected. In clinical trials, this could result in new therapies being deemed as failure, despite having a clinically significant improvement.

Instrument-based technologies such as the laser flare-cell photometer and OCT have shown potential for assessing anterior chamber cells in several studies. These instrument-based measures are less operator-dependent and carry advantages such as objectivity. However, their performance in these domains has not been compared in a systematic way. The field of ophthalmic imaging is continuing to expand rapidly; therefore, it is timely to undertake a systematic review to examine the evidence for instrument-based measures of intraocular inflammation.

This systematic review will explore the range of technologies available for measuring AC cells to estimate their level of reliability and correlation with clinical grading systems. The findings from this review would contribute to the process of validating instrument-based methods for guiding treatment decisions in clinical care and for measuring outcome in uveitis clinical trials.

Abbreviations
AC: Anterior chamber; MEDLINE: Medical Literature Analysis and Retrieval System Online; OCT: Optical Coherence Tomography; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis; QUADAS2: Quality Assessment of Diagnostic Accuracy Studies; SUN: Standardisation of Uveitis Nomenclature

Ethical approval and consent to participate
Not applicable.

Funding

Appendix

Table 2 MEDLINE sample search strategy

1	Exp Uveitis/
2	Uveiti*. Ti, ab.
3	1 or 2
4	Anterior chamber. Ti, ab.
5	Aqueous humour. Ti, ab.
6	Aqueous humor. Ti, ab.
7	4 or 5 or 6
8	Cell*. Ti, ab.
9	3 and 7 and 8

Availability of data and materials
Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Authors' contributions
XL, AKD and DJM conceived the review and participated in the review design, registration and write-up. AKD and PK provided expert clinical context in uveitis, ophthalmic imaging and instrument-based outcomes in ophthalmology to guide the review design. ALS participated in the review design and write-up. AKD is a guarantor of the review protocol. All authors read and approved the final manuscript.

Consent for publication
Not applicable.

Competing interests
None.

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

Author details
[1]Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [2]Academic Unit of Ophthalmology, Institute of Inflammation and Ageing, University of Birmingham, College of Medical and Dental Sciences, Birmingham, UK. [3]Institute of Child Health, University College London, London, UK. [4]NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK. [5]Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK. [6]Centre for Rare Diseases, Birmingham Health Partners, Institute of Translational Medicine, Birmingham, UK.

References
1. Durrani OM, Meads CA, Murray PI. Uveitis: a potentially blinding disease. Ophthalmologica. 2004;218:223–36.
2. Williams GJ, Brannan S, Forrester JV, Gavin MP, Paterson-Brown SP, Purdie AT, et al. The prevalence of sight-threatening uveitis in Scotland. Br J Ophthalmol. 2007;91:33–6.
3. Rothova A, Suttorp-van Schulten MS, Frits Treffers W, Kijlstra A. Causes and frequency of blindness in patients with intraocular inflammatory disease. Br J Ophthalmol. 1996;80:332–6.
4. Jabs DA. Standardization of uveitis nomenclature for reporting clinical data. Results of the first international workshop. Am J Ophthalmol. 2005;140:509–16.
5. Hogan MJ, Kimura SJ, Thygeson P. Signs and Symptoms of Uveitis*. Am J Ophthalmol. Elsevier Inc; 1959;47:155–170.
6. Schlaegel T. Essentials of uveitis. Boston: Little, Brown, Inc; 1967.

7. Nussenblatt RB, Whitcup SM. Uveitis: fundamentals and clinical practice 3rd. Philadelphia: Mosby; 2004.
8. Simpkin AL, Vyas JM, Armstrong KA. Diagnostic reasoning: an endangered competency in internal medicine training. Ann Intern Med. 2017;167:507–9.
9. Foster CS, Vitale AT, editors. Diagnosis and Treatment of Uveitis. Philadelphia: W B Saunders; 2002.
10. Li Y, Lowder C, Zhang X, Huang D. Anterior chamber cell grading by optical coherence tomography. Invest Ophthalmol Vis Sci. 2013;54(1):258-65. Published 2013 Jan 9. https://doi.org/10.1167/iovs.12-10477.
11. Igbre AO, Rico MC, Garg SJ. High-speed optical coherence tomography as a reliable adjuvant tool to grade ocular anterior chamber inflammation. Retina. 2014;34:504–8.
12. Sharma S, Lowder CY, Vasanji A, Baynes K, Kaiser PK, SK. S. Automated analysis of anterior chamber inflammation by spectral-domain optical coherence tomography. p. 1464–70.
13. Sawa M, Tsurimaki Y, Tsuru T, Shimizu H. New quantitative method to determine protein concentration and cell number in aqueous in vivo. Jpn J Ophthalmol. 1988;32:132–42.
14. Ladas JG, Wheeler NC, Morhun PJ, Rimmer SO, Holland GN. Laser flare-cell photometry: methodology and clinical applications. Surv Ophthalmol. 2005; 50:27–47.
15. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015.
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010;8:336–41.
17. Liu X, Moore DJ DA. Instrument-based tests for measuring anterior chamber (AC) cells in uveitis: a systematic review [Internet]. 2017 [cited 2018 Jul 2]. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=84156
18. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Quadas-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–36.
19. Wong IG, Nugent AK, Vargas-Martín F. The effect of biomicroscope illumination system on grading anterior chamber inflammation. Am J Ophthalmol. 2009;148:516–520.e2.
20. Kempen JH, Ganesh SK, Sangwan VS, Rathinam SR. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. Am J Ophthalmol. 2008;146:813–8.

Instrument-based Tests for Measuring Anterior Chamber Cells in Uveitis: A Systematic Review

Xiaoxuan Liu, Ameenat L. Solebo, Livia Faes, Sophie Beese, Tasanee Braithwaite, Matthew E. Round, Jesse Panthagani, Aditya U. Kale, Thomas W. McNally, Didar Abdulla, Pearse A. Keane, David J. Moore & Alastair K. Denniston

Taylor & Francis
Taylor & Francis Group

Check for updates

ORIGINAL ARTICLE

Instrument-based Tests for Measuring Anterior Chamber Cells in Uveitis: A Systematic Review

Xiaoxuan Liu, MBChB [1], Ameenat L. Solebo, BSc, MBBS, FRCOphth, PhD[2,3], Livia Faes, MSc[2,4], Sophie Beese, MPH[5], Tasanee Braithwaite, FRCOphth, MPH[6,7], Matthew E. Round, MBChB[8], Jesse Panthagani, BSc, MBBS, FRCS, FRCOphth[6], Aditya U. Kale, BSc, MBChB[1,6], Thomas W. McNally, BSc[1,6], Didar Abdulla, MD[6], Pearse A. Keane, MD, FRCOphth[2], David J. Moore, PhD[5], and Alastair K. Denniston, PhD, MRCP, FRCOphth[1,2,6,9]

[1]*Academic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK,* [2]*NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK,* [3]*Institute of Child Health, University College London, London, UK,* [4]*Department of Ophthalmology, Cantonal Hospital Lucerne, Lucerne, Switzerland,* [5]*Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK,* [6]*Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK,* [7]*Moorfields Eye Hospitals NHS Foundation Trust, London, UK,* [8]*Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK, and* [9]*Centre for Rare Diseases, Institute of Translational Medicine, Birmingham Health Partners, Birmingham, UK*

ABSTRACT

Purpose: New instrument-based techniques for anterior chamber (AC) cell counting can offer automation and objectivity above clinician assessment. This review aims to identify such instruments and its correlation with clinician estimates.

Methods: Using standard systematic review methodology, we identified and tabulated the outcomes of studies reporting reliability and correlation between instrument-based measurements and clinician AC cell grading.

Results: From 3470 studies, 6 reported correlation between an instrument-based AC cell count to clinician grading. The two instruments were optical coherence tomography (OCT) and laser flare-cell photometry (LFCP). Correlation between clinician grading and LFCP was 0.66–0.87 and 0.06–0.97 between clinician grading and OCT. OCT volume scans demonstrated correlation between 0.75 and 0.78. Line scans in the middle AC demonstrated higher correlation (0.73–0.97) than in the inferior AC (0.06–0.56).

Conclusion: AC cell count by OCT and LFP can achieve high levels of correlation with clinician grading, whilst offering additional advantages of speed, automation, and objectivity.

Keywords: Anterior chamber cells, aqueous humor, aqueous humour, diagnostic test, laser flare-cell photometry, optical coherence tomography, systematic review, uveitis

Uveitis, an umbrella term describing inflammatory ocular conditions, is a significant cause of blindness worldwide.[1–3] Anterior uveitis describes inflammation affecting the anterior chamber (AC), which is predominantly characterized by AC cells and flare, where disruption of the blood-aqueous barrier results in leakage of inflammatory blood constituents into the aqueous humor.

Detection and monitoring of disease activity is crucial for rationalizing medical therapy, which is particularly important because therapeutic interventions for uveitis carry risks of significant adverse ocular and systemic side effects; these include cataract raised intraocular pressure and opportunistic infection. The Standardization of Uveitis Nomenclature (SUN) Working Group proposed the now preferred clinical AC cell grading system.[4] In this, an observer aims a 1 mm^2 light beam through the AC and counts the number of illuminated cells visible. The cell count is then placed into one of six grades in the SUN grading system (Table 1). Prior to SUN, a number of alternative systems existed that quantified cells in a similar way.[5–9]

Multiple limitations of this system are recognized. First, it is prone to bias due to reliance on subjective estimation of an observer. Although instructions dictate that cell counting should be carried out in one moment in time, in reality, this is a near-impossible task, especially at higher grades where cell counts exceed 30–40 cells/mm^2. Second, the SUN grading system uses a non-linear, non-continuous scale with large steps between grades. Changes in inflammatory activity within one grade may go undetected, especially in the higher grades. Third, it relies upon the presence of an ophthalmic clinician trained in slit-lamp biomicroscopy, and therefore limits disease monitoring to a hospital setting. Consequently, delivery of uveitis care in other health-care settings such as remote screening and community-based monitoring has not been feasible.

Instrument-based techniques such as laser flare-cell photometry (LFCP), and more recently anterior segment optical coherence tomography (AS-OCT), have shown potential for objectively quantifying AC cells. LFCP became available in 1988 and uses the light scattering properties of AC particles to quantify the concentration of inflammatory materials in the aqueous humor. It has been primarily validated as a tool for measuring AC flare,[10] the cloudy appearance given to the aqueous during inflammation, however several models also have the ability to count AC cells. AS-OCT provides cross-sectional scans of the AC and can capture cells in aqueous humor as hyper-reflective dots. Given the drive towards objective, quantitative assessment of disease status, a systematic examination of the evidence for such technologies is timely.[11,12] This review aims to identify all instrument-based tools for counting AC cells and evaluate their correlation with clinician grading systems.

METHODS

This review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement.[13] The methodology was specified in advance and protocol registered with PROSPERO (CRD42017084156).[14,15]

Eligibility Criteria

We included studies that described one or more instrument-based methods for counting AC cells in patients with uveitis (index tests) in comparison to a clinician grading system (through slit-lamp examination). We also included studies reporting test reliability (e.g., intra or inter-observer reliability and/or repeatability). We did not place restrictions on age, gender, ethnicity, underlying etiology or disease activity status. Animal studies and studies involving only healthy participants, single case reports, commentaries, opinion articles, and pictorial articles were excluded.

The primary outcome was the level of correlation between index tests and clinician grading. The secondary outcome was intra/inter-observer reliability and repeatability of the index test.

Search Methods for Identifying Studies

We combined free text terms and index terms reflecting the pathological finding of interest 'cells' and 'anterior chamber' or 'aqueous humor', and the disease context 'uveitis'. The search strategy was adapted to match the index terms in different

TABLE 1. Clinician grading scales used in each study.

Grade	Previously published grading systems			Number of cells in each grade, by study					
	SUN	Hogan	BenEzra	Igbre	Invernizzi	Sharma	Li	Ohara	Tugal-Tutkun
0	<1	0	<5	<1	<1	<1	0	0	<5
0.5	1–5	-	-	1–5	1–5	1–5	1–5	1–4	-
1	6–15	5–10	5–10	6–15	6–15	6–15	6–10	5–10	5–10
2	16–25	10–20	11–20	16–25	16–25	16–25	11–20	11–30	11–20
3	26–50	21–50	21–50	26–50	26–50	26–50	21–50	31+	21–50
4	50+	50+	50+	50+	50+	50+	50+	-	50+
5	-	-	hypopyon	-	-	-	-	-	hypopyon

TABLE 2. Study Characteristics.

Author	Year	Study Design	No. of participants	No. of eyes	Gender, no. of eyes (%)	Mean age, years (range)	Aetiological classification, no. of eyes (%)
Ohara	1989	Prospective	124	127	44 (35%) male 80 (65%) female	NR (12–76)	Sarcoidosis 53 (43%), Behçet's 14 (11%), VKH 6 (5%) Bilateral ARN 3 (2%), other 14 (11%), unknown 34 (27%)
Tugal-Tutkun	2008	Prospective	232	153	124 (53%)male 108 (47%) female	Active ocular Behcets, 28 (NR) Inactive ocular Behcets, 29 (NR) Non-ocular Behcet's group, 31 (NR) Healthy control, 27 (NR)	Active ocular Behcets 54 (35%) Inactive ocular Behcets 53 (23%) Non-ocular Behcets 78 (34%) Health control 47 (20%)
Li	2013	Prospective	35	66	NR	NR	Non-granulomatous 30 (39%) Granulomatous 12 (16%) Quiescent 16 (21%) Healthy control 19 (25%)
Igbre	2014	Retrospective	41	78	12 (29%) male 29 (71%) female	48 (10–83)	Non-granulomatous anterior uveitis 9 (23%), sarcoidosis 6 (16%), HLA-B27 5 (12%), Panuveitis 4 (10%), intermediate uveitis 3 (7%), granulomatous anterior uveitis 2(6%), uveitis glaucoma hyphema syndrome 1 (2%), BCR 1 (2%), HSV 1 (2%), JIA 1 (2%), multifocal choroiditis 1 (2%), scleritis 1 (2%), herpetic keratouveitis 1 (2%), pars planitis 1 (2%), VKH 1 (2%), sympathetic ophthalmia 1 (2%), unknown 2 (6%),
Sharma	2015	Prospective	76	114	Single line scan: 16 (32%) male 34 (68%) female Volume scan: 8 (27%) male 22 (73%) female	43 (12–94)	Line scan: Idiopathic 34 (41%), HLA B27 17 (21%), JIA 12 (15%), Sarcoidosis 10 (12%), VKH 6 (7%), HSV 2 (2%), Behçet's 2 (2%) Volume scan: Idiopathic 10 (32%), HLA-B27 7 (23%), JIA 7 (23%), Sarcoidosis 4 (13%), VKH 1 (3%), Psoriatic arthritis 2 (6%).
Invernizzi	2017	Prospective	122	237	NR	Healthy controls 42 (NR) Inactive uveitis 43 (NR) Active uveitis 45 (NR)	Healthy controls 70 (30%) Inactive uveitis 97 (40%) Active uveitis 70 (30%)

NR – not reported, ARN - acute retinal necrosis, VKH - Vogt-Koyanagi-Harada, BCR - birdshot chorioretinopathy, HSV - herpes simplex virus, JIA - juvenile idiopathic arthritis).

TABLE 2. *(Continued)*

Author	Index test	Manufacturer and model	Image acquisition settings	Position scanned	Area	Volume	Image analysis software	Level of automation
Ohara	LFCP	FC-1000, KOWA (Kowa Company, Ltd, Tokyo, Japan)	Final reading is the average of 5 repeated readings.	NR	NA	0.075 mm³	None	Semi-automated
Tugal-Tutkun	LFCP	FC-2000, KOWA (Kowa Company, Ltd, Tokyo, Japan)	Seven readings are obtained. The highest and lowest values are discarded and mean of the remaining five readings is the final reading.	NR	NA	0.5mm³	None	Semi-automated
Li	OCT	Time domain OCT prototype system (Carl Zeiss Meditec Inc., Dublin, CA). A time-domain OCT system with an axial resolution of 17 μm and transverse resolution of 45 μm.	Two concentric circular scans with diameters of 2 and 4 mm. Inner and outer circular scans consisted of 256 and 512 axial scans, respectively. Each AC grading scan was divided into three regions: central (inner circular scan), superior (superior semicircle of the outer circular scan), and inferior (inferior semicircle of the outer circular scan).	Inferior, central and superior	2 concentric circular scans of 2 and 4 mm diameter	NR	Unspecified custom software developed by authors	Automated
Igbre	OCT	AS-OCT (Visante OCT, Zeiss Meditec Dublin, CA	Between 4 and 8 high-resolution corneal cross-sectional scans (10mm across each).	NR	4 to 8 lines of 10 mm width	NA	Image pre-processing 30% desaturation (Photoshop; Adobe Systems, San Jose, CA)	Manual
Sharma	OCT	The RTVue-100/CAM (Optovue Inc., Fremond, CA) with corneal adapter	Line scan: 6mm single B-scan at the central cornea. 3D volume scan: 6mm³ consisting of 512 B-scans each with 128 A-scans.	Central	One 6 mm line	Line scan: NA 3D volume scan: 6mm³	Custom software developed by authors: *ImageIQ* (Cleveland Ohio)	Automated for 3D scans, manual for line scan
Invernizzi	OCT	Casia SS-1000 OCT device (Tomey Corporation, Nagoya, Japan).	Two 6mm cross-sectional scans with a depth of 2048 A/B scans	Central	Two 6mm lines	NA	Built in software (Tomey Link Exam Viewer Version 7F.2)	Manual

NR – not reported, NA – not applicable, LFCP – laser flare-cell photometry, OCT – optical coherence tomography, SS – Swept source

Author	Index test	Clinical Grading System	Observer	No. eyes in each clinical grade
Ohara	LFCP	Unspecified	NR for index test One observer for reference test	NR
Tugal-Tutkun	LFCP	BenEzra	One observer for index test One observer for reference test (different observers)	Grade 0 (7) Grade 0.5 to 2 (26)* Grade 3–4 (17)* Grade 5 (hypopyon) 0

Author	Index test	Clinical Grading System	Grade distribution	Observer
Li	OCT	Modified Hogan	Grade 0.5 (6 non-granulomatous, 4 granulomatous), Grade 1 (10 non-granulomatous, 3 granulomatous), Grade 2 (2 non-granulomatous, 2 granulomatous), Grade 3 (8 non-granulomatous, 2 granulomatous), Grade 4 (4 non-granulomatous, 0 granulomatous)	NR for index test; One observer for reference test
Igbre	OCT	SUN	Grade 0 (22), Grade 0.5 (20), Grade 1 (13), Grade 2 (13), Grade 3 (1), Grade 4 (0)	One observer for index test; One observer for reference test (same observer)
Sharma	OCT	SUN	Grade 0 (24), Grade 0.5 (5), Grade 1 (16), Grade 2 (15), Grade 3 (11), Grade 4 (12)	One observer for index test; One of two senior uveitis specialists for reference test
Invernizzi	OCT	SUN	Grade 0 (6), Grade 0.5 (24), Grade 1 (19), Grade 2 (15), Grade 3 (3), Grade 4 (3)	One observer for index test; NR for reference test

*Tugal-Tutkun et al. reported number of eyes in grades 0.5–2 and 3–4 combined.

Author	Index test	Clinical Grading System	No. eyes included in correlation analysis n (description)	Correlation Coefficient (95% CI) Spearman's r
Ohara	LFCP	Not specified	NR	$r = 0.66$ (NR)
Tugal-Tutkun	LFCP	BenEzra	54 (active uveitis group)	$r= 0.87$ (NR)
Li	OCT	Modified Hogan	30 (Non-granulomatous group); 12 (granulomatous group)	Non-granulomatous: Central AC $r= 0.73$ (0.55–0.88), Superior AC $r= 0.75$ (0.56–0.88), Inferior AC $r= 0.56$ (0.36–0.81). Granulomatous: Central AC $r= 0.74$ (CI NR; estimated as 0.29–0.92*), Superior AC $r= 0.73$ (CI NR; estimated as 0.27–0.92*), Inferior AC $r= 0.06$ (CI NR; estimated as −0.53–0.61*)
Igbre	OCT	SUN	69 (all included subjects)	$r= 0.74$ (0.62 to 0.83)
Sharma	OCT	SUN	83 eyes (single line scan group); 31 eyes (volume scan group)	Single line scan $r = 0.97$ (CI NR; estimated as 0.95–0.98*), Manual volume scan $r= 0.78$ (0.60–0.90), Automated volume scan $r= 0.75$ (0.55–0.88)
Invernizzi	OCT	SUN	70 (active uveitis group)	$r = 0.94$ (0.91–0.96)

databases (**supplementary materials**). Database searches were carried out in MEDLINE, Embase, Cochrane Controlled Register of Trials (CENTRAL), Centre for Reviews and Dissemination Database (Health Technology Assessments and the Database of Abstracts and Reviews of Effects), Clinicaltrials. gov, WHO International Clinical Trials Registry Platform (ICTRP portal), British Library's ZETOC, Conference Proceedings Citation Index (Web of Science), British Library Ethos, ProQuest and OpenGrey. We searched all databases from inception to 22 March 2018, with no date or language restrictions. We manually searched citations of review articles and included studies to identify additional relevant articles.

Study Selection

Two reviewers independently screened studies at each stage. Disagreements were resolved through discussion and input from a third reviewer.

Data Collection

Two reviewers extracted data independently using a pre-specified data extraction sheet. The data included population characteristics (number of participants, gender, age, underlying etiology), index test characteristics (technology, manufacturer, model, image acquisition settings, area sampled and software automation), clinician grading (name of grading system used, number of patients in each grade) and outcome (correlation coefficient, inter/intra-observer reliability). Cell counting analysis was recorded as fully automated, semi-automated or manual. For the clinician grading, we extracted how each grade was defined and whether any modifications were made to validated clinical grading systems. We contacted three authors for further information[16–18], all of whom responded and one provided further data (confidence intervals) which was not reported in the original paper.[16]

Risk of Bias Assessment

Two reviewers independently assessed risk of bias using the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2).[19] We adapted each element in QUADAS-2 to address the review question. Specifically, we explored potential sources of bias arising from the index test and clinician grading procedures: whether the test protocols were determined _a priori_ and standardized for all participants, and whether observers were blinded to test measurements.

Data Synthesis and Analysis

For each outcome, studies were grouped by index test technology and then by choice of clinician grading tool. For each technology, we tabulated the evidence and provided a narrative synthesis. Where authors modified clinician grading systems, these were considered separately from the validated versions (Table 2). Where confidence intervals for correlation coefficients were not reported, we estimated them using sample size and correlation coefficient and presented this on a forest plot. All statistical analysis was performed using Stata Statistical Software (Release 15. College Station, TX: StataCorp LP.)

RESULTS

Results of the Search

The study selection process is summarized in the PRISMA flow diagram (Figure 1). The searches from database conception to 22 March 2018 yielded 3470 bibliographic records after de-duplication. Of these, 3432 were excluded upon screening of titles and abstracts. The large number of exclusions is due to the unrestrictive nature of our search strategy, which did not specify any index test terms, and the small number of published studies that made comparisons between an index test and clinician grading. The remaining 38 articles were obtained in full text for further scrutiny and a further 32 articles were excluded. The reasons for exclusion were missing or incomplete reporting of clinician grading system (n = 13), the target disease not being uveitis (n = 15) and no correlation/reliability outcome reported (n = 4). Six unique studies met the eligibility criteria and were included (Table 1).

Methodological Quality of the Included Studies

Using QUADAS-2, one study was identified as having unclear risk of bias for patient selection due to the exclusion of patients with posterior synechiae (**supplementary figure**), which is known to affect LFCP readings.[20] Another study had an unclear risk of bias in the index test domain as it was unclear whether observers were blinded to the clinician grading.[21] One study had a high risk of poor applicability due to patient selection, as only patients with Behcet's disease were included.[20] We graded all studies as having unclear risk of bias in the reference test domain due to previously mentioned concerns around the reliability of subjective clinician grading.

Patients' Characteristics and Design Features

The six studies enrolled 775 eyes (from 630 subjects) and dated from 1989 to 2017.[16–18,20–22] Participants were recruited prospectively in all studies, except for Igbre et al. who used existing clinical data. All comparisons between index test and clinician grading were done in a cross-sectional manner. Gender was reported in 5 out of 6 studies, in which 47.8% (n = 301) participants were male. The mean age was 45.2 years (range 27.0–81.0 years). Four studies included mixed etiologies, one study did not report underlying etiology,[16] and one included only Behcet's disease patients.[20] The underlying etiologies across all studies included non-granulomatous uveitis, sarcoidosis, HLA-B27-associated uveitis, unspecified panuveitis, unspecified intermediate uveitis, pars planitis, acute retinal necrosis, granulomatous uveitis, juvenile idiopathic arthritis, Behcet's disease, multifocal choroiditis, sympathetic ophthalmia, Vogt-Koyanagi-Harada disease, uveitis glaucoma hyphema syndrome, birdshot chorioretinopathy, herpes simplex virus-associated uveitis, herpetic keratouveitis, idiopathic uveitis and unknown cause. Three studies included healthy controls,[16,17,20] but only analyses where uveitis patients separately reported were included in the correlation analyses of this review.

Clinical Grading Systems

Three studies used the SUN grading system as a comparator,[16,18,22] one study used the scoring system described by BenEzra et al. in 1991,[9] one study used a modified version of the 1959 Hogan system,[5,17] and one study used an unspecified clinical grading system.[21] Upon contacting the author, the justification for modifying the Hogan grading system was due to the uveitis specialist's preference.[17] The differences between the grading systems are outlined in Table 1. Four studies reported the number of subjects with each clinical AC cell grade,[16–18,22] one study combined grades[20] (for example, "26 subjects had grades 0.5 to 2") and one did not report this.[21] Sources of variation include number of cells seen in each grade (particularly in grades 1 and 2), the addition of a "0.5+" grade in the SUN grading system, the inclusion of a grade 5 to account for presence of hypopyon by the BenEzra system, and the lack of a specified slit beam size in the Hogan and BenEzra systems (SUN grading specifies 1 mm²).[5,9]

Instruments for Measuring AC Cells

Six studies were included for analysis.[16–18,20–22] All six studies compared the measurements of AC cells on an instrument to a clinical grading system and reported the correlation coefficient. No studies reported reliability for instrument-based grading. We identified two instrument-based technologies for quantifying AC cells: OCT and LFCP.

Optical Coherence Tomography

Four studies reported correlation between OCT and a clinical grading system.[16-18,22] Three studies[16,18,22] used commercially available OCT machines and one used a prototype system.[17] The scanning protocols (including the scan settings, position, area, and volume scanned) were unique in each study. Li et al. used a time-domain OCT system (Carl Zeiss Meditec Inc., Dublin, CA), with an axial resolution of 17 microns and axial depth of 8 mm, to capture concentric cross-sections of the central AC in 35 uveitis patients (66 eyes). The Visante AS-OCT (Zeiss Meditec Dublin, CA) was used by Igbre et al. to capture 4–8 cross-sectional images in 41 patients (78 eyes), but the chamber position, area and volume scanned were not reported.[22] Sharma et al. captured line scans and 6 mm³ scans at the central cornea using the RTVue-100/CAM (Optovue) in 76 patients (114 eyes).[18] Invernizzi et al. used the swept source Casia SS-1000 OCT device (Tomey Corporation, Nagoya, Japan) to capture two 6 mm cross-sectional scans of the AC in 167 uveitic eyes and 70 healthy eyes.[16] Two studies used manual cell counting[16,22] and one automated this,[17] whilst the fourth study used both methods.[18] For the two studies using automated cell counting, algorithms were developed *de novo* for study purposes and are not openly available.[16,18]

Laser Flare-Cell Photometry

Two studies reported correlation between LFCP and a clinician grading system.[20,21] In both studies, the LFCPs were manufactured by KOWA (Kowa Company, Tokyo, Japan), but the models differed; FC-1000[20] and FC-2000.[20] All flare measurements were calculated automatically using the machine's built-in function. As per the manufacturer's recommendations, the observer took several readings, discarded the highest and lowest values, before averaging the final values to derive an average cell count measurement and a standard deviation. Neither study reported the position and area/volume of aqueous scanned.
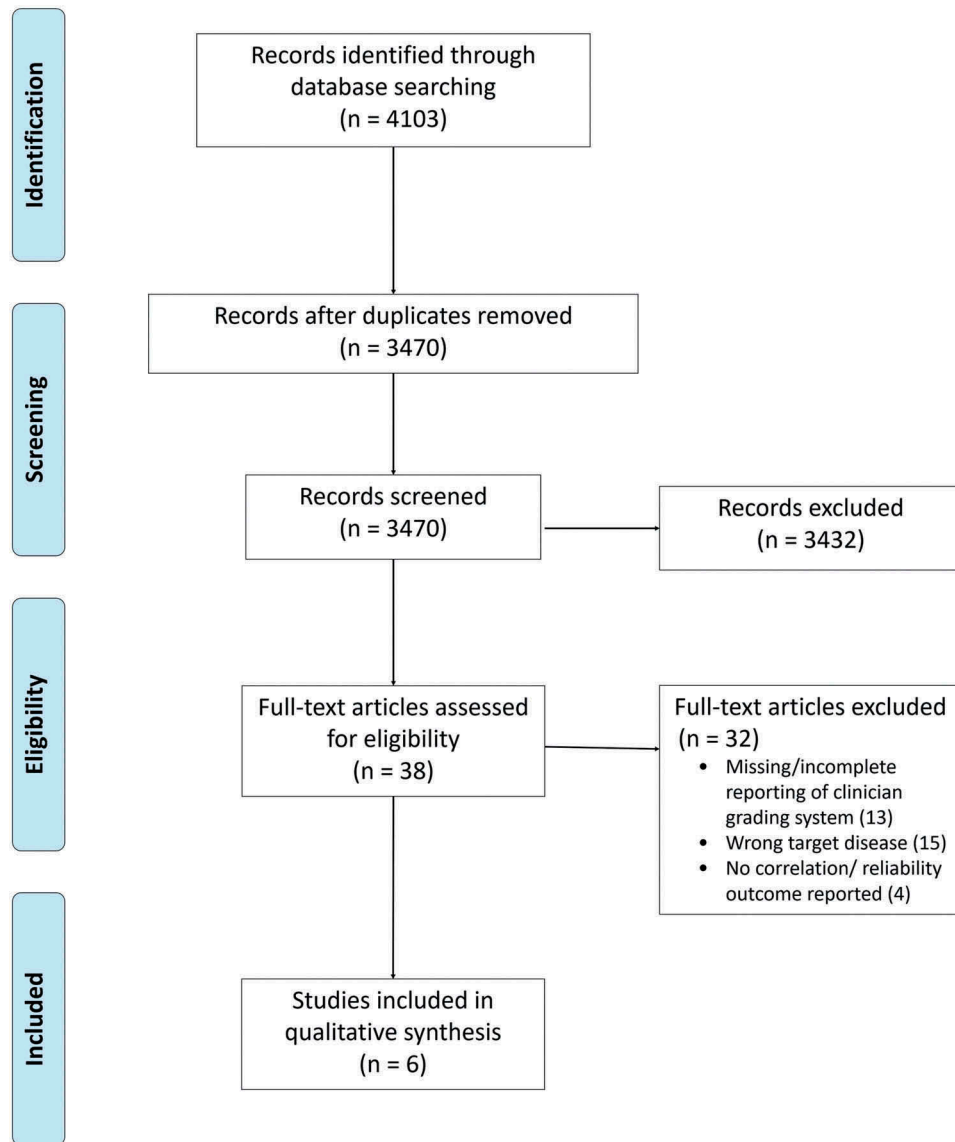
FIGURE 1. PRISMA flow diagram.

Correlation between Index Tests and Clinician Grading Systems

All six studies reported a correlation coefficient between the index test and a clinical grading system, using Spearman's *r*. The level of correlation between index tests and clinician grading systems is shown using a forest plot (Figure 2).

For the time-domain OCT devices, the correlations were reported to be 0.74 (95% CI 0.62–0.83) in the Visante device (Zeiss Meditec, Dublin, CA),[22] and up to 0.75 for the prototype Zeiss device, depending on position of the scan[17] (highest correlation *r* = 0.75 for superior AC and lowest correlation *r* = 0.06 for inferior AC). For the newer spectral-domain or swept source OCT devices, which unlike the time domain models, have a faster acquisition time and maximal axial imaging resolution smaller than the normal range of white cell width (10–17 microns),[23] higher correlation values were reported (0.97, *p* < .0001 for RTVue-100/CAM, Optovue,[18] and 0.94, *p* < .0001 for the Casia SS-1000 OCT device, Tomey Corporation.[16]) There was no apparent association between the level of automation of OCT images analysis and the correlation with AC cell count.

OCT can also acquire volume scans by repeating densely placed single line scans. All four studies used single line scans at different positions across the anterior chamber. Sharma et al. additionally compared single line scans to 3D cubic volume scans of 6 mm³, and found the single line scans to have higher correlation with the clinical grading than the volume scans (0.94 for single line and 0.74–0.77 for volume scan).[18]

For the LFCP, two studies reported correlation with clinician grading (*r* =0.66[20] and *r* =0.87[19]). The KOWA FC-2000, which scans a larger volume of

aqueous (0.5 mm^3) than the FC-1000 (0.075 mm^3), achieved a higher level of correlation ($r = 0.87$[19]).

Study Heterogeneity

There was considerable heterogeneity between the methodology and populations described by the three studies which shared a common comparator (SUN grading).[16,18,22] Due to the differences in scan acquisition parameters (varying sized scan areas and levels of automation) and distribution of AC cell severity in the study populations (as measured by clinician grading), we did not consider the index test measurements to be directly comparable by meta-analysis.

DISCUSSION

This is the first systematic review to evaluate instrument-based technologies for counting AC cells in uveitis. We found two technologies for this purpose: OCT and LFCP.

When these technologies were used in a relatively consistent way, with precisely specified measurement and scanning protocols, we found strong correlation with the SUN grading system ($r = 0.74$–0.97). However, the range of correlation for instrument-based measurements versus clinician grading ranged from 0.06 to 0.97. Included studies demonstrated a higher correlation coefficient achieved by OCT than LFCP. However, the inconsistent use of clinical comparators across studies prevented us from making direct comparisons between the technologies.

Performance and Limitations of Measures of AC Cells

Studies of instrument-based cell counting using OCT versus clinician grading reported correlations of $r = 0.06$–0.97, and for LFCP $r = 0.66$–0.87. The variation in correlations seen between studies of the same platform may arise due to several important factors which may impact instrument-based measures only, human clinical measures only, or both.

Factors Affecting Instrument-based Measures

Some variation in the correlation between studies may suggest that not all instrument-based measures of the same technology are equal, and that performance may be affected by the model and technique used. Newer models of OCT have higher resolution (enabling improved discrimination of cells) and faster acquisition time (overcoming the effects of missing or double-counting moving cells).

Factors Affecting the Performance of Human-based Clinical Measures

Some variation in the level of correlation may be unrelated to the technology, but rather reflect poor reliability of the clinician-based method. In addition to the well-recognized generic limitations of subjectivity and imprecision,[24,25] we noted some specific variations in choice of clinician grading systems used across studies. Two studies published after the 2005 SUN Workshop used non-SUN grading systems,[17,20] and one made a custom modification by adding a 0.5 grade to the pre-SUN Hogan system.[17] The reasons for this are unclear. It is unlikely that preference for one grading system over another is based on perceptions of superiority, as all clinical grading systems share the same issues around subjectivity. Additional factors that were not always recorded in these studies but are known to impact the reliability of the clinical measure are the experience of the clinician, and number of observers independently scoring each AC.[25]

Factors Affecting the Performance Of Both Instrument-based and Human-based Measures

Factors such as patient selection may affect both instrument-based and human measures. For example, including patients with corneal opacity is likely to reduce performance of both measures due to reduced cell discrimination, although there is some evidence it may impact OCT measures less.[26] Our review found higher levels of correlation for scans involving a smaller area of the central AC. This difference may arise from several factors including:

Areas sampled: Li et al. reported weaker correlation between clinical grading and OCT scans taken in the inferior, compared to the middle or superior AC. Li and colleagues suggested there may be an unequal distribution of AC cells from the superior to inferior parts of the AC, and a poorer correlation when comparing the middle AC (captured by clinician grading) and inferior AC (captured by OCT) could be expected. They hypothesized that smaller and lighter cells may be carried by the aqueous circulation to superior parts of the AC, whereas larger and heavier cells in the AC may accumulate at the bottom.[17]

Acquisition time: Increased acquisition time may allow floating AC cells to move through the aqueous during successive raster scans resulting in over- or under-counting of cells. Newer OCT models with higher acquisition speeds are unlikely to be affected by this problem; however, time-domain OCT models and various other operator and patient factors (such as poor fixation, opacities, and reflections) may affect time required for scan acquisition.

Study (Grouped by Technology)	Clinical Grading System		Correlation (95% CI)	Sample Size
Laser Flare Cell Meter				
Ohara (1989)	Unspecified		0.66 (0.55, 0.75)*	127
Tugal-Tutkun (2008)	BenEzra		0.87 (0.79, 0.92)*	54
OCT				
Circumferential Scans				
Li [NG, Central] (2013)	Hogan (Modified)		0.73 (0.50, 0.86)	30
Li [NG, Superior] (2013)	Hogan (Modified)		0.75 (0.53, 0.87)	30
Li [NG, Inferior] (2013)	Hogan (Modified)		0.56 (0.25, 0.77)	30
Li [G, Central] (2013)	Hogan (Modified)		0.74 (0.29, 0.92)*	12
Li [G, Superior] (2013)	Hogan (Modified)		0.73 (0.27, 0.92)*	12
Li [G, Inferior] (2013)	Hogan (Modified)		0.06 (-0.53, 0.61)*	12
Line Scans				
Igbre (2014)	SUN		0.74 (0.61, 0.83)	69
Sharma (2015)	SUN		0.97 (0.95, 0.98)*	83
Invernizzi (2017)	SUN		0.94 (0.90, 0.96)	70
Volume Scans				
Sharma [Automated] (2015)	SUN		0.78 (0.59, 0.89)	31
Sharma [Manual] (2015)	SUN		0.75 (0.54, 0.87)	31

FIGURE 2. Forest plot of correlation coefficients reported by all included studies between index test measurements versus clinician grading, grouped by index test technology and clinician grading system.

Strengths and Limitations of the Review

The strength in this study lies in its systematic approach of reviewing all publications of instrument-based tool for AC cell counting with clinician slit-lamp based grading system. Our search strategy was designed to have high sensitivity for such studies and we searched a broad range of databases, including conference proceedings, dissertation databases, and the grey literature. Our limitations include the assumption that clinician grading, the current gold standard, is an appropriate reference standard for comparison. Our review cannot answer the question of whether an instrument-based measure is more accurate than clinician grading. However, other advantages are apparent, including capture of a larger area of AC and the ability to automate the cell counting process, whilst maintaining a good correlation with the clinician-based method.

Limitations Due to Gaps in the Evidence

First, due to the small number of included studies and heterogeneity in study design, it was not possible to provide pooled estimates of correlation coefficients. It was also not possible to make direct comparisons between OCT and LFCP due to the non-standardized use of comparators. Second, there would be value in evaluating the techniques across different subgroups to ensure generalisability (i.e., subgroup analysis by different etiological groups, between active and inactive disease and by age group and gender) but none of the current studies reported enough subgroup data. Third, imaging protocols for each study were variable. All studies for OCT acquired line scans, however total area of aqueous captured differed in each study. This might not have been an issue had the number of cells been reported per area/volume of aqueous. However, all studies reported absolute total number of cells observed. Future standardization of the output metrics generated, including cell count per unit of aqueous, is needed. This is essential for reliable comparison between devices such as monitoring a patient over time between different health settings, where multiple devices may be used.

Clinical Relevance and Impact

This review found that instrument-based tools can achieve high correlation with clinical grading. As discussed, earlier differences in design across studies preclude reliable head-to-head comparison of the two instrument-based techniques, but it is likely that OCT will become the dominant technology for cell counting as the LFCP models offering cell count have been discontinued after the FC-2000. In addition, OCT can be automated and performed without the

need for a skilled clinician. Implementing this technology in routine clinical care could potentially offer more quantitative, objective, long-term monitoring of anterior uveitis. These technologies could also permit task-shifting away from a small number of clinical experts to disease monitoring delivered by technicians. This also carries implications for future care delivery models, opening the possibility of remote monitoring and community-based care.

Future studies should consider more explicit reporting of patient, eye and ocular disease characteristics to permit meaningful comparison of methods and devices. Controlled studies, including healthy individuals recruited from the full age range will also be important to capture any non-pathological changes in the permeability of the blood-aqueous barrier, which develops with age. It will also be necessary for devices to demonstrate discriminant validity, correctly identifying AC cellular activity resulting from uveitis, from red blood cells or pigmented iris endothelial cells. Prospective longitudinal studies of patients with quiescent and active inflammation are needed to determine the minimum clinically important difference and inform consensus around diagnostic thresholds.[27]

Based on our review of the literature we would propose that key industry standards that need to be defined in order to support cross-device comparison include: (1) unit of measurement (e.g., cells per mm^3); (2) volume and location within the AC that is sampled; (3) clear reporting of any custom analysis software, including image pre-processing, thresholds set for identifying image features as cells (such as brightness of pixels), discarding of spurious findings and the degree of manual input required. In addition, all studies that seek to validate such techniques should report: (1) population characteristics (including disease etiology and distribution of disease severity within the cohort); (2) internal validity measures (such as test–retest reliability and inter/intra-rater reliability in the case of non-automated techniques), and (3) confidence intervals for all reported performance metrics.

CONCLUSION

Instrument-based technologies such as OCT and LCFP offer objectivity and automation to the assessment of AC cells in uveitis, and in a controlled setting can demonstrate high correlation with the current clinical standard. OCT is likely to become the dominant technology for cell counting and is suitable for the wide-scale deployment that would be necessary for it to become the new standard. However, before this is possible, there is a need for consensus around measurement standards for such instruments that would enable cross-device comparison to support reliable longitudinal measurement for patients in the real world.

DISCLOSURE OF INTEREST

AUTHOR CONTRIBUTIONS

XL lead reviewer, manuscript drafting, manuscript reviewing. ALS second reviewer, manuscript drafting, manuscript reviewing. LF data analysis, manuscript drafting, manuscript reviewing. SB, AK, TM, DA data collection, manuscript drafting, manuscript reviewing. TB, MR, JP manuscript reviewing. PAK, DJM, AKD project conception, supervision, manuscript reviewing.

SUPPLEMENTARY MATERIAL

Supplemental data for this article can be accessed on the publisher's website.

ORCID

Xiaoxuan Liu   http://orcid.org/0000-0002-1286-0038

REFERENCES

1. Durrani OM, Meads CA, Murray PI. Uveitis: A potentially blinding disease. *Ophthalmologica*. 2004;218:223–236. doi:10.1159/000078612.

2. Williams GJ, Brannan S, Forrester JV, et al. The prevalence of sight-threatening uveitis in Scotland. *Br J Ophthalmol*. 2007;91:33–36. doi:10.1136/bjo.2006.101386.

3. Rothova A, Suttorp-van Schulten MS, Frits Treffers W, Kijlstra A. Causes and frequency of blindness in patients with intraocular inflammatory disease. *Br J Ophthalmol*. 1996;80:332–336. doi:10.1136/bjo.80.4.332.

4. Jabs DA, Nussenblatt RB, Rosenbaum JT; Group, S. of U. N. (sun) W. Standardization of uveitis nomenclature for reporting clinical data. Results of the first international workshop. *Am J Ophthalmol*. 2005;140:509–516. doi:10.1016/j.ajo.2005.03.057.

5. Hogan MJ, Kimura SJ, Thygeson P. Signs and symptoms of uveitis. I. Anterior uveitis. *Am J Ophthalmol*. 1959;47:155–170. doi:10.1016/s0002-9394(14)76546-8.

6. Schlaegel T. *Essentials of Uveitis*. Tokyo, Japan: Little, Brown, Inc; 1967.

7. Nussenblatt RB, Whitcup SM. *Uveitis: Fundamentals and Clinical Practice 3rd*. Mosby; 2004. Available at: https://www.uk.elsevierhealth.com/uveitis-9781437706673.html#panel1.

8. Foster CS. (charles S. & Vitale, A. T). *Diagnosis and Treatment of Uveitis*. 900). New Dehli, India: W.B. Saunders; 2002.

9. BenEzra D, Forrester JV, Nussenblatt RB, Tabbara K, Timonen P. *Uveitis Scoring System*. Berlin Heidelberg: Springer; 1991. Available at: https://www.springer.com/gb/book/9783540549574.

10. Sawa M. Laser flare-cell photometer: principle and significance in clinical and basic ophthalmology. *Jpn J Ophthalmol*. 2017;61:21–42. doi:10.1007/s10384-016-0488-3.

11. Denniston AK, Keane PA, Srivastava SK. Biomarkers and surrogate endpoints in uveitis: the impact of quantitative imaging. *Invest Ophthalmol Vis Sci*. 2017;58:BIO131–BIO140. doi:10.1167/iovs.17-21788.

12. Holland GN, Eydelman MB, Cunningham B, Chambers WA. Food and drug administration procedures for new drug and device approvals. *Am J Ophthalmol*. 2016;161:1–3. doi:10.1016/j.ajo.2015.11.022.

13. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6: e1000097. doi:10.1371/journal.pmed.1000097.

14. Liu X, Moore DJ, Denniston AK Instrument-based tests for measuring anterior chamber (AC) cells in uveitis: a systematic review. *PROSPERO* (2018). Available at: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=84156. (Accessed: 16th February 2019)

15. Liu X, Solebo AL, Keane PA, Moore DJ, Denniston AK. Instrument-based tests for measuring anterior chamber cells in uveitis: a systematic review protocol. *Syst Rev*. 2019;8:30.

16. Invernizzi A, Marchi S, Aldigeri R, et al. Objective quantification of anterior chamber inflammation: measuring cells and flare by anterior segment optical coherence tomography. *Ophthalmology*. 2017;124:1670–1677. doi:10.1016/j.ophtha.2017.05.013.

17. Li Y, Lowder C, Zhang X, Huang D. Anterior chamber cell grading by optical coherence tomography anterior chamber cell grading by OCT. *Invest Ophthalmol Vis Sci*. 2013;54:258–265. doi:10.1167/iovs.12-10477.

18. Sharma S, Lowder CY, Vasanji A, Baynes K, Kaiser PK, Srivastava SK. Automated analysis of anterior chamber inflammation by spectral-domain optical coherence tomography. *Ophthalmology*. 2015;122:1464–1470. doi:10.1016/j.ophtha.2015.02.032.

19. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–536. doi:10.7326/0003-4819-155-8-201110180-00009.

20. Tugal-Tutkun I, Cingü K, Kir N, Yeniad B, Urgancioglu M, Gül A. Use of laser flare-cell photometry to quantify intraocular inflammation in patients with Beḥet uveitis. *Graefes Arch Clin Exp Ophthalmol*. 2008;246:1169–1177. doi:10.1007/s00417-008-0823-6.

21. Ohara K, Okubo A, Miyazawa A, et al. Aqueous flare and cell measurement using laser in endogenous uveitis patients. *Jpn J Ophthalmol*. 1989;33:265–270.

22. Igbre AO, Rico MC, Garg SJ. High-speed optical coherence tomography as a reliable adjuvant tool to grade ocular anterior chamber inflammation. *Retina*. 2014;34:504–508. doi:10.1097/IAE.0b013e31829f73bd.

23. Young B, O'Dowd G, Woodford P. *Wheater's Functional Histology: A Text and Colour Atlas*. Philipelphia, PA: Churchill Livingstone; 2013.

24. Kempen JH, Ganesh SK, Sangwan VS, Rathinam SR. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. *Am J Ophthalmol*. 2008;146:813–818. doi:10.1016/j.ajo.2008.06.004.

25. Wong IG, Nugent AK, Vargas-Martín F. The effect of biomicroscope illumination system on grading anterior chamber inflammation. *Am J Ophthalmol*. 2009;148:516–520.e2. doi:10.1016/j.ajo.2009.04.027.

26. Agarwal A, Ashokkumar D, Jacob S, Agarwal A, Saravanan Y. High-speed optical coherence tomography for imaging anterior chamber inflammatory reaction in uveitis: clinical correlation and grading. *Am J Ophthalmol*. 2009;147:413–416.e3. doi:10.1016/j.ajo.2008.09.024.

27. Reitsma JB, Rutjes AWS, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol*. 2009;62:797–806. doi:10.1016/j.jclinepi.2009.02.005.

Appendix

MEDLINE Sample Search Strategy

1	Exp Uveitis/
2	Uveiti*. Ti, ab.
3	1 or 2
4	Anterior chamber. Ti, ab.
5	Aqueous humour. Ti, ab.
6	Aqueous humor. Ti, ab.
7	4 or 5 or 6
8	Cell*. Ti, ab.
9	3 and 7 and 8

QUADAS-2 Risk of Bias Assessment

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Ohara, 1989	☺	?	?	☺	☺	☺	☺
Tugal-Tutkun, 2008	?	☺	?	☺	☹	☺	☺
Li, 2013	☺	☺	?	☺	☺	☺	☺
Igbre, 2014	☺	☺	?	☺	☺	☺	☺
Sharma, 2015	☺	☺	?	☺	☺	☺	☺
Invernizzi, 2017	☺	☺	?	☺	☺	☺	☺

☺ Low Risk ☹ High Risk ? Unclear Risk

Chapter 3: Instrument-based tests for measuring anterior chamber flare in uveitis

This chapter explores potential instrument-based technologies for measuring AC flare. It consists of a published protocol and systematic review which aimed to identify instruments used to measure AC flare in uveitis. This is the second systematic review in a series of three identifying potential technologies for measuring the key inflammatory variables, currently quantified using the SUN grading system. This review evaluates the level of correlation between the instrument's measurements with clinician assessment (through the use of grading systems such as the SUN grading system) and/or protein concentration measured through aqueous paracentesis. As with the AC cells systematic review, the instrument's reliability is also assessed.

The protocol for the systematic review was registered on PROSPERO (CRD42017084167) and is published at: *McNally and Liu, et al. (2019). Instrument-based tests for quantifying aqueous humour protein levels in uveitis: a systematic review protocol. Syst Rev 8, 287.*

The report for the systematic review is published at: *Liu et al. (2020). Non-invasive Instrument-Based Tests for Quantifying Anterior Chamber Flare in Uveitis: A Systematic Review. Ocular Immunology and Inflammation, 1–9.*

Systematic Reviews

PROTOCOL   **Open Access**

Instrument-based tests for quantifying aqueous humour protein levels in uveitis: a systematic review protocol

Thomas W. McNally[1†], Xiaoxuan Liu[1,2,3†] ![ORCID], Sophie Beese[4], Pearse A. Keane[3], David J. Moore[4] and Alastair K. Denniston[1,2,3,5*]

Abstract

Background: Inflammation in anterior uveitis is characterised by breakdown of the blood-ocular barrier, which allows leakage of blood constituents of higher molecular weight into the aqueous humour. In routine clinical care, increase in aqueous protein levels can be observed at the slit lamp as 'flare' and the severity can be graded using various clinical grading systems, of which the Standardization of Uveitis Nomenclature (SUN) grading system is most commonly used. Alternative instrument-based technologies are available, which can detect aqueous protein levels in an objective and quantifiable way. This review will identify instruments capable of measuring anterior chamber inflammation in this way, their level of reliability, and how well the measurements correlate with clinical grading and/or actual aqueous protein concentration.

Methods: Standard systematic review methodology will be used to identify, select and extract data from studies that report the use of any instrument-based technology in the assessment of aqueous protein levels. Searches will be conducted through bibliographic databases (MEDLINE, EMBASE and Cochrane Library), clinical trial registries and the grey literature. No restrictions will be placed on language or year of publication. The outcomes of interest are the level of correlation between identified instrument-based test measurements, clinical grading and/or actual aqueous protein concentration, as well as the reliability of each index test identified. Study quality assessment will be based on QUADAS2. Correlation and reliability outcomes will be pooled and meta-analysed if appropriate.

Discussion: The assessment of inflammation in anterior chamber protein levels currently relies on crude and subjective clinical examination. The findings of this review will identify non-invasive technologies which show good correlation with actual protein concentration, which could be used in routine clinical practice for objective monitoring of AC inflammation.

Systematic review registration: PROSPERO CRD42017084167. Study screening stage has just been completed.

Keywords: Systematic review, Uveitis, Anterior chamber flare, Intraocular inflammation, Aqueous humour, Proteins, Blood-ocular barrier, Monitoring test, Diagnostic test, Optical coherence tomography, Laser flare photometry, Laser flare meter, Tyndallometry

---

* Correspondence: a.denniston@bham.ac.uk
†Xiaoxuan Liu and Thomas W McNally contributed equally to this work.
[1]Academic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK
[2]Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
Full list of author information is available at the end of the article

Background

Uveitis, a significant cause of blindness worldwide with a global prevalence of 38–114.5 per 100,000 [1, 2], describes a group of conditions characterised by intraocular inflammation [3, 4]. Although uveitis can occur at any age [3], it commonly affects the working-age group and therefore has a substantial socioeconomic impact [5].

Anterior uveitis is characterised by inflammation of the anterior uveal tract. The acute symptoms are primarily pain, redness and photophobia, but vision can be affected. With adequate treatment, these symptoms are reversible; however, permanent and vision-threatening complications (such as posterior synechiae formation, cataract and secondary glaucoma) can develop in the presence of prolonged inflammatory activity.

Inflammation in the anterior chamber (AC) can be observed as changes of the aqueous humour. The aqueous humour is a clear fluid in the anterior chamber of the eye that provides nutrition, removes excretory products of metabolism, transports neurotransmitters, stabilises the ocular structure and contributes to the regulation of the homeostasis of ocular structures [6]. In health, the aqueous humour is well isolated from the blood by blood-aqueous barriers including the endothelial cells of the iris capillaries and retinal vessels and non-pigmented ciliary body epithelium. The aqueous is mostly water, but also contains electrolytes, carbohydrates, glutathione, urea, oxygen, carbon dioxide, and proteins [6, 7]. In the absence of inflammation, only small proteins are present in low concentrations in the aqueous. Inflammation causes breakdown of the blood-aqueous barriers and results in the release of inflammatory cells and proteins of higher molecular weight into the eye [8, 9]. The accumulation of proteins in the aqueous humour can be observed as flare or Tyndall effect, which is an optical phenomenon of cloudiness in the aqueous humour due to increased protein content.

The ability to accurately measure the concentration of proteins in the aqueous is clinically important for detecting acute inflammatory episodes and assessing response to treatment [10]. There are several established methods for quantifying aqueous protein levels, including clinical examination and laser flare photometry. However, with the exception of aqueous paracentesis (invasive sampling of the aqueous humour), existing methods are surrogate measures relating to the change in optical properties of the aqueous humour when protein levels are increased. Each method has its limitations, and it is unclear whether measurements from different methods agree with each other. The following section discusses the pros and cons of existing measures for aqueous protein concentration.

Invasive quantification of aqueous protein sampling: aqueous paracentesis

Currently, the only means of directly measuring the protein content of the aqueous is by aqueous paracentesis and analysis in a laboratory setting [11, 12]. While this method provides the most accurate quantification of aqueous constituents, it is not used routinely in clinical practice as it is invasive and carries risk of sight-threatening complications [13]. Consequently, its use is mainly limited to the research setting, and non-invasive methods are used preferably in clinical practice.

Non-invasive quantification of aqueous protein level
Clinical examination

The current clinical standard for AC flare grading was defined at the Standardisation of Uveitis Nomenclature (SUN) working group consensus meeting in 2005 [4]. Prior to this, several systems existed for grading AC flare [14–17]. Slit-lamp biomicroscopy enables a clinician to grade increasing protein concentrations in the aqueous humour due to back-scattering of light emitted from a slit-lamp. The SUN system grades flare according to the observer's ability to visualise details of the iris and lens behind the aqueous. Grades range from 0, which corresponds to no visible flare, to 4+, which corresponds to intense flare (Table 1).

Grading of flare based on clinical examination is subjective and subject to high intra- and inter-observer variability [18]. In addition, it is a non-continuous scale, leading to large steps in disease activity between categories. Detection and monitoring of inflammatory changes in anterior uveitis is critically important for clinical management, decision making and clinical trials investigating therapeutic agents [19]. The need for a more robust and reproducible method of measurement is well-recognised within the uveitis community. Newer imaging techniques have become available in recent years; therefore, it is timely to

Table 1 The Standardisation of Uveitis Nomenclature (SUN) working group grading scheme for anterior chamber flare [4]

Grade	Flare	Description
0	None	
1+	Faint	Barely detectable
2+	Moderate	Iris and lens details clear
3+	Marked	Iris and lens details hazy
4+	Intense	Fixed coagulated aqueous with considerable fibrin

Source: Ref [4]

McNally *et al. Systematic Reviews* (2019) 8:287

Page 3 of 8

carry out a review of whether they can be purposed for this unmet need and evaluate how they compare to the current standard.

Laser flare and cell photometry

In 1988, Sawa et al. first described laser flare photometry (LFP) as a new method to precisely determine protein concentration in the aqueous humour [20]. LFP is a rapid and non-invasive technique which quantifies AC flare by projecting a laser beam through the anterior chamber and measuring the amount of back-scattered light. Particles present in the aqueous reflect photons of light, which are then measured by an inbuilt photomultiplier. The intensity of back-scattered light is proportional to the concentration and size of proteins in the aqueous chamber. Most models of LFP are capable of measuring flare, but the technology can also be used to count AC cells. However, assessment of AC cells using LFP is less accurate as inflammatory cells cannot be differentiated from other large particles such as pigment particles, debris or red blood cells [9].

Optical coherence tomography (OCT)

OCT is an imaging technique which has become available more recently. It uses coherent light to capture high resolution 2- and 3-dimensional images of structures within the eye. It is a fast and non-invasive method of obtaining high precision images and measurements of ocular structure [21]. Anterior segment OCT (AS-OCT) can provide cross-sectional imaging of the AC. Recently, OCT has demonstrated the ability to detect higher light reflectivity signals on AS-OCT during active inflammation [21].

Aim

The primary aim of this review is to identify all available non-invasive, instrument-based techniques with the potential to quantify protein levels in the aqueous as a measure of AC inflammation. We will evaluate the level of correlation of these measurements with clinical grading using slit lamp examination and/or actual aqueous protein concentration. Where reported, we will also evaluate the level of reliability of each test.

A secondary aim of the review is to evaluate the correlation between clinical grading and protein concentration measured by laboratory analysis of aqueous samples.

The following questions are proposed:

- Which index tests have the potential to quantify AC inflammation in uveitis?

- What is the level of correlation between the index test and clinical grading?
- What is the level of correlation between the index test and aqueous protein concentration?
- What is the level of repeatability of each test?
- What is the level of correlation between clinical grading using the slit lamp and aqueous protein concentration?

Methods

Protocol

This protocol has been designed in accordance with the guidelines of Preferred Reporting Items for Systematic Review and Meta-Analysis Protocol (PRISMA-P) [22]. The systematic review will be reported in accordance to the PRISMA guidelines [23].

Systematic review registration

This systematic review has been registered on PROSPERO (CRD42017084167) [24].

Searches

Databases

Our search strategy will include the following areas of interest: anatomical location ('anterior chamber'), the pathological finding ('flare') and the disease context ('uveitis'). To achieve optimal sensitivity, no search terms will be applied for the 'technologies/tests'. For bibliographic databases, free text and index terms will be combined for each search element where possible. A sample search strategy for MEDLINE is included in the Appendix 1. We will search:

- Bibliographic databases of published studies.
  - MEDLINE (Ovid), 1946 to present
  - Embase (Ovid), 1947 to present
  - Cochrane Central Register of Controlled Trials (CENTRAL), database inception to present
  - Centre for Reviews and Dissemination database (Health Technology Assessments and the Database of Abstracts and Reviews of Effects), database inception to present
- Registers of clinical trials
  - Clinicaltrials.gov. www.clinicaltrials.gov
  - WHO International Clinical Trials Registry Platform (ICTRP portal). www.who.int/ictrp.
- Abstract and conference proceedings
  - British Library's ZETOC.
  - Conference proceedings Citation Index (Web of Science).
- Dissertations, theses
- British library Ethos
- ProQuest. www.proquest.com
- Grey literature

- OpenGrey. www.opengrey.eu

No restrictions will be placed on year or language of publication. The literature search results will be entered onto EndNote (Clarivate Analytics, Philadelphia, PA) to facilitate removal of duplicates, study selection, recording decisions and references. References to other works will be considered for inclusion.

Selection criteria

Participants/population
Those with evidence of anterior chamber flare and/or a diagnosis of uveitis, irrespective of active or inactive inflammation, will be included. There will be no restrictions on age, gender, ethnicity, or underlying aetiology or anatomical subtype. Studies with only healthy participants and studies on animals will not be included.

Index test
Any study reporting one or more non-invasive, instrument-based technology with the potential for quantifying AC flare or aqueous protein measurements will be included. To address the secondary outcome (clinical grading versus laboratory measurements of aqueous proteins), any clinical grading system can be considered the index test.

Reference test
The reference test should be either the current clinical standard—clinician-based AC flare grading, or the laboratory standard—measurements of aqueous protein from aqueous paracentesis.

For the clinical grading versus laboratory analysis, the reference test must be a laboratory-based technique for measuring aqueous protein concentration of extracted aqueous samples.

Primary outcomes
The level of correlation (correlation coefficient) of index test measurements with clinical grading using slit lamp examination and/or actual aqueous protein concentration.

Secondary outcome
Intra/inter-observer reliability and repeatability of an index test (kappa statistic) and the level of correlation between clinical grading and actual protein concentration (clinical grading as the index test and protein concentration as the reference test).

Type of study
There will be no restrictions on study design; however, evaluation of correlation between index test and clinical grading using slit-lamp examination/ aqueous protein measurements require both tests to be done in a cross-sectional manner. If multiple time points are included, we will include results from all time points, but comparisons between tests will be cross-sectionally. Only those studies where measurements are taken within a reasonable time point (within 24 h of each other) will be included. Case reports involving only one subject, commentaries, opinion articles and pictorial articles will not be included.

Selection process
Titles and abstracts of studies will be screened for relevance to the review, to remove obviously irrelevant studies Two independent reviewers will carry out study selection and reach consensus by discussion or referral to a third reviewer.

Full text of potentially relevant articles will be retrieved and assessed for inclusion in the review against the full selection criteria.

Data extraction
A standardised data extraction form in Microsoft Excel (Microsoft, Washington, US) will be used to extract data from included studies. The extraction process will be carried out by two independent reviewers with referral to a third reviewer if necessary. Information extracted from all studies will include:

- Study characteristics
  - Title, authors, publication year, journal and language
  - Sample size
  - Study design
  - Index test used
    1. Manufacturer and model (including resolution, default settings)
    2. Measurements acquisition protocol
    3. Measurement analysis protocol
  - Clinical grading system measurements
    1. Clinical grading system used and whether any modifications were made
    2. Number of observers
  - Aqueous protein concentration
    1. Context and reason for aqueous extraction
    2. Aqueous extraction protocol
    3. Aqueous analysis protocol
- Patients' characteristics
  - Age, gender and ethnicity
  - Underlying aetiologies (anatomical subtype, aetiological classification)
  - Active or inactive disease

McNally *et al. Systematic Reviews*   (2019) 8:287

Page 5 of 8

- If the study involves a therapeutic intervention: treatment details (indication, drug, dosage, route, subject pre/post treatment status, length of follow-up and number of time points suitable for review question)
- Outcomes and findings
  - Data will be extracted in preparation for four separate analyses:
    1. Evaluation of correlation between index tests and actual aqueous protein concentration: The correlation coefficient will be directly extracted.
    2. Evaluation of correlation between index tests and clinical grading systems: The correlation coefficient reported will be directly extracted.
    3. Evaluation of correlation between the clinical grading system and aqueous protein concentration: The correlation coefficient will be directly extracted.

    If no correlation coefficient is reported, Index and reference test measurements will be extracted and, provided they are matched, used to calculate the correlation coefficient. If the two measurements are not matched, we will contact the authors for matched measurements.

    4. Evaluation of reliability and repeatability of an index test: Studies reporting intra- and inter-observer reliability will be analysed separately for assessment of repeatability. The reported kappa values will be extracted for intra-observer reliability, inter-observer reliability or both.

Cross-sectional measurements may be nested within longitudinal studies whose aims are not primarily to compare performance or correlation between two tests. In this situation, measurements at each time point will be extracted and analysed as individual cross-sectional comparisons.

Quality assessment

All included studies will be assessed for quality using elements of the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS2) [25]. Although QUADAS2 is designed for comparison of diagnostic accuracy, rather than agreement of test measurements, it is the most suitable existing tool for evaluating risk of bias in diagnostic tests. We have modified elements of the existing QUADAS2 signalling questions to suit the aims of this systematic review (see Appendix 2). For example, the original QUADAS2 framework includes a question

regarding whether thresholds were used and pre-specified. We have removed this item as we are not aiming to assess diagnostic accuracy. Similarly, we have supported the QUADAS2 with additional items such as the addition of the question 'were index test acquisition and analysis parameters determined a priori and consistent for all study participants?' Existing questions regarding blinding during test interpretation, applicability to the review question, and flow and timing of tests are kept as they remain important for this review.

Assessment will be carried out at the study level. Two independent reviewers will carry out quality assessment and refer to a third reviewer if needed. For studies assessing correlation between two tests, the four risks of bias domains below will be rated as low, high or unclear.

For studies investigating the reliability of a single index test, additional considerations will be made for:

1. Intra-observer reliability studies: the conditions under which the index test was performed should be reported and standardised.
2. Inter-observer reliability studies: any differences between observer characteristics, such as seniority and experience, should be reported.

Data synthesis

Studies will be included in four groups for data synthesis, one for each outcome: correlation between index test and the clinical grading, correlation between index test and aqueous protein concentration, correlation between the SUN grading system and laboratory measurement of aqueous protein concentration, and reliability of index test.

For each outcome, a narrative synthesis of tabulated evidence will be conducted and, where possible, supported with a meta-analysis.

1. Evaluation of correlation between index test and the SUN grading system
   These studies will be grouped by the type of technology used for the index test (i.e. LFP, OCT). If the data permits, correlation coefficients between each index test versus clinical grade will be compared and pooled for meta-analysis using a random effects model.
2. Evaluation of correlation between index test and laboratory measurement of aqueous protein concentration
3. Evaluation of correlation between the SUN grading system and laboratory measurement of aqueous protein concentration
4. Evaluation of reliability of an index test

These studies will be grouped by type of technology as above. We will analyse intra- and inter-observer reliability separately for each technology. If the data permits, reliability values (such as kappa values) will be compared and pooled for meta-analysis.

Studies will be assessed for clinical and methodological homogeneity to determine whether it is possible appropriately pool data for meta-analysis. The I^2 statistic will be used to quantify heterogeneity across studies in each meta-analysis. Correlation co-efficients will be normalised using Fisher's Z transformation for meta-analysis and back transformed for inference. Meta-analysis of kappa statistics will be performed using an inverse-variance weighted random effects model with standard errors estimated from reported 95% confidence intervals. All data will be reported narratively, regardless of whether data pooling and meta-analysis are possible or not.

Depending on the study data, appropriate sensitivity analyses will be carried out if there is significant heterogeneity between studies. Subgroup analyses for anatomical and aetiological subtypes of uveitis, index test technology and experience of graders (for clinical AC flare grading) will be considered to explore sources of heterogeneity. It is expected that only a small number of studies will meet inclusion criteria, which will limit our capacity to carry out meaningful subgroup analyses.

Discussion and potential impact

The assessment of inflammation in anterior uveitis currently relies on imperfect clinical methods. An increase in aqueous protein concentration is a detectable change which occurs in the presence of inflammatory breakdown of the blood ocular barrier. Laboratory measurement of aqueous protein concentration in aqueous samples obtained through paracentesis is a good gold standard test but carries too many risks to make it practical for monitoring patients. Thus, clinicians must resort to surrogate markers for aqueous protein levels such as 'flare'.

Whilst the SUN grading system marked a significant effort towards unifying the method for assessing uveitic inflammation, it continues to rely on clinical examination and the subjective appearance of aqueous 'clarity'. Many factors can also affect an observer's ability to observe flare including the slit-lamp optics, degree of illumination, ambient conditions and the observer's level of expertise [26].

The potential for instrument-based techniques for assessing AC inflammation carries significant advantages over clinical grading systems because they are objective, less operator dependent and produce a continuous numerical value that is precise at even low levels of inflammation. Objective measures of inflammation for monitoring AC cells and vitreous haze have been proposed, and OCT measurement of central macular thickness is already established in routine clinical practice [19, 21, 27, 28]. With the use of automated analysis software, these techniques could allow comprehensive objective disease monitoring, potentially even in a virtual care model.

To date, the use of non-invasive instrument-based techniques for assessing AC flare has not been assessed in a systematic way. The need for more objective measures of disease activity is well-recognised within the uveitis community, and given the rise in new imaging technologies over recent years which have the potential to meet this demand, it is timely to carry out a review of how they compare to the current standard. This systematic review will identify all technologies available for quantifying 'flare' and their level of reliability and correlation with laboratory measurement of aqueous protein concentration and/or clinical grading. The findings of this review could contribute to the validation process of instrument-based methods for monitoring inflammatory activity and guiding treatment decisions in uveitis.

Appendix 1

Table 2 MEDLINE search strategy

Number	Search
1	Anterior Chamber. Ti,ab.
2	Aqueous. Ti,ab.
3	Anterior segment. Ti,ab.
4	1 or 2 or 3
5	Flare*. Ti,ab,
6	Photon. Ti,ab.
7	Photons. Ti,ab.
8	Protein. Ti,ab.
9	Proteins. Ti,ab.
10	Photometry. Ti,ab.
11	Fluorophotometry. Ti,ab.
12	Tyndal* Ti,ab.
13	5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14	Exp Uveitis/
15	Uveiti*. Ti,ab.
16	Inflamm*. Ti,ab.
17	Blood aqueous barrier. Ti,ab.
18	14 or 15 or 16 or 17
19	4 and 13 and 18

McNally *et al. Systematic Reviews*   (2019) 8:287

Page 7 of 8

Appendix 2

Table 3 Modified elements of the QUADAS2 signalling questions used for quality assessment

Domain	Patient selection	Index test*	Reference standard	Flow and timing
Description	Describe methods of patient selection: Describe included patients (prior testing, presentation, intended use of index test and setting):	Describe the index test and how it was conducted and interpreted:	Describe the reference standard and how it was conducted and interpreted:	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram): Describe the time interval and any interventions between index test(s) and reference standard:
Signalling questions (yes/no/unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided? Did the study avoid inappropriate exclusions? *i.e. Participants may be excluded if justified in terms of interference with index test measurement (corneal opacities preventing visualisation of anterior structures).*	*Were index test acquisition and analysis parameters determined a priori and consistent for all study participants?*	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard? Did all patients receive the same reference standard? *i.e. Was the reference test conducted in the same way for each patient?* *For slit lamp examination: observer, slit lamp settings* *For laboratory protein measurements: method of aqueous extraction, sample storage and analysis.* Were all patients included in the analysis?
Risk of bias: High/low/ unclear	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: High/low/ unclear	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Italics denote signalling questions added by the authors for this systematic review.
*In index test, the signalling question regarding whether thresholds were used and pre-specified is not applicable for this review and was therefore removed

Author details
[1]Academic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK. [2]Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [3]NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, Birmingham, UK. [4]Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK. [5]Centre for Rare

McNally *et al. Systematic Reviews*   (2019) 8:287

Page 8 of 8

Diseases, Institute of Translational Medicine, Birmingham Health Partners, Birmingham, UK.

Author details
[1]Academic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK. [2]Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [3]NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, Birmingham, UK. [4]Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK. [5]Centre for Rare Diseases, Institute of Translational Medicine, Birmingham Health Partners, Birmingham, UK.

References
1. Williams GJ, Brannan S, Forrester JV, Gavin MP, Paterson-Brown SP, Purdie AT, et al. The prevalence of sight-threatening uveitis in Scotland. Br J Ophthalmol. 2007;91:33–6.
2. Rothova A, Suttorp-van Schulten MS, Frits Treffers W, Kijlstra A. Causes and frequency of blindness in patients with intraocular inflammatory disease. Br J Ophthalmol. 1996;80:332–6.
3. Durrani OM, Meads CA, Murray PI. Uveitis: a potentially blinding disease. Ophthalmologica. 2004;218:223–36.
4. Jabs DA, Nussenblatt RB, Rosenbaum JT; Standardization of Uveitis Nomenclature (SUN) Working Group. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. Am J Ophthalmol. 2005;140:509–16.
5. Lardenoye CWTA, van Kooij B, Rothova A. Impact of macular edema on visual acuity in uveitis. Ophthalmology. 2006;113:1446–9.
6. Goel M, Picciani RG, Lee RK, Bhattacharya SK. Aqueous humor dynamics: a review. Open Ophthalmol J. 2010;4:52–9.
7. Murthy KR, Rajagopalan P, Pinto SM, Advani J, Murthy PR, Goel R, et al. Proteomics of human aqueous humor. OMICS. 2015;19:283–93.
8. Shah SM, Spalton DJ, Taylor JC. Correlations between laser flare measurements and anterior chamber protein concentrations. Investig Ophthalmol Vis Sci. 1992;33:2878–84.
9. Tugal-Tutkun I, Herbort CP. Laser flare photometry: a noninvasive, objective, and quantitative method to measure intraocular inflammation. Int Ophthalmol. 2010;30:453–64.
10. Tugal-Tutkun I, Cingu K, Kir N, Yeniad B, Urgancioglu M, Gul A. Use of laser flare-cell photometry to quantify intraocular inflammation in patients with Behcet uveitis. Graefes Arch Clin Exp Ophthalmol. 2008;246:1169–77.
11. Saari KM, Guillén-Monterrubio OM, Hartikainen J, Hämäläinen MM, Taskinen K. Measurement of protein concentration of aqueous humour in vivo: correlation between laser flare measurements and chemical protein determination. Acta Ophthalmol Scand. 1997;75:63–6.
12. Davis JL, Ruiz P Jr, Shah M, Mandelcorn ED. Evaluation of the reactive T-cell infiltrate in uveitis and intraocular lymphoma with flow cytometry of vitreous fluid (an American Ophthalmological Society thesis). Trans Am Ophthalmol Soc. 2012;110:117–29.
13. Cheung CMG, Durrani OM, Murray PI. The safety of anterior chamber paracentesis in patients with uveitis. Br J Ophthalmol. 2004;88:582–3.
14. Nussenblatt RB, Whitcup SM. Uveitis: fundamentals and clinical practice 3rd. Philadelphia: Mosby; 2004.
15. Hogan MJ, Kimura SJ, Thygeson P. Signs and symptoms of uveitis. Am J Ophthalmol. 1959;47:155–70.
16. Schlaegel T. Essentials of uveitis. Boston: Little, Brown, Inc; 1967.
17. Foster CS, Vitale AT. Diagnosis and treatment of uveitis. Philadelphia: WB Saunders; 2002.
18. Kempen JH, Ganesh SK, Sangwan VS, Rathinam SR. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. Am J Ophthalmol. 2008;146:813–8.
19. Denniston AK, Keane PA, Srivastava SK. Biomarkers and surrogate endpoints in uveitis: the impact of quantitative imaging. Invest Ophthalmol Vis Sci. 2017;58:BIO131–40.
20. Sawa M, Tsurimaki Y, Tsuru T, Shimizu H. New quantitative method to determine protein concentration and cell number in aqueous in vivo. Jpn J Ophthalmol. 1988;32:132–42.
21. Invernizzi A, Marchi S, Aldigeri R, Mastrofilippo V, Viscogliosi F, Soldani A, et al. Objective quantification of anterior chamber inflammation: measuring cells and flare by anterior segment optical coherence tomography. Ophthalmology. 2017;124:1670–7.
22. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1.
23. Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Int J Surg. 2010;8:336–41.
24. Liu X, McNally T, Beese S, Keane P, Moore D, Denniston AK. PROSPERO Instrument-based tests for measuring anterior chamber (AC) flare in uveitis: a systematic review. PROSPERO. 2018. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=84167.
25. Whiting PF, Rutjes AW, Westwood ME, Mallet S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529.
26. Wong IG, Nugent AK, Vargas-Martín F. The effect of biomicroscope illumination system on grading anterior chamber inflammation. Am J Ophthalmol. 2009;148:516–520.e2.
27. Sharma S, Lowder CY, Vasanji A, Baynes K, Kaiser PK, Srivastava SK. Automated analysis of anterior chamber inflammation by spectral-domain optical coherence tomography. Ophthalmology. 2015;122:1464–70.
28. Montesano G, Way CM, Ometto G, Ibrahim H, Jones PR, Carmichael R, et al. Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis. Sci Rep. 2018;8:1648.

Publisher's Note

Non-invasive Instrument-Based Tests for Quantifying Anterior Chamber Flare in Uveitis: A Systematic Review

Xiaoxuan Liu, Thomas W. McNally, Sophie Beese, Laura E. Downie, Ameenat L. Solebo, Livia Faes, Syed Husain, Pearse A. Keane, David J. Moore & Alastair K. Denniston

ORIGINAL ARTICLE

Check for updates

Non-invasive Instrument-Based Tests for Quantifying Anterior Chamber Flare in Uveitis: A Systematic Review

Xiaoxuan Liu, MBChB [a,b], Thomas W. McNally[b], Sophie Beese, MPH[c], Laura E. Downie, PhD [d], Ameenat L. Solebo, PhD[e], Livia Faes, MD[f,g], Syed Husain[b], Pearse A. Keane, MD[f,h], David J. Moore, PhD[c], and Alastair K. Denniston, PhD [a,b,f,h]

aOphthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; bAcademic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK; cInstitute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK; dDepartment of Optometry and Vision Sciences, The University of Melbourne, Parkville, Australia; eInstitute of Child Health, University College London, London, UK; fNIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; gEye Clinic, Cantonal Hospital of Lucerne, Lucerne, Switzerland; hHealth Data Research UK, London, UK

ABSTRACT

Purpose: Anterior chamber (AC) flare is a key sign for anterior uveitis. New instrument-based techniques for measuring AC flare can offer automation and objectivity. This review aims to identify objective instrument-based measures for AC flare.
Methods: In this systematic review, we identified studies reporting correlation between instrument-based tests versus clinician AC flare grading, and/or aqueous protein concentration, as well as test reliability.
Results: Four index tests were identified in 11 studies: laser-flare photometry (LFP), optical coherence tomography, ocular flare analysis meter (OFAM) and the double-pass technique. The correlation between LFP and clinician grading was 0.40–0.93 and 0.87–0.94 for LFP and protein concentration. The double-pass technique showed no correlation with clinician grading and insufficient information was available for OFAM.
Conclusion: LFP shows moderate to strong correlation with clinician grading and aqueous protein concentration. LFP could be a superior reference test compared to clinician AC flare grading for validating new index tests.

Anterior uveitis describes inflammatory-mediated breakdown of the blood-aqueous barrier with resultant leakage of blood constituents into the aqueous humor. Clinically this is characterized by anterior chamber (AC) cells and flare. AC flare is an important clinical marker of inflammation and has been shown to be the predominant sign in syndromes such as childhood chronic anterior uveitis.[1,2] To measure the true extent of blood-aqueous barrier breakdown requires sampling of the aqueous humor through paracentesis using a needle inserted into the AC, and measurement of the protein concentration in that sample. Whilst this invasive test provides the most accurate quantification of aqueous constituents, it is not feasible for repeated measurement in the context of disease monitoring. The more common approach is to observe this change using slit-lamp biomicroscopy, as 'flare', an appearance of haziness of the aqueous humor. Flare can be graded using semi-quantitative scales, of which the Standardization of Uveitis Nomenclature (SUN) grading system is most commonly used.[3] The SUN system measures flare according to the observer's ability to visualize details of the iris and lens behind the aqueous. Grades range from 0, which corresponds to no visible flare, to +4, which

corresponds to intense flare (Table 1). Although this method is subjective, quantifying aqueous inflammation this way is widely accepted as a clinical standard and is used to inform treatment decisions.[4,5] It is recognized that non-invasive and objective methods for measuring aqueous inflammatory change would significantly improve clinical assessment of anterior uveitis.

Instrument-based techniques, such as laser flare photometry (LFP) have been available for the last 20 years but have not been widely adopted. This is despite the evidence supporting the validity and clinical utility of LFP.[6–9] More recently, newer imaging techniques such as anterior segment optical coherence tomography (AS-OCT) have also demonstrated the potential to quantify AC flare.[10] Given the need for an objective non-invasive method for assessing aqueous inflammation, a systematic examination of the evidence for such technologies is timely.[11] This review aims to identify all instrument-based tools for measuring aqueous humor inflammation in uveitis and to evaluate their correlation with laboratory measurements of aqueous protein concentration and/or slit-lamp-based clinician grading systems.

Grade	Flare	Description
0	None	No alteration to iris and lens visualization
1+	Faint	Barely detectable
2+	Moderate	Iris and lens details clear despite discernible haze
3+	Marked	Iris and lens details hazy
4+	Intense	Fixed coagulated aqueous with considerable fibrin

Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop.[3]

Methods

This review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement.[12] The methodology was specified in advance iregistered with PROSPERO (CRD42017084167).[13,14] The primary aim of this review was to evaluate all non-invasive, instrument-based methods for measuring aqueous humor inflammation and their level of correlation with, 1) the gold standard reference test: analysis of protein concentration in aqueous samples, and/or 2) the clinical reference test: slit-lamp-based AC flare grading performed by a clinician. We accepted both as reference tests in recognition that aqueous paracentesis is rarely performed, and clinician grading is widely used as the basis for final clinical decision-making in practice. A secondary aim was to identify studies which also reported the reliability of index tests and compare the reliability of different tests.

Search strategy

We combined free text terms and index terms reflecting the pathological finding of interest, 'flare' or 'proteins' and 'anterior chamber' or 'aqueous humor', and the disease context 'uveitis' or 'inflammation' (search strategy available in Supplementary Materials). Database searches were carried out in MEDLINE, Embase, Cochrane Controlled Register of Trials (CENTRAL), Center for Reviews and Dissemination Database (Health Technology Assessments and the Database of Abstracts and Reviews of Effects), Clinicaltrials.gov, WHO International Clinical Trials Registry Platform (ICTRP portal), British Library's ZETOC, Conference Proceedings Citation Index (Web of Science), British Library Ethos, ProQuest and OpenGrey. We searched all databases from inception to 07 August 2019, with no date or language restrictions. We manually searched citations of review articles and included studies to identify additional relevant articles.

Study selection

Two reviewers independently assessed study eligibility, and disagreements were resolved through consensus, or referral to a third reviewer if needed. Studies were eligible if they described one or more instrument-based methods for measuring aqueous humor protein levels (index test) and compared its measurements to actual aqueous protein concentration and/or clinician grading (reference tests). We also included studies which additionally reported test reliability. The primary outcome of interest was the level of correlation between index tests and either of the two reference tests. The

secondary outcome was intra/inter-observer reliability of the index test. We did not exclude studies based on subject age, gender, ethnicity, underlying etiology or disease activity status. Animal studies and studies involving only healthy participants, single case reports, commentaries, and opinion articles were excluded.

Data extraction

Two reviewers independently extracted data using a pre-specified data extraction sheet. Two texts were translated from Chinese into English. We contacted two authors for further information and both responded.[15,16] If only individual patient data were reported, we used this information to calculate the correlation coefficient.

Risk of bias assessment

Two reviewers independently assessed the risk of bias, in studies comparing correlation between two tests, using a modified version of the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2).[17] We pre-specified adaptations to the original QUADAS-2 signaling questions to address the review question. For example, one signaling question was added in the index test section on whether index test protocols were determined *a priori* and standardized for all participants.

Data analysis

For each outcome, studies were grouped by the reference test against which comparisons were made: aqueous protein concentration or clinician grading. For each index test, we tabulated the extracted information and provided a narrative synthesis of methodological characteristics and index tests evaluated. Where confidence intervals for correlation coefficients were not reported, we estimated them using sample size and correlation coefficient and presented this on a forest plot. All statistical analyses were performed using Stata Statistical Software (Release 15. College Station, TX: StataCorp LP.)

Results

Results of the search

The study selection process is summarized in the PRISMA flow diagram (Figure 1). The search yielded 3741 unique bibliographic records after the removal of duplicates. Of these, 3631 were excluded based on screening of titles and abstracts. The large number of excluded articles is owing to the unrestrictive nature of our search strategy, which was deliberately designed to not include any index test terms, to ensure full capture of all relevant studies. The remaining 110 articles were obtained in full text for further scrutiny and a further 99 articles were excluded. The reasons for exclusion were due to not matching the criteria for outcome (n = 52), population (n = 24), study design (n = 19) and for the lack of an appropriate reference test (n = 4). Eleven unique studies met the eligibility criteria and were included (Tables 2–5).
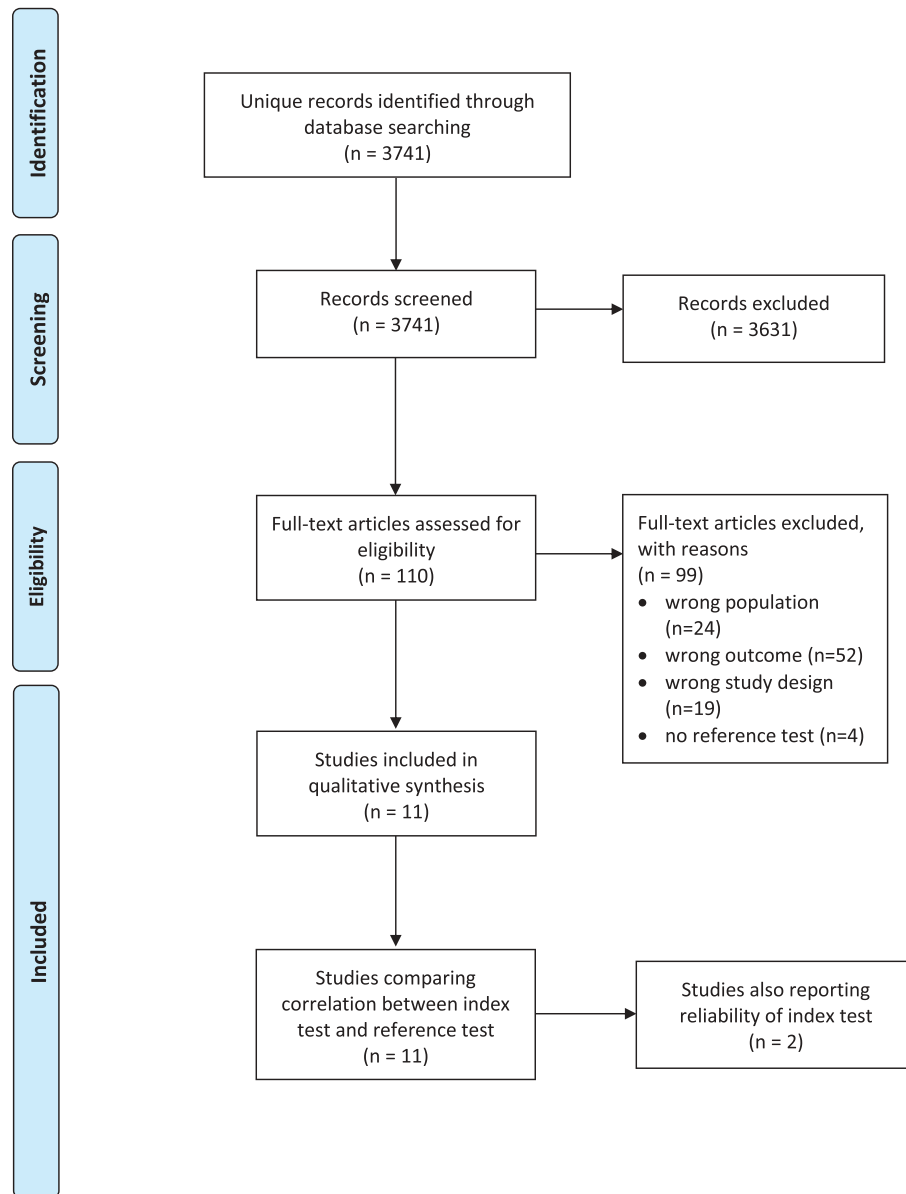
Figure 1. PRISMA flow diagram.

Participants' characteristics and study design features

The 11 studies included a total of 876 participants (at least 1016 eyes; one study did not report the number of eyes[18]), and dated from 1989 to 2017.[10,18–27] Study characteristics are summarized in Tables 2–5. One study used retrospectively collected routine clinical care data[20], whilst all other studies collected data prospectively. Gender was reported in 8 of the 11 studies, 44% of participants (n = 360) were male. Age was reported in 8 of the 11 studies[10,18,20–22,25–27], ranging from 12 to 86 years. Three studies included mixed etiologies, including sarcoidosis, Behcet's disease, Vogt–Koyanagi–Harada disease, acute retinal necrosis, lyme disease, progressive outer retinal necrosis, retinal vasculitis, herpes zoster ophthalmicus, and Fuch's heterochromic cyclitis (FHC).[18,19,24] Five studies did not report specific uveitis entities and instead reported anatomical classification or disease activity (active, inactive and healthy controls).[10,23,25–27] Two studies did

not report any etiological classifications.[21,22] One study included only eyes with FHC.[20]

Methodological quality of the included studies

A summary of the risk of bias assessment for the included studies is presented in Supplementary Figure 1. Eight out of eleven studies did not report how subjects were recruited into the study[19–22,24–27] and had an unclear risk of bias regarding participant selection. Given the known limitations of the clinician grading system, all studies which utilized clinician grading as the reference test (n = 8) were marked as unclear due to concerns around disease misclassification. It was unclear in five studies[19–22,25] whether the index test was interpreted without knowledge of the reference tests and vice versa. One study was

Table 2. Study characteristics.

Author	Index test	Data collection	No. of subjects	No. of eyes	Gender (%)	Age, years	Etiological classification, no. of eyes (%)
Ohara et al.[18]	LFP	Prospective	124	NR	44 (35%) male 80 (65%) female	NR (range 12–76)	Sarcoidosis 53 (43%), Behcet's disease 14 (11%), VKH 6 (5%), ARN 3 (2%), Other 14 (11%), Unknown 34 (27%)
Chiou et al.[19]	LFP	Prospective	17	17	NR	NR	ARN 5 (29%), Lyme disease 4 (23%), Progressive outer retinal necrosis 2 (12%), Anterior uveitis 1 (6%), Panuveitis 1 (6%), Retinal vasculitis 1 (6%), HZO 1 (6%), FHC 1 (6%), Behcet's disease 1 (6%)
Shah et al.[24]	LFP	Prospective	22	22	NR	Mean age: Normal 71 (SD 10) FHC 53 (SD 7) Uveitis 68 (SD 6)	FHC 5 (22.5%), Non-uveitic eyes undergoing routine cataract surgery 12 (55%), Chronic uveitis undergoing cataract/glaucoma surgery 5 (22.5%)
Yang et al.[26]	LFP	Prospective	162	Uveitis (194) Healthy eyes (52)	57 (52%) male 53 (48%) female	Mean 35 (range 3–66)	Anterior uveal inflammation 110 (68%), healthy controls 52 (32%)
Fang et al.[20]	LFP	Retrospective	40	47	15 (38%) male 25 (62%) female	Mean 34 (SD 10)	FHC 47 (100%)
Zhou et al.[27]	LFP	Prospective	129	171	68 (53%) male 61 (47%) female	Mean 42 (range 14–66)	Anterior uveitis 87 (51%), Intermediate uveitis 20 (12%), Posterior uveitis 64 (37%)
Konstantopoulou et al.[21]	LFP	Prospective	75	110	23 (31%) male 52 (69%) female	Median 42 (IQR 31–54)	NR
Shoughy et al.[25]	LFP	Prospective	20	20	13 (65%) male 7 (35%) female	Mean 52 (range 17–86)	Anterior non-granulomatous uveitis 5 (25%), VKH 4 (20%), FHC 1 (5%), Non-uveitis 10 (50%)
Invernizzi et al.[10]	LFP OCT	Prospective	122	237	102 (43%) male 135 (57%) female	Mean age: Control 42 (SD 14) Inactive uveitis 43 (SD 19) Active uveitis 45 (SD 22)	Healthy control 70 (30%), inactive uveitis 97 (40%), active uveitis 70 (30%)
Lam et al.[22]	OFAM	Prospective	121	90	38 (42%) male 52 (58%) female	Mean age: Active uveitis 44 No uveitis 46	NR
Nanavaty et al.[23]	Double-pass technique	Prospective	44	56	NR	NR	Anterior uveitis 24 (43%) Intermediate uveitis 9 (16%) Posterior uveitis 10 (18%) Panuveitis 13 (23%)

NR: not reported, OFAM: ocular flare analysis meter, OCT: optical coherence tomography, LFP: laser flare photometry, FHC: Fuch's heterochromic cyclitis. ARN: acute retinal necrosis, VKH: Vogt–Koyanagi–Harada disease, JIA: juvenile idiopathic arthritis, CMV: cytomegalovirus, HZO: herpes zoster ophthalmicus.

identified as having a high risk of bias for patient flow as readings were unsuccessful in 31 of the 121 included subjects.[22] Another study had high applicability concerns as the entire patient cohort was eyes with FHC.[20]

Clinical reference test: slit-lamp-based clinician grading

Nine out of 11 studies[10,18,20–23,25–27] compared an index test with AC flare grading based on clinician slit-lamp examination. Six studies[10,20–23,25] used the SUN grading system and three studies[18,26,27] did not specify a standardized grading system. Six studies reported the number of eyes at each AC flare grade.[10,21,22,25–27] Three of these studies included patients in all four grades of severity, two studies had eyes with each grade except grade 4, and one study included eyes with grades 0.5 and grade 1 of AC flare only. Three studies did not report the number of patients in each grade of AC flare.[18,22,23]

Laboratory reference test: aqueous protein concentration

Three out of 11 studies compared an index test with aqueous protein concentration.[19,24,25] Two studies[24,25] used paracentesis samples taken from individuals with uveitis prior to routine cataract surgery, one study[19] used diagnostic paracentesis samples for eyes with uveitis. Two studies[19,24] included IgG and albumin concentrations, one of which also measured total protein, and the other study[25] did not specify which proteins were measured.

Instruments for measuring AC cells

Four different classes of index tests that fit the description of non-invasive imaging techniques were identified. The majority of studies evaluated the use of LFP (9 studies). LFP devices included various models produced by KOWA, including the FC-1000, FC-2000, FC-500 and the only two models which are currently commercially available, the FM-600 and FM-700. Most studies reported taking between 3 and 7 repeated measurements at each

Table 3. Index test characteristics.

Author	Index test	Manufacturer and model	Image acquisition settings	Area/volume	Image analysis software
Ohara et al.[18]	LFP	Kowa FC-1000	An average of 5 readings taken through dilated pupil	Per 0.075mm³	Built in software only
Chiou et al.[19]	LFP	Kowa FC-1000	NR	Per 0.075mm³	Built in software only
Shah et al.[24]	LFP	Kowa FC-1000	5 averaged measurements Measurements where BG reading >15% was discarded,	0.3 x 0.5 mm	Built in software only
Yang et al.[26]	LFP	Kowa FC-2000	5 averaged measurements taken per eye	0.3 x 0.5 mm	Built in software only
Fang et al.[20]	LFP	Kowa FC-2000	An average of 3 measurements. Measurements with artifacts are discarded.	0.3 x 0.5 mm	Built in software only
Zhou et al.[27]	LFP	Kowa FM-600	7 consecutive measurements taken, highest and lowest values discarded, and remaining measurements averaged.	0.3 x 0.5 mm	Built in software only
Konstantopoulou et al.[21]	LFP	Kowa FC-500	7 measurements are acquired. The highest and lowest values are discarded, and remaining measurements averaged.	NR	Built in software only
Shoughy et al.[25]	LFP	Laser flare photometry (model NR)	NR	NR	NR
Invernizzi et al.[10]	LFP	Kowa FM-700	7 averaged consecutive measurements	0.3 x 0.5 mm	Built in software only
	OCT	Casia SS-1000 OCT device (Tomey Corporation, Nagoya, Japan)	Two 6mm line scans, each consisting of 2048 A scans.	200 x 200 pixel area	Custom software. A ratio of brightness value between aqueous and air anterior to the cornea is derived to produce an aqueous-to-air relative intensity.
Lam et al.[22]	OFAM	Custom OFAM device	Single measurement on an undilated eye	NR	Custom software
Nanavaty et al. [23]	Double-pass technique	The Optical Quality Analysis System (OQAS II; Visionmetrics, Terrassa, Spain)	3 averaged measurements taken from dilated pupils.	NR	Built in software only

NR: not reported, OFAM: ocular flare analysis meter, OCT: optical coherence tomography, LFP: laser flare photometry.

observation and taking the averaged value, as is the usual procedure according to LFP instructions. The sampling volume was reported as 0.075 mm³ for the FC-1000 (mean anterior chamber volume being approximately 145 mm³)[28] and sampling area was reported to be 0.3 mm by 0.5 mm for the FC-2000, FC-600 and FC-700. All measurements were derived using the built-in software of the LFP. One study evaluated a swept-source OCT device (Casia SS-1000, Tomey Corporation, Japan), taking two 6 mm cross-sectional scans in the anterior chamber. The AS-OCT images were then used to derive an image brightness ratio between the aqueous and air anterior to the cornea, using custom software, producing an 'aqueous-to-air' relative intensity (ARI).[10] One study used a custom-built ocular flare analysis meter (OFAM)[22] and one study used an optical quality analysis system (OQAS II, Visionmetrics, Terrassa, Spain) based on the double-pass technique, a technique measuring the amount of ocular scatter caused by the presence of flare.[23]

Index test reliability

Only two studies reported index test reliability. Invernizzi et al. reported an intraclass correlation of 0.78 for the OCT-derived ARI index measurement, and Shah et al. reported a coefficient of variation of 7.3% for the Kowa FC-1000 LFP. Nanavaty et al.

performed a reproducibility study, however, these were on healthy pseudophakic eyes, rather than uveitis eyes.

Correlation between index tests and the clinical reference test: slit-lamp-based clinician grading

Six studies reported correlation between an index test and clinician grading of AC flare (five studies using the LFP and one study using the optical quality analysis system).[18,20,23,25–27] The total number of eyes included in these six studies was 478. Various statistical methods including Kendall's, Spearman's and Pearson's correlation coefficients were used. The level of correlation between the LFP and clinician grading ranged from 0.40 to 0.93. The one study using the optical quality analysis system reported a Pearson's r2 of 0.0048.[23] Although the OFAM and OCT devices were compared against SUN grading, no correlation coefficient was reported.

Correlation between index tests versus the laboratory reference test: aqueous protein concentration

Three studies reported correlation between an index test and aqueous protein concentration, all of which used LFP as the index test.[19,24,25] Shah et al. and Shoughy et al.

Table 4. Index test versus reference tests.

Author	Index test	Clinical grading system	Manufacturer and model	Observers	No. of eyes in each clinical grade	No. of eyes included in correlation analysis	Statistical test	Correlation Coefficient (95% CI)
Ohara et al.[18]	LFP	NR	Kowa FC-1000	One observer for all clinical assessments	NR	127[a]	Spearman r	0.76
Yang et al.[26]	LFP	NR	Kowa FC-2000	NR	Grade 0, 74; Grade 1, 98; Grade 2 18; Grade 3, 2; Grade 4, 2	194	Spearman r	0.75
Fang et al.[20]	LFP	SUN	Kowa FC-2000	One observer for all clinical assessments	Only grade 0.5 and grade 1 eyes were included	47	Kendall r	0.40
Zhou et al.[27]	LFP	NR	Kowa FM-600	NR	**Anterior uveitis** Grade 1, 48; Grade 2, 35; Grade 3, 2; Grade 4, 2 **Intermediate uveitis** Grade 1, 10; Grade 2, 10	87 (anterior uveitis) 20 (intermediate uveitis)	Pearson r	0.86 (anterior) 0.87 (intermediate group)
Konstantopoulou et al.[21]	LFP	SUN	Kowa FC-500	Two observers independently performed clinical grading	Grade 0, 5; Grade 1, 74; Grade 2, 29; Grade 3, 2; Grade 4, 0	110	NR	NR
Shoughy et al.[25]	LFP	SUN	Laser flare photometry (model NR)	Two observers independently performed clinical grading	Grade 0, 5; Grade 1, 3; Grade 2, 1; Grade 3, 1; Grade 4, 0	10	Spearman r	0.93
Invernizzi et al.[10]	OCT	SUN	Casia SS-1000 OCT device	NR	Grade 0, 32; Grade 1, 21; Grade 2, 15; Grade 3, 2	70	NR	NR
	LFP	SUN	Kowa FM700	NR	Grade 0, 32; Grade 1, 21; Grade 2, 15; Grade 3, 2	70	NR	nr
Lam et al.[22]	OFAM	SUN	Custom OFAM device	One observer for all clinical assessments	NR	NR	NR	NR
Nanavaty et al.[23]	Double-pass technique	SUN	The Optical Quality Analysis System	Two independent observers for clinical assessment Double-pass technique by a technician who was blinded to clinical assessment	NR	56	Pearson r^2	0.0048

NR: not reported, OFAM: ocular flare analysis meter, OCT: optical coherence tomography, LFP: laser flare photometry, ARI: aqueous-to-air relative intensity [ARI] index, CC: correlation coefficient.
[a]Including repeat visits of same eye.

Table 5. Index test versus aqueous protein concentration.

Author	Index test	Protein concentration test	Proteins measured	Protein concentration range	No. of eyes included in correlation analysis	Statistical test	Correlation coefficient (95% CI)
Chiou et al.[19]	Kowa FC-1000 LFCP	Diagnostic paracentesis	IgG Albumin	NR	17 (IgG) 10 (albumin)	NR	0.87 (IgG) 0.94 (albumin)
Shah et al.[24]	Kowa FC-1000 LFCP	Paracentesis before routine cataract surgery	Total protein Albumin IgG	Normal IgG 0.3–4 mg/dl Normal Alb 3.1–14 mg/dl Normal Total protein 14–45 mg/dl FHC IgG <2 - 8 mg/dl FHC Alb 6–36 mg/dl FHCs Total protein 14–51 mg/dl Uveitis IgG 6-50mg/dl Uveitis Alb 48–290 mg/dl Uveitis Total protein 62–388 mg/dl	22	Pearson r	0.90 (albumin) 0.88 (IgG)
Shoughy et al.[25]	LFP (Model NR)	Paracentesis before routine cataract surgery	NR	Normal 10–35.48 mg/dl Anterior non-granulomatous uveitis, 10–1490 mg/dl FHC, 23.95 mg/dl[a] VKH, 30–600 mg/dl	10	Pearson r	0.99

NR: not reported, LFP: laser flare photometry, CC: correlation coefficient, FHC: Fuch's heterochromic cyclitis, VKH: Vogt–Koyanagi–Harada syndrome.
[a]One subject only.

included non-uveitic eyes in the correlation analysis, whereas Chiou et al. included eyes with endogenous uveitis only. The total number of eyes included across all three studies was 59. One study[24] calculated a Pearson's r and another study[19] did not report which statistical test was used. In the last study, individual patient data was reported so we derived the correlation between the two tests for uveitic eyes using Pearson's r.[25] The level of correlation
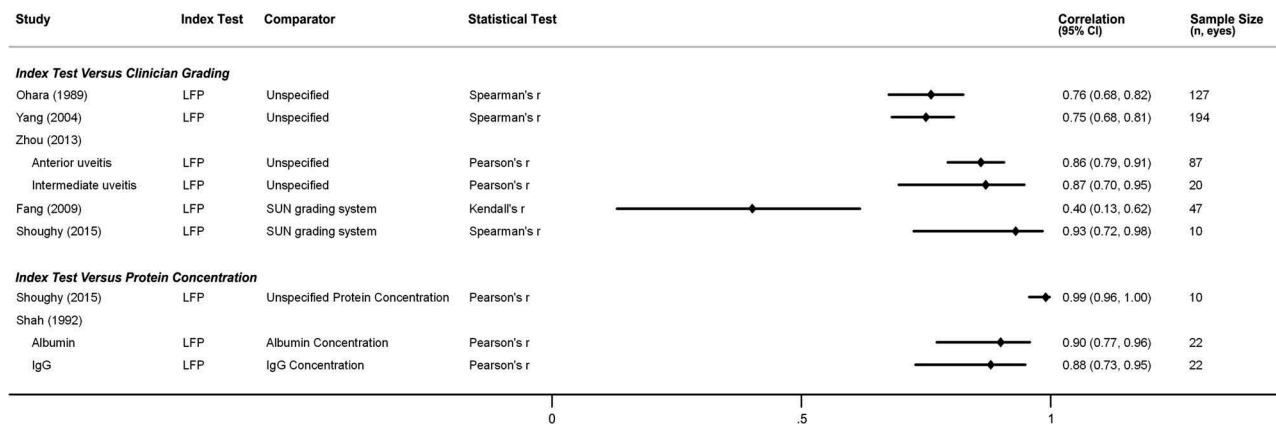
Figure 2. Level of correlation between LFP and clinician grading and aqueous protein concentration.

between index test measurements and protein concentrations ranged from 0.87 to 0.99. From the limited data, there were no apparent associations between the type of protein and the level of correlation.

The forest plot showing correlation between LFP and the two reference tests: clinician grading and aqueous protein concentration is shown in Figure 2. None of the included studies reported confidence intervals to correlation coefficients and those shown in the forest plot were estimated using sample size and correlation coefficient.

Study heterogeneity

There was considerable heterogeneity between the methodology and populations across the studies. These characteristics were wide ranging particularly in regard to the various device models used, the distribution of disease subtype and severity, and the statistical tests used to calculate correlation. Given this level of heterogeneity, any meta-analysis of correlation coefficients would be inappropriate.

Discussion

This is the first systematic review to evaluate all non-invasive instrument-based tests for quantifying aqueous humor inflammation. We found four non-invasive index tests: OCT, LFP, OFAM, and the double-pass technique (using the OQAS). Of all the index tests, we found LFP to have the strongest evidence base, with good correlation with clinician grading and even better correlation with aqueous sample protein concentration. However, only a small number of studies provided sufficient information to support this finding and incomplete reporting and inconsistent methodology of included studies meant we were unable to pool estimates of correlation between index and reference tests.

Our review found a strong correlation between the LFP and clinician grading of AC flare, as well as aqueous protein concentration, in most studies. There was no apparent relationship between the device model and level of correlation, and the current available model (FM-600), validated in one study, was reported to have good correlation ($r = 0.86$–0.87) with clinician grading. An interesting finding was that LFP measurements showed a stronger correlation with protein concentration (r= 0.87–0.99) than with clinician grading ($r = 0.40$–0.93). If aqueous paracentesis was assumed to be the true gold standard, then this finding would suggest LFP is a more accurate marker of aqueous protein levels than clinician grading. Shoughy et al. was the only study which reported all three methods at the individual eye level, allowing us to calculate correlation between all three methods. Their measurements showed a higher correlation between aqueous protein concentration and LFP ($r = 0.99$) than aqueous protein concentration and clinician grading ($r = 0.93$); however, this is based on only 10 eyes (5 of which were grade 0 by SUN grading).

The OFAM is a newer device utilizing the Rayleigh scattering effect, which the authors suggest to have a higher sensitivity to smaller molecules than the Tyndall effect used by LFP. Although the authors report significant differences in OFAM measurement in eyes with grade 1 and 2, when compared with grade 0, the device could not differentiate between grades 1 and 2. The double-pass technique using the OQAS showed poor correlation with AC flare grading, but the authors reported a significant correlation with AC cells ($r = 0.87$) and significantly more ocular scatter in eyes with anterior uveitis than intermediate and posterior subtypes.

Invernizzi et al. reported the only evaluation of OCT for quantifying AC flare. Although they did not report a correlation coefficient between OCT or LFP compared against clinician grading, they showed that OCT-derived ARI index significantly increased with each grade of AC flare. Similarly, LFP readings in their study significantly increased with each grade of AC flare, with the exception of grade 0 to grade 1. When comparing the two index tests, LFP and ARI index, they found a moderate correlation ($r = 0.61$). OCT has the added advantage of sampling a larger theoretical volume of AC compared to LFP and is fast and convenient to acquire. Additionally, the counting of AC cells using anterior segment OCT has been described by Invernizzi et al. and others and has been shown to be an automatable process. Therefore, OCT has the added advantage of offering a comprehensive all-in-one multi-faceted assessment of AC inflammation.[29]

Strengths and limitations of the review

This review represents the first systematic evaluation of technologies for measuring AC flare or aqueous inflammation.

Previous reviews have summarized the level of validation for LFP; however, this is the first review to consider all technologies, including newer imaging modalities such as OCT. Our search strategy was designed to be highly sensitive and included a broad range of databases, including conference proceedings, dissertation databases, and the gray literature. However, our review also has a number of limitations. We only included studies reporting correlation between tests or reliability of tests; therefore, any other methods of test comparison (such as those demonstrating a significant difference in index test measurements across SUN grades) were not included. This is because our original intention was to choose a commonly reported metric which would enable comparison between index tests. Whilst correlation is useful, it is limited to demonstrating agreement and non-inferiority to the comparator. From correlation, it is not possible to determine if the index test is more accurate than the reference test. This is an important consideration when the clinician grading system is the reference test, where the index test may in fact provide a more accurate measurement.

Limitations of the evidence

First, due to the small number of included studies and heterogeneity in study design, it was not possible to provide pooled estimates of correlation or reliability. Second, authors sometimes reported correlation coefficients estimated from a mixed cohort of uveitic and healthy eyes. Where the two groups could be separated, or where individual patient data was available, we extracted and calculated correlation coefficients from uveitic eyes only. However, this was not possible in all studies and non-uveitic eyes were included in the original analysis as grade 0. Third, in two studies, correlation was derived from eyes with two clinical grades of AC flare only (i.e. SUN grade 1 and grade 2 only). In these cases, there is an applicability concern around whether the study population adequately represents the target population and the correlation coefficient is also less meaningful.

Clinical relevance and impact

Anterior chamber flare is an important measure in the assessment of uveitic inflammation. This review finds that instrument-based tools such as the LFP can achieve good agreement with widely accepted clinical and laboratory reference tests and may be a more accurate marker of true aqueous protein concentration than clinical grading. Despite being available for some time, the LFP has not been widely adopted in clinical care. This may be due to practical reasons, as using the LFP involves taking up to seven readings with a degree of manual input, including discarding spurious readings and computing an average value. Additionally, the LFP is sensitive to ambient lighting and therefore requires a completely dark room, and acquisition can be more difficult for eyes with posterior synechiae and a shallow AC. Such practical considerations may outweigh the added clinical value of such a device. However, we would argue that in a clinical trial context, accuracy and reliability should take precedence and the LFP could be adopted in place of AC flare grading. Unfortunately, despite

its availability for the last 20 years, LFP has not been widely adopted as the preferred test for quantifying AC flare as an outcome measure in clinical trials.[30] Other available instrument-based methods for quantifying AC cells, vitreous haze, retinal vascular leakage and choroidal thickness, such as through the use of OCT, have also not been adopted in clinical trials.[31]

We would also suggest that for the validation of newer imaging methods, the LFP may be a more appropriate reference test in comparison to clinician grading, when aqueous protein concentration is not available. However, this is based on findings from a small number of eyes in a single study. We would suggest that a well-conducted study on a larger cohort of eyes, comparing AC flare grading using the slit-lamp, aqueous paracentesis and LFP, would be helpful for confirming whether LFP can replace clinician grading as the clinical reference test.

Conclusion

Instrument-based tests have the potential to offer more objectivity to measurements of AC inflammation compared to the widely used clinician grading using slit-lamp. The validation of LFP is the most mature for this purpose and has shown a strong correlation with clinician grading of AC flare and more importantly with aqueous protein concentrations. Although LFP may not be widely adopted in clinical practice, it may have value as a non-invasive reference test with which to validate emerging technologies for measuring AC flare.

Author contributions

XL: lead reviewer, manuscript drafting, manuscript reviewing. TWM: second reviewer, data extraction manuscript drafting, manuscript reviewing. SB: data extraction, manuscript drafting, manuscript reviewing. LF, LED, ALS, SH: manuscript drafting, manuscript reviewing. PAK, DJM, AKD project conception, supervision, manuscript reviewing.

ORCID

Xiaoxuan Liu 🔟 http://orcid.org/0000-0002-1286-0038
Laura E. Downie 🔟 http://orcid.org/0000-0002-1596-2259
Alastair K. Denniston 🔟 http://orcid.org/0000-0001-7849-0087

References

1. Holland GN. A reconsideration of anterior chamber flare and its clinical relevance for children with chronic anterior uveitis (an

American Ophthalmological Society thesis). *Trans Am Ophthalmol Soc.* 2007;105:344.

2. Davis JL, Dacanay LM, Holland GN, Berrocal AM, Giese MJ, Feuer WJ. Laser flare photometry and complications of chronic uveitis in children. *Am J Ophthalmol.* 2003;135(6):763–771. doi:10.1016/S0002-9394(03)00315-5.

3. Jabs DA, Nussenblatt RB, Rosenbaum JT. Group S of UN (sun) W. Standardization of uveitis nomenclature for reporting clinical data. Results of the first international workshop. *Am J Ophthalmol.* 2005;140:509–516.

4. Agrawal R, Keane PA, Singh J, Saihan Z, Kontos A, Pavesio CE. Comparative analysis of anterior chamber flare grading between clinicians with different levels of experience and semi-automated laser flare photometry. *Ocul Immunol Inflamm.* 2016;24 (2):184–193. doi:10.3109/09273948.2014.990042.

5. Kempen JH, Ganesh SK, Sangwan VS, Rathinam SR. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. *Am J Ophthalmol.* 2008;146 (6):813–818. doi:10.1016/j.ajo.2008.06.004.

6. Herbort CP, Tugal-Tutkun I. The importance of quantitative measurement methods for uveitis: laser flare photometry endorsed in Europe while neglected in Japan where the technology measuring quantitatively intraocular inflammation was developed. *Int Ophthalmol.* 2017;37 (3):469–473. doi:10.1007/s10792-016-0253-0.

7. Sawa M. Laser flare-cell photometer: principle and significance in clinical and basic ophthalmology. *Jpn J Ophthalmol.* 2017;61 (1):21–42. doi:10.1007/s10384-016-0488-3.

8. Tappeiner C, Heinz C, Roesel M, Heiligenhaus A. Elevated laser flare values correlate with complicated course of anterior uveitis in patients with juvenile idiopathic arthritis. *Acta Ophthalmol.* 2011;89(6):e521-e527. doi:10.1111/j.1755-3768.2011.02162.x.

9. Bernasconi O, Papadia M, Herbort CP. Sensitivity of laser flare photometry compared to slit-lamp cell evaluation in monitoring anterior chamber inflammation in uveitis. *Int Ophthalmol.* 2010;30(5):495–500. doi:10.1007/s10792-010-9386-8.

10. Invernizzi A, Marchi S, Aldigeri R, et al. Objective Quantification of Anterior Chamber Inflammation: measuring cells and flare by anterior segment optical coherence tomography. *Ophthalmology.* 2017;124(11):1670–1677. doi:10.1016/j.ophtha.2017.05.013.

11. Denniston AK, Keane PA, Srivastava SK. Biomarkers and surrogate endpoints in uveitis: the impact of quantitative imaging. *Invest Ophthalmol Vis Sci.* 58;6:BIO131–BIO140. doi:10.1167/iovs.17-21788.

12. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg.* 2010;8 (5):336–341. doi:10.1016/j.ijsu.2010.02.007.

13. Liu X, Moore D, Denniston AK. PROSPERO instrument-based tests for measuring anterior chamber (AC) flare in uveitis: a systematic review. PROSPERO. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=84167

14. McNally TW, Liu X, Beese S, Keane PA, Moore DJ, Denniston AK. Instrument-based tests for quantifying aqueous humour protein levels in uveitis: a systematic review protocol. *Syst Rev.* 2019;8 (1):287.

15. Oshika T, Nishi M, Mochizuki M, et al. Quantitative assessment of aqueous flare and cells in uveitis. *Jpn J Ophthalmol.* 1989;33 (3):279–287.

16. Tugal-Tutkun I, Cingü K, Kir N, Yeniad B, Urgancioglu M, Gül A. Use of laser flare-cell photometry to quantify intraocular inflammation in patients with Beḥet uveitis. *Graefes Arch Clin Exp Ophthalmol.* 2008;246(8):1169–1177. doi:10.1007/s00417-008-0823-6.

17. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009.

18. Ohara K, Okubo A, Miyazawa A, Miyamoto T, Sasaki H, Oshima F. Aqueous flare and cell measurement using laser in endogenous uveitis patients. *Jpn J Ophthalmol.* 1989;33:265–270.

19. Chiou AG, Florakis GJ, Herbort CP. Correlation between anterior chamber IgG/albumin concentrations and laser flare photometry in eyes with endogenous uveitis. *Ophthalmologica.* 1998;212 (4):275–277. doi:10.1159/000027306.

20. Fang W, Zhou H, Yang P, Huang X, Wang L, Kijlstra A. Aqueous flare and cells in Fuchs syndrome. *Eye.* 2009;23(1):79–84. doi:10.1038/sj.eye.6702991.

21. Konstantopoulou K. Del'Omo R, Morley AM, et al. A comparative study between clinical grading of anterior chamber flare and flare reading using the Kowa laser flare meter. *Int Ophthalmol.* 2015;35 (5):629–633. doi:10.1007/s10792-012-9616-3.

22. Lam DL, Axtelle J, Rath S, et al. A rayleigh scatter-based ocular flare analysis meter for flare photometry of the anterior chamber. *Transl Vis Sci Technol.* 2015;4(6):7. doi:10.1167/tvst.4.6.7.

23. Nanavaty MA, Stanford MR, Sharma R, Dhital A, Spalton DJ, Marshall J. Use of the double-pass technique to quantify ocular scatter in patients with uveitis: A pilot study. *Ophthalmologica.* 2010;225(1):61–66. doi:10.1159/000316690.

24. Shah SM, Spalton DJ, Taylor JC, et al. Correlations between laser flare measurements and anterior chamber protein concentrations. *Invest Ophthalmol Vis Sci.* 1992;33(10):2878–2884.

25. Shoughy SS, Elkum N, Tabbara KF. Aqueous protein level and flare grading. *Acta Ophthalmol.* 2015;93(2):e173-e174. doi:10.1111/aos.12522.

26. Yang P-Z, Wang H, Huang X-K, et al. [Quantitative determination of aqueous flare and cells in the eyes of patients with inflammation of anterior uvea]. *Zhonghua Yan Ke Za Zhi.* 2004;40 (8):510–513.

27. Zhou Y, Jing X-H, Li X-Y. Application value of laser flare cell meter in uveitis. *Guoji Yanke Zazhi.* 2013;13(9):1775–1777.

28. Wang N, Wang B, Zhai G, Lei K, Wang L, Congdon N. A method of measuring anterior chamber volume using the anterior segment optical coherence tomographer and specialized software. *Am J Ophthalmol.* 2007;143(5):879–881. doi:10.1016/j.ajo.2006.11.051.

29. Liu X, Solebo AL, Faes L, et al. Instrument-based tests for measuring anterior chamber cells in uveitis: a systematic review. *Ocul Immunol Inflamm.* 2019;1–12. doi:10.1080/09273948.2019.1640883.

30. Denniston AK, Holland GN, Kidess A, et al. Heterogeneity of primary outcome measures used in clinical trials of treatments for intermediate, posterior, and panuveitis. *Orphanet J Rare Dis.* 2015;10(1):97. doi:10.1186/s13023-015-0318-6.

31. Herbort CP, Tugal-Tutkun I, Neri P, Pavésio C, Onal S, LeHoang P. Failure to integrate quantitative measurement methods of ocular inflammation hampers clinical practice and trials on new therapies for posterior uveitis. *J Ocul Pharmacol Ther.* 2017;33(4):263–277. doi:10.1089/jop.2016.0089.

Page left intentionally blank

Page intentionally left blank

Appendix

QUADAS-2 Risk of Bias Assessment

Risk of Bias assessment using QUADAS-2

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Chiou, 1998	?	?	?	?	☺	☺	☺
Ohara, 1989	☺	☺	☺	☺	☺	☺	?
Shah, 1992	?	☺	☺	☺	☺	☺	☺
Yang, 2004	?	☺	☺	☺	☺	☺	?
Fang, 2009	?	☺	?	?	☹	☺	?
Nanavaty, 2011	☺	☺	☺	☺	☺	☺	?
Zhou, 2013	?	?	☺	☺	?	☺	?
Lam, 2015	?	?	?	☹	☺	☺	?
Shoughy, 2015	?	?	?	?	☺	☺	☺
Konstantopolou, 2015	?	☺	?	☺	☺	☺	?
Invernizzi, 2017	☺	?	☺	☺	☺	☺	?

☺ Low Risk ☹ High Risk ? Unclear Risk

MEDLINE Sample Search Strategy

1	Anterior chamber.ti,ab.
2	aqueous.ti,ab.
3	Anterior segment.ti,ab.
4	1 or 2 or 3
5	Flare*.ti,ab.
6	photon.ti,ab.
7	photons.ti,ab.
8	protein.ti,ab.
9	proteins.ti,ab.
10	photometry.ti,ab.
11	fluorophotometry.ti,ab.
12	5 or 6 or 7 or 8 or 9 or 10 or 11
13	Exp Uveitis/
14	Uveiti*.ti,ab.
15	Inflamm*.ti,ab.
16	Blood aqueous barrier.ti,ab.
17	13 or 14 or 15 or 16
18	4 and 12 and 17

Chapter 4: Instrument-based tests for measuring vitreous inflammation in uveitis

This chapter explores potential instrument-based technologies for measuring vitreous haze. It consists of a published protocol and systematic review which aimed to identify instruments used to measure vitreous inflammation in uveitis. This is the last systematic review in a series of three identifying potential technologies for measuring the key inflammatory variables, currently quantified using the SUN grading system. This review evaluates the level of correlation between the instrument's measurements and clinician assessment (through the use of grading systems such as the SUN grading system). As with the other reviews in this series, the instrument's reliability is also assessed.

The protocol for the systematic review was registered on PROSPERO (CRD42017084168).

The report for the systematic review is published at: *Liu, et al. (2020). Noninvasive Instrument-based Tests for Detecting and Measuring Vitreous Inflammation in Uveitis: A Systematic Review. Ocular Immunology and Inflammation, 1–12.*

Noninvasive Instrument-based Tests for Detecting and Measuring Vitreous Inflammation in Uveitis: A Systematic Review

Xiaoxuan Liu , Benjamin TK Hui , Christopher Way , Sophie Beese , Ada Adriano , Pearse A Keane , David J Moore & Alastair K Denniston

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

✪ OPEN ACCESS ⟳ Check for updates

Noninvasive Instrument-based Tests for Detecting and Measuring Vitreous Inflammation in Uveitis: A Systematic Review

Xiaoxuan Liu, MBChB [ID][a,b,c], Benjamin TK Hui[a], Christopher Way, MBChB[d], Sophie Beese, MPH[e], Ada Adriano, MSc[e], Pearse A Keane, MD[f], David J Moore, PhD[e], and Alastair K Denniston, PhD [ID][a,b,f,c]

aOphthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; bAcademic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham UK; cHealth Data Research UK, London, UK; dMusgrove Park Hospital, Taunton and Somerset NHS Foundation Trust, Taunton, UK; eInstitute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham UK; fNIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, UK

ABSTRACT

Purpose: This systematic review aims to identify instrument-based tests for quantifying vitreous inflammation in uveitis, report the test reliability and the level of correlation with clinician grading.
Methods: Studies describing instrument-based tests for detecting vitreous inflammation were identified by searching bibliographic databases and trials registers. Test reliability measures and level of correlation with clinician vitreous haze grading are extracted.
Results: Twelve studies describing ultrasound, optical coherence tomography (OCT), and retinal photography for detecting vitreous inflammation were included: Ultrasound was used for detection of disease features, whereas OCT and retinal photography provided quantifiable measurements. Correlation with clinician grading for OCT was 0.53–0.60 (three studies) and for retinal photography was 0.51 (1 study). Both instruments showed high inter- and intra-observer reliability (>0.70 intraclass correlation and Cohen's kappa), where reported in four studies.
Conclusion: Retinal photography and OCT are able to detect and measure vitreous inflammation. Both techniques are reliable, automatable, and warrant further evaluation.

Vitreous inflammation, or vitritis, is a clinical manifestation commonly found in posterior-segment involving uveitis. It is the hall-mark of intermediate uveitis, but is also common in panuveitis and may occur in posterior uveitis.[1,2] Infiltration of the vitreous body with inflammatory cells and proteinaceous exudates gives a characteristic hazy appearance, reducing the clarity of structures behind it (the optic disc and retinal vessels) during fundoscopy.[3] The clinical standard for measuring vitreous haze has been the National Eye Institute vitreous haze (NEI VH) scale since the Standardization of Uveitis Nomenclature (SUN) Workshop in 2005.[1] Prior to the SUN workshop, three grading systems existed.[2,4,5] The NEI VH scale is a 6-point grading system for estimating the vitreous clarity as seen through indirect ophthalmoscopy and is also referred to as the National Institute for Health (NIH) or Nussenblatt scale.[1,2] The clinician's estimate is compared to a standardized set of photographs and given a score of 0, +0.5, +1, +2, +3, or +4 (Table 1). This grading system has been the widely accepted standard for clinical assessment in routine care and for assessing disease outcomes in clinical trials.[6–8] It has been adopted as part of composite measures of disease outcome for uveitis, alongside other markers of inflammation such as anterior chamber cells/flare, central macular thickness, visual function, and quality of life.[9,10]

However, there are drawbacks to clinician grading. Firstly, this method is subjective with only moderate interobserver agreement, even when assessed by experienced uveitis specialists.[11,12] Secondly, the grading scale is non-continuous and non-linear, with large steps between each grade. Lastly, the system is poorly discriminatory for low levels of vitreous inflammation, where the need for sensitive detection of inflammatory activity to allow early clinical intervention, is greatest.[13]

More recently, measuring vitreous inflammation using instrument-based systems such as imaging devices has been proposed as a solution to some of these challenges. Several instruments, including fundus photography, ultrasound, and optical coherence tomography (OCT) have been used to visualize the vitreous body. These instrument-based methods have the theoretical advantage of being objective and automatable, and the changes detectable by each could be employed as surrogate measures of vitreous inflammation. This systematic review aims to identify all non-invasive, instrument-based tools (hereon referred to as index tests) with the ability to detect and measure vitreous inflammation in uveitis, and report the level of correlation between index tests and clinician grading, as well as the index tests' reliability.

Grade	Description
0	No evidence of vitreal haze
Trace/0.5+	Slight blurring of optic disc margin
1+	Obscured view but definition to optic nerve head and retinal vessels
2+	Obscured view but definition to retinal vessels
3+	Optic nerve head visualized but borders are very blurry
4+	Obscured fundal view

[a]Nussenblatt et al. Standardization of Vitreal inflammatory Activity in Intermediate and Posterior Uveitis. Ophthalmology. 1985;92(4). Adopted with minor modifications by Jabs et al. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. Am J Ophthalmol. 2005;140(3).

Methods

This systematic review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement.[14] The methodology was specified in advance and the protocol registered with PROSPERO (CRD42017084168).[15] Our search seeks to identify all index tests for detecting and quantifying vitreous inflammation. Where index tests were compared against a clinician grading system, the level of correlation was extracted. Any evaluation of test reliability, such as intra- and inter-observer reliability was also extracted.

Search strategy

We combined free text terms and index terms reflecting the pathological finding of interest, 'vitreous haze' or 'vitritis' and the disease context 'uveitis,' 'inflammation,' 'blood-retinal barrier,' and 'leak' where possible (search strategy available in **Supplementary Materials**). Database searches were carried out in MEDLINE, Embase, Cochrane Controlled Register of Trials (CENTRAL), Center for Reviews and Dissemination Database (Health Technology Assessments and the Database of Abstracts and Reviews of Effects), Clinicaltrials.gov, WHO International Clinical Trials Registry Platform (ICTRP portal), British Library's ZETOC, Conference Proceedings Citation Index (Web of Science), British Library Ethos, ProQuest and OpenGrey. We searched all databases from inception to December 4, 2019, with no date or language restrictions. We manually searched citations of review articles and included studies to identify additional relevant articles.

Study selection

Two reviewers independently assessed study eligibility and resolved disagreements by consensus or by referral to a third reviewer. Studies were eligible if they described one or more index tests for detecting and measuring vitreous inflammation. Studies were not excluded based on the basis of subject age, gender, ethnicity, underlying etiology, or disease activity status. Animal studies and studies involving only healthy participants, single case reports, commentaries, and opinion articles were excluded.

Data extraction

Two reviewers independently extracted data using a pre-specified data extraction sheet and resolved any discrepancies through consensus and referral to a third reviewer when needed. Data extracted included study design, population characteristics and disease phenotype, details of the index and reference tests, and outcomes relating to correlation between the two tests and test reliability. The full list of extracted items can be found in Table 2.

Risk of Bias Assessment

Relevant features of the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2) were used to assess for bias in the studies. The assessment considered patient selection (if the patients receiving the index and reference tests were representative of uveitis patients and the spectrum of uveitic subtypes), index test (if the index test was interpreted without knowledge of the reference test), reference test (if the reference test was interpreted without knowledge of the index test) and flow and timing (if all patients received both tests within an appropriate time interval – within same day assessment was deemed sufficient to ensure the inflammatory status of the eye had not changed). Not all elements of QUADAS-2 were applicable. For example, "whether the reference standard is likely to correctly quantify the target disease (vitreous inflammation)" would be marked unclear for all studies, due to the known poor reliability of clinician grading. As QUADAS-2 is only applicable for studies comparing an index test to reference test, the assessment was only carried out in studies evaluating correlation between the two tests and not in studies evaluating index test reliability.

Data analysis

For each index test, we tabulated the extracted information and provided a narrative synthesis of methodological characteristics and index tests evaluated. Studies which compared index test measurements with a reference test (such as clinician grading) and reported a correlation coefficient were included in the analysis. In these studies, where confidence intervals for correlation coefficients were not reported, correlation coefficients were normalized using Fisher's Z transformation for meta-analysis and back transformed and presented on a forest plot for visualization only. All statistical analyses were performed using Stata Statistical Software (Release 15. College Station, TX: StataCorp LP). Meta-analysis was not performed for test correlation or reliability due to heterogeneity between studies.

Results

Results of the Search

The study selection process is summarized in the PRISMA flow diagram (Figure 1).

The search yielded 7122 unique bibliographic records after removal of duplicates. Of these, 7100 were excluded based on screening of titles and abstracts. The large number of excluded records was due to the unrestrictive nature of the search strategy,

Table 2. Study characteristics.

Study	Study Design	Technology	Inclusion criteria	Exclusion Criteria	n Subjects	n Eyes	Age	n Male (%)	Etiology, n Eyes (%)
Oksala (1977)	Prospective	USS	AAU	nr	25	25	mean nr range 13–62	nr	nr
Haring (1988)	Prospective	UBM	Intermediate uveitis	nr	13	26	mean 32 (13–75 yrs)	4 (31)	nr
Doro (2005)	Prospective	UBM	Idiopathic typical or suspicious intermediate uveitis	nr	5	7	mean 17 (range 12–25)	2 (40)	idiopathic 7 (100)
Davis (2010)**	Retrospective	Retinal photography	Eyes with uveitis	Cataracts and prominent fundus pathology	nr	97	nr	nr	nr
Passaglia (2018)**	Retrospective	Retinal photography	Uveitis images from a clinical trial library, with varying degrees of vitreous haze and minimal/absent corneal and lens opacities	nr	nr	120	nr	nr	nr
Madow (2011)	Retrospective	Retinal photography	Patients with intermediate uveitis, posterior uveitis, or panuveitis from the MUST trial.[a]	Unreadable photographs, digital photographs, and those not yet received at the center by Feb 2009*	142	265	nr	nr	nr
Keane (2014)***	Retrospective	OCT	1) Eyes with vitreous haze secondary to intermediate, posterior, or panuveitis;2) Eyes with uveitis but with no evidence of vitreous haze; 3) Eyes without evidence of intraocular inflammation or vitreoretinal disease	nr	60	60	Uveitis with vitreous haze mean 49 (SD 18); Uveitis without vitreous haze, mean 44 (SD 14); Healthy controls, mean 67 (SD 8)	nr	Uveitis with haze: idiopathic 17 (57), BCR 5 (17), Toxoplasma 2 (7), sarcoidosis 2 (7), other 4 (12); Uveitis without haze: idiopathic 6 (50), BCR 2 (17), other 4 (33); Healthy 18 (100).
Zarranz-Ventura (2016)	Retrospective	OCT	Intermediate uveitis, posterior uveitis, or panuveitis	ERM preventing adequate transmission of light to the RPE, severe anatomic disruption of retinal integrity preventing adequate delineation of the RPE for analysis.	105	105	Uveitis eyes with vitritis mean 44 (SD 18); Uveitis eyes with no vitritis mean 47 (SD 15)	nr	Behcet's disease 24 (23); BCR 22 (21); Sarcoidosis 14 (13); Non-differentiated 11 (10); Pars planitis 8 (8); VKH 6 (6); MS 4 (4); Other 16 (15)
Sreekantam (2017)	Retrospective	OCT	Eyes with uveitic CMO & having STTA	nr	22	22	47 (range 23–74 yrs)	5 (23)	Idiopathic 14 (63), Sarcoidosis 4 (17), TINU syndrome 1 (5), Behcet's disease 1 (5), Reiter syndrome 1 (5), VKH disease 1 (5)

(Continued)

Table 2. (Continued).

Study	Study Design	Technology	Inclusion criteria	Exclusion Criteria	n Subjects	n Eyes	Age	n Male (%)	Etiology, n Eyes (%)
Keane (2015)***	Retrospective	OCT	1) Eyes with vitreous haze secondary to intermediate, posterior, or panuveitis;2) Eyes with uveitis but with no evidence of vitreous haze; 3) Eyes without evidence of intraocular inflammation or vitreoretinal disease	nr	60	60	Uveitis with vitreous haze mean 49 (SD 18); Uveitis without vitreous haze mean 44 (SD 14); Healthy controls mean 67 (SD 8)	nr	Uveitis with haze 30 (100): idiopathic 17 (57), BCR 5 (17), Toxoplasma 2 (7), sarcoidosis 2 (7), other 4 (12);Uveitis without haze 12 (100): idiopathic 6 (50), BCR 2 (17), other 4 (33); Healthy 18 (100);
Coric (2019)	Retrospective	OCT	Patients with uveitis and multiple sclerosis and healthy controls	**MS patients with uveitis:** - Poor scan quality- Relapse or corticosteroid treatment in month prior to baseline assessment- Pregnancy- previous neurological/ psychiatric disorder- Drug or alcohol abuse- MRI abnormalities not consistent with **MSControls (as for MS patients with uveitis), plus:** - Family member with MS (first or second degree of consanguinity)- Significant MRI abnormalities	375	nr	Mean (SD): MS 52 (10) controls 49 (8)	Sex (F:M) MS 195:95 Controls 53:32	multiple sclerosis 290 (77)
Mahendradas (2017)	Prospective	OCT (using enhanced vitreous imaging technique)	Uveitis with vitreous cells	nr	33	59	Mean 43 (range 12–72)	11 (33)	Idiopathic Panuveitis 6; Intermediate uveitis 5; VKH 5TB 3; Sarcoidosis 3, Multifocal retinitis 3, Toxoplasmosis 2Posterior uveitis 2, Serpiginous choroiditis 1Idiopathic retinal vasculitis 1Sympathetic Ophthalmia 1Dengue retinitis 1

nr: not reported, MUST: Multicenter Uveitis Steroid Treatment, USS: ultrasound, UBM: ultrasound biomicroscopy, OCT: optical coherence tomography, AAU: acute anterior uveitis, BCR: birdshot chorioretinitis, TINU: tubular interstitial nephritis associated uveitis, VKH: Vogt Kayanagi Harada, MS: multiple sclerosis, ERM: epiretinal membrane, RPE: retinal pigmented epithelium, TB: tuberculosis, STTA: subtenon triamcinolone acetate, CMO: cystoid macular edema.ªKempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. Multicenter Uveitis Steroid Treatment Trial Research Group. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. Am J Ophthalmol. 2010; 149(4):550–561.

*Image data was from the MUST trial (NCT00132691), which excluded participants with inadequately controlled diabetes, Participants with uncontrolled glaucoma, advanced glaucomatous optic nerve injury, a history of scleritis; presence of an ocular toxoplasmosis scar and HIV infection or other immunodeficiency disease for which corticosteroid therapy would be contraindicated according to best medical judgment.

** Passaglia 2018 and Davis 2010 developed grading systems using the same set of fundus images.

*** Keane 2014 and Keane 2015 used the same image dataset for both studies.

102

Table 2. (Continued).

Study	Technology	Manufacturer model	Acquisition/Image processing protocol	Area/volume	Analysis software	Automation
Index test characteristics						
Oksala (1977)	USS	Model 7100 a (Kretztechnik, Austria)	"The transducer was pressed against the sclera and the beam aimed at the vitreous from several different directions"	Whole axial length	na	Manual
Haring (1988)	UBM	UBM 840, (Zeiss-Humphrey, San Leandro, CA, USA)	UBM using 50 MHz probe. Gain was 80 dB: approximate spatial resolution 50 μm and penetration depth 5 mm. "During each examination radial scanning of the anterior chamber angle/ciliary body region and the pars plana and peripheral retina was conducted in all clock hours, but with emphasis on the lower circumference of the eye."	nr	na	Manual
Doro (2005)	UBM	Model P45 (Paradigm Medical Industries, Salt Lake City, Utah, USA): 50-MHz probe and Cinescan S (Quantel Medical, Clermont-Ferrand, France): 20-MHz immersion open probe.	nr	nr	na	Manual
Davis (2010)**	Retinal photography	30° Zeiss fundus camera model FF4 (Carl Zeiss Meditec, Inc, Pleasanton, California, USA) with a Nikon film camera (Nikon Instruments Inc, Melville, New York, USA)	Photos modified by the application of a Bangerter occlusion filter. Films were then digitized with Nikon film scanner at 24-bit color and resolution of 300 dpi and images saved as TIF format	30° photograph	na	Manual
Passaglia (2018)**	Retinal photography	nr	Images stored as TIF format cropped to an area of 512 by 512 pixels centered on the macula	na	Custom software	Fully automated
Madow (2011)	Retinal photography	nr	Images digitized with Nikon Coolscan film scanner at 300 dpi and saved as TIF format	nr	na	Manual
Keane (2014)***	OCT	Heidelberg SPECTRALIS OCT	3–5 B-scans passing through the foveal central subfield	nr	OCTOR custom software	Semi-automatic
Zarranz-Ventura (2016)	OCT	Cirrus HD-OCT (Carl Zeiss Meditec, Dublin, California, USA)	"Macular Cube" protocol: 128 horizontal B-scans. This covers a 6 by 6 mm area.	36 mm²	OCTOR custom software	Semi-automatic
Sreekantam (2017)	OCT	Heidelberg SPECTRALIS OCT	Volume scan images of 20° by 20°,containing a minimum of 25 OCT B-scans on TruTrack Active and Auto Rescan follow up modes active	20° by 20°	OCTOR custom software	Semi-automatic
Keane (2015)***	OCT	Heidelberg SPECTRALIS OCT	Volume scans centered on the fovea.	nr	VITAN custom software	Fully automated
Coric (2019)	OCT	Heidelberg SPECTRALIS OCT	Macular volume scan centered around the fovea (20x20°, 512 A-scans, 49 B-scans, vertical alignment, automatic real time 16)	20° by 20°	VITAN custom software	Fully automated
Madendradas (2017)	OCT (using enhanced vitreous imaging technique)	Heidelberg SPECTRALIS OCT	Four sets of horizontal and vertical 9 mm B scans with ART 100. Enhanced Depth Imaging: position of scan shifted to lower half of screen Combined Depth Imaging: position of scan shifted to middle of screen Enhanced Vitreous Imaging: +2 diopters added to the focus	na	na	Manual

nr: not reported, USS: ultrasound, UBM: ultrasound biomicroscope, OCT: optical coherence tomography, ART: automated real-time averaging.
** Passaglia 2018 and Davis 2010 developed grading systems using the same set of fundus images.
*** Keane 2014 and Keane 2015 used the same image dataset for both studies.

Table 2. (Continued).

Index test correlation with clinical grading and reliability

Study	Technology	Automation	Comparator	Eyes per grade (n)	Correlation (LCI, UCI)	Reliability test	Reliability test result
Oksala (1977)	USS	Manual	none	na	na	nr	na
Haring (1988)	UBM	Manual	Slit lamp fundoscopy	na	na	nr	na
Doro (2005)	UBM	Manual	none	na	na	nr	na
Davis (2010)**	Retinal photography	Manual	none	na	na	ICC	0.88 (Interobserver)
Passaglia (2018)**	Retinal photography	Fully automatic	NIH scale (6 point photographic scale), Miami scale (9 level photographic scale)	nr	Exact agreement: Cohen's K = 0.61 (NIH scale) and 0.67 (Miami scale), Within-one level agreement: K = 0.78 (NIH scale) and 0.82 Miami scale Within-two level agreement: K = 0.80 (NIH scale) and 0.84 (Miami scale).	nr	na
Madow (2011)	Retinal photography	Manual	NEI VH scale	0 (85) 1+ (135) 2+ (35) 3+ (9) 4+(1)	(correlation test nr) r = 0.51	ICC	0.87 (Interobserver); 0.84–0.93 (Intraobserver);
Keane (2014)***	OCT	Semi-automatic	NEI VH Grade	0, 12 0.5+, 4 1+, 13 2+, 10 3+, 3 4+, 0	Spearman's r = 0.57	BA 95% limits of agreement between two graders. Where, "uveitis with vitreous haze" median measurement = 0.150 (IQR = 0.135) "uveitis with no VH/ normal" Median measurement = 0.0767 (IQR = 0.048)	All: 0.035 Uveitis with vitreous haze (grade 0.5+ and above) 0.045; Healthy control (grade 0): 0.023;
Zarranz-Ventura (2016)	OCT	Semi-automatic	NEI VH Grade	0, 54 0.5+, 21 1+, 18 2+, 9 3+, 3 4+, 0	Spearman's r = 0.53	nr	na
Sreekantam (2017)	OCT	Semi-automatic	none	na	na	nr	na
Keane (2015)***	OCT	fully automated	NEI VH Grade	0, 12 0.5+, 4 1+, 13 2+, 10 3+, 3 4+, 0	Spearman's r: Vitreous:RPE signal ratio = 0.59Vitreous:RPE textural ratio = 0.60	nr	na
Coric (2019)	OCT	Fully automatic	none	na	na	nr	na
Mahendradas (2017)	OCT (using enhanced vitreous imaging technique)	Manual	Each image graded for posterior vitreous visualization: 0 = not visible 1 = barely visible 2 = clearly visible	na	na	interobserver agreement (Cohen k) of visualization scoring	Conventional OCT: 0.771 EDI 0.732 CDI 0.722 EVI 0.743

na: not applicable, nr: not reported, USS: ultrasound, UBM: ultrasound biomicroscope, OCT: optical coherence tomography, NIH: national institute of health, NEI: national eye institute, ICC: intraclass correlation, BA: bland-altman, VH: vitreous haze, IQR: interquartile range, EDI: enhanced depth imaging, CDI: combined depth imaging, EVI: enhanced vitreous imaging

** Passaglia 2018 and Davis 2010 developed grading systems using the same set of fundus images.

*** Keane 2014 and Keane 2015 used the same image dataset for both studies.

LCI (lower confidence interval) and UCI (upper confidence interval) was not reported in all studies.
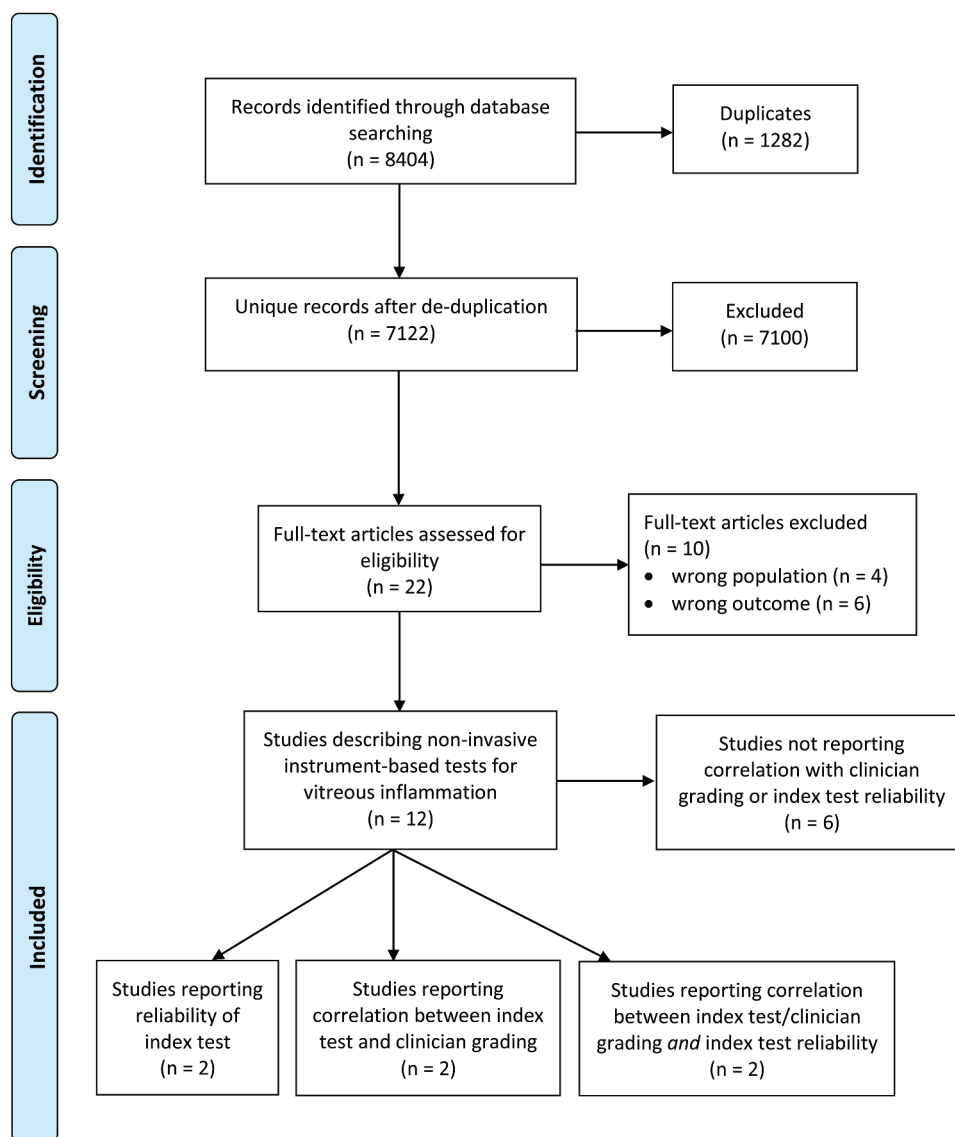
Figure 1. PRISMA flow diagram.

which was deliberately wide, to ensure full capture of all potentially relevant technologies. The remaining 22 articles were reviewed in full text and further 10 articles were excluded. The reasons for exclusion were due to not matching the criteria for outcome (n = 6) or target population (n = 4). Twelve articles were included; two studies compared an index test with a clinician grading system, two reported test reliability, and two did both. Six studies described index tests but did not report correlation with clinician grading or index test reliability (Table 2).

Participants' characteristics and study design

The 12 studies included a total of at least 840 participants[16-25] (two studies did not report the number of participants[26,27]) and at least 846 eyes (one study did not report the number of eyes.[23]) The studies were published between 1977 and 2019. Four studies were conducted prospectively[16,17,24,25] and eight were retrospective.[18-,23-,26-,27] Only five studies reported

gender, with 29% of participants (n = 149) being male.[16,17,22-24] The age of participants in studies ranged from 12 to 75 years.[16,17,19-25] Five studies included mixed etiologies, including toxoplasmosis, sarcoidosis, Behcet's disease, Birdshot chorioretinopathy, pars planitis, Vogt-Koyanagi-Harada disease, multiple sclerosis, tubulointerstitial nephritis and uveitis syndrome, Reiter's syndrome, multifocal retinitis, serpiginous choroiditis, idiopathic retinal vasculitis, sympathetic ophthalmia, and Dengue retinitis.[19–22,24] Two studies were narrower in their inclusion criteria with one study including intermediate uveitis only[17] and the other including patients with multiple sclerosis only[23]. Five studies did not specify the underlying etiology of participants.[16,25-27]

Clinical reference test

Six out of 12 studies did not compare an index test against a comparator.[17,22-26] One study compared ultrasound biomicroscopy against qualitative features on slit-lamp

105

fundoscopy.[16] Four studies compared OCT[19-21] and retinal photography[27] against the NEI VH scale only, and one study compared retinal photography against both NEI VH scale and a photographic scale called the Miami scale (described in next section).[27]

Instruments for detecting and quantifying vitreous inflammation

Three types of technologies with the ability to detect and quantify vitreous inflammation were identified from the 12 studies: ultrasound, retinal photography, and OCT.

Three studies employed ultrasound. One study used an A-scan instrument, model 7100A (Kretztechnik, Austria) with a transducer of 6 MHz/S mm[25] and two studies used ultrasound biomicroscopy (the UBM 840 (Zeiss-Humphrey, San Leandro, CA, USA) with a 50 MHz probe in one study[16] and the Model P45 (Paradigm Medical Industries, Salt Lake City, Utah, USA) with a 50 MHz probe plus the Cinescan S (Quantel Medical, Clermonth-Ferrand, France) with a 20 MHz immersion open probe in another study.[17]) The images in all three studies were interpreted manually and qualitatively, by the operator, in real-time.

Three studies used retinal photography. Davis et al. developed a 9-point scale using calibrated Bangerter filters to blur fundus photographs, originally acquired using 30⁰ Zeiss fundus camera model FF4 (Carl Zeiss Meditec Inc, Pleasanton, California, USA) with a Nikon film camera (Nikon Instruments Inc, Melville, New York, USA).[26] This 9-point scale is known as the Miami scale and is designed to be a reference for manual clinician grading of fundus photographs. The authors tested the use of this reference scale using film fundus photographs from an imaging archive (unspecified camera and system). Madow et al. used fundus photographs originally acquired as color film slides for the MUST trial[12] and digitized them using Nikon Coolscan film scanner (Nikon, Inc, Melville, New York, USA) at 300 dpi and saved as TIFF format.[18] Madow et al. used the Miami scale developed by Davis et al. to grade the severity of vitreous haze in these photographs.[18] Passaglia et al. applied an automated retinal photography analysis software to grade fundus photographs from a clinical trial library (unspecified source, camera, and system) according to the NEI VH and Miami VH scales.[27]

Six studies used OCT. Five studies used the Heidelberg SPECTRALIS OCT[19,20,22-24] and one used the Cirrus HD-OCT (Carl Zeiss Meditec, Dublin, California, USA).[21] Two studies used the same semi-automated image analysis technique (custom OCTOR software),[19,22] two used the same fully automated image analysis technique (custom VITAN, which employs the same principles of pixel intensity as OCTOR, requires no manual input other than confirmation of the selected vitreous area)[20,23] and one study used manual analysis of OCT images using a subjective observer-based grading system consisting of grades 0–2, where grade 0 was 'not visible,' grade 1 was 'barely visible,' and grade 2 was 'clearly visible.'[24]

Index test reliability

Four studies reported index test reliability using varying methodologies. Davis et al. reported an intraclass correlation (ICC) of 0.88 between two observers grading fundus photographs against the 9-point Miami scale.[26] Madow et al. reported an inter-observer ICC of 0.87 and an intra-observer ICC of between 0.84 and 0.93 against the Miami scale.[18] Keane et al. used Bland–Altman plots to assess interobserver variability and reported a median 95% limits of agreement (LoA) of 0.0353 for all OCTs, 0.0450 in OCTs of uveitic eyes with vitreous haze and 0.0226 for OCTs of healthy eyes or uveitic eyes without vitreous haze. They reported the variance ratio (F statistic) as non-significant between groups, suggesting the measurement variance was similar in eyes with and without vitreous inflammation.[19] Mahendradas et al. reported interobserver agreement as Cohen's kappa >0.7 for all four tested techniques (standard OCT, enhanced vitreous imaging, enhanced depth imaging, and combined depth imaging).[24]

Correlation between index tests and the clinical reference test: Slit-lamp based clinician grading

Four studies reported correlation between an index test and clinician grading of vitreous inflammation (three studies using OCT[19-21] and one study using retinal photography.[18]) All studies reporting correlation used the NEI VH scale as a comparator. The total number of participants included in these four studies was 307 (430 eyes). Spearman's r was used by all studies except by Madow et al. to measure the association between index test measurements and the NEI VH scale. The level of correlation between OCT measurements and the NEI VH scale using the semi-automated OCTOR software was 0.53–0.57,[19,21] whereas for the fully automated VITAN software correlation was marginally higher at 0.59–0.60.[20,28] Both studies by Keane et al., reporting the use of OCTOR and VITAN, used the same retrospective dataset of images. The level of correlation between manual grading of retinal photographs (using the Miami scale) versus clinician examination (using the NEI VH scale) was reported as $r = 0.51$. The correlation between index tests and the NEI VH scale are shown in Figure 2. None of the four studies reported confidence intervals for correlation coefficients and those shown in the forest plot were estimated using sample size and correlation coefficient. Passaglia et al. measured agreement between automated fundus photography grading (using the Miami scale) and clinician grading, rather than correlation. They report exact agreement, agreement within one level and agreement within two levels of 0.61, 0.78, and 0.80, respectively, against clinician grading using the NIH scale and 0.67, 0.82, and 0.84, respectively, against the clinician grading using the Miami scale.[27]

Risk of Bias Assessment

The patient cohorts in the correlation studies were a mixture of uveitis etiologies with a low risk of spectrum bias, except in the retinal photography study by Madow et al., where the risk was not assessable as the underlying etiology was not reported.18
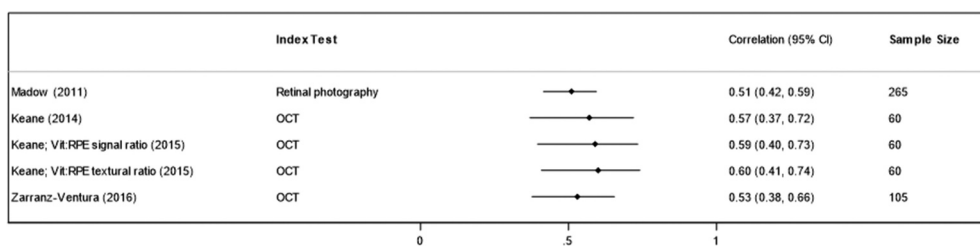
Figure 2. Level of correlation between index tests and clinician grading (SUN/NEI/Nussenblatt vitreous haze scale). RPE: retinal pigmented epithelium, OCT: optical coherence tomography. *Keane 2015 uses VITAN, an automated version of the previously published OCTOR software. Two variations of the same technique are presented: vitreous:RPE signal ratio and vitreous:RPE textural ratio.** Keane 2014 and Keane 2015 used the same cohort of patients for both studies.

Other than Madow et al., all studies used automated/semi-automated systems to quantify vitreous haze; therefore, it was assumed there was no potential influence from knowledge of the clinician grading. All studies used previously recorded clinician grading (from clinical care or clinical trials data), therefore there was no possibility that the reference test could have been influenced by the index test, which was conducted afterward. Madow et al. did not report whether the fundus photograph readers were blinded to the clinician grading results. Although the time interval between index and reference tests were not explicitly reported by any of the studies, it is presumed that clinician grading and the images acquired were performed on the same visit in all studies, even if image analysis for index tests were done at a later date.

Study heterogeneity

After accounting for overlap between studies in terms of similar imaging techniques and duplicated patient cohorts, there was considerable heterogeneity between the methodology and populations across the included studies. Given this level of heterogeneity, we have not performed any meta-analysis of correlation or test reliability for index tests.

Discussion

This is the first systematic review for identifying noninvasive instrument-based tests for detecting and measuring vitreous inflammation in uveitis. Three technologies were found: ultrasound, retinal photography, and OCT. Ultrasound remains primarily for qualitative assessment of features in the vitreous body and has not been shown to quantify inflammation. Retinal photography and OCT have demonstrated the most potential as methods for quantifying vitreous inflammation through automated and semi-automated means of image analysis. However, only 12 studies have been undertaken and even fewer provided sufficient evidence on test reliability or correlation with clinician grading.

Davis et al. and Madow et al. reported good interobserver reliability (ICC>0.84) and moderate correlation ($r = 0.51$) of manual grading using retinal photography (assessed using the Miami scale).[18,26] This photographic method introduces two advantages beyond the traditional indirect biomicroscopic approach (assessed against the NEI VH scale). Firstly, it captures an adequate view of the fundus and removes the variability introduced by the level of the indirect biomicroscopy

skills of the examiner. Secondly, it is based on a 9-point scale rather than the 6-point NEI VH scale, allowing smaller differences to be captured between grades. The automated retinal photography technique applied by Passaglia et al. brings added objectivity beyond the direct biomicroscopic assessment of the NEI VH scale or the original subjective photograph-to-photograph comparison of the Miami grading. On the other hand, the OCT-based technique utilizes signal intensity detected in the vitreous, to derive a measure of light reflectivity as a continuous variable. The ability to detect vitreous reflectivity on a continuous scale means the OCT-based method may potentially offer sensitivity to even smaller, but potentially clinically significant, changes in vitreous inflammation.

Whilst automation of image analysis may improve reliability, we did not find that it consistently improves correlation with clinician grading. The fully automated VITAN OCT algorithm was tested on the same dataset as the semi-automated OCTOR algorithm and showed marginally higher correlation when compared to the NEI VH grade ($r = 0.60$ versus 0.57).[20] Manual grading of retinal photography showed moderate correlation ($r = 0.51$)[18] when compared to the NEI VH scale, similar to the moderate agreement reported for fully automated photographic grading (Cohen's K = 0.61).[27]

Strengths and limitations of the review

This review represents the first systematic evaluation of technologies for measuring vitreous inflammation in uveitis. The search strategy was designed to be highly sensitive, using a broad range of databases, including conference proceedings, dissertation databases and the grey literature. This review also has several limitations. An issue in undertaking systematic reviews of correlation between assessment methods is the absence of an adequate specific tool for assessing risk of bias in studies. We have used relevant elements of the QUADAS-2 tool for risk of bias assessment in test accuracy studies for the correlation studies only, where one test was being compared against another. However, this method of assessing risk of bias could not be applied to include studies which only evaluated one test (i.e. for index test reliability). Second, although we included all studies reporting instruments with the potential to detect and measure vitreous inflammation, the data extraction and analysis were focused on test reliability or correlation with the clinical standard. As a result, two studies that provide evidence of the clinical validity and value of new techniques were not discussed in detail.[22,23] These include Sreekantam

et al.'s study which reported a highly statistically significant reduction of OCT-based vitreous signal (using OCTOR) pre- and post-injection of sub-tenon's triamcinolone in patients with uveitic macular edema, demonstrating the potential clinical utility of this technique for detecting treatment response and its potentially superior sensitivity for measuring change over the clinician based grading system; however, this study did not include NEI VH scale as a comparator.[22] Coric et al. also explored whether a difference in vitreous intensity could be detected in patients with multiple sclerosis versus healthy controls, but did not find a measurable difference; again this study did not include NEI VH scale as a comparator.[23] Additional imaging techniques such as ultra-wide field fundus photography (using the Optos ultra-wide field camera) have also been used to detect presence and absence of vitreous haze through manual observation.[29] Third, the focus of this review was on correlation with the reference test. Whilst correlation is helpful in early validation, it is limited to demonstrating agreement and non-inferiority to the comparator. From correlation, it is not possible to determine if the index test is more accurate than the reference test. To determine accuracy, a more reliable reference test than the NEI VH scale is required, such as the use of invasive vitreous sampling to determine the level of protein and cellular infiltrates in the vitreous. Due to risks involved, it is unlikely that vitreous sampling will be ethically justifiable in routine practice. In the absence of a reliable reference test, future work could compare the ability of index tests versus clinician grading to detect changes in inflammation, such as the pre- and post-treatment comparison Sreekantam et al. conducted.[22] The ability to demonstrate higher sensitivity to small changes as well as stronger association with other inflammatory markers (such as central macular thickness) and visual function, would provide further evidence of accuracy in measuring the true disease state.

Limitations of the evidence

Firstly, due to the small number of included studies and heterogeneity in study design, meta-analyses of correlation or reliability were not possible. Several studies were conducted by the same author groups and presented sequential updates of the same technique using different approaches to image analysis, including automation.[20,27] Most studies used retrospectively collected images, with several applying newer analysis techniques to the same image set. Incomplete reporting and varying methodology of the included studies also meant we were unable to pool estimates of correlation between index and reference tests. Secondly, authors sometimes reported correlation coefficients estimated from a mixed cohort of uveitic and healthy eyes. With the exception of Madow et al., where only uveitic eyes were included in the study, all other studies reporting correlation coefficients were a mixture of healthy and uveitic eyes.[19–21] It was not possible to separate the two cohorts as correlation was reported at an aggregated level in all cases. On the other hand, all studies reporting intra/inter-observer reliability included uveitic eyes only. Thirdly, of those studies that reported NEI VH grading, no patients for OCT and only one patient for retinal photography had grade 4+.[18] It could be that in dense vitreous haze, neither OCT nor photography can successfully acquire a usable image and such cases could have been excluded on the basis of poor image quality. However, it is unclear how those index tests performed in the most severe grades of vitreous inflammation.

Clinical relevance and impact

Of the instrument-based tests identified, OCT and retinal photography are presented with the most supporting evidence in this review. Both instruments offer the attractiveness of being technologies already widely available in ophthalmic clinics. Additionally, both techniques can be combined with automated image analysis techniques. OCT additionally offers a measurement which can be continuous and it has also been shown to be sensitive to respond to treatment.[22] At this stage there are only a few reports identified for either technology and these reports were mostly retrospective studies with small numbers of subjects. As noted earlier very few patients with severe vitritis are included in these studies, and it is difficult to draw conclusions on the validity of both instruments in the most severe levels of inflammation. It could be argued that, where inflammation is obviously detectable through clinical examination, there is less additional value of quantification by a noninvasive imaging technique. However, clearly, the ideal scenario is to have a technique that is sensitive to changes at both ends of the scale, including detecting worsening or improvement in severe inflammation.

Another major consideration is around generalizability of the study findings in the presence of ocular co-pathology. Of particular concern is media opacity such as cataract, which may cause a similar hazy appearance on fundoscopy and which could degrade image quality on both retinal photography and OCT. Given cataracts are a major complication of chronic intraocular inflammation and ocular steroid therapy, many patients with posterior uveitis have cataracts.[30] In the included studies of this review, only Davis et al. reported the exclusion of subjects with cataracts.[26] Zarranz-Ventura et al. assessed the use of OCT of patients with uveitis, which also included pseudophakia and patients who had undergone vitrectomy. They demonstrated no observable difference in the measurement for each of these groups compared to phakic and non-vitrectomised eyes, respectively.[21]

An important area for future work is to evaluate the relationship between instrument-based measures and visual function. Sreekantam et al. reported a correlation coefficient of 0.70 between VIT/RPE-relative intensity and visual acuity, a stronger correlation than was demonstrated when the same OCTOR technique was compared to the NEI VH grading by *Keane* et al. ($r = 0.60$).[19,22] However, this is not a direct comparison due to different subjects in each study. No other studies explored the association between the index test measurements and visual acuity or any other measure of visual function. Whilst the relationship of visual function to inflammatory activity is complex, often being delayed and indirect, it is worthy of exploration. These tests will be of greatest value if their use enables better control of inflammation, such that vision is maintained in the immediate and long term. It is worth noting that the importance of demonstrating clinical validity through association with visual function was

emphasized by regulatory bodies at the American Uveitis Society workshop at the University of California Los Angeles (UCLA) in March 2019 on *Objective Measures of Intraocular Inflammation for Use in Clinical Trials.*[31]

If the unreliability of the current reference standard is limiting the evaluation and adoption of emerging techniques, are there any other techniques we should consider as a reference test? As previously discussed, direct sampling of vitreous is unlikely to be ethically justifiable unless it is being done for diagnostic purposes. More invasive tests to quantify vitreous inflammation also exist but are largely unused. Vitreous fluorophotometry is an intravenous fluorescein-based imaging technique which can quantify leakage of dye from the blood-retinal-barrier (BRB) and has been used in the assessment of inflammation of the posterior segment.[32] Vitreous fluorophotometry measures leakage by measuring the degree of fluorescence throughout the eye at defined axial points before and after the intravenous injection of fluorescein. It relies on the principle that the amount of leakage is proportional to the degree of BRB breakdown. However, due to its invasive nature, vitreous fluorophotometry is rarely performed and for the most part, has been used as an experimental technique rather than for clinical care.[33] Nonetheless, it is worth considering that invasive tests like fluorophotometry may be more direct measures of inflammatory activity and may serve as better reference tests with which to validate newer noninvasive tests. Assuming invasive approaches are not undertaken, evidence supporting new techniques and eventual adoption as a 'reference standard' is likely to depend on demonstrating high test reliability, strong association with other evidence of inflammation (such as macular thickness, presence of vitreous cells and other vitreous inflammatory infiltrates, presence of retinal vasculitis and vascular leakage and new active inflammatory lesions), and association with visual function (recognizing that this may not be direct or immediate).

Conclusion

Non-invasive instrument-based tests for measuring vitreous inflammation have the potential to improve reliability and speed compared to clinician grading using indirect ophthalmoscopy. Retinal photography and OCT are two promising technologies with the potential to quantify vitreous inflammation; however, further evidence beyond the proof-of-concept studies identified by this review are required to demonstrate clinical utility. Further evaluation in prospective studies should explore association with other measures of posterior-segment inflammation as well as visual function.

Author contributions

XL, CW, BH: data extraction, manuscript drafting, manuscript reviewing. SB, AA: data extraction, manuscript reviewing. PAK, DJM, AKD project conception, supervision, manuscript reviewing.

Declaration of interest

XL and AKD receive a proportion of their funding from the Wellcome Trust, through a Health Improvement Challenge grant (200141/Z/15/Z).

ORCID

Xiaoxuan Liu MBChB http://orcid.org/0000-0002-1286-0038
Alastair K Denniston PhD http://orcid.org/0000-0001-7849-0087

References

1. Jabs DA, Nussenblatt RB, Rosenbaum JT. Group S of UN (sun) W. Standardization of uveitis nomenclature for reporting clinical data. Results of the first international workshop. *Am J Ophthalmol.* 2005;140:509–516.
2. Nussenblatt RB, Palestine AG, Chan CC, Roberge F. Standardization of vitreal inflammatory activity in intermediate and posterior uveitis. *Ophthalmology.* 1985;92(4):467–471. doi:10.1016/S0161-6420(85)34001-0.
3. Forrester J. Uveitis: pathogenesis. *Lancet.* 1991;338(8781):1498–1501. doi:10.1016/0140-6736(91)92309-P.
4. Foster CS, Charles S, Vitale AT. *Diagnosis and Treatment of Uveitis.* Philadelphia, PA: W.B. Saunders; 2002:900.
5. Kimura SJ, Thygeson P, Hogan MJ. Signs and symptoms of uveitis. II. Classification of the posterior manifestations of uveitis. *Am J Ophthalmol.* 1959;47(5):171–176. doi:10.1016/S0002-9394(14)78240-6.
6. Denniston AK, Holland GN, Kidess A, et al. Heterogeneity of primary outcome measures used in clinical trials of treatments for intermediate, posterior, and panuveitis. *Orphanet J Rare Dis.* 2015;10(1):97. doi:10.1186/s13023-015-0318-6.
7. Lowder CMDP, Belfort RJMD, Lightman SMD, et al. Dexamethasone intravitreal implant for noninfectious intermediate or posterior uveitis. *Arch Ophthal.* 2011;129(5):545. doi:10.1001/archophthalmol.2010.339.
8. Kempen JH, Van Natta ML, Altaweel MM, et al. Factors predicting visual acuity outcome in intermediate, posterior, and panuveitis: the multicenter uveitis steroid treatment (MUST) trial. *Am J Ophthalmol.* 2015;160(6):1133–1141.e9. doi:10.1016/j.ajo.2015.09.017.
9. Jaffe GJ, Dick AD, Brézin AP, et al. Adalimumab in patients with active noninfectious uveitis. *N Engl J Med.* 2016;375(10):932–943. doi:10.1056/NEJMoa1509852.
10. Nguyen QD, Merrill PT, Jaffe GJ, et al. Adalimumab for prevention of uveitic flare in patients with inactive non-infectious uveitis controlled by corticosteroids (VISUAL II): a multicentre, double-masked, randomised, placebo-controlled phase 3 trial. *Lancet.* 2016;388(10050):1183–1192. doi:10.1016/S0140-6736(16)31339-3.
11. Hornbeak DM, Payal a, Pistilli M, et al. Interobserver agreement in clinical grading of vitreous haze using alternative grading scales. *Ophthalmology.* 2014;121(8):1643–1648. doi:10.1016/j.ophtha.2014.02.018.
12. Kempen JH, Ganesh SK, Sangwan VS, Rathinam SR. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. *Am J Ophthalmol.* 2008;146(6):813–818. doi:10.1016/j.ajo.2008.06.004.
13. Denniston AK, Keane PA, Srivastava SK. Biomarkers and surrogate endpoints in uveitis: the impact of quantitative imaging. *Invest Ophthalmol Vis Sci.* 2017;58(6):BIO131–BIO140. doi:10.1167/iovs.17-21788.
14. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses:

the PRISMA statement. *Int J Surg.* 2010;8(5):336–341. doi:10.1016/j.ijsu.2010.02.007.

15. Liu X, Moore D, Denniston A. Instrument-based tests for measuring vitreous inflammation in uveitis: a systematic review. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=84168. Published February 19, 2018. Accessed April 1, 2020.

16. Haring G, Nolle B, Wiechens B. Ultrasound biomicroscopic imaging in intermediate uveitis. *Br J Ophthalmol.* 1998;82(6):625–629. doi:10.1136/bjo.82.6.625.

17. Doro D, Manfre a, Deligianni V, Secchi AG. Combined 50- and 20-MHz frequency ultrasound imaging in intermediate uveitis. *Am J Ophthalmol.* 2006;141(5):953–955. doi:10.1016/j.ajo.2005.11.048.

18. Madow B, Galor a, Feuer WJ, Altaweel MM, Davis JL. Validation of a photographic vitreous haze grading technique for clinical trials in uveitis. *Am J Ophthalmol.* 2011;152(2):170–176.e1. doi:10.1016/j.ajo.2011.01.058.

19. Keane PA, Karampelas M, Sim DA, et al. Objective measurement of vitreous inflammation using optical coherence tomography. *Ophthalmology.* 2014;121(9):1706–1714. doi:10.1016/j.ophtha.2014.03.006.

20. Keane PA, Balaskas K, Sim DA, et al. Automated analysis of vitreous inflammation using spectral-domain optical coherence tomography. *Transl Vis Sci Technol.* 2015;4(5):4. doi:10.1167/tvst.4.5.4.

21. Zarranz-Ventura J, Keane PA, Sim DA, et al. Evaluation of objective vitritis grading method using optical coherence tomography: influence of phakic status and previous vitrectomy. *Am J Ophthalmol.* 2016;161:172–180e4. doi:10.1016/j.ajo.2015.10.009.

22. Sreekantam S, Macdonald T, Keane PA, Sim DA, Murray PI, Denniston AK. Quantitative analysis of vitreous inflammation using optical coherence tomography in patients receiving sub-Tenon's triamcinolone acetonide for uveitic cystoid macular oedema. *Br J Ophthalmol.* 2017;101(2):175–179. doi:10.1136/bjophthalmol-2015-308008.

23. Coric D, Ometto G, Montesano G, et al. Objective quantification of vitreous haze on optical coherence tomography scans: no evidence for relationship between uveitis and inflammation in multiple sclerosis. *Eur J Neurol.* 2020;27(1):144–e3. doi:10.1111/ene.14048.

24. Mahendradas P, Madhu S, Kawali A, et al. Enhanced vitreous imaging in uveitis. *Ocul Immunol Inflamm.* 2017;27(1):1–7.

25. Oksala A. Ultrasonic findings in the vitreous body in patients with acute anterior uveitis. *Acta Ophthalmol.* 1977;55(2):287–293. doi:10.1111/j.1755-3768.1977.tb01310.x.

26. Davis JL, Madow B, Cornett J, et al. Scale for photographic grading of vitreous haze in uveitis. *Am J Ophthalmol.* 2010;150(5):5. doi:10.1016/j.ajo.2010.05.036.

27. Passaglia CL, Arvaneh T, Greenberg E, Richards D, Madow B. Automated method of grading vitreous haze in patients with uveitis for clinical trials. *Transl Vis Sci Technol.* 2018;7(2):10. doi:10.1167/tvst.7.2.10.

28. Nanavaty MA, Stanford MR, Sharma R, Dhital a, Spalton DJ, Marshall J. Use of the double-pass technique to quantify ocular scatter in patients with uveitis: a pilot study. *Ophthalmologica.* 2010;225(1):61–66. doi:10.1159/000316690.

29. Dickson D, Agarwal a, Sadiq MA, et al. Assessment of vitreous haze using ultra-wide field retinal imaging. *J Ophthalmic Inflamm Infect.* 2016;6(1):1.doi:10.1186/s12348-016-0105-0.

30. Jones NP. The manchester uveitis clinic: the first 3000 patients, 2: uveitis manifestations, complications, medical and surgical management. *Ocul Immunol Inflamm.* 2015;23(2):127–134. doi:10.3109/09273948.2014.968671.

31. American Uveitis Society UCLA Stein Eye Institute. *UCLA/AUS workshop on objective measures of intraocular inflammation.*

32. Spalton DJ. Ocular fluorophotometry. *Br J Ophthalmol.* 1990;74 (7):431–432. doi:10.1136/bjo.74.7.431.

33. Cunha-Vaz JG. Vitreous fluorophotometry recordings in posterior segment disease. *Graefes Arch Clin Exp Ophthalmol.* 1985;222(4–-5):241–247. doi:10.1007/BF02133688.

Appendix

MEDLINE Sample Search Strategy

1	Exp Uveitis/
2	Uveiti*.ti,ab.
3	Inflamm*.ti,ab.
4	Leak*.ti,ab
5	Blood retinal barrier.ti,ab.
6	1 or 2 or 3 or 4 or 5
7	vitreous.ti,ab.
8	Vitreous haze.ti,ab.
9	vitritis.ti,ab.
10	7 or 8 or 9
11	6 and 10

Chapter 5: Developing an OCT-based technique for measuring vitreous inflammation

One of the key markers of inflammation in uveitis (particularly posterior segment-involving uveitis) is vitreous haze. As presented in **Chapter 4**, OCT is a promising technique for objectively quantifying vitreous haze. **Chapter 5** and **Chapter 6** describe the development and validation pathway of an OCT-based imaging technique, developed by our group, for measuring vitreous inflammation. **Chapter 5** provides a brief synopsis of the development pathway, including publications prior to my joining the team in 2017, as well as studies I contributed to between 2017-2020 (**Sections 5.2**). In **Section 5.2**, I briefly summarise the methods and findings of each study (published papers which I co-authored are included in the **Appendix**). **Chapter 6** presents the prospective clinical study (OCTAVE) where I was involved in the study design, undertook the ethics and sponsorship applications and led on all recruitment, follow-up, collection of data, and data analysis.

5.1 Background to technique

5.1.1 Proof of concept

The analysis of vitreous inflammation using OCT was a technique first described by Keane *et al* in 2014. The image analysis system (OCTOR) allowed delineation of the vitreous boundaries and summation of pixel-level intensity in that space as a surrogate measure of vitreous haze.(Sadda *et al.*, 2007, 2010) The underlying principles of this technique are based on the assumption that in higher levels of vitreous inflammation, inflammatory exudates leaked into the

vitreous body would cause higher levels of light scatter. As OCT imaging is based on the detection of light reflectivity, higher levels of light scatter would result in increased signal intensity detectable by OCT. The earlier technique involved two steps, first a manual boundary annotation is performed by an observer to mark out the vitreous space (between the uppermost edge of the B scan and the internal limiting membrane) and the retinal pigmented epithelium (RPE). A mean pixel intensity value is calculated for each area and a ratio is derived, termed the Vit:RPE relative intensity index.

In this first study, Keane and colleagues retrospectively derived the Vit:RPE relative intensity index from macular OCTs of 30 patients with uveitic vitreous haze and 30 patients (12 uveitis patients and 18 non-uveitis patients) without vitreous haze. A significant difference was detected when comparing eyes with uveitic vitreous haze versus eyes without vitreous haze (p = 0.0001) and between each grade of vitreous haze (p = 0.001). The correlation between NEI vitreous haze grading and Vit:RPE relative intensity index was moderate (Spearman's r = 0.566, p = 0.0001). Similarly, moderate correlation was shown with visual acuity (r = 0.573, p = 0.0001), AC cell grading (r = 0.613, p = 0.0001) and AC flare grading (r = 0.385, p = 0.003).

5.1.2 Different device and different patients

The proof-of-concept study was initially conducted using the Heidelberg SPECTRALIS OCT only. To test generalisability of the approach across different OCT devices, a subsequent study was carried out on a Cirrus spectral-domain OCT system (Cirrus HD-OCT; Carl Zeiss Meditec, Dublin, California, USA) on a separate cohort of uveitis patients.(Zarranz-Ventura *et al.*, 2016) In this retrospective study, macular OCT images from 105 patients from a different hospital and with various clinical grades of vitreous haze also demonstrated moderate correlation with the NEI VH scale (Spearman's r = 0.535, P<0.001). In addition, this study explored the effects of

phakic status (69 phakic and 32 pseudophakic) and vitrectomy (92 non-vitrectomised eyes and 13 with previous pars plana vitrectomy) on the Vit:RPE relative intensity index. No significant differences were detected across eyes of different phakic status or vitrectomy status, both across all levels of NEI VH grading and within each level of grading. However only eyes with grades 0, 0.5 and 1 were included in these sub analyses.

5.1.3 Automation

Although OCTOR automatically quantifies pixel intensity, it required up to 5-10 minutes per B scan for an experienced grader to manually segment the vitreous and RPE regions. This was acknowledged as a major limitation to the platform for real-world clinical use. In 2015, Keane *et al* published an automated platform called VITreous ANalysis software (VITAN), which automatically segments RPE and selects a rectangular area of vitreous.(Keane *et al.*, 2015) The only manual input required is a verification that the selected vitreous area is appropriate, after which the system proceeds to calculate the Vit:RPE ratio. This updated approach not only removed subjectivity arising from manual annotation, but also reduced the analysis time from 5-10 minutes to 2 seconds. The study uses the same retrospective cohort of uveitis patients as those included in the Keane 2014 paper and achieved comparable correlation with NEI VH grading as the manual method (r = 0.585 for the automated method and r = 0.566 for the previous manual method).

5.1.4 Detection of treatment response to local steroid injection

In 2016, Sreekantam *et al* reported the ability of the Vit:RPE relative intensity index to detect treatment response in a retrospective analysis of patients with uveitic cystoid macular oedema (CMO) receiving sub-Tenon's triamcinolone acetonide (STTA).(Sreekantam *et al.*, 2017) The pre-treatment and post-treatment OCTs were analysed using the OCTOR algorithm in the same

way as Keane et al. In this consecutive cohort of 22 patients, a significant reduction in the Vit:RPE relative intensity index was detectable post-treatment (p < 0.001). A moderate correlation (r = 0.534) between the change in Vit:RPE relative intensity index and change in the central macular thickness was also found. Although this study included only a small number of patients, it was the first longitudinal study to explore within-individual change of the measurement in response to treatment. Even though only 22 subjects were included, a highly significant difference was detectable by the OCT method.

5.2 Refining the OCT technique

Chapters 5.2 – 5.4 describes test development work on the OCT technique which I contributed to, where we aimed to refine the OCT technique and optimise the image acquisition protocol, prior to commencement of the OCTAVE study in **Chapter 6**. In these next sections, the methods and results of each study is summarised in more detail and the published papers are included in the **Appendix**.

5.2.1 Defining optimum scan protocol

I contributed to the data analysis, manuscript drafting and reviewing of this paper, which was published in *Montesano, G, et al. (2018). Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis. Sci Rep 8, 1648.* **(Appendix)**

5.2.1.1 Introduction

Previous retrospective studies have been reliant on standard-of-care macula OCT scans where a number of imaging settings were set at the default configuration. In the next stage, through prospectively collected scans, there was an opportunity to explore different image acquisition configurations, to see whether the technique could be further refined.

OCT image acquisition includes a number of settings which can be changed by the operator at the point of acquisition, which may affect quality and resolution of the OCT scan. For example, on the Heidelberg SPECTRALIS OCT device, the operator is required to manually bring the retina into focus, by adjusting a dial and making a judgement of when the retina has come into focus. This is a manual setting which has to be adjusted during almost every acquisition, by the operator. There is also the ability to adjust the Automated Real Time (ART) number. This is a function of the SPECTRALIS OCT device which combines multiple images captured in the same location in order to increase the signal-to-noise ratio. As the ART function is increased, the number of images acquired and averaged to derive the final image is increased, to a maximum number selected by the operator. As a result, faint signals are increased from the noise floor and the contrast between anatomical structures is enhanced.(Aumann *et al.*, 2019) Other common adjustable functions include positioning of the retina in the acquisition window and the number and configuration of B scans acquired.

This study sought to define the optimum operator settings to achieve the most reliable Vit:RPE relative intensity index measurement. We assessed the effects of several operator dependent settings on measurement variability.

5.2.1.2 Summary of methods

This was a prospective study involving 15 healthy subjects who underwent a series of repeated macular OCT scans using the Heidelberg SPECTRALIS OCT device. One eye of each participant underwent 10 different acquisition protocols, repeated 3 times each **(Table 8)**. All scans were acquired using the SPECTRALIS OCT device, using a 30-degree lens, in a 7 B scan volume configuration. Image analysis was carried out using the automated VITAN software as previously described to derive the Vit/RPE relative intensity ratio.

Table 8. Scanning protocols for testing different setting combinations used. Each scan protocol was repeated 3 times.(Montesano *et al.*, 2018)

Acquisition Protocol	Vertical position of retina with the scan	ART Level	Focus
1	Middle	100	In focus
2	Middle	50	In focus
3	Middle	25	In focus
4	Middle	12	In focus
5	Middle	6	In focus
6	Middle	100	+5
7	Middle	100	+10
8	Middle	100	-5
9	Middle	100	-10
10	Bottom	100	In focus

Vertical position referred to the position of the retina in the live acquisition window (or the *Z* plane). Adjustment of this position has a direct effect on how much vitreous volume is captured (i.e. retina positioned at the more inferior part of the window maximises the amount of vitreous captured). ART level can be set between OFF (1) - 100 by the operator (i.e. no image averaging - 100 images averaged). Focus is adjusted using a dial during acquisition and judged by the operator to bring the retina in focus in infrared image. 'In focus' is the default position and refers to when focus is at the level of the internal limiting membrane (i.e. retinal vessels are at the plan of focus), '+5' and '+10' dioptres is a positive shift of the focus from this default position and vice versa '-5' and '-10' is the opposite adjustment.

Linear mixed models were employed to assess:

1. the effects of different settings (ART level, focus and positioning) on the Vit/RPE relative intensity ratio and,

2. the variability of the measurement at the three levels:

a) Within acquisition: between the 7 B scans within each volume scan,

b) Within eye: between the 3 repeated volumes within each eye and

c) Between eye: between the 15 participant eyes.

5.2.1.3 Summary of results

5.2.1.3.1 ART level

Changing the ART level did not have a significant effect on the Vit/RPE relative intensity index (p=0.08) or the within volume (p = 0.308) and within eye (p=0.869) measurement variability. A moderate effect was found on the variability across subjects (p = 0.005). Pairwise comparison showed that ART 100 scans showed higher variability than ART 6 (p = 0.032, 3.99 fold increase) and ART 25 (p = 0.004, 5.41 fold increase) **(Table 9)**.

Table 9. Effect of the ART value on the Vit/RPE relative intensity index (reproduced from Montesano et al)

ART	Global mean ratio	Within scan variability	Intra - subject variability	Inter - subject variability
6	0.043	0.010	0.006	0.008
12	0.043	0.009	0.007	0.010
25	0.043	0.010	0.007	0.007
50	0.039	0.010	0.007	0.012
100	0.048	0.011	0.007	0.016*

On this basis, we determined that lower ART values are likely to produce the most reliable measurement.

5.2.1.3.2 Focus

Adjusting the focus to 10+ dioptres and 5+ dioptres caused 21 (6%) and 7 (3%) OCT volume acquisitions respectively to fail. This is a function of the SPECTRALIS device which will not

acquire the image unless sufficient quality is achieved. Image analysis by VITAN failed in 504/1575 B scans where the image was brought out of focus (32% failure rate).

We found that adjusting the focus had a significant effect on the Vit:RPE relative intensity index (p<0.001). All adjustments on focus (whether positive or negative dioptres) caused the Vit/RPE relative intensity index to increase. The measurement variability at all three levels were also increased (all p<0.001) for all adjustments in focus, except for the -5 dioptre adjustment **(Table 10)**.

Table 10. Effect of focus on the Vit/RPE relative intensity index (reproduced from Montesano et al)

Focus	Global mean ratio	Within scan variability	Intra - subject variability	Inter - subject variability
−10 D	0.162**	0.019**	0.015**	0.036
−5 D	0.088**	0.013	0.012*	0.029
In focus	**0.048**	**0.011**	**0.007**	**0.023**
+5 D	0.192**	0.032**	0.022**	0.040*
+10 D	0.200**	0.027**	0.033**	0.036

On this basis, we determined that setting the focal plane to the level of the retina is likely to produce the most reliable measurement.

5.2.1.3.3 Positioning

Lastly, positioning the retina at the bottom of the acquisition window significantly increased the Vit:RPE relative intensity index (p<0.001) and the variability at all three levels (p<0.01 for all levels) compared to positioning the retina in the middle of the acquisition window. On this basis, we determined that retina positioning at the middle of the acquisition window is likely to produce the most reliable measurement.

5.2.1.4 Conclusions

This study explored the extent to which operator dependent factors may impact the reliability of the Vit:RPE relative intensity index. Through this investigation, we were able to identify the operator settings which produced the most reliable measurements in healthy controls. We therefore defined a standardised image acquisition protocol, which is defined as:

A Macular OCT volume, where:

1. The retina is positioned in the middle of the acquisition window
2. The ART level is set at 9 (chosen pragmatically due to the minimal effect ART was shown to have on the measurement and because ART 9 is the default setting in most SPECTRALIS protocols)
3. The focal plane is set at the level of the retina.

5.2.2 Refining the OCT technique: minimum number of B scans required

I contributed to the data analysis and manuscript drafting and reviewing of this paper, which is published as *Terheyden, et al (2021). Automated quantification of posterior vitreous inflammation: optical coherence tomography scan number requirements. Sci Rep 11, 3271* **(Appendix)**.

5.2.2.1 Introduction

As the vitreous body is neither homogeneous, nor stationary, it follows that Vit:RPE relatively intensity measurements from B scans in different locations may vary. Various materials distributed in the vitreous humour, including degenerative material, inflammatory cells and exudates can cause variations in OCT reflectivity from B scan to B scan. Retinal OCT scans also only capture a relatively small area of vitreous, compared to the entire vitreous body.

Depending on the configuration of the scan pattern, it is possible to change the theoretical volume captured by OCT. Increasing the length of B scans (i.e. increasing the number of A scans) and the number of B scans will increase the overall area represented in the volume. The axial depth of each B scan is fixed and cannot be changed on most OCT devices. It might therefore seem reasonable to increase the B scan length and number as much as possible, to maximise the amount of vitreous included in one acquisition.

However, the theoretical volume has to be balanced by a number of other considerations. Increasing the length and number of B scans will increase the length of time required for each acquisition. This time demand increases the likelihood of movement artefact and can impact on quality of scan. Similarly, should the speed of acquisition between sequential B scans be slower than movement of the vitreous body, then opacities could theoretically be double counted (e.g. a vitreous floater which moves at the same speed as B scan acquisition could theoretically be captured by all B scan acquisitions). A longer scan time increases the burden on patients, especially those who naturally have trouble fixating on a spot for extended periods of time, meaning its use in these patients will be limited. Although time required for scan acquisition is a theoretical concern, its likelihood is minimal as the speed of acquisition for most OCT volume scans nowadays is a matter of seconds.

In the earlier work by Keane *et al.* and Zarranz-Ventura *et al*. macular volumes consisted of 3 to 5 and 128 B scans respectively.(Keane *et al.*, 2014; Zarranz-Ventura *et al.*, 2016) Our later test reliability study in healthy eyes used macular volumes with 7 B scans.(Montesano *et al.*, 2018) However, investigation of optimum number of B scans required has not been formally conducted. Therefore in this next study, we sought to define the minimum number of B scans required to achieve reliable Vit:RPE relative intensity index measurements.

5.2.2.2 Summary of methods

In this retrospective study, 49 macular OCT scans from 49 patients with uveitis were included. All OCT scans consisted of 19 B scans covering theoretical areas of 20°×15°. Alternative B scans per volume were selected and the average Vit:RPE relative intensity ratio of these 9 B scans were considered the 'reference' measurement for each volume. The measurements of five alternative configurations, consisting of 1, 3, 5 and 7 central B scans and 3 widely distributed scans, were compared to the reference measurement from 9 B scans **(Figure 8)**.

The intraclass correlation coefficient (ICC) between all B scan Vit:RPE relative intensity measurements per volume was calculated. The median absolute difference and limits of agreement for each configuration was compared to the reference measurement.

Figure 8. Illustration of an en face infrared image linked to an OCT volume scan that consists of 19 B-scans. *The white dots indicate which scans (green lines) have been included in the different sub selections.* (Terheyden *et al.*, 2021)
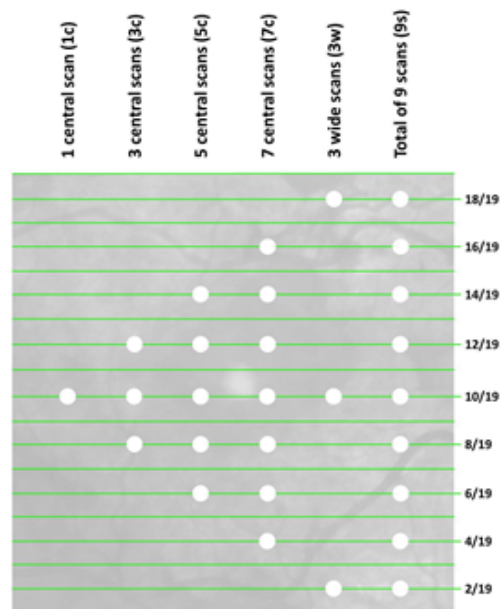
5.2.2.3 Summary of results

The ICC of the Vit:RPE relative intensity index across 9 B-scans for the included participants was 0.923 (95% confidence interval 0.886 – 0.952), indicating high agreement between B scan measurements.

We found the median Vit:RPE relative intensity index across all B scans was 0.029 (interquartile range 0.032), ranging from 0.0026 to 0.394. Compared to the reference (mean Vit:RPE relative intensity index of 9 B scans), the mean absolute difference and limits of agreement was smallest when 7 central scans are included and largest when only a single central scan is used. However, 3 wide B scans showed higher agreement than 5 central B scans, suggesting distribution across the whole theoretical area is potentially more important than number of B scans **(Table 11)**.

Table 11. Mean absolute difference and limits of agreement between the reference standard (mean measurement of 9 B-scans) and each configuration.(Terheyden *et al.*, 2021)

Scan sub selection	Median Absolute Difference from reference standard (IQR)	Limits of agreement compared to reference standard
1 central scan (1c)	0.006 (0.009)	[-0.039;0.037]
3 central scans (3c)	0.005 (0.011)	[-0.033;0.032]
5 central scans (5c)	0.004 (0.009)	[-0.028;0.026]
7 central scans (7c)	0.001 (0.004)	[-0.009;0.009]
3 wide scans (3w)	0.003 (0.005)	[-0.014;0.013]

5.2.2.4 Conclusions

This study provided evidence that the Vit:RPE relative intensity index is consistent between B scans in a volume. Where scan workload is problematic, 3 widely distributed B scans can be used to produce a reliable measurement, similar to including 5-7 central B scans. Similarly, where speed of scan acquisition is critical (such as when dealing with subjects with poor fixation), 3 widely distributed B scans could be sufficient.

5.2.3 Refining the OCT technique: Moving to raw images

In all previously mentioned studies, the image analysis technique has been based on OCT image exports. The images used are exported in the standard format offered by the Heidelberg Eye Explorer software. The images produced are subject to a number of image intensity adjustments made by the software, which could potentially affect the derived Vit:RPE relative intensity index. For example, as OCT scans are used primarily to assess retinal structures, retinal structures (i.e. hyperreflective areas) are preferentially enhanced in order to make the retinal layers appear brighter. Similarly, faint reflectivity signals which could represent noise artefact are dampened and appear darker. These adjustments are embedded within the Heidelberg Eye Explorer software and not accessible to the user. However, without knowledge of these transformations, we could not know the true vitreous signal. This raised concerns that any imposed correction functions, intended to decrease noise and enhance perceived image quality (to the human eye), could either cancel out or enhance clinically relevant vitreous signals. This prompted us to explore the possibility of exporting Heidelberg OCT images in the raw format.

Through collaboration with Heidelberg Engineering, we were permitted access to the raw OCT exports and were able to utilise the uncorrected OCT signal within each B scan. With this functionality unlocked, a new custom image analysis software was developed (by Ometto G), called EQUIP (*Extended OCT Quantification in Uveitis to Inform Practice*). EQUIP is able to apply the same technique as previously described on raw OCT exports, which has intensity corrections and brightness transformations removed. Like OCTOR and VITAN, EQUIP first segments each B scan to isolate vitreous space and then automatically quantifies the average signal intensity within the vitreous space. However, EQUIP does not utilise the concept of Vit:RPE relative intensity ratio, but instead derives an averaged absolute reflectivity measure for

the vitreous space only. This change was made possible by moving to analysis of the raw OCT signal, but also motivated by the concern that in cases of RPE disruption (such as in the case of inflammatory macular oedema and choroidal neovascular membrane), the RPE measurement within the Vit:RPE ratio would be disproportionately affected.

In the latest version of EQUIP, raw images are exported in the proprietary Heidelberg OCT format ready for analysis. The image analysis summates the total intensity in the vitreous region and the total intensity in the entire B scan (which is taken as a representation of the total signal reaching the retina). The vitreous intensity is calculated as a proportion of the total B scan intensity. In simplistic terms, the measurement can be thought of as a fraction where the numerator is the total intensity of the vitreous area (red area) and the denominator is the total intensity of the B scan (red area plus green area) **(Figure 9).** The final output is a continuous measurement which has been normalised by logarithmic transformation and the measurement is recorded in arbitrary units. The measurement is a negative value, where a more negative value suggests lower image intensity and a more positive value suggests higher image intensity.

Figure 9. Segmented B scan (A) with region of interest being the superior purple area (B) representing the vitreous.
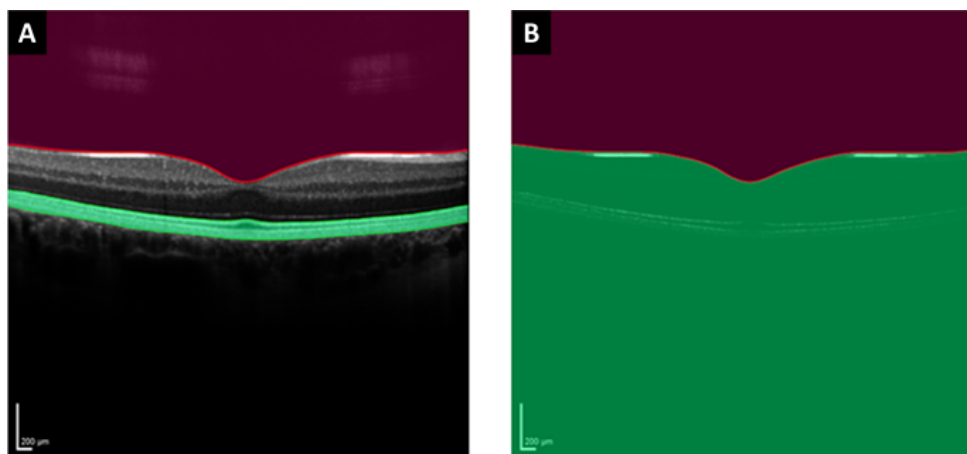
5.3 The final OCT technique protocol

Based on the work described in this chapter, a recommended acquisition protocol was defined for the prospective clinical study in **Chapter 6**. The final parameters were decided based on a balance between achieving the most reliable measurement and defining a scan protocol which would not be too onerous for the operator or the patient. Additional considerations included, whether the scan protocol could be tailored similarly to others which are routinely used for retinal disease. For example, the 'Fast Macular' preset on the Heidelberg SPECTRALIS OCT is a commonly used protocol for most medical retina clinics (an OCT volume consisting of 25 B scans set at ART 9) and which operators and patients are already familiar with. The intention of the final protocol is to achieve a balance of test reliability versus minimising the burden on operators, maximising acceptability to patients and allowing the best chances of adoption in practice.

The final imaging protocol **(Figure 10)** for OCTAVE was:

- Heidelberg SPECTRALIS OCT using a 30° lens;
- 7 20° horizontal B scans, centred on the macula;(Terheyden *et al.*, 2021)
- ART level set at 9;(Montesano *et al.*, 2018)
- Scan operators were instructed to position the retina in the middle of the scan acquisition window and focus the image at the level of the internal limiting membrane, as is conventional practice for OCT.(Montesano *et al.*, 2018)

Figure 10. Example of 7 B scan volume with en face infrared image (left) and the cross-sectional OCT is shown (right).

The most inferior B scan (position 1/7) is selected and shown in the OCT window (left).
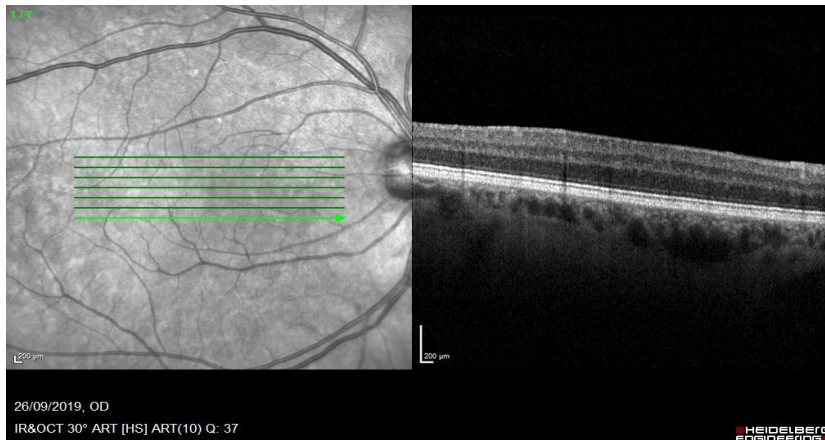
The scans acquired using this protocol would require exporting in the raw format and analysed using the custom EQUIP module.

Chapter 6: OCTAVE – a prospective study evaluating the OCT technique in healthy and uveitic eyes

Optical Coherence Tomography-Assisted Vitreous Evaluation (OCTAVE) sponsored by University Hospitals Birmingham NHS Foundation Trust (RRK6369) is a prospective observational study to validate the novel OCT-derived vitreous haze technique. It was conducted between January 2018 to September 2019, with the last patient recruited on 26th September 2019 and data locked on 17th February 2020. The OCTAVE study is supported by a Wellcome Trust Health Innovation Challenge Grant (200141/Z/15/Z).

OCTAVE has two key components: the evaluation of the OCT-derived vitreous haze technique in healthy eyes and evaluation in eyes with uveitis.

First, the technique is evaluated in healthy eyes. This part of the study assesses the test-retest variability of the measurement in healthy eyes and determines whether there is significant

variation between the B scans of each volume within the same eye. This part of the study informs what the expected mean and normal range of the vitreous intensity value is in healthy eyes.

Second, the technique is evaluated in uveitic eyes. The within eye test-retest variability is evaluated in uveitis eyes with varying levels of vitreous haze (as measured by the NEI vitreous haze grading system). This was done to assess the impact of uveitis (both active and quiescent) on the test reliability. The ability of the test to discriminate presence and absence of inflammation was assessed by comparing the OCT measurement in healthy eyes, uveitic eyes with vitreous haze and uveitic eyes without vitreous haze. Lastly, the association between the OCT-derived vitreous haze measurement and other clinically important measurements such as NEI vitreous haze grading, central macular thickness and visual acuity is assessed. The OCTAVE study report is currently in preparation for submission for peer-reviewed publication.

6.1 Evaluation of technique in healthy eyes

6.1.1 Introduction and aims

As highlighted earlier in **Section 1.5 (Evaluation of a new test)**, before seeking to evaluate the OCT technique in the uveitis (clinical validation in disease), it is important to ensure good reliability in healthy eyes in a controlled environment (technical validation and clinical validation in health). Initially, this would include assessment in healthy eyes using repeated measures on the same visit. Other factors may contribute to the overall technical variability, for example the variation introduced by using different physical OCT machines (of the same model) and different operators acquiring the scan.

The variability of vitreous intensity in different parts of the eye is also unknown. In particular, it is unclear how spatial distribution along the vertical axis affects the vitreous. It is possible that there is a gravitational effect on vitreous debris that a higher vitreous intensity signal is typically found in inferior parts of the OCT volume compared to the superior parts. Previous work using VITAN in healthy subjects has suggested variations in the Vit:RPE relative intensity index in different intraocular areas, with highest density seen in the inferior portion.(Barry *et al.*, 2016).

The aim of this study was to (1) define the expected normal range for vitreous intensity within healthy eyes, to inform interpretation of measurements in abnormal eyes, (2) estimate the reliability of the OCT measurements within healthy eyes, by measuring the test-retest variability between repeated scans in the same eye, and (3) measure the variability between individual B scans on a vertical axis of each volume is also assessed to explore whether there is a difference in spatial distribution of vitreous intensity.

6.1.2 Methods

Participant recruitment

In this study, healthy participants were prospectively recruited between January 2018 to December 2019 at University Hospitals Birmingham NHS Foundation Trust. Participants were hospital staff or family members/carers of patients attending the ophthalmology outpatient department. All participants underwent an ophthalmic examination to exclude the presence of any pathology. This protocol was approved by the NRES East Midlands Ethics Committee (Ref:14/EM/1163) and written consent was given by all subjects. This protocol adhered to the tenets of the declaration of Helsinki.

Image acquisition

Participants did not receive dilating drops as OCT scans can typically be acquired in undilated healthy eyes. OCT scans were performed using the Heidelberg SPECTRALIS OCT (Heidelberg Engineering, Heidelberg, Germany) using a 30 degree lens, as per the pre-specified protocol (7 horizontal 20 degree B scans, centred on the macula, covering a theoretical area of 20° by 5° (5.9mm by 1.5mm) and with the macula positioned in the middle of the B scan). For each participant, the right eye was scanned three consecutive times by the same operator. Each OCT image was exported in the raw format and processed using the custom EQUIP module to derive the vitreous intensity measurement.

Statistical Analysis

A multi-level mixed model was used for this analysis to account for the clustering (7 B scans per volume and 3 repeated volumes per subject). The intraclass correlation (ICC) at the within subject-eye and between subject-eye levels were calculated. To assess the effects of spatial distribution on vitreous intensity, we used the same model to assess the difference between vitreous intensity value across the 7 B scan positions, using the most inferior B scan as a reference. All statistical analysis was performed using Stata 16 (StataCorp LLC, Texas, USA).

6.1.3 Results

Forty-one eyes of 41 healthy subjects were included in the study. Subjects included 16 (39%) males and 25 (51%) females with a mean age of 48 years (SD $\pm$19). All participants underwent the complete scan protocol, resulting in 123 scans in total. The mean (SD) vitreous intensity value across all healthy eyes was -25.04 ($\pm$1.77; 95% CI of -24.5 to -25.5) units **(Figure 11)**.

Figure 11. Distribution of OCT vitreous intensity measurements in 41 healthy subjects.
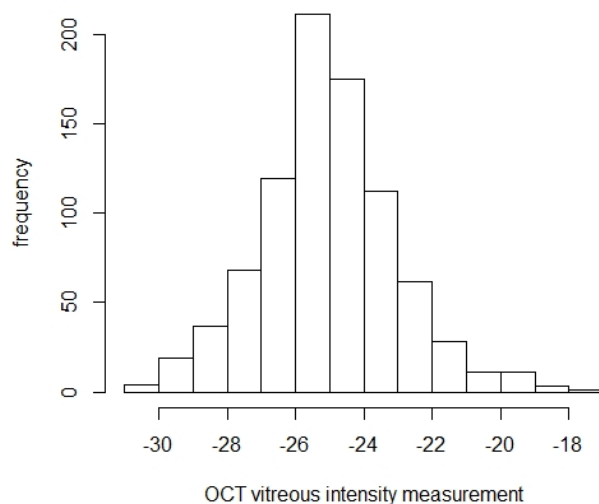
In the test-retest variability analysis of three repeated scans in each eye, the within-eye ICC was 0.79 (95% CI:0.71-0.85) and the between-patient eye ICC was 0.75 (95% CI:0.65-0.86). When comparing vitreous intensity measurements across the 7 B scan positions **(Figure 10)**, a significant difference was detectable between the middle B scans: position 4 ($p<0.005$), position 5 ($p<0.005$) and position 6 ($p<0.05$) in comparison to the most inferior B scan (position 1) **(Table 12)**.

Table 12. Mean vitreous intensity in the 7 different B scan positions in healthy eyes.

B scan position	Mean vitreous intensity (SD)
7	-25.0 (1.98)
6*	-25.1 (1.85)
5**	-25.3 (1.99)
4**	-25.3 (2.22)
3	-25.1 (1.95)
2	-24.7 (1.83)
1 (reference)	**-24.8 (2.07)**

1 = most inferior B scan, 7 = most superior B scan.
*$p<0.05$, **$p<0.005$ within the multilevel mixed model when compared to reference position; all other positions were non-significant.

6.1.4 Discussion

This study assessed test-retest variability of the vitreous haze measurement in healthy eyes and the effect of spatial position of B scans on the vitreous haze measurements. It found that in a controlled environment, with a single operator taking repeated images on the same visit, within-eye ICC was 0.79 (considered high reliability) and between-eye reliability was 0.75 (considered moderate reliability) in healthy controls.(Portney, Watkins and Others, 2009) Position of individual B scans can affect the measurement, with central B scans showing significantly lower vitreous intensity measurements compared to the most inferior B scan, whereas the most superior B scan was not significantly different.

This study showed that the current OCT technique demonstrates good reliability. There are several possible explanations as to why B scan position can affect the measurement. There may be a true biological variability in the composition of the vitreous, or it could be due to technical limitations of the imaging technique. The focal point of OCT volume is typically set with the central B scan positioned centrally at the macula (i.e. in the position of the 4th B scan). With the natural curvature of the retina and the globe, B scans further away from the central point are further away from the focal point and therefore are more out of focus. Increased noise artefact in the image attributable to poorer image quality may be misinterpreted as high vitreous intensity by the EQUIP algorithm.

This study has several limitations. First, the sample size is relatively small and was determined by published 'rule-of-thumb' guidance, rather than any formal sample size calculations, as there are no pre-existing estimates of the measurement variance.(McNeish and Stapleton, 2016) Second, this study only explored within/between subject-eye test-retest variability by one operator and on the same visit. In order to assess how the test reliability may be adversely affected by different operators, subsequent studies should include different operators, with varying degrees of experience, performing repeated scans on the same visit. Furthermore,

subjects should be scanned at different points in the day or on different days to determine the

biological variability of the measurement in the same patient over time. Third, whilst all scans

were successfully acquired with undilated pupils, it is unclear whether pupillary dilation would

have improved scan quality and/or changed the measurements. Future studies in healthy eyes

should take pre and post-dilation OCT scans to evaluate whether pupillary dilation has an effect

on the measurement.

This study provided evidence that the vitreous intensity measure showed good reliability in the

absence of any pathology. The mean and variation (SD) in healthy eyes were established,

which will inform interpretation of measurements in uveitic eyes. Different B scan positions can

give rise to variations in the measurement, with the central B scans showing lower vitreous

intensity levels than the peripheral B scans (both superior and inferior). In the study described in

Section 5.2.2 variability between 9 B scans of a volume was assessed. Although that analysis

found an ICC of 0.92 (95% CI 0.89 – 0.95) across all B scans within a volume, they also found

including a small number of peripheral B scans (3 wide B scans), was more likely to give a

similar measurement to the averaged measurement across all B scans, compared to if only the

central B scans were included (5 central B scans). Therefore, although the overall variability

across B scans was low, differences in peripheral B scan measurements may have been

disproportionately contributing to the variation. This would be in keeping with the results shown

in this study and can be explained by either true biological variation in the vitreous composition,

or by the optical effects of moving away from the focal plane. This would not be a problem if

focal measurements (i.e. single B scan) was sufficiently representative of the whole vitreous

volume, however the vitreous is not a homogenous structure and we would ideally want to

maximise the number of B scan to provide a representative global measure of the entire

vitreous body. Given the need to capture a wide area of vitreous and considering any optical

effects of peripheral scans is likely to affect all volumes in the same way, in subsequent work

the mean of all 7 B scans will be calculated to derive a single global vitreous intensity per eye.

The first part of the OCTAVE study demonstrated good within-eye reliability of the OCT

technique in healthy eyes. The mean and expected variability (SD) of the measurement in

healthy eyes was established, which will inform interpretation of measurements in the

subsequent study in uveitic eyes.

6.2 Evaluation of the technique in uveitic eyes

6.2.1 Introduction and aims

The first part of OCTAVE demonstrated that the OCT-based measurement of vitreous intensity

is reliable in healthy eyes. The second part of OCTAVE aims to determine whether the

measurement is similarly reliable in eyes with uveitis. This study also aims to determine the

clinical validity of the OCT-based measurement by assessing its relevance to clinically important

outcomes. First, the ability of the measurement to discriminate between healthy eyes and eyes

with uveitis is assessed. Then, the association between the OCT vitreous intensity and NEI

vitreous haze grading, as well as other clinically important variables in uveitis such as central

macular thickness and visual acuity, is assessed.

6.2.2 Methods

Participant inclusion

For this study, participants with uveitis and varying degrees of vitreous haze, including no

vitreous haze, are included. Patients who were attending the uveitis clinic at University Hospitals

Birmingham NHS Foundation Trust between January 2018 to December 2019 were

consecutively enrolled into the study. All patients with a diagnosis of posterior segment-involving uveitis were included. Exclusion criteria included those younger than 16 years and those where the diagnosis was subsequently confirmed to not be uveitis (including masquerade syndromes such as lymphoma). Healthy participants included in the analysis were the same cohort described in **Section 6.1**. All eligible participants were prospectively enrolled and informed consent given. This protocol was approved by the London-South East Research Ethics Committee (18/LO/1332). This protocol adhered to the tenets of the declaration of Helsinki.

Image acquisition

Dilated OCT macula scans were performed as described earlier using the Heidelberg SPECTRALIS OCT (Heidelberg Engineering, Heidelberg, Germany) using a 30 degree lens, as per the pre-specified protocol (7 horizontal 20 degree B scans, centred on the macula, covering a theoretical area of 20° by 5° (5.9mm by 1.5mm) and with the macula positioned in the middle of the B scan)**.** For each participant, the eye was scanned three consecutive times by an experienced OCT technician prior to clinical examination. The OCT technician did not have prior knowledge of the patient's disease history or clinical status. Each OCT image was exported in the raw format and processed using the custom EQUIP module to derive the vitreous intensity measurement.

Clinical examination

At each routine clinical visit, participants had visual acuity tested (Snellen chart at 6 metres) and underwent routine dilated ophthalmic examination by an experienced uveitis specialist. The disease locality, AC cell grade (SUN grading(Jabs *et al.*, 2005)), phakic status (recorded as phakic, pseudophakic or cataract), vitreous haze grade (NEI vitreous haze grading, assessed through indirect ophthalmoscopy) were recorded. All OCT analysis for the vitreous intensity was

carried out at the end of the study, therefore the uveitis specialist had no knowledge of the OCT-derived vitreous intensity measurement. Additionally, the OCT central subfield thickness (CST), as automatically calculated by Heidelberg Eye Explorer, was taken as a measure of macular thickness.

Statistical analysis

Test reliability: A multi-level mixed model was used for this analysis to account for the clustering (the average of 7 B scans per volume and 3 repeated volumes per subject). The intraclass correlation (ICC) at the within subject-eye and between subject-eye levels were calculated across the whole cohort and separated by NEI vitreous haze grade.

Ability to discriminate eyes with and without inflammation: The study cohort was split into three groups: healthy subjects (the study cohort from Section 6.1), uveitic eyes with no vitreous haze (NEI vitreous haze grade 0) and uveitic eyes with vitreous haze (NEI vitreous haze grades 0.5 and above). The vitreous intensity measure for 7 B scans in each volume was averaged to derive a single measurement per eye. A simple linear regression was fitted, with the OCT vitreous haze measurement modelled as the outcome and the three groups treated as a categorical independent variable, to assess differences between each group.

Association with NEI vitreous haze grading: To assess the association between the OCT vitreous intensity measurement with NEI vitreous haze grading, the mean difference of the OCT vitreous measurement at each NEI grade was calculated compared to the reference measurements from the healthy control group described in **Section 6.1**. The NEI grading was treated as a categorical independent variable and OCT vitreous measurement modelled as the

outcome. The Wald Chi Squared Test was used to assess for overall significance of difference between OCT intensity measurements in each NEI grade.

The effect of age, phakic status and AC cells on the OCT vitreous intensity measurement:
The effects of age, phakic status (grouped as cataract present or absent) and AC cells (SUN grade 0 versus SUN grade 0.5+ and above) on the OCT vitreous intensity measurement was evaluated in a multivariate analysis whilst adjusting for severity of inflammation as per the NEI vitreous haze grade (categorised as NEI VH grade 0, 0.5+, 1+ or 2+ and above). The OCT vitreous intensity measurement was modelled as the outcome and predictors were age, phakic status, AC cells and NEI vitreous haze grading.

Ability of OCT vitreous intensity to predict VA and CMT: To explore whether the OCT vitreous intensity measurement can predict VA and CMT, two separate multivariate analyses were conducted. In both analyses, OCT vitreous intensity and NEI vitreous haze grading are treated as predictors, with NEI vitreous haze grading grouped as grade 0, 0.5+, 1+ and 2+ and above. In the first multivariate analysis, which models VA as the outcome, age, phakic status (grouped as cataract present or absent) and CMT are included as predictors. VA was converted from snellen to logMAR (counting fingers (CF) and hand movements (HM) were converted to 2.1 and 2.4 logMAR, respectively).(Sparrow *et al.*, 2012) In the second multivariate analysis, where CMT is modelled as the outcome, AC cell grading (grouped as AC cells present/absent) and NEI vitreous haze grading are treated as predictors, with NEI vitreous haze grading grouped as grade 0, 0.5+, 1+ and 2+ and above.

One eye per participant was included in the analysis. As patients with active vitreous inflammation were uncommon, for all analyses, the eye with inflammatory activity (defined as NEI vitreous haze grade 0.5+ or presence of cystoid macular oedema, active chorioretinal lesions and active vasculitis) was purposefully selected to oversample this cohort. In the

absence of disease activity, or if both eyes showed active disease, the right eye was selected.

All statistical analysis was performed in Stata 16 (StataCorp LLC, Texas, USA).

6.2.3 Results

Participant characteristics

Seventy-eight subjects participated in this study, including 46 subjects with vitreous haze in one or both eyes and 32 without vitreous haze in either eye. 41 healthy subjects described in **Section 6.1** were also included in the analysis for comparison. One subject with grade 4+ vitreous haze (NEI VH scale) was excluded from the analysis as the vitreous density was so severe that an OCT image could not be acquired. The participant characteristics are summarised in **Table 13.** Vitreous intensity (SD) for participants without haze was -24.2 ($\pm$1.5) and for participants with haze was -22.8 ($\pm$1.9).

Table 13. OCTAVE Study: Characteristics of recruited patients with uveitis

	Uveitis without haze (32)	Uveitis with haze (45)
Sex		
Male (%)	11 (34%)	17 (38%)
Female (%)	21 (66%)	28 (62%)
Mean age (SD)	52 (+17)	51 (+18)
Ethnicity		
White	21	33
Asian/Asian British	6	10
Black/African/Caribbean/Black British	2	2
Mixed/Multiple ethnic groups	3	0
Aetiology		
Idiopathic	16	22
Behçet's disease	1	0
Birdshot chorioretinitis	0	3
Cytomegalovirus retinitis	0	1
Fuchs' heterochromic cyclitis	0	1
Herpes zoster ophthalmicus	2	2
HLA-B27 associated uveitis	2	2
Juvenile idiopathic arthritis	1	3
Multiple sclerosis	1	1
Tuberculosis	1	1
Multifocal choroiditis	1	0
Punctate inner choroidopathy	1	0
Retinal vasculitis	2	0
Sarcoidosis	3	8
Toxoplasmosis	1	1
Phakic status		
No cataract	12	17
Cataract	11	15
Pseudophakic	9	13
Anatomical subtype of uveitis		
Anterior	3	5
Intermediate	15	16
Posterior	4	7
Panuveitis	10	17

Measurement repeatability in uveitic eyes

The OCT vitreous intensity measurement showed high within-eye repeatability at all grades of vitreous haze in the uveitic eyes (ICC > 0.79 at all grades). **Table 14** shows the within-participant eye reliability separated by NEI vitreous haze grade. Only one eye was tested per participant.

Table 14. Within-participant eye ICC from the 3 repeated measurements in eyes with vitreous haze of varying severity.

NEI vitreous haze grade	Number of eyes	Within-participant eye ICC (95% CI)
0	32	0.83 (0.75 - 0.89)
0.5+	25	0.90 (0.83 - 0.94)
1+	13	0.89 (0.78 - 0.95)
2+	4	0.79 (0.46 - 0.94)
3+	3	0.93 (0.72 - 0.99)

Ability to discriminate between eyes with and without vitreous haze

There were 32 uveitic eyes with no vitreous haze (NEI vitreous haze grade 0) and 45 uveitic eyes with vitreous haze (NEI vitreous haze grades 0.5 and above). Compared to the 41 healthy eyes described in Section 6.1, both uveitis eyes with and without haze showed significantly different OCT vitreous intensities **(Table 15)**.

*Table 15. Mean OCT vitreous intensity score in eyes with and without vitreous haze. Individual patient-eye data is provided in supplementary figure 1 of the **Appendix**.*

Group	Number of eyes	Mean OCT vitreous intensity (SD)
Healthy eyes	41	-25.0 (+1.77)
Uveitic eyes - quiescent (NEI vitreous haze grade 0)	32	-24.2 (+1.46)*
Uveitic eyes - active (NEI vitreous haze grade 0.5 and above)	45	-22.8 (+1.87)**

*p<0.05, **p<0.005 in the simple linear regression

Association with NEI vitreous haze grading

In the uveitic eyes, the OCT vitreous haze measurement was significantly increased at NEI vitreous haze grades 0, 0.5, 1, 2 and 3 compared to measurements obtained from healthy controls ($p < 0.005$) **(Table 16).** Mean vitreous intensity (SD) for participants with grade 0 was -24.2 (+1.46), for grade 0.5+ was -22.7 (+2.0), for grade 1+ was -22.7 (+0.7), for grade 2+ was -22.7 (+0.7) and for grade 3+ was -22.5 (+1.3) **(Figure 12)**.

Table 16. Mean difference of OCT vitreous intensity at NEI vitreous haze grades 0.5+ to 3+ compared to grade 0.

NEI vitreous haze*	Number of Eyes	Mean difference above reference (95% CI)
Healthy eyes	41	Reference
0	32	1.01 (0.42, 1.61)*
0.5	25	2.47 (1.83, 3.11)**
1	13	2.61 (1.81, 3.42)**
2	4	2.52 (1.20, 3.85)**
3	3	2.28 (0.76 - 3.79)*

*p<0.05, **p<0.005 in the Wald Chi Squared Test

Figure 12. Box and whisker plot showing median (IQR) of OCT vitreous intensity measurements in each NEI vitreous haze grade. Individual patient-eye data is provided in supplementary figure 2 of the Appendix.
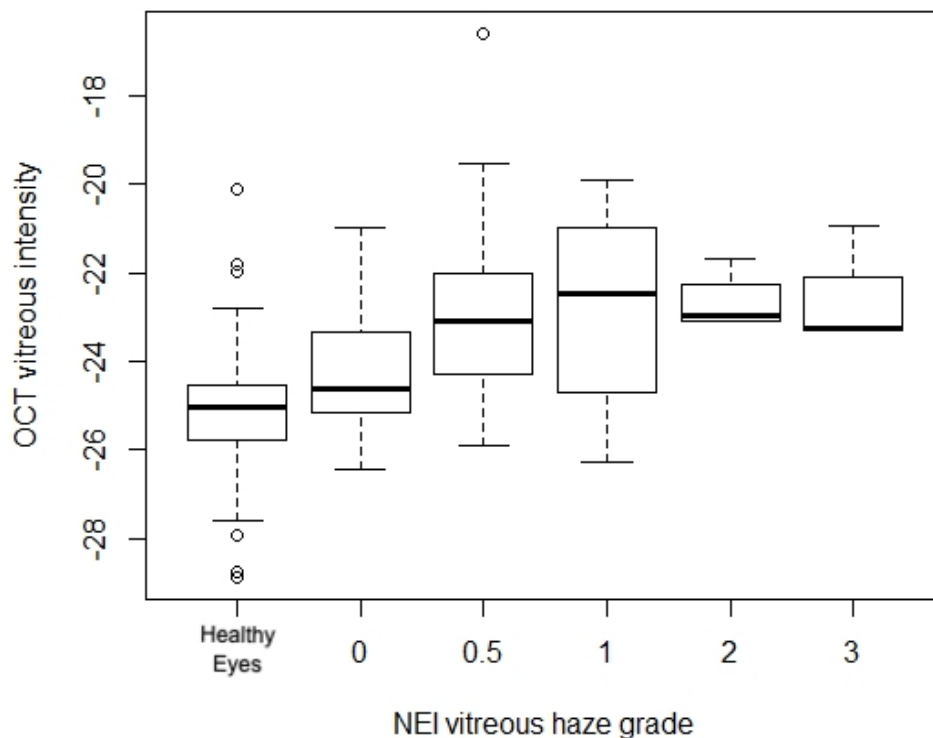
Minimum point is 1.5 x IQR below first quartile, and maximum point is 1.5 x IQR above first quartile. Circles denote outliers.

Effects of age, phakic status and AC cells on the OCT vitreous intensity measurement

In the multivariate analysis, the presence of cataract was shown to have a significant association with increased OCT vitreous measurement (p = 0.03) when adjusting for NEI vitreous haze grade **(Table 17; Figure 13)**.

Table 17. Effect of phakic status on OCT vitreous intensity (Only one eye for grades 2 and 3 where cataract was present).

NEI vitreous haze grade	Mean OCT vitreous intensity (SD)	
	Cataract not present	Cataract present
0	-24.4 (1.44)	-23.8 (1.48)
0.5	-23.2 (1.42)	-22.1 (3.14)
1	-23.1 (2.13)	-21.6 (1.33)
2	-23.0 (0.12)	-21.7 (-)
3	-22.1 (1.63)	-23.3 (-)

Figure 13. OCT vitreous intensity in eyes with and without cataract.

Point size indicates severity of vitreous haze as assessed by NEI vitreous haze scale (smallest = grade 0, largest = grade 3).
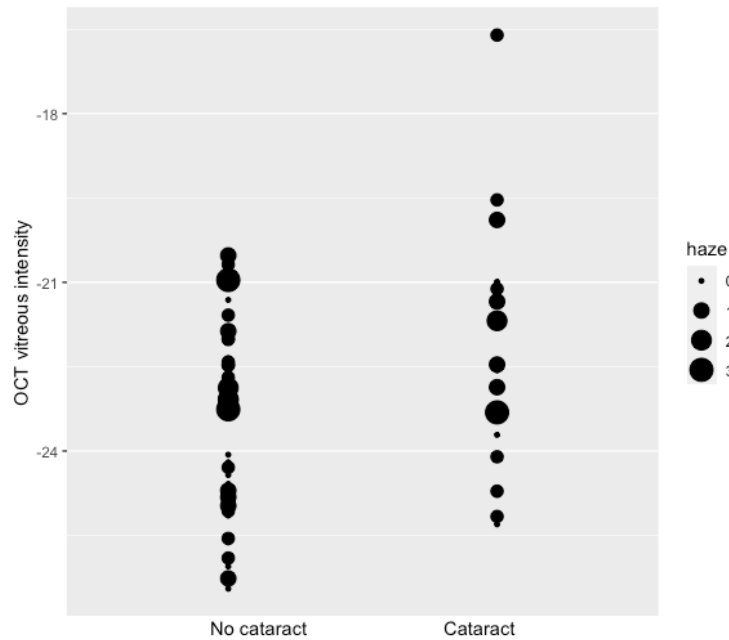
No significant association was detected between the OCT vitreous measurement and patient age ($p = 0.05$) or presence/absence of AC cells ($p = 0.88$).

*Table 18. Mean OCT Vitreous Intensity at difference AC cell Grades (SUN AC Cells Grading System). Individual patient-eye data is provided in supplementary figure 3 of **Appendix**.*

AC Cell Grade	Number of Eyes	Mean OCT vitreous intensity (SD)
0	49	-23.6 (1.65)
0.5	17	-23.1 (2.26)
1	2	-24.0 (1.09)
2	4	-23.9 (1.62)
3	3	-20.7 (0.75)
4	1	-21.0 (-)

Ability to predict VA and CMT

In the multivariate analysis where VA was modelled as the outcome, the OCT vitreous intensity measure showed a significant association with VA (p = 0.04) **(Figure 14)**, whereas no association could be found with the NEI vitreous haze (p = 0.17, p = 0.70 *and p* = 0.07 respectively for grades 0.5+, 1+ and 2+ and above). A strongly significant association was found between CMT and VA (p < 0.005). No association was found between phakic status and VA (p = 0.65).

Table 19. Mean visual acuity in different NEI Vitreous Haze Grades

NEI Vitreous Haze Grade	Number of Eyes	Mean LogMAR vision (nearest value, SD)
0	32	0.3 (0.5)
0.5	25	0.4 (0.6)
1	13	0.4 (0.7)
2	4	0.3 (0.3)
3	3	-0.5 (0.5)

Figure 14. Association between OCT vitreous intensity measurement with visual acuity (LogMAR).

Point size indicates severity of vitreous haze as assessed by NEI vitreous haze scale (smallest = grade 0, largest = grade 3).
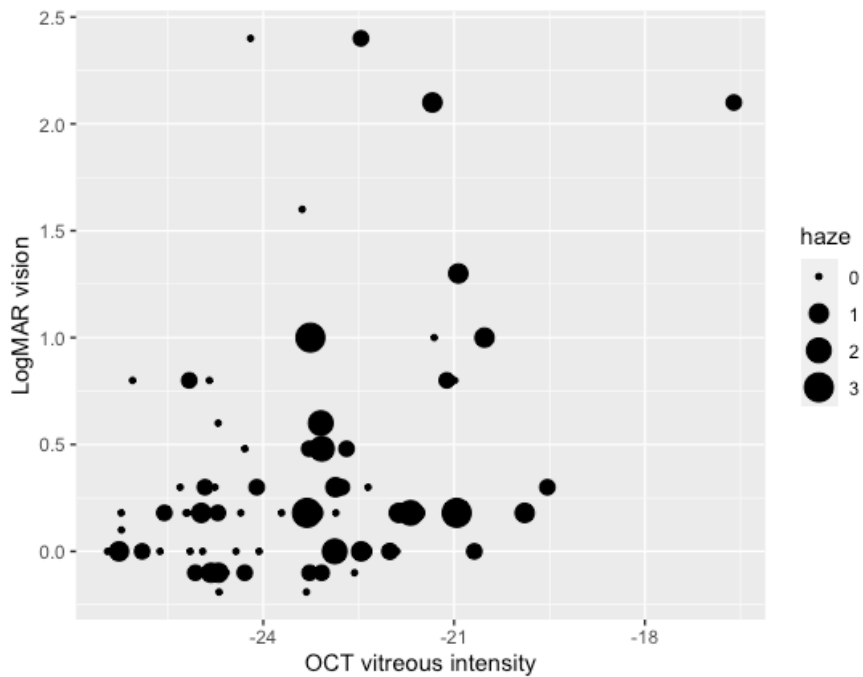
In the multivariate analysis where CMT was modelled as the outcome, neither the OCT vitreous intensity measure (p = 0.35) **(Figure 15)** or the NEI vitreous haze grades (p = 0.33 for grade 0.5+, p = 0.99 for grade 1+ and p = 0.04 for grade 2+ and above) were associated with CMT. Presence of AC cells also did not show association with CMT (p = 0.14).

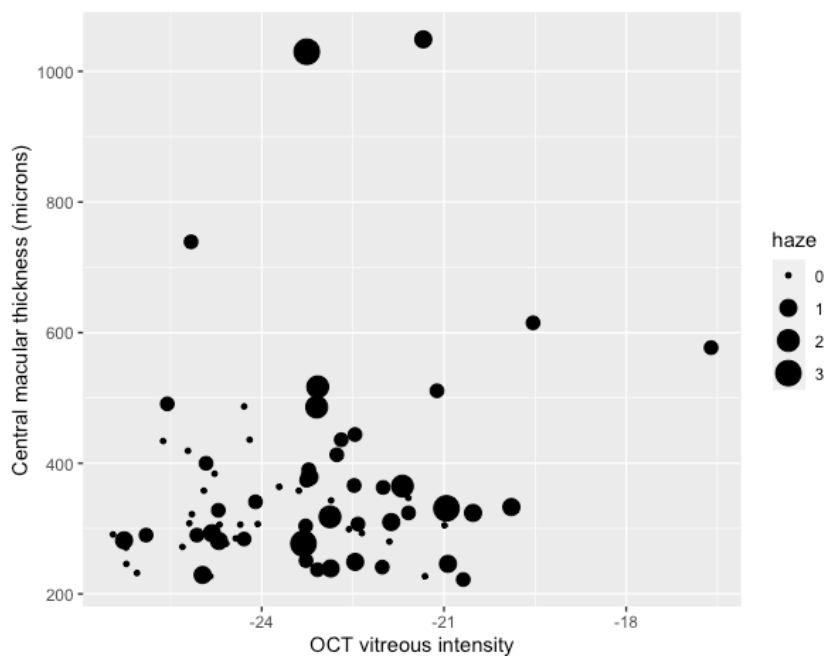*Figure 15. Association between OCT vitreous intensity measurement with central macular thickness.*

Point size indicates severity of vitreous haze as assessed by NEI vitreous haze scale (smallest = grade 0, largest = grade 3).

6.2.4 Discussion

This study was the first prospective clinical evaluation of the OCT vitreous intensity measurement in uveitis patients. Using the updated EQUIP system, we found the within subject test-retest variability of the measurement to be comparable with healthy eyes (ICC >0.79 in all grades of NEI vitreous haze compared to 0.79 in healthy eyes) and that the measurement was significantly different between healthy eyes, uveitic eyes without vitreous haze and uveitic eyes with vitreous haze. The OCT vitreous intensity also showed an association with NEI vitreous haze grading, with measurements in each increasing NEI grade showing significant incremental increase in OCT intensity. However, substantial overlaps in the OCT measurement were

observed between NEI vitreous haze grades. These overlaps may be attributable to poor signal-to-noise ratio from the OCT technique, but could also be attributable to the imprecision and poor reliability of clinician-based grading using the NEI vitreous haze scale.

The presence of cataract can increase the OCT vitreous measurement. This fits with the underlying principle of the OCT technique, which measures light scattering caused by vitreous debris and exudates. It is therefore expected that the presence of other light scattering media opacity such as cataracts can confound the measurement. Other anterior structures which may cause light scatter include AC cells, but in this study the presence of these were not shown to affect the OCT vitreous measurement. Age has also been associated with changes in the vitreous compositions, but was also not shown to be associated with the measurement.(Sebag, 1987)

In the multivariate analysis, we were able to detect a significant association between the OCT vitreous measurement and visual acuity, whereas none could be shown for NEI vitreous grading and VA. This association was significant even after adjusting to other potential confounders of reduced VA, including phakic status and CMT. This suggests the OCT vitreous measurement may be a better predictor of visual function than the NEI clinical grading system and is in keeping with the findings of Sreekantam *et al*, who reported a significant correlation between VA and OCT Vit:RPE relative intensity ratio.(Sreekantam *et al.*, 2017)

The OCT vitreous measurement did not demonstrate an association with CMT. This could be explained in several ways: 1) there may be no consistent biological association between the presence and severity of vitreous haze density and macular oedema, 2) whilst CMT and vitreous inflammation can co-exist, macular oedema may develop following some time after vitreous haze (or vice versa) and therefore the association cannot be detected cross-sectionally, 3) the resolution of vitreous haze may occur faster than CMT (or vice versa) and the window of opportunity to detect both signs at their peak manifestation is relatively small. To understand the

time course relationship between vitreous haze and macular oedema (or indeed any other manifestation of inflammation), a longitudinal study with close timepoints during active flares is needed.

Strengths and weaknesses of the study

The main strength of this study is that it is the first prospective evaluation of the OCT vitreous measurement in a real-world setting. In comparison to previous retrospective studies, a number of potentially confounding factors could be controlled for, such as ensuring the same uveitis specialist was performing the ophthalmic examination in a standardised way and assessing all vitreous haze grades as per the NEI vitreous haze scale. Additionally, the scan protocol was predefined and based on prior evidence of the most reliable technique, instead of previous retrospective studies which accepted routine care OCT scan protocols. The association between Vit:RPE vitreous intensity (using OCTOR) and CMT and VA have previously been explored by Sreekantam *et al,* however this study was also the first to consider its predictive properties in the context of other potential confounders such as age, phakic status and AC cells. There are also several limitations of this study which should be noted. First, the number of eyes with significant vitreous inflammatory activity was small. As the study setting was a routine follow-up uveitic clinic, where the majority of patients were being monitored for stable disease, active uveitic inflammation was uncommon. Despite attempts to increase the proportion of eyes with higher levels of vitreous haze by purposely sampling the inflamed eye in each patient, the majority of included eyes had vitreous haze grade 0 to 1+ and only 7 out of 77 eyes had grade 2+ and above. Second, only one experienced uveitis specialist performed the clinical assessment, including assessing the NEI vitreous haze grade. This was because only one uveitis specialist experienced in vitreous haze grading was available for the entire duration of the study. Other clinicians were present, such as ophthalmology specialist trainees and clinical

research fellows, but were not present for the full duration of the study and deemed insufficiently experienced with the assessment to provide an accurate grade. Third, several relevant markers of inflammation, such as AC flare, were not recorded or quantified in the OCTAVE study. These were not included in the study protocol but would have provided a more complete clinical picture of the inflammatory disease state.

This second part of the OCTAVE study found that the test measurement is highly reliable in uveitic patients as well as in healthy eyes. It also provides evidence that the OCT vitreous intensity measurement can differentiate between eyes with and without vitreous haze. Moreover, an association could be observed between the OCT vitreous measurement and the NEI vitreous haze grade. However, there is significant variability of OCT vitreous intensity measurements within each NEI grade, which can be interpreted as poor signal:noise ratio of the OCT measurement (i.e. measurement variability is caused by noise, which confounds the OCT signal causing changes in the measurement unrelated to vitreous inflammation). Potential causes of noise which can disrupt the OCT measurement include non-inflammatory features such as tear film irregularities, inadequate pupillary dilation, cataracts; Or inflammatory causes such as keratic precipitates, AC cells and flare, posterior sycheciae and vitreous cells. Or, it could be due to poor reliability of the clinical grading system (i.e. differences in the OCT measurement reflect true differences in inflammatory activity, but the NEI grading system is unreliable and prone to misclassifications.) It is difficult to interpret these results in the absence of a reliable reference test. Longitudinal studies which capture responsiveness of the OCT measurement to changes in disease state, such as changing from a state of active to inactive inflammation, would provide further supporting evidence of its clinical validity.

The second part of the OCTAVE study demonstrated similarly good within-eye reliability of the OCT technique in uveitic eyes as healthy eyes. Comparison with the NEI vitreous haze grading, despite its limitations, suggests the OCT technique is sensitive to increased vitreous haze.

Moreover, it showed a stronger association with visual function (VA) than the NEI vitreous haze

grading system.

Chapter 7: Conclusions

7.1 Summary of Thesis

Being able to measure inflammatory activity is crucial for managing therapeutic decisions in uveitis, to reduce inflammatory activity and damage and to avoid sight loss. To do so requires the availability of suitable tests, which can accurately and reliably detect inflammatory activity. The current standard measures, which are based on clinician estimates from clinical examination, are subjective, unreliable and insensitive to small changes in disease activity. Their ability to inform changes in disease status is limited both in clinical care and as outcome measures in therapeutic trials.

Instrument-based markers of disease activity carry several potential advantages, including greater sensitivity, objectivity and the ability to automate the provision of key clinically useful information, such as disease trajectory over time (such as the example for chorioretinal lesions shown in **Section 1.3**). Instrument-based markers may therefore provide a more suitable measure of disease activity than the current clinician-based methods.

The three systematic reviews conducted in **Chapters 2 – 4** identified the current state-of-the-art technologies with the potential to objectively quantify AC cells, AC flare and vitreous haze. In the first review for AC cells, AS-OCT and the laser flare-cell photometer (LFP) were identified as potential candidate instruments, with AS-OCT showing the strongest evidence for producing measurements which were highly correlated with SUN AC grading (r = 0.06 - 0.97). The second review for AC flare identified four potential technologies (AS-OCT, LFP, OFAM and the double pass technique). The LFP was supported by the strongest evidence, demonstrating high correlation with SUN grading (r = 0.86 - 0.87) and even higher correlation with aqueous protein concentration sampled by aqueous paracentesis (r = 0.87 - 0.99). Three potential instruments were identified in the third review for quantifying vitreous inflammation (ultrasound, OCT and

retinal photography). Of these, OCT-based measurements showed moderate correlation with

NEI vitreous haze grading (r = 0.60), as did retinal photographic grading (r = 0.51).

Of the inflammatory markers explored in each review, vitreous haze is currently the clinical

measure in greatest need of an instrument-based equivalent, given its recognition as a key

marker of disease in posterior segment-involving uveitis. Through the OCTAVE study, I have

advanced the development of our OCT-based technique for quantifying vitreous intensity

(EQUIP) through prospective clinical validation. This study evaluated the technique both in the

healthy state and the disease state and explored the effects of ocular co-pathology such as

cataract and AC cells on the measurement. The OCTAVE study also explored the ability of the

measurement to predict visual function and central macular thickness.

In this conclusion chapter, I will use the evidence for EQUIP as an illustrative example of the

test development pathway for instrument-based measures in uveitis **(Figure 16)**, evaluating its

progress against the stages of test development, and highlighting common principles which will

be applicable to the general field of test evaluation in uveitis.

Figure 16. Pathway to validation of a new test, applied to the OCT vitreous measurement EQUIP

A pathway to validation of a new test was described in **Section 1.5 (Evaluation of a new test)**. This figure marks in this pathway, the steps which were completed by the OCTAVE study (green tick) and those which remain partially completed (yellow arrow).
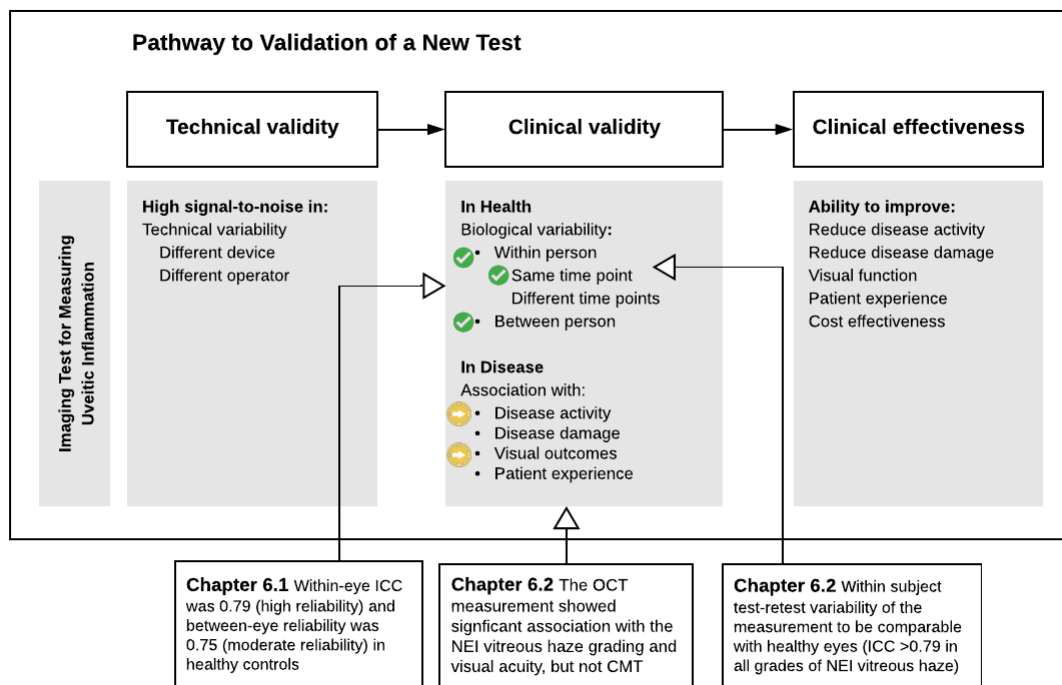
7.2 Challenges for the validation of tests in uveitis

The pathway to validation of a new test specifies the demonstration of three main qualities: technical validity, clinical validity and clinical effectiveness. The findings from this thesis highlight several barriers to validation which are common to all tests for measuring inflammatory activity in uveitis, not just for vitreous inflammation.

Demonstrating clinical validity in the absence of a reference test

As demonstrated through the three systematic reviews on instruments-based tests for AC cells, AC flare and vitreous haze. The lack of an appropriate reference test is the main challenge for

validating new tests for uveitis. If it were tolerated and ethical, the best reference test for vitreous inflammation would be sampling of vitreous, such as through vitreous paracentesis and quantification of inflammatory proteins and cytokines in the laboratory. As shown in **Chapter 3**, aqueous paracentesis has been carried out in test evaluation studies for measuring AC flare and has produced convincing evidence for the clinical validity of the laser flare photometer. Obtaining samples of the vitreous is more difficult than for aqueous humour and taking small samples is not typically performed for diagnosis, as it is for the aqueous. Instead, a pars plana vitrectomy may be performed, such as for diagnosis of lymphoma, and carries with it significant risks such as endophthalmitis, retinal tears and detachments, cataract progression and hypotony.(Hwang, Yeh and Bergstrom, 2014; Mura and Barca, 2014) Other invasive tests, such as vitreous fluorophotometry may provide physiological evidence of blood-retinal-barrier breakdown, but are not available in most clinical settings and carry risks of adverse reaction to the injected dye.(Raines, 1988)

Achieving sufficient spectrum of disease and sample size

Another major challenge in all of uveitis research, not just in test evaluation, is that uveitis is both uncommon and heterogeneous; indeed, when broken down into its individual syndromes, most of these are individually rare. These individual syndromes are generally grouped anatomically (for example into posterior segment-involving uveitis), but this does make the significant assumption that the test will perform similarly across all syndromes. Furthermore, the episodic nature of many of these uveitis syndromes means that only a minority of patients may be active in the duration of a study. For example, in the OCTAVE study there were only a small number of patients with significant vitreous haze (NEI vitreous haze grades 1+ and above). This means not only is the overall sample size typically small, but unless strict inclusion criteria are applied, there is often a spectrum bias towards milder disease too. Conducting single centre

studies of sufficient power is a challenge and large, expensive, multi-centre trials may be required to enrol sufficient participants with active disease.

Demonstrating clinical effectiveness

Once the technical and clinical validity of a new test has been proven, there remains a need to demonstrate clinical effectiveness. So far, neither our OCT-based vitreous measurement nor any of the methods identified in the three systematic reviews have demonstrated the ability to improve clinical outcomes. At this later stage of test evaluation, consideration needs to be given as to how the test will be used in the clinical context. For example, the test may be used as a diagnostic or a monitoring tool, in which case thresholds for positive disease diagnosis needs to be determined for the former and appropriate monitoring intervals, as well as thresholds for clinically meaningful change, for the latter. Tests for measuring inflammatory activity could be used for diagnosis and monitoring, and this is the case for current measures of disease activity (the NEI vitreous haze scale). The test could be used in a secondary care setting (hospital eye clinics) or in the community and therefore the way results are presented and interpreted may need to be adjusted. For example, in hospital the test may be used as one component of assessment combined with a slit-lamp examination by an ophthalmologist, as opposed to in the community where the test may be performed in isolation and a clinical decision is made on the test alone.

7.3 Technical limitations of the OCT-based approach

There are several potential technical limitations to the EQUIP solution which may contribute to the variability seen in the OCTAVE study, which are inherent to its design and may limit its validity. These limitations are discussed separately to the OCTAVE study as they could not

155

necessarily be improved by study design but may inform technical improvements for the OCT image analysis technique going forward.

7.3.1 Limited volume of posterior vitreous sampled

Only a small part of the vitreous is sampled by the current scanning method. The theoretical area sampled by 20° B scan is 5.9mm by 1.5mm (including an area of vitreous, the retina and of the choroid). This represents only a small volume of the entire vitreous (estimated total volume 4.6-4.9ml)(Azhdam, Goldberg and Ugradar, 2020) and most of the anterior vitreous is unaccounted for. The distribution of vitreous haze along the axial length of the eye is not well understood, other than in certain uveitic entities (such as toxoplasmosis) which present with patches of vitreous haze directly overlying areas of focal chorioretinitis(Foster and Vitale, 2002). Focal patches of vitreous haze could be missed by the current EQUIP approach, which currently focuses at the macula, in order to avoid distortion caused by refractive properties in non-central areas and achieve the most focused image. Increasing the area/volume scanned in the axial plane could be achieved by increasing the number of B scans in the volume. Other potential methods of increasing the area captured could involve utilising widefield imaging (using the 55 degree lens as opposed to the 30 degree lens) or capturing multiple volume scans in different directions of gaze to access more peripheral areas, such the approach described by Barry *et al.*(Barry *et al.*, 2016) However, these solutions do not improve the volume capture in the axial direction. The depth of OCT B scans on the SPECTRALIS OCT device is fixed and it is currently not possible to increase this. Improving the depth of vitreous captured is likely going to require exploration using alternative devices.

7.3.2 The effects of anterior structures on the OCT signal

In the same way that anterior vitreous is not captured by the current method, neither are other anterior segment ocular structures, which may disrupt the OCT signal and confound the measurement but without being recognised as such. For example, media opacity such as cataracts are known to have a light scattering effect on OCT and causes degradation of the image quality.(van Velthoven *et al.*, 2006; Kok *et al.*, 2013) The effects have even been documented to vary depending on the type of cataract, with posterior cortical cataracts being associated with worse quality scans than nuclear sclerotic.(van Velthoven *et al.*, 2007) Similar effects are likely caused by other anterior segment opacifications along the scan axis, such as keratic precipitates, AC cells, AC flare, posterior synechiae, vitreous cells and vitreous floaters. The current measurement acquired from the EQUIP technique assumes that all signal degradation is caused by vitreous opacity and cannot adjust for the impact of anterior structures.

## 7.3.3 Confounding effects of retinal co-pathology

The current EQUIP analysis method considers the vitreous intensity as a proportion of the intensity from the entire B scan. This means changes in reflectivity caused by the retina and choroid can affect the overall vitreous OCT intensity measure (see **Section 5.2.3**) for how the measurement is calculated). The measurement may therefore be affected by retinal co-pathology which may be inflammatory (such as CMO, chorioretinal infiltrates, choroidal neovascular lesions) or non-inflammatory (such as epiretinal membrane and drusen). The degree of change in signal caused by such co-pathology would require further investigation, but may be difficult to determine, particularly with inflammatory co-pathology which would likely impact on both the vitreous intensity and the overall B scan intensity.

7.4 Future work on the OCT vitreous intensity measure

This section describes planned future work which aims to improve those parts of the validation pathway which are currently 'partially validated' **(Figure 17)**.

Figure 17. Planned future work for the OCT vitreous haze technique set in the context of the test validation pathway.

Each box refers to the subchapter heading for future work.
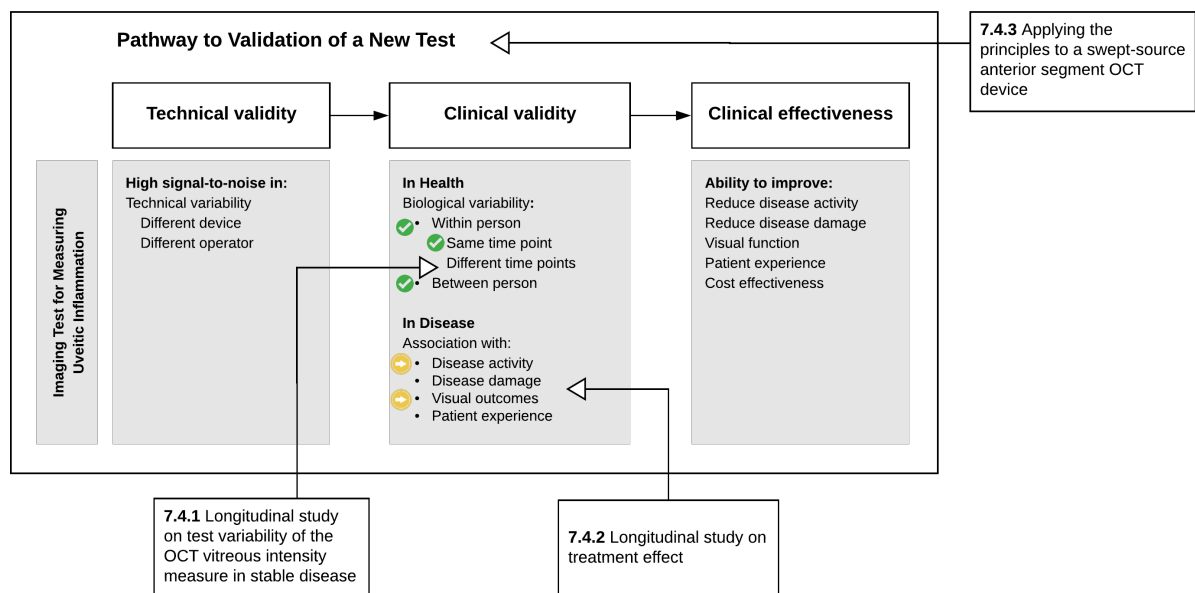
7.4.1 Longitudinal study on test variability of the OCT vitreous intensity measure in stable disease

The first part of the OCTAVE study tested technical and biological variability of the OCT vitreous intensity measurement in the same visit but did not establish biological variability at different time points in the same day or on different days. This is an important future investigation to be done in healthy eyes and in stable uveitic eyes. First the day-to-day biological variability should be established, to help interpret a subsequent study in stable uveitic eyes. It is important that

future work is investigated in this order, as variations in the uveitic eyes will need to be interpreted with the assumption that the variability is not due to true fluctuations in inflammatory activity. For both healthy and uveitic eyes, an ideal longitudinal investigation will consist of 1) a frequent series of tests (such as twice daily) over a short period of time (several weeks), and 2) an evaluation of measurement stability over time in patients with presumed quiescent disease. This investigation in healthy eyes will be helpful for understanding the expected variability of the measurement over time in the absence of inflammation. The investigation in uveitic eyes will inform expected baseline variability in quiescent disease, or it may reveal low level fluctuations in clinically quiescent disease and allow detection of subclinical inflammatory activity.

7.4.2 Longitudinal study on treatment effect

As discussed throughout this thesis, one of the biggest challenges to test evaluation in uveitis is the lack of a suitable reference test. To demonstrate clinical validity of a new test, it is inadequate to simply compare the new measure with other measures. Being able to show the ability of an instrument to detect a change in disease status would provide additional evidence that the test has clinical validity.

Future work should include longitudinal studies which evaluate whether the OCT vitreous haze measurement is sensitive to a change in disease state. The ideal study design would be to prospectively measure OCT vitreous intensity before and after treatment, with the outcome being significance of change detectable by the OCT vitreous intensity measure compared to the NEI vitreous haze scale. One potential approach to do this is to nest a silent test evaluation study within a therapeutic trial. For example, in the case of EQUIP, most clinical trials in uveitis already collect the necessary outcomes at each visit: OCT macula scans, clinician-based measures of inflammatory activity (including NEI vitreous haze scale, AC cells/flare, presence/absence of new inflammatory chorioretinal lesions and other markers of improving or

worsening disease). The addition of an observational analysis which would support the clinical validation of a new test or biomarker would be relatively easy to do, without negatively impacting upon the trial conduct. This study can also be done retrospectively on existing clinical trial data, providing the tests under evaluation were conducted in a standardised way.

7.4.3 Applying the principles to a swept-source anterior segment OCT device

One of the major limitations to the current approach is that it only considers the posterior part of the vitreous. The current sampled area includes only a small volume of vitreous directly anterior to the macula. This means any changes in the anterior vitreous, which may disrupt the OCT signal, is not captured and cannot be adjusted for. If such changes were captured, the impact of such changes on the vitreous intensity measurement could be quantified and adjusted for.

In October 2018, a new swept-source anterior segment OCT device was released - *Anterion* (Heidelberg Engineering, Heidelberg, Germany). This OCT device has an additional feature which produces measurements of a continuous A scan through the entire axial length of the eye (OCT B scans are constructed by multiple adjacent A scans). This feature was intended for axial length measurement, by measuring the distance of the A scan peak signal from the corneal surface (front of eye) to the retina, for planning of cataract surgery. However, we are now exploring whether the use of this single A scan can be repurposed for measuring the reflectivity signal from the front to back of the eye, including the axial length of the vitreous and structures anterior to it. Using this function, it could be possible to quantify the OCT signal intensity degradation anterior to the vitreous, correct for this, and produce a signal intensity measure which more accurately reflects the vitreous. From early 2019, we began to add the *Anterion* as an additional investigation and images for 43 patients with uveitis have been acquired to date.

However, extraction of the raw A scan data from the *Anterion* device has been a technical barrier to analysing these images so far. We have been working with Heidelberg Engineering team to develop new software which would allow us to extract the raw OCT data needed from the device and refine the imaging technique.

7.5 Concluding remarks on test evaluation in uveitis

7.5.1 Challenges to deployment of tests in uveitis

Providing the technical and clinical validity of a new test has been proven, there remains a number of key considerations around test deployment. In **Section 1.5.2 (Choosing a suitable test)**, these were summarised under the heading 'practicality', which could be expanded into usability, acceptability and cost **(Table 20)**.

Table 20. Practicality Attributes of a Clinical Test

Usability	Is the test safe and easy to conduct? Is the test easy to interpret? Can the test results assist decision making?
Acceptability	Is the test acceptable to the users (including operator, patient and clinician)?
Cost effectiveness	Is the cost of the test justified? Is the cost of the test sustainable?

A good example of where clinical validity of a test has been proven, but it has fallen short in deployment is the laser flare photometer. As discussed in **Chapter 3**, despite a large body of evidence demonstrating its clinical validity, the laser-flare photometer is not commonly used in the uveitis clinic. The reasons for this is unclear from the published literature, but our experience

(and that of many other uveitis experts anecdotally) is the impractical nature and time-costliness to acquire measurements (the requirement of a dark room, needing to capture 7 measurements and manually discarding outliers) makes the device practically difficult to implement in the real world. Additionally, the laser flare photometer only measures one component of inflammation, whereas instruments like OCT can be repurposed for different disease measures. OCT is already routinely performed for disease monitoring in uveitis, as well as most other retinal diseases, therefore its usability, acceptability and cost effectiveness are already established by standard of care.

Creating new clinical pathways and early detection through disease monitoring

Another advantage of automated disease quantification using imaging is the potential for testing in community and remote settings. In the modern healthcare system, the frequency of disease monitoring can be defined less by what is optimal for the patient, and more by resource limitations (numbers of specialists, availability of clinic space). This has been amplified since the COVID-19 pandemic which has significantly reduced the capacity of clinics in hospital settings. Automated disease monitoring using imaging could theoretically be delivered by an imaging technician alone. This opens up the possibility of community-based monitoring which could be carried out more regularly, in a way that is more accessible and convenient to the patient. Given that this new monitoring regime is a possibility, it is necessary to consider what an appropriate monitoring regime looks like. The interval between monitoring visits should ideally be short enough that clinically significant changes in disease state (i.e. activity at or above the level where damage may occur) wouldn't be missed, yet long enough to be practical and acceptable for the patient. Longitudinal studies on test variability over time, as proposed in **Section 7.4**, will provide some of the necessary evidence to inform what an appropriate monitoring interval may be. Ultimately, change in disease activity (from an inactive to clinically

significant active state) may be too rapid to be detected. For example, if real changes in disease activity develop in a matter of days, it would not be practical for patients to attend OCT scans every few days for routine monitoring. In this case, the value of routine monitoring may only be for detecting signs of activity/damage in between visits which may inform clinical judgement around treatment effectiveness or whether maintenance treatment is required. The detection of acute activity is therefore limited to 'as needed' consultations and is reliant upon patient detection. Until a solution for home monitoring becomes available, early detection of uveitis flares may simply not be possible.

7.5.2 Next phase of test validation

Although the evidence supporting instrument-based measures of uveitis inflammation identified and generated throughout this thesis covers only part of the pathway to validation, it is worth noting that this body of literature is both comparatively larger and offers more promising evidence of validity than the few published studies evaluating clinician-based assessment. The alternative solution to outdated systems like the NEI vitreous haze grading scale need not be perfect to bring about improvement to clinical care, but just needs to be better than the current methods.

A major challenge in this field has been the lack of an appropriate gold standard reference test, with which new tests can be benchmarked against and proven to be valid. Instead, we will need to rely upon a combination of association with other signs of inflammation, clinically important outcomes (visual function or patient reported outcomes) and critically, its sensitivity to change in disease state. To overcome the additional challenges in the context of uveitis (the relative rarity of disease subtypes, particularly posterior-segment involving uveitis, the heterogeneity between and within uveitic entities and the poorly understood relationships between disease activity,

disease damage and clinically significant changes in function), longitudinal data in large cohorts of uveitic patients will be required to achieve sufficient power in these observations. This will require a concerted effort in future uveitis therapeutic trials, to begin prospective collection of observational data using the most promising instrument-based techniques.

References

ADVISE (2020) *Adalimumab vs. Conventional Immunosuppression for Uveitis Trial - ClinicalTrials.gov*. Available at: https://clinicaltrials.gov/ct2/show/record/NCT03828019 (Accessed: July 13, 2020).

Agrawal, R. *et al.* (2016) "Choroidal Vascularity Index (CVI)--A Novel Optical Coherence Tomography Parameter for Monitoring Patients with Panuveitis?," *PloS one*, 11(1), p. e0146344.

Agrawal, R. V. *et al.* (2010) "Current approach in diagnosis and management of anterior uveitis," *Indian journal of ophthalmology*. Medknow, 58(1), pp. 11–19.

Almeida, D. R. P. *et al.* (2015) "Comparison of retinal and choriocapillaris thicknesses following sitting to supine transition in healthy individuals and patients with age-related macular degeneration," *JAMA ophthalmology*, 133(3), pp. 297–303.

American Uveitis Society/UCLA Stein Eye Institute (no date) "UCLA/AUS Workshop on Objective Measures of Intraocular Inflammation," in.

Aojula, A. *et al.* (2018) "Segmentation error in spectral domain optical coherence tomography measures of the retinal nerve fibre layer thickness in idiopathic intracranial hypertension," *BMC ophthalmology*, 17(1), p. 257.

Aumann, S. *et al.* (2019) "Optical Coherence Tomography (OCT): Principle and Technical Realization," in Bille, J. F. (ed.) *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics*. Cham (CH): Springer.

Azhdam, A. M., Goldberg, R. A. and Ugradar, S. (2020) "In Vivo Measurement of the Human Vitreous Chamber Volume Using Computed Tomography Imaging of 100 Eyes," *Translational vision science & technology*, 9(1), p. 2.

Barry, R. J. *et al.* (2016) "Assessment of eccentric vitreous density in healthy subjects using optical coherence tomography," *Investigative ophthalmology & visual science*. The Association for Research in Vision and Ophthalmology, 57(12), pp. 4231–4231.

Bodaghi, B. *et al.* (2001) "Chronic severe uveitis: etiology and visual outcome in 927 patients from a single center," *Medicine*, 80(4), pp. 263–270.

Bonfioli, A. A. *et al.* (2005) "Intermediate uveitis," *Seminars in ophthalmology*, 20(3), pp. 147–154.

Bossuyt, P. M. *et al.* (2006) "Comparative accuracy: assessing new tests against existing diagnostic pathways," *BMJ* , 332(7549), pp. 1089–1092.

Braithwaite, T. *et al.* (2019) "The use of patient-reported outcome research in modern ophthalmology: impact on clinical trials and routine clinical practice," *Patient related outcome measures*, 10, pp. 9–24.

Browning, D. J. *et al.* (2004) "Comparison of the clinical diagnosis of diabetic macular edema with diagnosis by optical coherence tomography," *Ophthalmology*, 111(4), pp. 712–715.

Campbell, J. P. *et al.* (2012) "Wide-field retinal imaging in the management of noninfectious posterior uveitis," *American journal of ophthalmology*, 154(5), pp. 908-911.e2.

Caspi, R. R. (2011) "Understanding autoimmune uveitis through animal models. The Friedenwald Lecture," *Investigative ophthalmology & visual science*, 52(3), pp. 1872–1879.

Chen, J. *et al.* (2013) "Use of optical coherence tomography and electroretinography to evaluate retinal pathology in a mouse model of autoimmune uveitis," *PloS one*. Public Library of Science (PLoS), 8(5), p. e63904.

Choma, M. *et al.* (2003) "Sensitivity advantage of swept source and Fourier domain optical coherence tomography," *Optics express*, 11(18), pp. 2183–2189.

Chu, C. J. *et al.* (2016) "Multimodal analysis of ocular inflammation using the endotoxin-induced uveitis mouse model," *Disease models & mechanisms*, 9(4), pp. 473–481.

Costa, R. A. *et al.* (2006) "Retinal assessment using optical coherence tomography," *Progress in retinal and eye research*, 25(3), pp. 325–353.

Daniel, E. *et al.* (2010) "Mycophenolate mofetil for ocular inflammation," *American journal of ophthalmology*, 149(3), pp. 423-32.e1–2.

Darrell, R. W., Wagener, H. P. and Kurland, L. T. (1962) "Epidemiology of uveitis. Incidence and prevalence in a small urban community," *Archives of ophthalmology*, 68, pp. 502–514.

David, R. *et al.* (1992) "Diurnal intraocular pressure variations: an analysis of 690 diurnal curves," *The British journal of ophthalmology*, 76(5), pp. 280–283.

Denniston, A. K. *et al.* (2011) "Endogenous cortisol and TGF-beta in human aqueous humor contribute to ocular immune privilege by regulating dendritic cell function," *The journal of immunology*. The American Association of Immunologists, 186(1), pp. 305–311.

Denniston, A. K. *et al.* (2012) "Aqueous humor suppression of dendritic cell function helps maintain immune regulation in the eye during human uveitis," *Investigative ophthalmology & visual science*, 53(2), pp. 888–896.

Denniston, A. K. *et al.* (2015) "Heterogeneity of primary outcome measures used in clinical trials of treatments for intermediate, posterior, and panuveitis," *Orphanet journal of rare diseases*, 10(1), p. 97.

Denniston, A. K. and Dick, A. D. (2013) "Systemic therapies for inflammatory eye disease: Past, Present and Future," *BMC ophthalmology*, 13(1), p. 18.

Denniston, A. K., Keane, P. A. and Srivastava, S. K. (2017) "Biomarkers and Surrogate Endpoints in Uveitis: The Impact of Quantitative Imaging," *Investigative ophthalmology & visual science*, 58(6), pp. BIO131–BIO140.

Deschenes, J. *et al.* (2008) "International Uveitis Study Group (IUSG): clinical classification of

uveitis," *Ocular immunology and inflammation*. Informa UK Limited, 16(1), pp. 1–2.

Durrani, O. M. *et al.* (2004) "Degree, duration, and causes of visual loss in uveitis," *The British journal of ophthalmology*, 88. doi: 10.1136/bjo.2003.037226.

Ehlers, J. P. *et al.* (2017) "Automated quantitative characterisation of retinal vascular leakage and microaneurysms in ultra-widefield fluorescein angiography," *The British journal of ophthalmology*, 101(6), pp. 696–699.

El-Asrar, A. M. A. *et al.* (2010) "A clinical approach to the diagnosis of retinal vasculitis," *International Ophthalmology*, pp. 149–173. doi: 10.1007/s10792-009-9301-3.

Enoch, J. *et al.* (2019) "Evaluating Whether Sight Is the Most Valued Sense," *JAMA ophthalmology*. doi: 10.1001/jamaophthalmol.2019.3537.

Fardeau, C. *et al.* (2016) "Uveitic macular edema," *Eye*. Springer Nature, 30(10), pp. 1277–1292.

FDA-NIH Biomarker Working Group (2016) *BEST (Biomarkers, EndpointS, and Other Tools) Resource*. Food and Drug Administration (US).

Ferrante di Ruffano, L. *et al.* (2012) "Assessing the value of diagnostic tests: a framework for designing and evaluating trials," *BMJ* , 344, p. e686.

Forooghian, F. *et al.* (2008) "Evaluation of time domain and spectral domain optical coherence tomography in the measurement of diabetic macular edema," *Investigative ophthalmology & visual science*, 49(10), pp. 4290–4296.

Forrester, J. V., Kuffova, L. and Dick, A. D. (2018) "Autoimmunity, autoinflammation, and infection in uveitis," *American journal of ophthalmology*, 189, pp. 77–85.

Foster, C. S. (charles S. and Vitale, A. T. (2002) *Diagnosis and treatment of uveitis*. W.B. Saunders, p. 900.

Foster C, S. and Vitale Albert, T. (2013) *Diagnosis and Treatment of Uveitis*. Philadelphia: Jaypee Brothers Medical Publishers (P) Ltd, p. 1290.

Frank, R. N. *et al.* (2004) "Temporal variation in diabetic macular edema measured by optical coherence tomography," *Ophthalmology*, pp. 211–217. doi: 10.1016/j.ophtha.2003.05.031.

Freddo, T. F. (2013) "A contemporary concept of the blood-aqueous barrier," *Progress in retinal and eye research*. Elsevier BV, 32, pp. 181–195.

Friedman, D. S. *et al.* (2013) "Risk of elevated intraocular pressure and glaucoma in patients with uveitis: results of the multicenter uveitis steroid treatment trial," *Ophthalmology*, 120(8), pp. 1571–1579.

Gangaputra, S. *et al.* (2009) "Methotrexate for ocular inflammatory diseases," *Ophthalmology*, 116(11), pp. 2188–98.e1.

Garcia-Martin, E. *et al.* (2011) "Fourier-domain OCT in multiple sclerosis patients: reproducibility and ability to detect retinal nerve fiber layer atrophy," *Investigative ophthalmology & visual*

science, 52(7), pp. 4124–4131.

Giani, A. *et al.* (2012) "Aligning scan locations from consecutive spectral-domain optical coherence tomography examinations: a comparison among different strategies," *Investigative ophthalmology & visual science*, 53(12), pp. 7637–7643.

Glasziou, P. P., Irwig, L. and Aronson, J. K. (2008) *Evidence-Based Medical Monitoring: From Principles to Practice*. John Wiley & Sons.

Gueudry, J. and Muraine, M. (2018) "Anterior uveitis," *Journal francais d'ophtalmologie*, 41(1), pp. e11–e21.

Gupta, B. *et al.* (2009) "Diurnal variation of macular oedema in CRVO: prospective study," *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie*, 247(5), pp. 593–596.

Harthan, J. S. *et al.* (2016) "Diagnosis and treatment of anterior uveitis: optometric management," *Clinical optometry*, 8, pp. 23–35.

Hogan, M. J., Kimura, S. J. and Thygeson, P. (1959) "Signs and symptoms of uveitis. I. Anterior uveitis," *American journal of ophthalmology*, 47(5 Pt 2), pp. 155–170.

Holland (2015) *NEI/FDA Workshop in Clinical Trial End-points for Inflammatory Eye Diseases*.

Holland, G. N. (2007) "A reconsideration of anterior chamber flare and its clinical relevance for children with chronic anterior uveitis (an American Ophthalmological Society thesis)," *Transactions of the American Ophthalmological Society*. American Ophthalmological Society, 105, p. 344.

Hornbeak, D. M. *et al.* (2014) "Interobserver agreement in clinical grading of vitreous haze using alternative grading scales," *Ophthalmology*, 121(8), pp. 1643–1648.

Huang, D. *et al.* (1991) "Optical Coherence Tomography," *Science*, 254(5035), pp. 1178–1181.

Hwang, C. S., Yeh, S. and Bergstrom, C. S. (2014) "Diagnostic vitrectomy for primary intraocular lymphoma: when, why, how?," *International ophthalmology clinics*, 54(2), pp. 155–171.

Invernizzi, A. *et al.* (2017) "Objective Quantification of Anterior Chamber Inflammation: Measuring Cells and Flare by Anterior Segment Optical Coherence Tomography," *Ophthalmology*, 124(11), pp. 1670–1677.

Jabs, D. A. *et al.* (2005) "Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop," *American journal of ophthalmology*. United States, 140(3), pp. 509–516.

Jaffe, G. J. *et al.* (2016) "Adalimumab in Patients with Active Noninfectious Uveitis," *The New England journal of medicine*, 375(10), pp. 932–943.

Jaffe, G. J. *et al.* (2019) "Effect of an injectable fluocinolone acetonide insert on recurrence rates in chronic noninfectious uveitis affecting the posterior segment: Twelve-month results,"

Ophthalmology. Elsevier BV, 126(4), pp. 601–610.

Jiang, A. *et al.* (2020) "Repeatability of automated leakage quantification and microaneurysm identification utilising an analysis platform for ultra-widefield fluorescein angiography," *The British journal of ophthalmology*, 104(4), pp. 500–503.

Jones, N. P. (2015) "The Manchester Uveitis Clinic: The first 3000 patients, 2: Uveitis Manifestations, Complications, Medical and Surgical Management," *Ocular immunology and inflammation*, 23(2), pp. 127–134.

Kakinoki, M. *et al.* (2009) "Comparison of macular thickness between Cirrus HD-OCT and Stratus OCT," *Ophthalmic surgery, lasers & imaging: the official journal of the International Society for Imaging in the Eye*, 40(2), pp. 135–140.

Karampelas, M. *et al.* (2015) "Quantitative analysis of peripheral vasculitis, ischemia, and vascular leakage in uveitis using ultra-widefield fluorescein angiography," *American journal of ophthalmology*, 159(6), pp. 1161-1168.e1.

Karim, R. *et al.* (2013) "Interventions for the treatment of uveitic macular edema: a systematic review and meta-analysis," *Clinical ophthalmology* , 7, pp. 1109–1144.

Keane, P. A. *et al.* (2014) "Objective measurement of vitreous inflammation using optical coherence tomography," *Ophthalmology*, 121(9), pp. 1706–1714.

Keane, P. A. *et al.* (2015) "Automated Analysis of Vitreous Inflammation Using Spectral-Domain Optical Coherence Tomography," *Translational vision science & technology*, 4(5), p. 4.

Kempen, J. H. *et al.* (2008) "Interobserver Agreement in Grading Activity and Site of Inflammation in Eyes of Patients with Uveitis," *American journal of ophthalmology*, 146(6), pp. 813–818.

Kempen, J. H. *et al.* (2013) "Fluorescein angiography versus optical coherence tomography for diagnosis of uveitic macular edema," *Ophthalmology*. United States: Elsevier Inc. (360 Park Avenue South, New York NY 10010, United States), 120(9), pp. 1852–1859.

Kempen, J. H., Gewaily, D. Y. and Cw, N. (2016) "Systemic immunosuppressive therapy for eye diseases (SITE) research group. Remission of intermediate uveitis: incidence and predictive factors."

Kim, J. S. *et al.* (2015) "Clinical trials in noninfectious uveitis," *International ophthalmology clinics*, 55(3), pp. 79–110.

Kim, J. S. *et al.* (2016) "Enhanced Depth Imaging Optical Coherence Tomography in Uveitis: An Intravisit and Interobserver Reproducibility Study," *American journal of ophthalmology*, 164, pp. 49–56.

Kok, P. H. B. *et al.* (2013) "The relationship between the optical density of cataract and its influence on retinal nerve fibre layer thickness measured with spectral domain optical coherence tomography," *Acta ophthalmologica*, 91(5), pp. 418–424.

Konstantopoulou, K. *et al.* (2015) "A comparative study between clinical grading of anterior

chamber flare and flare reading using the Kowa laser flare meter," *International Ophthalmology*, pp. 629–633. doi: 10.1007/s10792-012-9616-3.

Kotaniemi, K. *et al.* (2005) "Uveitis in young adults with juvenile idiopathic arthritis: a clinical evaluation of 123 patients," *Annals of the rheumatic diseases*, 64(6), pp. 871–874.

Krebs, I. *et al.* (2011) "Quality and reproducibility of retinal thickness measurements in two spectral-domain optical coherence tomography machines," *Investigative ophthalmology & visual science*, 52(9), pp. 6925–6933.

Kumar, P. *et al.* (2015) "Analysis of Anterior Chamber Inflammation in Pediatric Uveitis by Spectral Domain Optical Coherence Tomography," *Investigative ophthalmology & visual science*, 56(7), p. 6182.

Leder, H. A. *et al.* (2013) "Ultra-wide-field retinal imaging in the management of non-infectious retinal vasculitis," *Journal of ophthalmic inflammation and infection*, 3(1), p. 30.

Ledingham, J. *et al.* (2017) "BSR and BHPR guideline for the prescription and monitoring of non-biologic disease-modifying anti-rheumatic drugs," *Rheumatology* , 56(12), p. 2257.

Lee, R. W. *et al.* (2014) "Autoimmune and autoinflammatory mechanisms in uveitis," *Seminars in immunopathology*. Springer Science and Business Media LLC, 36(5), pp. 581–594.

Leite, M. T. *et al.* (2011) "Comparison of the diagnostic accuracies of the Spectralis, Cirrus, and RTVue optical coherence tomography devices in glaucoma," *Ophthalmology*, 118(7), pp. 1334–1339.

Leitgeb, R., Hitzenberger, C. and Fercher, A. (2003) "Performance of fourier domain vs. time domain optical coherence tomography," *Optics express*, 11(8), pp. 889–894.

Leung, C. K.-S. *et al.* (2009) "Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study," *Ophthalmology*, 116(7), pp. 1257–63, 1263.e1–2.

Liu, X., Kale, A. U., *et al.* (2019) "Optical coherence tomography (OCT) in unconscious and systemically unwell patients using a mobile OCT device: a pilot study," *BMJ open*, 9(11), p. e030882.

Liu, X., Kale, A., *et al.* (2019) "Test-retest Variability of Optical Coherence Tomography (OCT) and OCT Angiography in a Mobile OCT Module and level of Agreement with Conventional Table-top OCT," *Investigative ophthalmology & visual science*. The Association for Research in Vision and Ophthalmology, 60(9), pp. 1853–1853.

Lowder, C. M. D. P. *et al.* (2011) "Dexamethasone Intravitreal Implant for Noninfectious Intermediate or Posterior Uveitis," *Archives of ophthalmology*, 129(5), p. 545.

Mackenzie, P. J. *et al.* (2007) "Sensitivity and specificity of the optos optomap for detecting peripheral retinal lesions," *Retina* , 27(8), pp. 1119–1124.

Macular Edema Ranibizumab v. Intravitreal Anti-inflammatory Therapy Trial - ClinicalTrials.gov (NCT02623426) (2020). Available at: https://clinicaltrials.gov/ct2/show/NCT02623426

(Accessed: July 14, 2020).

Madhusudhan, S., Keane, P. A. and Denniston, A. K. (2016) "Adjunctive use of systematic retinal thickness map analysis to monitor disease activity in punctate inner choroidopathy," *Journal of ophthalmic inflammation and infection*, 6(1), p. 9.

Mallett, S. *et al.* (2012) "Interpreting diagnostic accuracy studies for patient care," *BMJ* , 345, p. e3999.

Mansberger, S. L. *et al.* (2017) "Automated Segmentation Errors When Using Optical Coherence Tomography to Measure Retinal Nerve Fiber Layer Thickness in Glaucoma," *American journal of ophthalmology*, 174, pp. 1–8.

Mary, L. T. and Turgeon, E. D. (2005) "Clinical Hematology: Theory and Procedures." Boston: Lippincott Williams & Wilkims.

Matas, J. *et al.* (2019) "Predictors for functional and anatomic outcomes in macular edema secondary to non-infectious uveitis," *PloS one*, 14(1), p. e0210799.

McCormick, E. (2020) *Specsavers launches national OCT campaign*, *Association of Optometrists*. Available at: https://www.aop.org.uk/ot/industry/high-street/2020/02/04/specsavers-launches-national-oct-campaign (Accessed: July 11, 2020).

McNeish, D. and Stapleton, L. M. (2016) "Modeling clustered data with very few clusters," *Multivariate behavioral research*. Informa UK Limited, 51(4), pp. 495–518.

Montesano, G. *et al.* (2018) "Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis," *Scientific reports*, 8(1), p. 1648.

Mori, K. *et al.* (2012) "Montage images of spectral-domain optical coherence tomography in eyes with idiopathic macular holes," *Ophthalmology*, 119(12), pp. 2600–2608.

Mrejen, S. and Spaide, R. F. (2013) "Optical coherence tomography: imaging of the choroid and beyond," *Survey of ophthalmology*, 58(5), pp. 387–429.

Muller, W. A. (2013) "Getting leukocytes to the site of inflammation," *Veterinary pathology*. SAGE Publications, 50(1), pp. 7–22.

Mura, M. and Barca, F. (2014) "25-Gauge Vitrectomy," in, pp. 45–53.

Nguyen, Q. D. *et al.* (2016) "Adalimumab for prevention of uveitic flare in patients with inactive non-infectious uveitis controlled by corticosteroids (VISUAL II): a multicentre, double-masked, randomised, placebo-controlled phase 3 trial," *The Lancet*, 388(10050), pp. 1183–1192.

Nussenblatt, R. B. *et al.* (1985) "Standardizatlon of Vitreal inflammatory Activity in Intermediate and Posterior Uveitis," *Ophthalmology*, 92(4), pp. 467–471.

Nussenblatt, R. B. and Whitcup, S. M. (2004) *Uveitis: fundamentals and clinical practice 3rd*. Philadelphia: Mosby.

Ometto, G. *et al.* (2020) "Merging Information From Infrared and Autofluorescence Fundus Images for Monitoring of Chorioretinal Atrophic Lesions," *Translational vision science &*

technology, 9(9), p. 38.

Pasadhika, S. *et al.* (2009) "Azathioprine for ocular inflammatory diseases," *American journal of ophthalmology*, 148(4), pp. 500-509.e2.

Patel, P. J. *et al.* (2009) "Segmentation error in Stratus optical coherence tomography for neovascular age-related macular degeneration," *Investigative ophthalmology & visual science*, 50(1), pp. 399–404.

Pecen, P. E. *et al.* (2017) "Peripheral Findings and Retinal Vascular Leakage on Ultra-Widefield Fluorescein Angiography in Patients with Uveitis," *Ophthalmology Retina*. S.K. Srivastava, Cole Eye Institute, Cleveland Clinic, 9500 Euclid Avenue, i30, Cleveland, OH 44195, United States. E-mail: srivass2@ccf.org, 1(5), pp. 428–434.

Pichi, F. *et al.* (2020) "Optical coherence tomography diagnostic signs in posterior uveitis," *Progress in retinal and eye research*, 75, p. 100797.

Polito, A. *et al.* (2007) "Effect of posture on the diurnal variation in clinically significant diabetic macular edema," *Investigative ophthalmology & visual science*, 48(7), pp. 3318–3323.

Portney, L. G., Watkins, M. P. and Others (2009) *Foundations of clinical research: applications to practice*. Pearson/Prentice Hall Upper Saddle River, NJ.

Qian, R. *et al.* (2019) "Ocular anterior chamber blood cell population differentiation using spectroscopic optical coherence tomography," *Biomedical optics express*. The Optical Society, 10(7), pp. 3281–3300.

Raines, M. F. (1988) "Vitreous fluorophotometry: a review," *Journal of the Royal Society of Medicine*, 81(7), pp. 403–406.

Rao, N. (2013) "Uveitis in developing countries," *Indian journal of ophthalmology*, 61(6), p. 253.

Reznicek, L. *et al.* (2014) "Megahertz ultra-wide-field swept-source retina optical coherence tomography compared to current existing imaging devices," *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie*, 252(6), pp. 1009–1016.

Rodriguez, A. *et al.* (1996) "Referral patterns of uveitis in a tertiary eye care center," *Archives of ophthalmology*, 114(5), pp. 593–599.

Rosenbaum, J. T. and Dick, A. D. (2018) "The Eyes Have it: A Rheumatologist's View of Uveitis," *Arthritis & rheumatology (Hoboken, N.J.)*, 70(10), pp. 1533–1543.

Rose-Nussbaumer, J. *et al.* (2015) "Aqueous cell differentiation in anterior uveitis using fourier-domain optical coherence tomography," *Investigative ophthalmology & visual science*, 56(3), pp. 1430–1436.

Rothova, A. *et al.* (1996) "Causes and frequency of blindness in patients with intraocular inflammatory disease," *The British journal of ophthalmology*, 80(4), pp. 332–336.

Saag, K. G. *et al.* (1994) "Low dose long-term corticosteroid therapy in rheumatoid arthritis: an

analysis of serious adverse events," *The American journal of medicine*, 96(2), pp. 115–123.

Sadda, S. R. *et al.* (2006) "Errors in retinal thickness measurements obtained by optical coherence tomography," *Ophthalmology*, 113(2), pp. 285–293.

Sadda, S. R. *et al.* (2007) "Error correction and quantitative subanalysis of optical coherence tomography data using computer-assisted grading," *Investigative ophthalmology & visual science*, 48(2), pp. 839–848.

Sadda, S. R. *et al.* (2010) "Impact of scanning density on measurements from spectral domain optical coherence tomography," *Investigative ophthalmology & visual science*, 51(2), pp. 1071–1078.

Schlaegel, T. (1967) *Essentials of Uveitis*. Boston: Little, Brown, Inc.

Schmidt-Erfurth, U. and Waldstein, S. M. (2016) "A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration," *Progress in retinal and eye research*, 50, pp. 1–24.

Sebag, J. (1987) "Age-related changes in human vitreous structure," *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie*, 225(2), pp. 89–93.

Sharma, S. *et al.* (2015) "Automated Analysis of Anterior Chamber Inflammation by Spectral-Domain Optical Coherence Tomography," *Ophthalmology*. Elsevier Inc, 122(7), pp. 1464–1470.

Sparrow, J. M. *et al.* (2012) "The Cataract National Dataset electronic multi-centre audit of 55 567 operations: risk indicators for monocular visual acuity outcomes," *Eye*, pp. 821–826. doi: 10.1038/eye.2012.51.

Sreekantam, S. *et al.* (2017) "Quantitative analysis of vitreous inflammation using optical coherence tomography in patients receiving sub-Tenon's triamcinolone acetonide for uveitic cystoid macular oedema," *The British journal of ophthalmology*, 101(2), pp. 175–179.

Sugar, E. A. *et al.* (2011) "Identifying a clinically meaningful threshold for change in uveitic macular edema evaluated by optical coherence tomography," *American journal of ophthalmology*, 152(6), pp. 1044-1052.e5.

Suhler, E. B. *et al.* (2018) "Safety and Efficacy of Adalimumab in Patients with Noninfectious Uveitis in an Ongoing Open-Label Study: VISUAL III," *Ophthalmology*. doi: 10.1016/j.ophtha.2017.12.039.

Suhler, E. B. *et al.* (2020) "Long-term safety and efficacy of adalimumab in patients with noninfectious intermediate uveitis, posterior uveitis, or panuveitis," *Ophthalmology*. Elsevier BV. doi: 10.1016/j.ophtha.2020.10.036.

Tallouzi, M. O. *et al.* (2020) "Outcomes important to patients with non-infectious posterior segment-involving uveitis: a qualitative study," *BMJ Open Ophthalmology*. BMJ Specialist Journals, 5(1), p. e000481.

Tangelder, G. J. M. *et al.* (2008) "Precision and reliability of retinal thickness measurements in

foveal and extrafoveal areas of healthy and diabetic eyes," *Investigative ophthalmology & visual science*, 49(6), pp. 2627–2634.

Terheyden, J. H. *et al.* (2021) "Automated quantification of posterior vitreous inflammation: Optical coherence tomography scan number requirements," *Sci Rep*, 11(1), p. 3271.

Thorne, J. E. *et al.* (2019) "Periocular Triamcinolone vs. Intravitreal Triamcinolone vs. Intravitreal Dexamethasone Implant for the Treatment of Uveitic Macular Edema: The PeriOcular vs. INTravitreal corticosteroids for uveitic macular edema (POINT) Trial," *Ophthalmology*. Elsevier BV, 126(2), pp. 283–295.

Tugal-Tutkun, I. and Herbort, C. P. (2010) "Laser flare photometry: A noninvasive, objective, and quantitative method to measure intraocular inflammation," *International ophthalmology*. Springer Netherlands, 30(5), pp. 453–464.

Uji, A. and Yoshimura, N. (2015) "Application of extended field imaging to optical coherence tomography," *Ophthalmology*, 122(6), pp. 1272–1274.

van Velthoven, M. E. J. *et al.* (2006) "Influence of cataract on optical coherence tomography image quality and retinal thickness," *The British journal of ophthalmology*, 90(10), pp. 1259–1262.

van Velthoven, M. E. J. *et al.* (2007) "Recent developments in optical coherence tomography for imaging the retina," *Progress in retinal and eye research*, 26(1), pp. 57–77.

Wakefield, D. and Chang, J. H. (2005) "Epidemiology of uveitis," *International ophthalmology clinics*, 45(2), pp. 1–13.

Walton, R. C. and Ashmore, E. D. (2003) "Retinal vasculitis," *Current opinion in ophthalmology*, 14(6), pp. 413–419.

Wong, I. Y., Koizumi, H. and Lai, W. W. (2011) "Enhanced depth imaging optical coherence tomography," *Ophthalmic surgery, lasers & imaging: the official journal of the International Society for Imaging in the Eye*, 42 Suppl, pp. S75-84.

Zarranz-Ventura, J. *et al.* (2016) "Evaluation of Objective Vitritis Grading Method Using Optical Coherence Tomography: Influence of Phakic Status and Previous Vitrectomy," *American journal of ophthalmology*, 161, pp. 172-180e4.
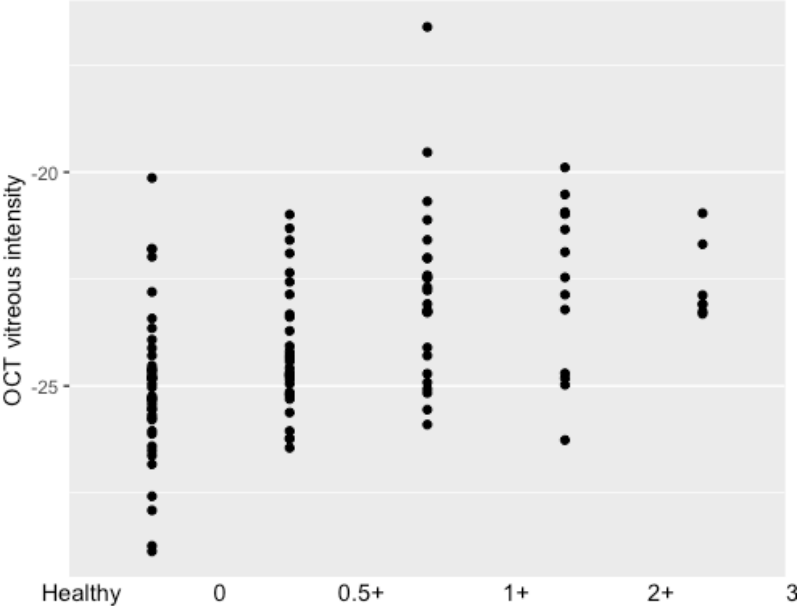
Appendix

Supplementary Figures

Supplementary Figure 1. Individual participant OCT vitreous intensity score eyes with and without vitreous haze
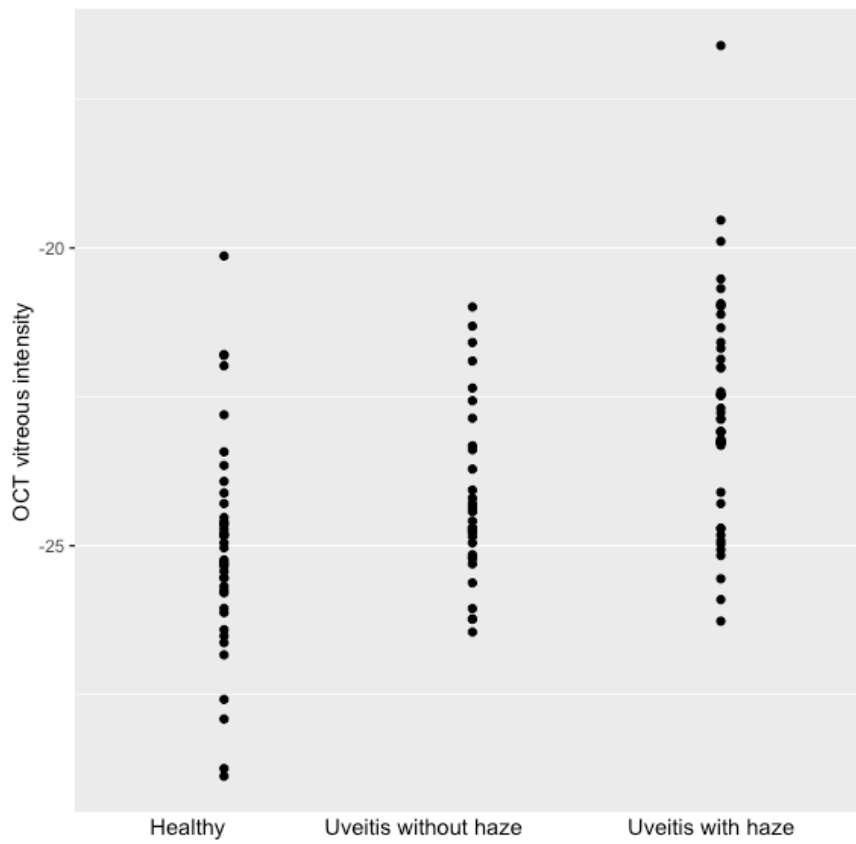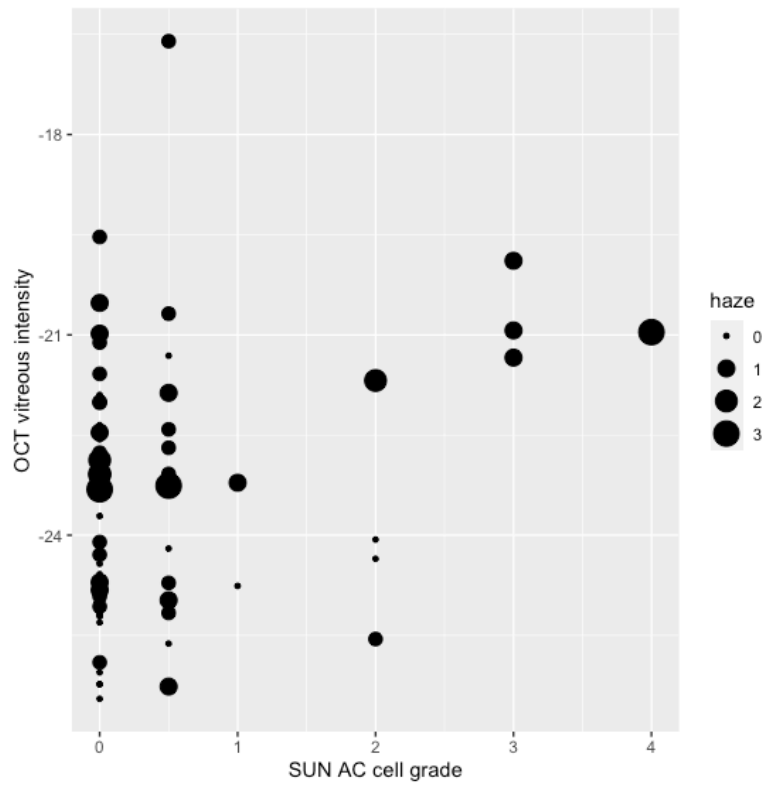
Supplementary Figure 2. Individual participant OCT vitreous intensity score eyes with and without vitreous haze

Supplementary Figure 3. Individual participant OCT vitreous intensity score grouped by AC cell grade

List of publications related to the thesis

Papers included in Chapters 2-4:

1. Liu X, *et al.* (2019). Instrument-based tests for measuring anterior chamber cells in uveitis: a systematic review protocol. Syst Rev. 22;8(1):30.

2. Liu X, *et al.* (2020) Instrument-based Tests for Measuring Anterior Chamber Cells in Uveitis: A Systematic Review. Ocul Immunol Inflamm. 17;28(6):898-907.

3. McNally TW and Liu X, *et al.* (2019) Instrument-based tests for quantifying aqueous humour protein levels in uveitis: a systematic review protocol. Syst Rev. 2019 Nov 26;8(1):287.

4. Liu X, *et al.* (2020) Non-invasive Instrument-Based Tests for Quantifying Anterior Chamber Flare in Uveitis: A Systematic Review. Ocul Immunol Inflamm. 7:1-9.

5. Liu X, *et al.* (2020) Noninvasive Instrument-based Tests for Detecting and Measuring Vitreous Inflammation in Uveitis: A Systematic Review. Ocul Immunol Inflamm. 2020 Oct 6:1-12.

Papers included in Appendix:

6. Montesano, G, et al. (2018). Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis. Sci Rep 8, 1648.

7. Ometto G, *et al.* (2019). ReLayer: a Free, Online Tool for Extracting Retinal Thickness From Cross-Platform OCT Images. *Transl Vis Sci Technol.* 8(3):25.

8. Terheyden, J. H. *et al.* (2021) "Automated quantification of posterior vitreous inflammation: Optical coherence tomography scan number requirements," *Sci Rep*, 11(1), p. 3271.

9. Ometto G, *et al.* (2020) Merging Information From Infrared and Autofluorescence Fundus Images for Monitoring of Chorioretinal Atrophic Lesions. *Transl Vis Sci Technol*. 9(9):38.

SCIENTIFIC REP**O**RTS

OPEN

Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis

G. Montesano[1], C. M. Way[2,3], G. Ometto[1], H. Ibrahim[2,3], P. R. Jones[1,4], R. Carmichael[3], X. Liu[2,3], T. Aslam[5,6,7], P. A. Keane[4], D. P. Crabb[1] & A. K. Denniston[2,3,4]

Detection and evaluation of inflammatory activity in uveitis is essential to the management of the condition, and yet continues to be largely dependent on subjective clinical measures. Optical coherence tomography (OCT) measurement of vitreous activity is an alternative to clinical vitreous haze scoring and has passed a number of early validation studies. In this study we aimed to evaluate the impact of 'operator factors' on the variability of the technique as part of the validation process, and to help evaluate its suitability for 'real world' use. Vitreous haze index was calculated as a ratio between the reflectivity of the vitreous and of the outer retina in each scan. Different scanning conditions were tested and their effect on the measurement is reported. Our results show that the 'quantitative imaging' technique of OCT-measured vitreous activity had good reliability in normal subjects under a range of 'real world' conditions, such as when the operator changes the averaging value. The technique was however vulnerable to highly inaccurate focussing or abnormal downward displacement of the image. OCT-based quantification of vitreous activity is a promising alternative to current subjective clinical estimates, with sufficient 'tolerance' to be used in routine clinical practice as well as clinical trials.

Uveitis is a group of diseases characterized by intraocular inflammation which collectively are a major cause of blindness worldwide[1–4]. One core objective in diagnosis and treatment is the correct identification and measurement of inflammatory activity[5,6]. This assessment has major impact both on routine clinical practice and on endpoint definition in clinical trials. Traditionally, the National Eye Institute (NEI) system for grading of vitreous haze has been the major disease activity endpoint for trials in posterior segment-involving uveitis, acknowledged by the United States Food and Drug Administration (FDA) and European Medicines Agency (EMA)[7,8]. However, the NEI system suffers from being (1) subjective, (2) noncontinuous, (3) poorly discriminatory at lower levels of inflammation, and (4) poorly sensitive in a clinical trial context[5,6,9]. A novel, automated method for the quantification of vitreous haze using optical coherence tomography (OCT) imaging was recently introduced[10] thus providing objective measurement of vitreous inflammation. The method was based on a previously published study[11], using a semi – automated implementation to correlate clinical vitreous haze scores in patients with uveitis and in healthy volunteers. The fully automated method was introduced to avoid the manual segmentation of OCT image sets by graders, a subjective and time-consuming step in the measuring process.

The new technique overcomes many of the well-known limitations of the NEI clinical score, and appears to be a major step forward in the drive towards sensitive objective endpoints for use in uveitis trials and to direct treatment decisions in routine clinical practice[12]. As part of its further validation it is important to determine what the

[1]City, University of London, Optometry and Visual Sciences, London, United Kingdom. [2]Academic Unit of Ophthalmology, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, United Kingdom. [3]Department of Ophthalmology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom. [4]NIHR Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute of Ophthalmology, London, United Kingdom. [5]Manchester Royal Eye Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom. [6]Faculty of Medical and Human Sciences, University of Manchester, Manchester, United Kingdom. [7]School of Built Environment, Heriot-Watt University, Edinburgh, United Kingdom. G. Montesano and C. M. Way contributed equally to this work. Correspondence and requests for materials should be addressed to A.K.D. (email: A.Denniston@bham.ac.uk)

Vertical position of retina within the scan	ART Level	Focus
Middle	100	In focus
Middle	50	In focus
Middle	25	In focus
Middle	12	In focus
Middle	6	In focus
Middle	100	+5
Middle	100	+10
Middle	100	−5
Middle	100	−10
Bottom	100	In focus

Table 1. Scanning protocol. Different setting combinations used during acquisition. Each scan with a specific setting has been repeated three times. ART = Automated Real Time.

potential limitations of this technique in the 'real world' – essentially what are the circumstances under which it would no longer be reliable. In general terms these can be considered as either 'operator factors' (dependent on how the technique is done) or 'patient factors' (intrinsic to the patient and their eye(s)).

In this report, we present a detailed analysis of the impact of 'operator factors' on the variability of the technique, with particular focus on the factors that can significantly affect the measure in healthy subjects where no inflammation is present. The experimental protocol was designed to test different scanning conditions using the Spectralis OCT (Heidelberg Engineering, Heidelberg, Germany). The analysis is aimed at the identification of the optimal acquisition settings that minimise the test-retest variability and changes in the measured value.

Methods

Scanning protocol. Fifteen volunteers with a refractive error within +5 and −5 dioptres (D) were recruited. All subjects underwent a complete ophthalmic examination by an experienced clinician (AD) to confirm the absence of any pathologies. This protocol was approved from the NRES East Midlands Ethics Committee (Ref: 14/EM/1163). Written informed consent was gathered from all subjects. This protocol adhered to the tenets of the declaration of Helsinki.

Macular OCT scans centred on the foveal pit and spanning 20 degrees horizontally were acquired from the right eye of each subject using a spectral domain OCT device (Spectralis SD – OCT, Heidelberg Engineering, Heidelberg, Germany) with a 30-degree lens. An experienced operator used 10 different acquisition settings, each repeated 3 times, to acquire 30 raster scans (7 sections per scan) from each subject. Five Automated Real Time (ART) levels and five focus levels of the retina in the infrared (IR) fundus image were used in the acquisition protocol, as shown in Table 1. The ART level indicates the number of images that are averaged to produce the image of a single section. The positioning of the retina was set to the middle of the scan. This choice was forced by the final application of the proposed methodology, aimed at the measurement of the vitreous haze in patients with uveitis where macular oedema can be present. In fact, the presence of oedema forces the positioning to the middle of the scan in order to capture whole thickness of the swollen retina. As an additional comparison, one acquisition setting included the bottom positioning, ART 100 and in focus.

Table 1 reports the different settings of the acquisition protocol in detail.

Image analysis. To calculate the Vitreous/RPE-relative intensity (VRI), each image was analysed with the VITreous ANalysis (VITAN) software[10], implemented in MATLAB (The MathWorks, Natick, MA, USA). Briefly, for each scan a morphological opening to segment the retina and RPE within the image was performed. Then, a vitreous patch was automatically generated based on this segmentation, excluding any retinal tissue (Fig. 1). The mean intensity of the vitreous patch and of the segmented RPE was measured and the ratio was calculated. The RPE intensity was used as a normalisation term, compensating for global reduction in the signal strength arising from diffused media opacities. The VITAN software then exported the VRI ratio, the vitreous mean intensity and the RPE mean intensity to a spreadsheet for analysis.

Statistical analysis. Linear mixed models were used to assess the effect of different settings on the VRI. ART level and focus were analysed separately, with the ratio as the response variable. Observations consisted of the ratio calculated from each image of the scan. Clustering of sections within the same raster scan and of different repetitions within the same subject was addressed using nested random effects[13]. Due to the discrete nature of the settings, ART level and focus were used as factors rather than continuous variables. The same analysis was used to analyse separately the Vitreous and RPE intensities to calculate the effect of different acquisition parameters on these two values.

A similar approch was employed for the analyses of the variability of the measured ratio at three different levels: within the same raster scan, within subjects (intra-subject) and across subjects (inter-subjects). In this approach, the residuals of the measurement represented the observations. At each level, residuals were calculated as the difference (1) between each measurement and the mean of the seven sections in the raster scan, (2) between each mean of the raster scan and the mean of the three-repeated acquisitions and (3) between the mean of the acquisitions in the single subject and the mean of acquisitions across all subjects. Then, the squared residuals were used to model the variability of the measure at each level (within the raster scan, within subjects and across

180
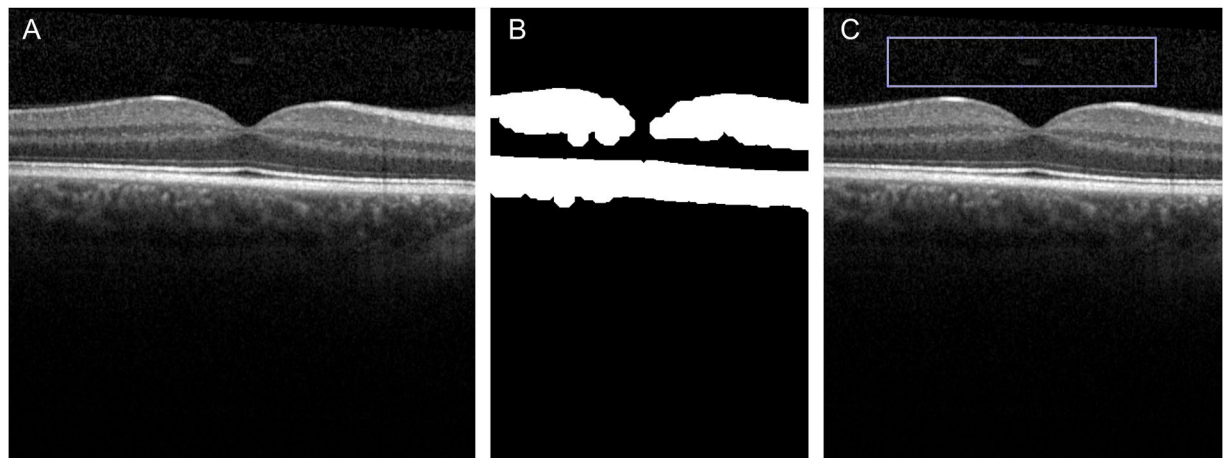
Figure 1. VITAN procedure. (**A**) Example original image. (**B**) Binary image of OCT scan automatically segmented to highlight retinal/RPE layers and cropped to isolate central areas. (**C**) Final automated area of capture overlaid onto original image for user approval.

ART	Global mean ratio	Within scan variability	Intra - subject variability	Inter - subject variability
6	0.043	0.010	0.006	0.008
12	0.043	0.009	0.007	0.010
25	0.043	0.010	0.007	0.007
50	0.039	0.010	0.007	0.012
100	0.048	0.011	0.007	0.016*

**Table 2.** Effect of ART on the VRI. $*P < 0.05$; $**P < 0.01$. Second column reports the estimated mean ratio value for different ART settings. No significant difference could be detected. Variability with different settings is reported as the square root of the estimate from the squared residuals model. The asterisk indicates the only significant difference ($p < 0.05$) that could be detected in pairwise comparisons between different settings (ART 100 showed and increased variability compared to the ART 6 and ART 25). ART = Automated Real Time.

subjects) while changing the value of the parameter of interest (ART level or focus). Assuming normality of the residuals, the squared residuals follow a chi-squared distribution, which is a special case of the Gamma distribution. Therefore, generalized linear models with a Gamma distributed error and a logarithmic link function were used to model the effect of the different settings on squared residuals. The variability was reported as the square root of the estimate obtained from the model of squared residuals.

When a significant effect was detected, pairwise comparisons were performed between different settings and a multiple test correction with the Tukey method was applied.

When failure of the VITAN algorithm could not provide the measurement from at least 3 of the 7 scans or from at least 2 of the 3 repetitions, the raster scan or the repetition was discarded from the analysis for the variability.

All analyses were performed in R (R Foundation for Statistical Computing, Vienna, Austria) and MATLAB.

Data availability statement. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Results

Twenty-one scans with the $+10\,D$ and 7 with $+5\,D$ settings could not be obtained due to difficulties in the acquisition. The VITAN algorithm failed to obtain the measurement in 46 out of 1575 theoretical scans from those with different ART level (3% failure rate) and in 504 out of 1575 with different focus (32% failure rate).

Effect of ART level on VRI value. In our set of images, the ART level had minimal non-significant effect on the VRI value (overall p - value = 0.08, values are reported in Table 2), with slightly higher values with ART 100.

Effect of ART level on VRI variability. Modelling the squared residuals according to different ART levels revealed no significant effects on the variability of different sections within each raster scan (overall p-value = 0.308) and within different raster scans on the same subject (intra-subject variability, overall p-value = 0.869). A moderate effect could be found on the variability across subjects (inter-subject variability, overall p-value = 0.005). In pairwise comparisons, the ART 100 yielded higher variability compared to ART 6 (p = 0.032, 3.99-fold increase) and ART 25 (p = 0.004, 5.41-fold increase). Estimates from the model for variability are reported in Table 2. Figure 2 shows a graphical depiction of these results with a box plot graph.
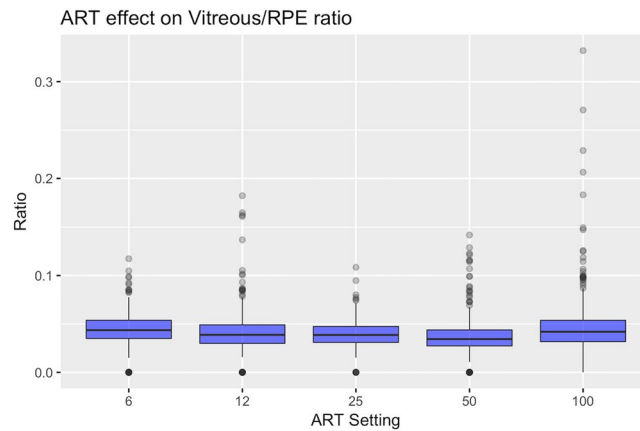
Figure 2. Effect of different ART settings on the VRI ratio. The box plot shows how different ART settings affect the mean VRI value and its variability. The ratio value did not show important variations, with slightly higher and more variable values with ART 100 (Refer to Table 2). The boxes extend from the 25th to the 75th. Outliers (black dots) are points more distant than. The whiskers extend 1.5 times the interquartile range from the box limits. Points exceeding this limits are flagged as outliers (black dots). ART = Automated Real Time.

Focus	Global mean ratio	Within scan variability	Intra - subject variability	Inter - subject variability
−10 D	0.162**	0.019**	0.015**	0.036
−5 D	0.088**	0.013	0.012*	0.029
In focus	**0.048**	**0.011**	**0.007**	**0.023**
+5 D	0.192**	0.032**	0.022**	0.040*
+10 D	0.200**	0.027**	0.033**	0.036

Table 3. Effect of focus on the VRI. *$P < 0.05$; **$P < 0.01$. The second column reports the estimated mean ratio value for different focus settings. All p – values have been calculated comparing each other level to the "In focus" condition (in bold). Variability with different settings is reported as the square root of the estimate from the squared residuals model. The asterisks indicate significant differences according to the legend at the bottom of the table.

Effect of Focus on VRI value. In contrast with the analysis of the ART level, the analysis of the focus showed that this parameter had a major effect on the VRI (overall p-value < 0.001). All acquisition out of focus (referred to the retinal IR image) increased the VRI significantly (Minimum difference + Standard Error: 0.039 ± 0.008; $p < 0.001$), with larger increase using positive offsets (0.14 ± 0.008 increase for +5 D and 0.15 ± 0.01 for +10 D). Results are reported in Table 3.

Effect of Focus on VRI variability. As shown in Fig. 3, Focus significantly affected variability at all levels (within scans, within subjects and across subjects). All settings that deviated from the optimal retinal focus caused a significant increase in within scan variability (all $p < 0.001$) except for the −5 D condition ($p = 0.08$). A significant increase in the within subject variability was observed in any focus offsets (all $p < 0.018$), while only the +5 D caused a significant increase in variability across subjects ($p = 0.038$). Among all settings, positive offsets caused the largest increase in variability compared to the in focus condition. Values are reported in Table 3.

Effect of vertical positioning on the VRI. When compared to standard (middle) positioning within the z-plane, relatively inferior positioning of the retinal image within the acquisition frame also significantly increased the VRI value (Estimated difference ± Standard Error: 0.114 ± 0.007; $p < 0.001$) and the variability at all levels (all p-values < 0.01).

Differential contribution of Vitreous and RPE intensity on the VRI. The individual variation of the two measured components of the ratio (Vitreous intensity and the RPE intensity) is reported in Table 4 for the acquisition parameters that significantly affected the measurement (i.e. focus and positioning). Variation is reported with the absolute difference and the percentage relative to the reference levels of each setting: 'in focus' for the focus and 'middle' for the positioning. For different settings of the focus, the major contribution to the variation in the ratio was due to changes in the vitreous intensity, particularly for positive offsets. Conversely, the increase in vitreous intensity observed with the bottom positioning was caused by both the increase of Vitreous intensity (the numerator of the ratio) and the reduction of RPE intensity (the denominator). Independently of their magnitude, all variations were statistically significant ($p < 0.05$).

Focus effect on Vitreous/RPE ratio
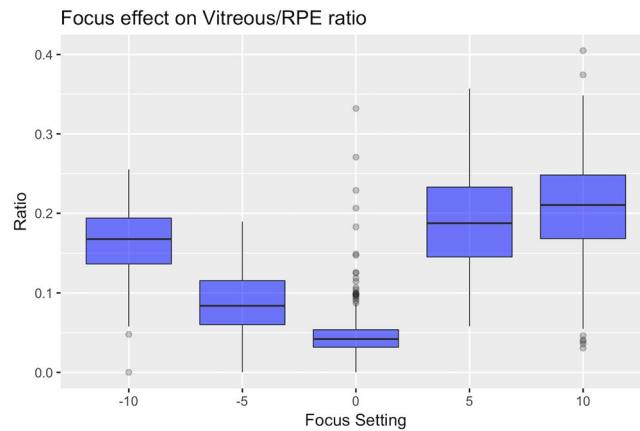
Figure 3. Effect of different focus settings on the VRI. The box plot shows how different focusing condition increase the mean VRI value and its variability compared to scans focused on the retina (denoted as 0 in the graph). The boxes extend from the 25th to the 75th. Outliers (black dots) are points more distant than. The whiskers extend 1.5 times the interquartile range from the box limits. Points exceeding this limits are flagged as outliers (black dots).

		Ratio		Vitreous		RPE	
		Difference	Perc. Change (%)	Difference	Perc. Change (%)	Difference	Perc. Change (%)
Focus	−10 D	0.12	239.78	23.948	262.79	16.91	9.22
	−5 D	0.04	82.56	8.955	98.27	21.51	11.72
	5 D	0.14	297.45	27.677	303.71	6.16	3.36
	10 D	0.15	315.49	31.065	340.89	14.26	7.77
Position	Bottom	0.11	70.30	19.051	67.64	−14.27	−8.43

Table 4. Different contributions to the VRI. The table reports the differences in VRI ratio values (column 3), vitreous intensity (column 5) and RPE intensity (column 7) compared to their respective reference levels (the "In focus" condition for focus and the "Middle" location for position). Columns 4, 6 and 8 report the same changes as percentage increase (or decrease if negative) from the reference level.

Discussion

Our previous work showed that the measurement of the VRI from OCT scans is correlated with the clinical score for vitreous haze in uveitis patients[11,14] and that it could be partially automated[10] and that it was highly sensitive to detecting treatment responses[15]. However, in order to assess whether the VRI can be used in routine clinical practice to detect pathological vitreous haze, it is crucial to study how this measurement can vary using different acquisition settings. This report investigates the extent to which 'operator factors' such as the effect of image averaging (ART level), defocussing and retinal positioning might impact the reliability of the technique. This is particularly important when considering a technique that is intended for use in everyday clinics and not just in the more controlled environment of a clinical trial.

The ART level is used to improve the quality of the images via averaging by increasing signal to noise ratio[16]. The analysis showed a mild, non-significant effect of image averaging ($p = 0.078$) on the ratio measurement (Table 2), with slightly higher values obtained using ART 100. The maximum difference obtained between estimated values was 0.0093. Such a difference is well below the observed increase with vitritis, reported in our previous retrospective analysis (difference in medians, Vitritis – Healthy group = 0.0733)[11]. ART 100 showed a higher inter-subject variability (overall p-value = 0.005), but only pairwise comparisons with the ART 6 and ART 25 were significant. No significant effect could be detected for the within-scan and intra-subject variability. These results are compatible with the fact that image averaging should make the vitreous intensity converge toward a mean value, with no major impact on the ratio value. However, averaging can occasionally smooth out sharp features[17] and change the textural properties of the vitreous. This effect could have an impact on the analysis of images from patients with uveitis by reducing the discriminability between diffuse haze and residual, small clumps of the vitreous in the absence of an active inflammatory processes. From a clinical perspective, it is important to notice that no significant differences could be detected across scans with lower ART values. In the clinical evaluation of macular oedema, a raster scan with the default ART (9) is acquired as a trade-off between image quality and acquisition speed. Results show that VRI can be safely calculated without changing the standard acquisition setting in clinical routine. This result could allow a retrospective application of the measurement, even in sets of OCT images that have not been acquired for this specific purpose.

Changes in the focus had a high impact on the VRI. Different OCT imaging of the vitreous can be obtained with different focusing[18,19]. Although vitreous details can be better imaged with anterior focusing (positive

offsets), a more accurate resolution of vitreous structures can falsely increase the VRI and fail to highlight the diffuse haze due to inflammation. This was well reflected by the increase in vitreous intensity observed when changing the focus and was more prominent when using positive offsets (Table 4). Measurement variability was also greatly affected by the focus and particularly by positive offsets. Increased variability across different sections of the same scan (within scan variability) can be explained by the presence of vitreous structures, varying in density as the scan location moves from the inferior to the superior part of the macular cube. This could have also been the cause of the overall greater variability in the ratio value on scan repetitions (possibly due to slight shifts in the location of the acquisition pattern each time) and in the inter-subject variability where only the $+5\,D$ offset was significantly different (possibly due to a better focusing on the vitreous and thus more affected by inter individual changes in the vitreous structure). Increased variability and vitreous values resulting from posterior focusing might be related to an increase in noise and a relative decrease in the signal to noise ratio, with a worse resolution of the RPE and of the vitreous signal.

Finally, the position of the retinal section within the scan also affected the ratio significantly, increasing the ratio value and the variability of the measure when displaced to the bottom. This change with the bottom positioning might constitute a limitation when imaging patients with important macular oedema, as the RPE is forcefully moved downward in the scan to accommodate the entire retinal thickness in the scan. As shown in Table 4, this increase might be due to the combined effect in the reduction of the RPE intensity with the bottom displacement (possibly a consequence of the known fading effect at the edges of the acquisition window) and to the increase of the vitreous intensity (due to an increase in the noise in the analysed vitreous patch).

This study forms part of the ongoing validation process for a 'quantitative imaging' approach to vitreous haze using OCT. We recognise that one of the limitations of this study is that it did not deal with all possible reliability factors but deliberately focused on 'operator factors' rather than 'patient factors'. 'Patient factors' include the effect of media opacities and ocular surface issues. Media opacities and tear film inhomogeneity are known factors affecting the quality and the signal to noise ratio in OCT scans[16,20], and might falsely increase the measured vitreous haze. An in-depth analysis of these aspects will be possible with a large cohort of normal subjects with a wide age range. Given its focus on 'operator factors' this study was undertaken on healthy controls, and so, unlike most of our previous studies, did not allow a discrimination analysis to investigate the ability of detecting vitritis in uveitis patients. Further investigation of variability and discriminative power of the method will be undertaken as part of a major validation study (OCTAVE) which will also evaluate the impact of increasing the volume sampled through alternative OCT acquisition protocols (eg wide-angle OCT and extra-macular OCT). Lastly, most OCT devices present Gamma-transformed images to increase the contrast of the retinal layers. However, this might not be the optimal condition for vitreous analysis. Measurements obtained from raw, unprocessed data might be more suitable in order to precisely quantify the signal intensity.

In conclusion, this study in healthy subjects suggests that the OCT-based VRI ratio is reasonably tolerant of 'operator factors' and would remain reliable if transferred from a clinical trial setting to the 'real world'. Additional validation studies are ongoing to evaluate the impact of 'patient factors' on reliability, and to assess repeatability and discrimination in a prospective cohort of patients with uveitis as part of the OCTAVE study.

References

1. Durrani, O. M., Meads, C. A. & Murray, P. I. Uveitis: a potentially blinding disease. *Ophthalmologica. Journal international d'ophtalmologie. International journal of ophthalmology. Zeitschrift fur Augenheilkunde* **218**, 223–236, https://doi.org/10.1159/000078612 (2004).
2. Wakefield, D. & Chang, J. H. Epidemiology of uveitis. *International ophthalmology clinics* **45**, 1–13 (2005).
3. Rothova, A., Suttorp-van Schulten, M. S., Frits Treffers, W. & Kijlstra, A. Causes and frequency of blindness in patients with intraocular inflammatory disease. *The British journal of ophthalmology* **80**, 332–336 (1996).
4. Nussenblatt, R. B. The natural history of uveitis. *International ophthalmology* **14**, 303–308 (1990).
5. Lin, P., Suhler, E. B. & Rosenbaum, J. T. The future of uveitis treatment. *Ophthalmology* **121**, 365–376, https://doi.org/10.1016/j.ophtha.2013.08.029 (2014).
6. Denniston, A. K. & Dick, A. D. Systemic therapies for inflammatory eye disease: past, present and future. *BMC ophthalmology* **13**, 18, https://doi.org/10.1186/1471-2415-13-18 (2013).
7. Nussenblatt, R. B., Palestine, A. G., Chan, C. C. & Roberge, F. Standardization of vitreal inflammatory activity in intermediate and posterior uveitis. *Ophthalmology* **92**, 467–471 (1985).
8. Jabs, D. A., Nussenblatt, R. B. & Rosenbaum, J. T., Standardization of Uveitis Nomenclature Working, G. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *American journal of ophthalmology* **140**, 509–516 (2005).
9. Madow, B., Galor, A., Feuer, W. J., Altaweel, M. M. & Davis, J. L. Validation of a photographic vitreous haze grading technique for clinical trials in uveitis. *American journal of ophthalmology* **152**, 170–176 e171, https://doi.org/10.1016/j.ajo.2011.01.058 (2011).
10. Keane, P. A. *et al. Automated Analysis of Vitreous Inflammation* Using Spectral-Domain Optical Coherence Tomography. *Translational vision science & technology* **4**, 4, https://doi.org/10.1167/tvst.4.5.4 (2015).
11. Keane, P. A. *et al.* Objective measurement of vitreous inflammation using optical coherence tomography. *Ophthalmology* **121**, 1706–1714, https://doi.org/10.1016/j.ophtha.2014.03.006 (2014).
12. Denniston, A. K., Keane, P. A. & Srivastava, S. K. Biomarkers and Surrogate Endpoints in Uveitis: The Impact of Quantitative Imaging. *Investigative ophthalmology & visual science* **58**, BIO131–BIO140, https://doi.org/10.1167/iovs.17-21788 (2017).
13. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, https://doi.org/10.18637/jss.v067.i01 (2015).
14. Zarranz-Ventura, J. *et al.* Evaluation of Objective Vitritis Grading Method Using Optical Coherence Tomography: Influence of Phakic Status and Previous Vitrectomy. *American journal of ophthalmology* **161**(172–180), e171–174, https://doi.org/10.1016/j.ajo.2015.10.009 (2016).
15. Sreekantam, S. *et al.* Quantitative analysis of vitreous inflammation using optical coherence tomography in patients receiving sub-Tenon's triamcinolone acetonide for uveitic cystoid macular oedema. *The British journal of ophthalmology* **101**, 175–179, https://doi.org/10.1136/bjophthalmol-2015-308008 (2017).
16. Podkowinski, D. *et al.* Impact of B-Scan Averaging on Spectralis Optical Coherence Tomography Image Quality before and after CataractSurgery. *Journal of ophthalmology* **2017**, 8148047, https://doi.org/10.1155/2017/8148047 (2017).

17. Gelfand, J. M., Nolan, R., Schwartz, D. M., Graves, J. & Green, A. J. Microcystic macular oedema in multiple sclerosis is associated with disease severity. *Brain: a journal of neurology* **135**, 1786–1793, https://doi.org/10.1093/brain/aws098 (2012).

18. Takahashi, A., Nagaoka, T. & Yoshida, A. Enhanced vitreous imaging optical coherence tomography in primary macular holes. *International ophthalmology* **36**, 355–363, https://doi.org/10.1007/s10792-015-0126-y (2016).

19. Pang, C. E., Freund, K. B. & Engelbert, M. Enhanced vitreous imaging technique with spectral-domain optical coherence tomography for evaluation of posterior vitreous detachment. *JAMA ophthalmology* **132**, 1148–1150, https://doi.org/10.1001/jamaophthalmol.2014.1037 (2014).

20. Stein, D. M. *et al.* Effect of corneal drying on optical coherence tomography. *Ophthalmology* **113**, 985–991, https://doi.org/10.1016/j.ophtha.2006.02.018 (2006).

Author Contributions

G.M. data analysis, manuscript drafting, manuscript reviewing. C.W. data collection, manuscript drafting, manuscript reviewing. G.O. data analysis, manuscript drafting, manuscript reviewing. H.I. data analysis, manuscript reviewing. P.J. supervision, manuscript reviewing. R.C. data collection, manuscript reviewing. X.L. data collection, manuscript reviewing. T.A. software development, data analysis, manuscript reviewing. P.K. conceptualization, supervision, manuscript reviewing. D.C. conceptualization, supervision, manuscript reviewing. A.D. conceptualization, supervision, manuscript reviewing.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

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

Article

ReLayer: a Free, Online Tool for Extracting Retinal Thickness From Cross-Platform OCT Images

Giovanni Ometto[1], Ismail Moghul[2], Giovanni Montesano[1,3,4], Andrew Hunter[5], Nikolas Pontikos[4,6], Pete R. Jones[1,4,6], Pearse A. Keane[4,6,7], Xiaoxuan Liu[8,9], Alastair K. Denniston[7,8,9], and David P. Crabb[1]

[1] Division of Optometry and Visual Science, School of Health Sciences, City, University of London, London, UK
[2] UCL Cancer Institute, University College London, London, UK
[3] University of Milan, School of Ophthalmology, Milan, Italy
[4] Moorfields Eye Hospital, London, UK
[5] School of Computer Science, University of Lincoln, Lincoln, UK
[6] Institute of Ophthalmology, University College London, London, UK
[7] NIHR Biomedical Research Centre (Moorfields Eye Hospital NHS Foundation Trust/University College London), London, UK
[8] Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[9] Centre for Translational Inflammation Research, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, UK

Purpose: To describe and evaluate a free, online tool for automatically segmenting optical coherence tomography (OCT) images from different devices and computing summary measures such as retinal thickness.

Methods: ReLayer (https://relayer.online) is an online platform to which OCT scan images can be uploaded and analyzed. Results can be downloaded as plaintext (.csv) files. The segmentation method includes a novel, one-dimensional active contour model, designed to locate the inner limiting membrane, inner/outer segment, and retinal pigment epithelium. The method, designed for B-scans from Heidelberg Engineering Spectralis, was adapted for Topcon 3D OCT-2000 and OptoVue AngioVue. The method was applied to scans from healthy and pathological eyes, and was validated against segmentation by the manufacturers, the IOWA Reference Algorithms, and manual segmentation.

Results: Segmentation of a B-scan took ≤ 1 second. In healthy eyes, mean difference in retinal thickness from ReLayer and the reference standard was below the resolution of the Spectralis and 3D OCT-2000, and slightly above the resolution of the AngioVue. In pathological eyes, ReLayer performed similarly to IOWA ($P = 0.97$) and better than Spectralis ($P < 0.001$).

Conclusions: A free online platform (ReLayer) is capable of segmenting OCT scans with similar speed, accuracy, and reliability as the other tested algorithms, but offers greater accessibility. ReLayer could represent a valuable tool for researchers requiring the full segmentation, often not made available by commercial software.

Translational Relevance: A free online platform (ReLayer) provides free, accessible segmentation of OCT images: data often not available via existing commercial software.

Introduction

Optical coherence tomography (OCT) allows for the acquisition of cross-section pictures of the retina (Fig. 1a). Since its invention, OCT images have rapidly become an established medical tool, supporting clinicians' diagnosis/decisions, and a fundamental resource in scientific research.[1] The key information provided by these pictures, also called b-scans, is the
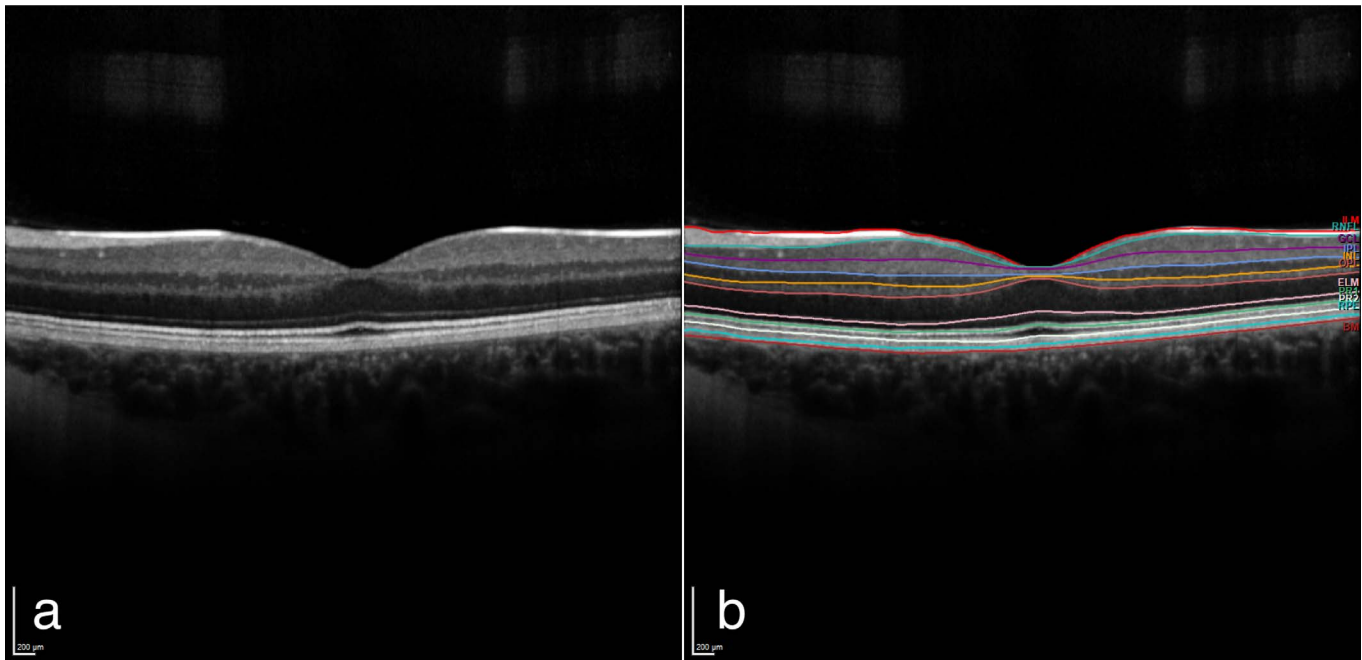
Figure 1. (a) A B-scan image from the test data set as exported from the Heidelberg Engineering Spectralis. (b) The same image scan with the manufacturer's segmentation of 11 layers obtained with the proprietary Heyex software and shown superimposed.

measurement of the thickness of retinal layers that is essential for detecting, monitoring, and guiding treatment for many eye conditions, including glaucoma, diabetic retinopathy, macular edema, age-related macular degeneration, macular hole, macular pucker, central serous retinopathy, and vitreous traction.[2] Currently, measurements can only be obtained using proprietary software and are not available for export or manipulation (Fig. 1b). This presents a limitation, particularly in scientific research, as the availability of this information is essential for the understanding of structural changes of the retina in eye-related pathologies.

To address this problem, segmentation algorithms for the layers in OCT images have been published[3–14] and some of these have been made freely available as software/code: for example, the IOWA Reference Algorithms v3.8.0[7] software, the Graph-Based Segmentation,[8] and the Retina Segmentation Toolbox.[9] However, open segmentation remains inaccessible to most clinicians and researchers due to lack of time, skills, and resources to run, compile, or replicate published algorithms/code.

ReLayer (https://relayer.online) is a free, online platform designed to provide a solution to the accessibility problem and to produce measurements that are as accurate as those from the proprietary software. This is achieved by introducing a novel,

cross platform, segmentation algorithm that is accessible via web browsers. The platform can be used simply by drag-and-dropping image files onto the web-interface (Fig. 2). The analysis is run on Matlab R2016a software (MathWorks, Natick, MA) installed on the server. Results are visualized graphically and are made available for download in comma-separated-value (.csv) format. ReLayer provides the segmentation of inner limiting membrane (ILM), retinal pigment epithelium (RPE), and inner segment/outer segment (ISOS) layers. The retinal thickness, calculated as the distance from ILM to the interface between the Bruch's membrane (BM) and RPE, is computed and visualized on the platform. Here we evaluate the performance of this prototype system, and compare speed, accuracy, and reliability against other available methods, in scans from different acquisition devices and in scans from patients and healthy volunteers.

Materials and Methods

Algorithm

The algorithm was designed to segment retinal layers from 6-mm-wide macular B-scans acquired with Heidelberg Engineering (Heidelberg, Germany) Spectralis and exported as .tiff image files, the default
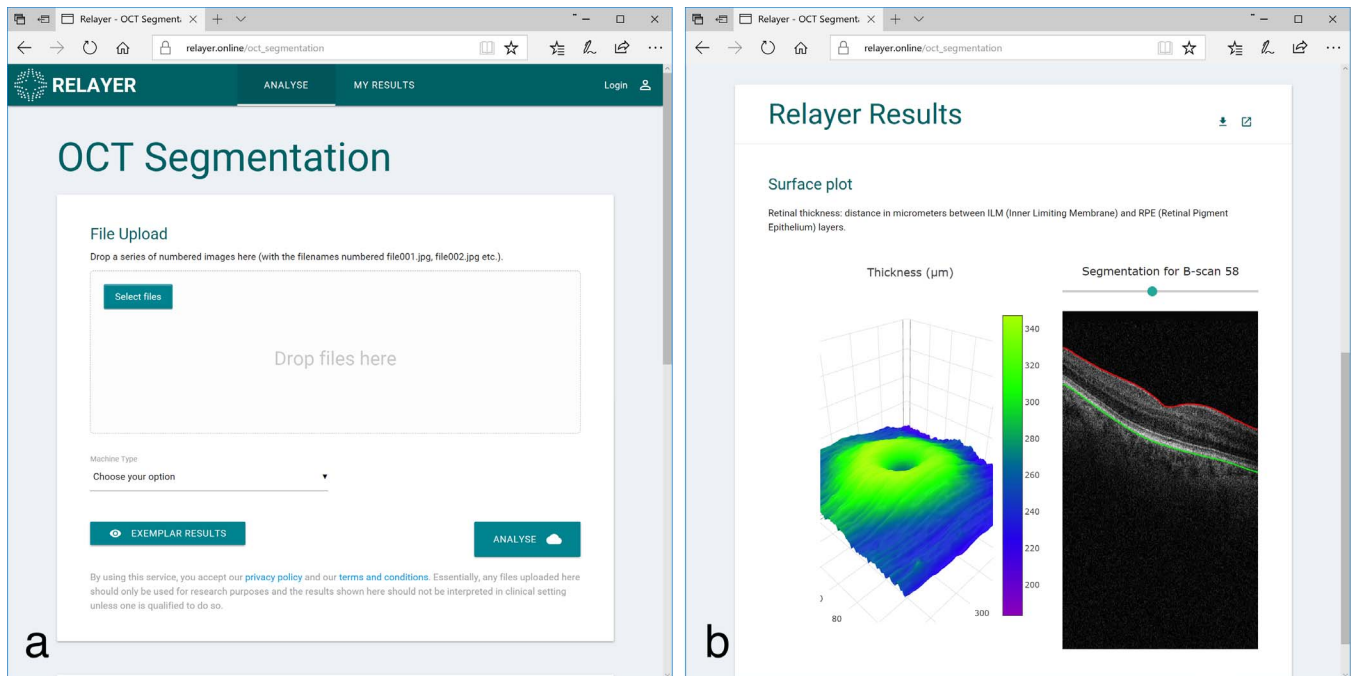
Figure 2. The web interface of ReLayer: (a) the main page showing the area dedicated to the drag-and-drop of the scans or the alternative browsing option and the button to launch the analysis. (b) The visualization of the results including the segmentation superimposed on the scans and the interactive, three-dimensional visualization of the retinal thickness.

export format. The algorithm was then adapted to process 6-mm-wide scans exported in the same image format from Topcon (Tokyo, Japan) 3D OCT-2000 and 3-mm-wide scans from OptoVue (Fremont, CA) AngioVue devices. This was achieved by resampling images from the two devices to match the axial and lateral micrometer-to-pixel ratio of those of the Spectralis. Exported images are 512 × 495 pixels (width × height) in size from the Spectralis, 512 × 855 from the 3D OCT-2000, and 640 × 304 from the AngioVue. Manufacturers report an approximate axial micrometer-to-pixel ratio of 3.87, 2.59, and 3.05 μm, respectively, for the three devices, and an axial resolution of 3.9, 5 to 6, and 5 μm.[15–17] For generalization, we describe the algorithm using micrometers when possible.

The algorithm was developed using Matlab R2016a software (MathWorks) with the Image Processing Toolbox. The algorithm sequentially attempts the identification of the three layers in a B-scan, in order: ILM, RPE, and ISOS. The segmentation of each layer is a two-step process that restricts the search space for the next layer (Fig. 3). In short, the first step is the detection of a line representing the initial guess for each layer, and is based on the detection of nodal points laying over horizontal edges in the image. The second step corrects each guess by

moving it closer to the edges showing in the image. This is obtained using a novel technique based on the active contour model.[18,19] If the input to the algorithm is a volume of multiple B-scans, the algorithm analyzes each B-scan sequentially.

Detection of the Initial Guess

The initial guess was obtained through the identification of 36 nodal points spanning the whole width of the scan and connected with linear interpolation. Due to the bright, linear, and quasi-horizontal appearance of the retinal layers in the scans, these points were selected from those of the horizontal edges, conventionally defined by the magnitude of vertically oriented gradients of intensity. To detect the edges, the image was preprocessed using Gaussian filtering (sigma = 3 pixels, kernel size = 6 pixels) to remove noise (Fig. 4a), and then the magnitude of the vertical component of the gradient was calculated using the Sobel gradient operator.[20] The result of this operation was a new image of the same size of the original one, where the value at each pixel was the magnitude of the vertical gradient at the corresponding location in the original image (Fig. 4b). Then, 36, 14-pixel wide columns (c_i, $i = \{1, \ldots, 36\}$), equally spaced and spanning the whole image-width, were selected. The left and right halves of the first and last
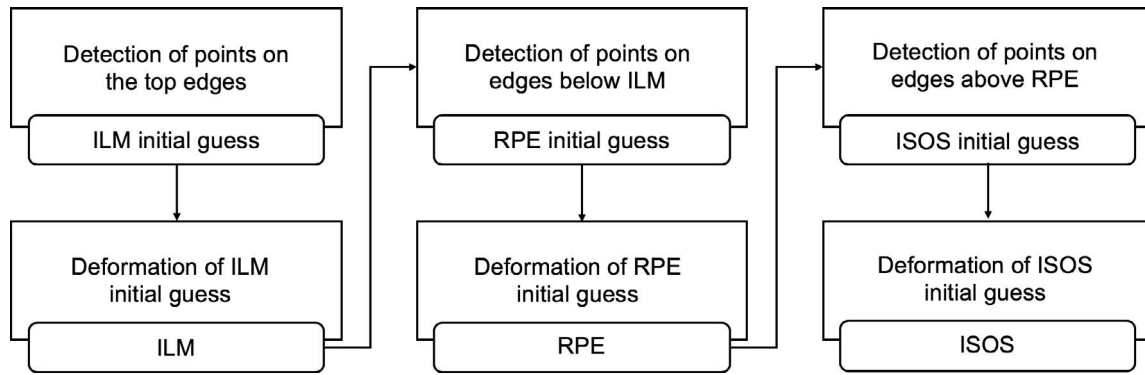
Figure 3. Flowchart of the sequence of operations performed by the algorithm.

columns, centered on the edges of the image, were discarded. Then, all values in each column were averaged across the rows to obtain 36 vertical profiles of the averaged gradient (vp_i) (Fig. 4c). The average was used to weaken the impact of localized, vertical gradients. These vertical profiles were analyzed to identify their peaking values. The peaking values were used in the selection of 36 points $p(i)$, $i = \{1, \ldots, 36\}$, centered in the middle of the respective column c_i and vertically located at the location of the peak. Peaks were identified separately for the initial guesses of the ILM, RPE, and ISOS, to obtain three sets of 36 nodal points: $p_{ILM}(i)$, $p_{RPE}(i)$ and $p_{ISOS}(i)$, respectively (Fig. 5). Of the two highest peaks in each vp_i, the one closer to the top edge of the image was selected as $p_{ILM}(i)$ (Fig. 5a). The closest peak to the bottom of the image, of those below the ILM and higher than half the highest peak below the ILM, was defined as $p_{RPE}(i)$ (Fig. 5c). To detect the points of the ISOS, each vp_i was multiplied by a gamma probability density function (g_{pdf}), with the origin shifted 20 µm above the RPE, oriented toward the top of the image and defined by the shape parameter k = 1.84 and scale parameter $\theta = 58$ µm. The resulting statistical mode of such g_{pdf} was approximately equal to 80 µm. The g_{pdf} was designed so that the multiplication $vp_i * g_{pdf}$ would strengthen the peaks of vp_i close to the peak of g_{pdf} and cancel out peaks below or closer than 20 µm to the RPE. Then, the points $p_{ISOS}(i)$ were selected as the highest peaks in the profiles $vp_i * g_{pdf}$ (Fig. 5e). If the algorithm could not identify any of these peaks, the relative points for the initial guess were discarded. Finally, the initial guesses for the ILM, RPE, and ISOS were obtained by linear interpolation of the identified nodal points $p_{ILM}(i)$, $p_{RPE}(i)$, and $p_{ISOS}(i)$.

Active Contour Model

The second step in the analysis was based on a modified version of the established technique known as "active contour model" or "snake,"[18] frequently used in computerized image analysis for the segmen-
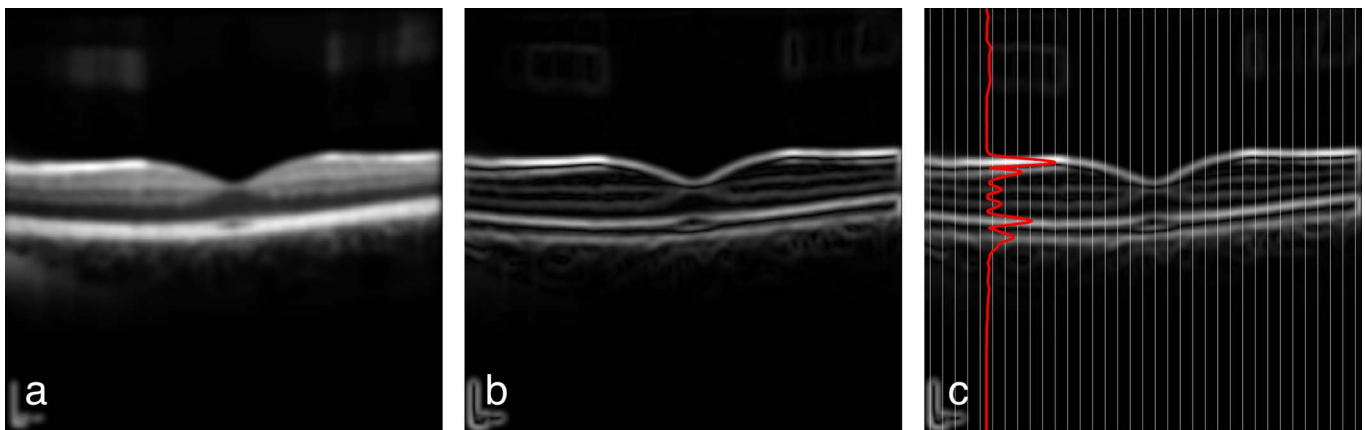
Figure 4. (a) Example B-scan processed with the Gaussian filter for noise removal; (b) processed with Sobel gradient operator for edge detection; (c) divided in 36 columns to obtain 36 vertical profiles of the of the averaged gradient. In *red*, the vertical profile vp_6 obtained averaging the values across the rows of the sixth column.
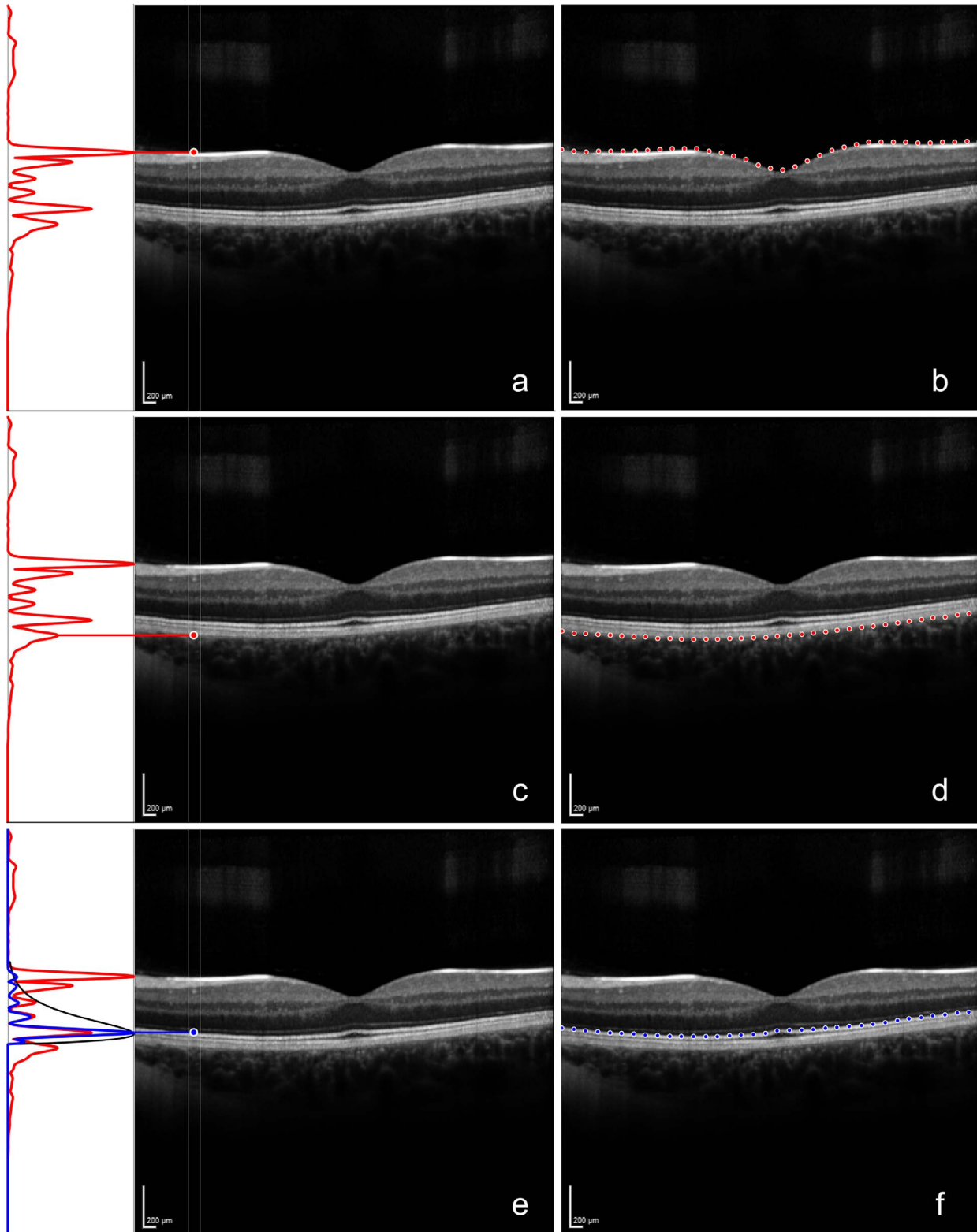
Figure 5. (a) The vertical profile vp_6 and the nodal point $p_{ILM}(6)$. The vertical coordinate of the point is identified by the high peak in vp_6 closest to the top of the image; (b) the 36 nodal points p_{ILM} of the initial guess for ILM. (c) The vertical profile vp_6 and the nodal point $p_{RPE}(6)$. The vertical coordinate of the point is identified by the high peak in vp_6 below ILM and closest to the bottom margin of the image; (d) the 36 nodal points p_{RPE} of the initial guess for RPE. (e) The vertical profile vp_6 (in *red*), the g_{pdf} starting 20 above the RPE (in *black*), the profile of $vp_i * g_{pdf}$ (in *blue*), and the point $p_{ISOS}(6)$. The vertical coordinate was obtained as the highest peak of $vp_i * g_{pdf}$; (f) the 36 nodal points p_{ISOS} of the initial guess for ISOS.

Table 1. Weights and Parameters of the Active Contours Models

	ILM	ISOS	RPE
k	0.8	0.8	0.4
w_{line}	0	0	0
w_{edge}	30	30	10
α	0.3	0.3	1
β	10^2	10^2	10^4

tation of contours in images. Briefly, the model iteratively deforms a line (the initial guess) so that it adheres to the boundaries in an image. This is achieved by solving an energy minimization problem governed by the image energy E_{image} that pulls the points of the line toward lines and edges in the image, and the internal energy of the line $E_{internal}$, interpretable as the stretching-bending capabilities of the line, which resists the deformation. This model has the advantage of being robust to noise and discontinued boundaries but requires priors, such as an initial guess and the weights/parameters defined by the user. The total energy of the contour E_{tot} that is minimized through the iterations is given by:

$$E_{tot} = \int_0^1 k \cdot E_{image}(v(s), w_{line}, w_{edge}) + E_{internal}(v(s), \alpha, \beta) ds,$$

where v_i, $i = \{1, \ldots, 512\}$ are the points of the contour across the width of the image, w_{line}, w_{edge} and k are user-assigned weights that define the impact of lines and edges in the calculation of E_{image}, and the impact of E_{image} in the calculation of E_{tot}, while α and β are parameters that control the amount of stretch and curvature of the contour. The proposed algorithm used a modified version of the conventional two-dimensional model that restricts the deformation of the line to the vertical dimension, by allowing the points in the line to move only upwards or downwards. This was achieved by removing horizontal components from the conventional formulation of the minimization problem, reducing its dimensionality and therefore the complexity of the computation. This modified, one-dimensional active contour model was used three times by the algorithm, one for each of the initial guesses. The parameters used are shown in Table 1. The active contour models of ILM and ISOS shared the same parameters. Different parameters were used for the RPE to make the contour "stiffer", allowing less sharp bends to reflect the cross-sectional morphology of the layer. At the end of the 50th

iteration, the deformed initial guess represented the final segmentation.

Evaluation

The method was tested against the manufacturer's segmentation for scans acquired from healthy eyes with the Spectralis and AngioVue. The segmentation from the Spectralis was obtained from .vol files exported from a version of the Heyex software (Heidelberg Engineering) enabled for RAW data export. The segmentation from the AngioVue was obtained from the .xml files exported from the device. The manufacturer's segmentation of scans from healthy eyes acquired with the Topcon device was not available, and the method was evaluated against a manual segmentation by an expert clinician (XL). The manual segmentation was aided by a custom tool created for the purpose with Matlab. The tool allowed the clinician to select points on the scans and interpolated them with polynomial fitting lines. The test scans were obtained from three data sets of volunteers and one data set of patients, each acquired with one of the three devices for previous studies. Volunteers underwent a visit from a clinician to exclude any pathology. The fourth data set included randomly selected patients with a range of known retinal pathology. The first data set included 48 raster scans acquired with Spectralis from 48 healthy eyes of 24 volunteers (protocol approved from the NRES East Midlands Ethics Committee, Ref: 14/EM/1163).[21] The second data set consisted of 18 raster scans acquired with the 3D OCT-2000 from 18 healthy eyes of nine volunteers (protocol approved from the City University of London Ethics Committee, Ref: OPT/PR/16-17/36). The third data set included 15 raster scans from 15 healthy eyes of 15 volunteers acquired with the AngioVue (protocol approved from the Humanitas Gavazzeni Hospital Ethics Committee, Ref: 253-17 GAV).[22] The fourth data set included 19 raster scans acquired with Spectralis from 19 eyes of 19 patients in presence of known retinal pathology (including macular edema, age-related macular degeneration, previous choroidal neovascularization) (protocol approved from the NRES East Midlands Ethics Committee, Ref: 14/EM/1163).[21] From each raster scan of a volunteer, a single B-scan was randomly selected and used for the evaluation. The two segmentations from the Spectralis and AngioVue underwent a check by a clinician (GM) to identify any visible errors and, with the manual segmentation of the Topcon scans, were used as the Reference Standard (RS) for testing. In the data set of patients with macular edema, one of the five central B-scans was randomly selected for the

Table 2. Distance Between the Proposed Segmentation and the RS, and Difference of the Calculated Thickness for the Three Devices

Acquisition Device	Depth Resolution, μm	MAD			
		Thickness (RPE-ILM)	ILM	RPE	ISOS
Heidelberg Engineering Spectralis	3.87	0.92 (0.32) pixels 3.59 (1.24) μm	0.89 (0.21) pixels 3.45 (0.83) μm	0.88 (0.32) pixels 3.41 (1.24) μm	1.03 (0.40) pixels 3.97 (1.55) μm
Topcon 3D OCT-2000	5–6	1.40 (0.92) pixels 3.63 (2.39) μm	0.85 (0.19) pixels 2.21 (0.51) μm	1.17 (0.48) pixels 3.02 (1.25) μm	1.38 (0.31) pixels 3.58 (0.80) μm
OptoVue AngioVue	5	2.02 (1.11) pixels 6.16* (3.41) μm	0.69 (0.10) pixels 2.11 (0.30) μm	2.04 (1.17) pixels 6.21* (3.60) μm	1.45 (0.37) pixels 4.41 (1.11) μm

Mean distance values above the resolution of the instrument are marked with an asterisk (*). SD values are reported in parenthesis.

analysis from each raster scan to capture structural changes. Here, the RS was obtained as the average of two manual segmentations by two clinicians (GM and XL). Areas of the scans where either clinician was unable to identify the layer were excluded from the analysis. On these scans, the RS was compared to the segmentation results of the proposed algorithm as well as to the segmentation from the manufacturer (Heidelberg Engineering) and from the IOWA Reference Algorithms. The full segmentation from IOWA was obtained from the XML files created during the segmentation. The mean absolute distance (MAD) was chosen to quantify the difference between RS and the tested segmentation for the three layers. The layer here defined as RPE was compared to the BM identified by the Spectralis to reflect the different notations. Retinal thickness, obtained as the distance between ILM, ISOS, and RPE (or BM in the Spectralis), was also measured. Finally, the micrometer-to-pixel ratios provided by the manufacturers were used to convert measurement in pixels to micrometers. This conversion allowed the evaluation of the MAD, revealing meaningful segmentation differences when the calculated distance was higher than the resolution of the device. MAD values for retinal thickness were calculated for each scan and compared across algorithms using mixed effects models. All statistical calculations were performed in R (R Foundation for Statistical Computing, Vienna, Austria).

Results

Mean time to segment a single B-scan was completed in 0.94 seconds (standard deviation [SD] 0.15) when running on a desktop computer with an Intel (Santa Clara, CA) Core i5-6500 CPU @ 3.20 GHz and 16 MB of RAM memory. The online processing may take longer as images need to be uploaded to the server and can vary depending on the workload of the server, the number of clients connected at the time of processing, and the size of the volume to be processed. For these reasons, the total execution time for a volume can range between approximately 30 seconds to 2 minutes.

In healthy volunteers, the clinical validation of the segmentation by the manufacturers did not identify any errors in the scans from the Spectralis, while some were identified in the segmentation of the RPE and ISOS layers from the AngioVue, but were considered minor. Table 2 shows the resulting MAD in pixels and in micrometers between layers segmented by ReLayer and RS for the three devices in scans from healthy eyes. The mean difference in calculated retinal thickness (SD) was 3.45 μm (0.83), 3.63 μm (2.39), and 6.16 μm (3.41), respectively for the Spectralis, 3D OCT-2000, and AngioVue. These values were lower than the resolution of the Spectralis and the 3D OCT-2000 (3.9 and 5–6 μm, respectively) and slightly above the declared resolution of the AngioVue (5 μm). This means that the mean difference of the proposed segmentation and the RS is negligible for the measurement of the retinal thickness from the first two devices under evaluation. Similarly, the comparison of individual layers revealed a mean difference below the resolution of all three instruments, with the exception of the RPE layer segmented on AngioVue scans, where the mean difference resulted slightly above its resolution. Results of the segmentation are shown for random samples of the three data sets in Figure 6. By inspection, no major differences could be identified between the proposed method and the segmentation used as the RS.
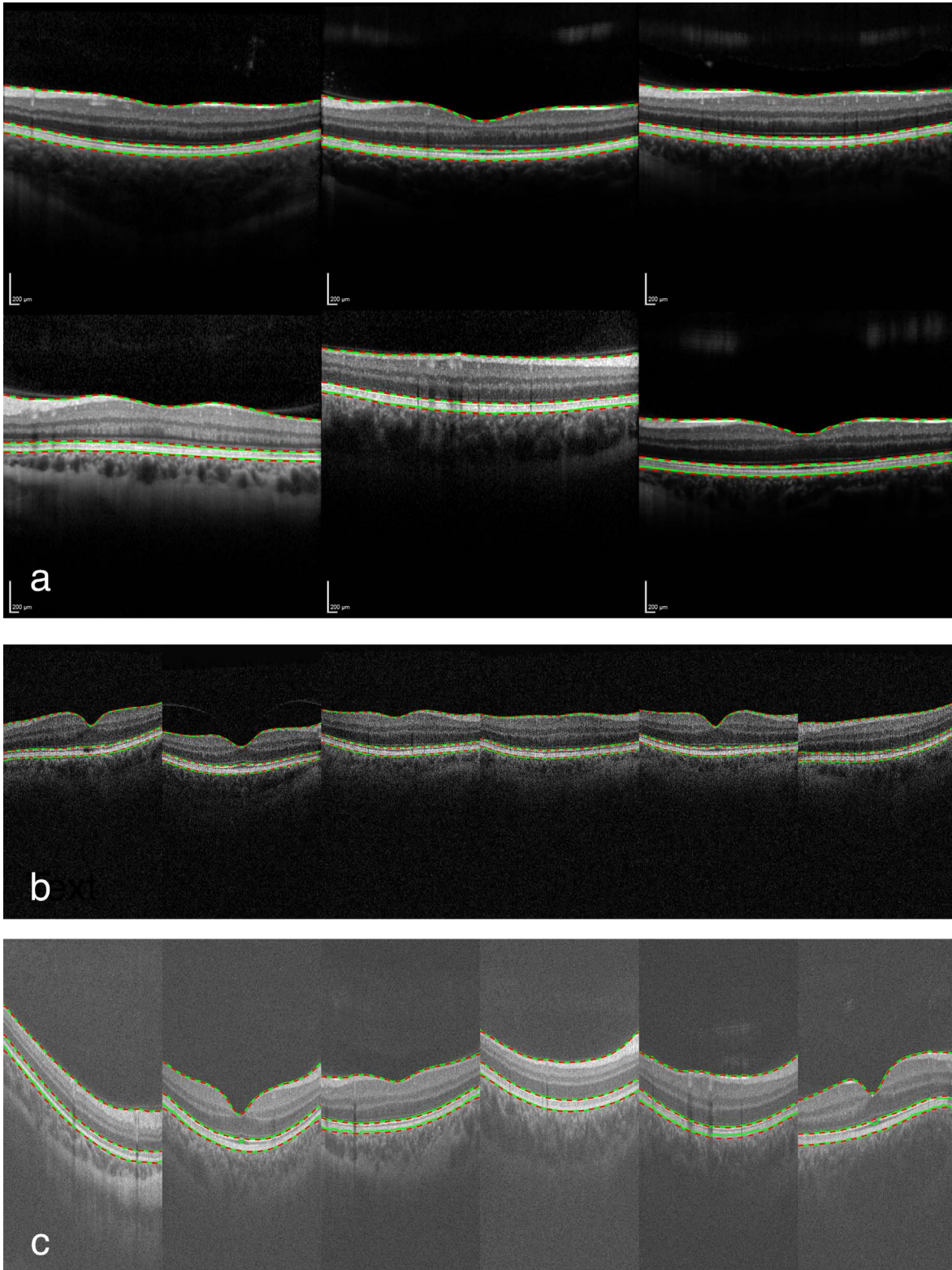
Figure 6. The segmentation results from RS (*green*) and from ReLayer (*dashed red*) in a random subset of the test images from healthy subjects for images acquired with: (a) Heidelberg Engineering "Spectralis," (b) Topcon 3D "OCT-2000," and (c) OptoVue "AngioVue."

Table 3. Distance Between the Three Segmentation Algorithms and the RS in the Data Set of Scans From Pathological Eyes

	MAD			
Algorithm	Thickness (RPE-ILM)	ILM	RPE	ISOS
ReLayer	2.18 (0.98) pixels	0.37 (0.15) pixels	0.75 (0.69) pixels	1.09 (1.60) pixels
	8.44* (3.80) μm	1.45 (0.59) μm	2.89 (2.69) μm	4.21* (6.19) μm
Heidelberg Engineering	4.07 (0.67) pixels	0.85 (0.19) pixels	1.56 (0.53) pixels	1.50 (0.43) pixels
	15.74* (2.61) μm	2.55 (0.64) μm	6.06* (2.06) μm	5.80* (1.68) μm
IOWA	2.23 (0.42) pixels	1.60 (0.14) pixels	1.57 (0.40) pixels	1.24 (1.02) pixels
	8.64* (1.64) μm	6.19* (0.55) μm	6.09* (1.56) μm	4.82* (3.93) μm

Mean distance values above the resolution of the instrument (3.97 μm) are marked with an asterisk. SD values are reported in parenthesis.

Table 3 shows the results of the comparison between ReLayer, Spectralis, and IOWA algorithms against the RS in scans acquired with Spectralis from eyes with pathology. All thickness measurements obtained with the three different segmentation algorithms showed an average difference with RS that was greater than the published resolution of the device (3.9 μm). All individual layers identified by IOWA deviated from the RS in a measure that was on average greater than the resolution of the device. The mean distance between the segmentation from Spectralis and RS was below the resolution and, therefore, negligible only for segmentation of the ILM. The mean distance between the segmentation from ReLayer and the RS was negligible for both ILM and RPE layers. However, calculated SD values showed a greater variation around the average distance between ReLayer and RS than between the other two algorithms and RS, particularly for the RPE and ISOS. MAD average value for the retinal thickness obtained with ReLayer on each scan showed no statistically significant differences from IOWA ($P = 0.973$, 0.05 ± 0.23 μm maximum error-difference). However, both ReLayer and IOWA showed significantly lower MAD values compared with Spectralis ($P < 0.001$, 1.88 ± 0.23 μm and 1.83 ± 0.23 μm maximum error-difference, respectively). Examples of the segmentation results from the three algorithms and RS are shown in Figure 7 on selected areas to highlight their behavior in presence of pathological changes. See Supplementary Figure S1 for the segmentation results on the whole data set.

Discussion

In healthy eyes, the segmentation from ReLayer was as accurate as that from the manufacturer in scans obtained from the Heidelberg Engineering Spectralis and as reliable as a manual segmentation by a clinical expert in scans from the Topcon 3D OCT-2000. The calculated thickness differed from the RS by 3.59 and 3.63 μm on average; these values are below the resolution of both instruments and compatible with measured repeatability of acquisition devices.[23] The mean difference of the segmentation, evaluated for individual layers, was also below the resolution, meaning that further improvements would not be beneficial. When compared with the segmentation from the OptoVue AngioVue, the thickness calculated by the proposed method differed by 6.16 μm, 1.16 μm above the published resolution of the device. However, small segmentation errors were noted by the clinician when evaluating the RS segmentation from the AngioVue. These imperfections could have contributed to this difference and represented a limitation in the assessment of the method for this particular device. The brighter appearance of scans taken with this device could have also affected the segmentation by slightly changing the behavior of the proposed algorithm, which is based on the gradient between brighter/darker areas in the image. However, this could be an instance where using the same cross-platform segmentation algorithm would be particularly beneficial, providing a more homogeneous approach to the task. In short, these results indicate that the proposed segmentation was correct and within expected levels of measurement variability of that obtained from the devices.[23]

The evaluation on scans from eyes with pathology revealed nonnegligible differences between all three segmentation methods and the RS. MAD of the proposed method from the RS was lower than that seen in Spectralis and IOWA. In addition, greater
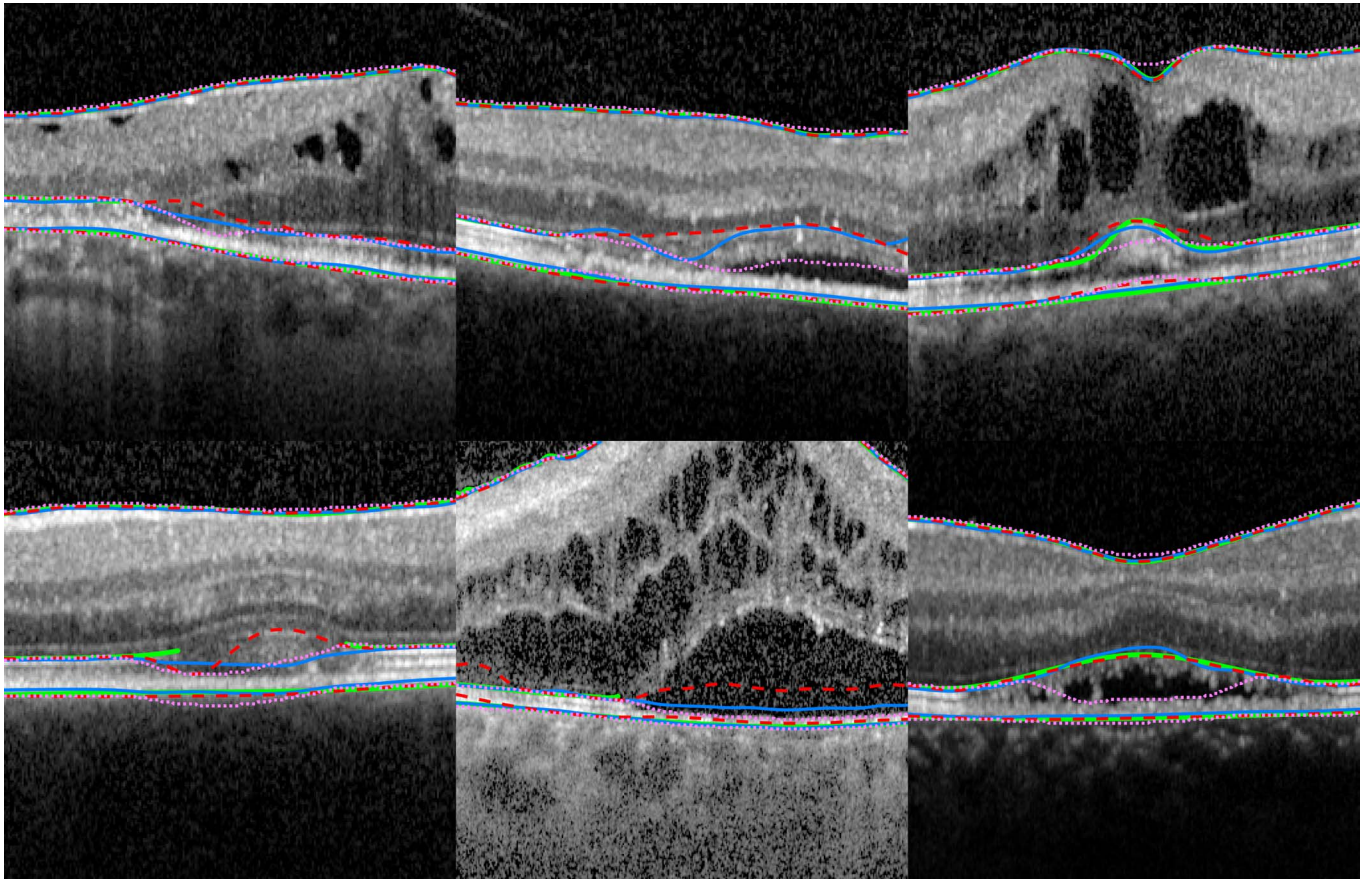
Figure 7. Segmentation results on six selected areas depicting structural changes from six of the pathological scans used for testing. The RS is shown in *green* where both clinicians could identify the layer. ReLayer segmentation is represented by a *dashed red line*, Heidelberg Engineering segmentation by a *blue line*, and IOWA segmentation by a *pink dotted line*.

variation demonstrated by greater SD values suggests that segmentation errors by ReLayer were larger in some areas of these scans. Errors by the proposed algorithm could be due to its design, which is based on several hard thresholds. Alternatively, errors could be caused by isolated edges extending across multiple A-scans and trapping the segmentation into local minima. These results reflect the different behaviors of the three methods rather than identifying one to be superior. Segmentation discrepancies with the RS larger than the resolution show that a single, generalizable algorithm capable of accurately segmenting layers in both healthy and pathological scans is still an open challenge. These findings support ideas of using disease-specific algorithms in the presence of ocular conditions.[6,24] This problem will be addressed by ReLayer with future introduction of variations of the algorithm to the platform, customized to address individual conditions. The comparison of the MAD average for the calculated retinal thickness showed a similar deviation from the RS for ReLayer and

IOWA, which was smaller than Spectralis. Notably, the maximum error-difference was below the resolution of the instrument.

Execution time is generally slower than the segmentation provided by the manufacturers and IOWA but is still fast enough to allow the analysis of raster scans in a clinical setting, for example during a patient's consultation. ReLayer is at the very early stages of its development and improvements are planned in the near future. Execution time can and will be improved considerably by translating the code into a compiled language. In addition, we plan to move our service to cloud computing, allowing for users all over the world to use the software at the same time, with no negative impact on performance

ReLayer is fully automatic, free, and has no requirements other than the access to a web browser. The intuitive drag-and-drop of the scans, the 3D visualization of the thickness profile, and the download of the coordinates of all segmented layers in .csv files, make results easily accessible. For these reasons,

we believe that ReLayer represents a useful tool to both researchers and clinicians. Future developments will include the segmentation of new layers, support for OCT scans from Carl Zeiss Meditec (Jena, Germany) Cirrus devices, and for wide-field scans. Finally, the platform will be upgraded to include disease-specific segmentation, to allow processing of multiple volumes with a single upload and to allow manual correction of the results.

Acknowledgments

Disclosure: **G. Ometto**, None; **I. Moghul**, None; **G. Montesano**, None; **A. Hunter**, None; **N. Pontikos**, None; **P.R. Jones**, None; **P.A. Keane**, Deepmind (C), Allergan (R), Bayer (R), Carl Zeiss Meditec (R), Haag-Streit (R), Heidelberg Engineering (R), Novartis (R), Topcon (R); **X. Liu**, None; **A.K. Denniston**, None; **D.P. Crabb**, Roche (F), CenterVue (C), Allergan (R), Santen (R)

References

1. Fujimoto J, Swanson E. The development, commercialization, and impact of optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2016;57:OCT1–OCT13.
2. American Academy of Ophthalmology. What conditions can OCT help to diagnose? Available at: https://www.aao.org/eye-health/treatments/what-does-optical-coherence-tomography-diagnose. Accessed December 1, 2018.
3. DeBuc DC. A review of algorithms for segmentation of retinal image data using optical coherence tomography. In: Ho P-G, ed. *Image Segmentation*. London, UK: InTech; 2011:15–54.
4. Lang A, Carass A, Hauser M, et al. Retinal layer segmentation of macular OCT images using boundary classification. *Biomed Opt Exp*. 2013; 4:1133–1152.
5. Lang A, Carass A, Sotirchos E, Calabresi P, Prince JL. Segmentation of retinal OCT images using a random forest classifier. *Proc SPIE Int Soc Opt Eng*. 2013:86690R.
6. Fang L, Cunefare D, Wang C, Guymer RH, Li S, Farsiu S. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed Opt Exp*. 2017;8:2732–2744.
7. Abràmoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. *IEEE Rev Biomed Eng*. 2010;3:169–208.
8. Teng P. Caserel–an open source software for computer-aided segmentation of retinal layers in optical coherence tomography images. *Zenodo*. 2013;10. https://doi.org/10.5281/zenodo.17893.
9. Rathke F, Schmidt S, Schnörr C. Probabilistic intra-retinal layer segmentation in 3-D OCT images using global shape regularization. *Med Image Anal*. 2014;18:781–794.
10. Belghith A, Bowd C, Medeiros FA, et al. Does the location of Bruch's membrane opening change over time? Longitudinal analysis using San Diego Automated Layer Segmentation Algorithm (SALSA). *Invest Ophthalmol Vis Sci*. 2016;57:675–682.
11. Belghith A, Bowd C, Medeiros FA, Weinreb RN, Zangwill LM. Automated segmentation of anterior lamina cribrosa surface: how the lamina cribrosa responds to intraocular pressure change in glaucoma eyes? *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. New York, NY: IEEE; 2015:222–225.
12. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmol Retina*. 2017;1:322–327.
13. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Med*. 2018; 24:1342.
14. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018; 172:1122–1131.
15. Heidelberg Engineering "Spectralis" datasheet. Available at: https://business-lounge.heidelbergengineering.com/gb/en/products/spectralis/spectralis/downloads/. Accessed December 1, 2018.
16. Topcon "3D OCT-2000" datasheet. Available at: http://www.topcon-medical.eu/files/EU_

Downloads/Products/3D_OCT-2000/3D_OCT_ 2000series_en.brochure.pdf. Accessed December 1, 2018.

17. OptoVue "AngioVue" datasheet. Available at: https://www.optovue.com/products/AngioVue-ifusion. Accessed December 1, 2018.

18. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *Int J Comp Vis*. 1988;1: 321–331.

19. Farsiu S, Chiu SJ, Izatt JA, Toth CA. Fast detection and segmentation of drusen in retinal optical coherence tomography images. *Ophthalmic Technologies XVIII*. Bellingham, WA: International Society for Optics and Photonics; 2008: 68440D.

20. Sobel I, Feldman G. A 3x3 isotropic gradient operator for image processing, presented at a talk at the Stanford Artificial Project. In: Duda R, Hart P, eds. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1968:271–272.

21. Montesano G, Way C, Ometto G, et al. Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis. *Sci Rep*. 2018;8:1648.

22. Allegrini D, Montesano G, Fogagnolo P, et al. The volume of peripapillary vessels within the retinal nerve fibre layer: an optical coherence tomography angiography study of normal subjects. *Br J Ophthalmol*. 2018;102:611–621.

23. Terry L, Cassels N, Lu K, et al. Automated retinal layer segmentation using spectral domain optical coherence tomography: evaluation of inter-session repeatability and agreement between devices. *PLoS One*. 2016;11:e0162001.

24. Chiu SJ, Izatt JA, O'Connell RV, Winter KP, Toth CA, Farsiu S. Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images. *Invest Ophthalmol Vis Sci*. 2012;53:53–61.

Page intentionally left blank

scientific reports

OPEN

Automated quantification of posterior vitreous inflammation: optical coherence tomography scan number requirements

Jan Henrik Terheyden[1]✉, Giovanni Ometto[2], Giovanni Montesano[2], Maximilian W. M. Wintergerst[1], Magdalena Langner[1], Xiaoxuan Liu[3,4], Pearse A. Keane[5], David P. Crabb[2], Alastair K. Denniston[3,4,5] & Robert P. Finger[1]✉

Quantifying intraocular inflammation is crucial in managing uveitis patients. We assessed the minimum B-scan density for reliable automated vitreous intensity (VI) assessment, using a novel approach based on optical coherence tomography (OCT). OCT volume scans centered on the macula were retrospectively collected in patients with uveitis. Nine B-scans per volume scan at fixed locations were automatically analyzed. The following B-scan selections were compared against the average score of 9 B-scans per volume scan as a reference standard: 1/3/5/7 central scans (1c/3c/5c/7c), 3 widely distributed scans (3w). Image data of 49 patients (31 females) were included. The median VI was 0.029 (IQR: 0.032). The intra-class-correlation coefficient of the VI across the 9 B-scans was 0.923. The median difference from the reference standard ranged between 0.001 (7c) and 0.006 (1c). It was significantly lower for scan selection 3w than 5c, p(adjusted) = 0.022, and lower for selection 7c than 3w, p(adjusted) = 0.003. The scan selections 7c and 3w showed the two highest areas under the receiver operating curve (0.985 and 0.965, respectively). Three widely distributed B-scans are sufficient to quantify VI reliably. Highest reliability was achieved using 7 central B-scans. Automated quantification of VI in uveitis is reliable and requires only few OCT B-scans.

Uveitis is a common inflammatory disease of the eye, accounting for 5–10% of visual impairment worldwide[1,2]. The disease affects the vascular layer (consisting of iris, ciliary body and choroid) of people who are frequently of working age[1,3]. Quantification of intraocular inflammation is crucial in managing patients with uveitis. To date the quantification of intraocular inflammation is mostly done semi-quantitatively by subjective clinical evaluation, which comes with a range of limitations common to subjective ratings[3–5]. Thus, several approaches have been developed to quantify vitreous intensity (VI) more objectively[6–9]. This includes quantification of vitreous inflammation based on optical coherence tomography (OCT) scans[8–12].

The developed algorithm for an automatic assessment of vitreous inflammation is based on the measurement of hyperreflective spots within the posterior vitreous included on macular OCT scans. As this parameter alone is prone to artefacts due to media opacities, a score relative to the retinal pigment epithelium has been established in previous studies and evaluated against the reference standard of the Standardization of Uveitis Nomenclature (SUN) clinical grading of vitreous haze[8–13].

The application of an OCT-based, automated algorithm for quantification of vitreous inflammation requires manual selection and a certain amount of manual post-processing steps of scans. For this reason, the number of scans should be limited to the minimum amount required for reliable quantification of VI to facilitate future employment in clinical routine and randomized controlled clinical trials. These applications include a potential use of the OCT-based parameter as a biomarker for therapeutic decisions, follow-up intervals and as a clinical

[1]Department of Ophthalmology, University Hospital Bonn, 53127 Bonn, Germany. [2]Division of Optometry and Visual Science, City, University of London, London, UK. [3]Academic Unit of Ophthalmology, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK. [4]Department of Ophthalmology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [5]NIHR Biomedical Research Centre At Moorfields Eye Hospital and UCL Institute of Ophthalmology, London, UK. ✉email: Jan.Terheyden@ukbonn.de; Robert.Finger@ukbonn.de
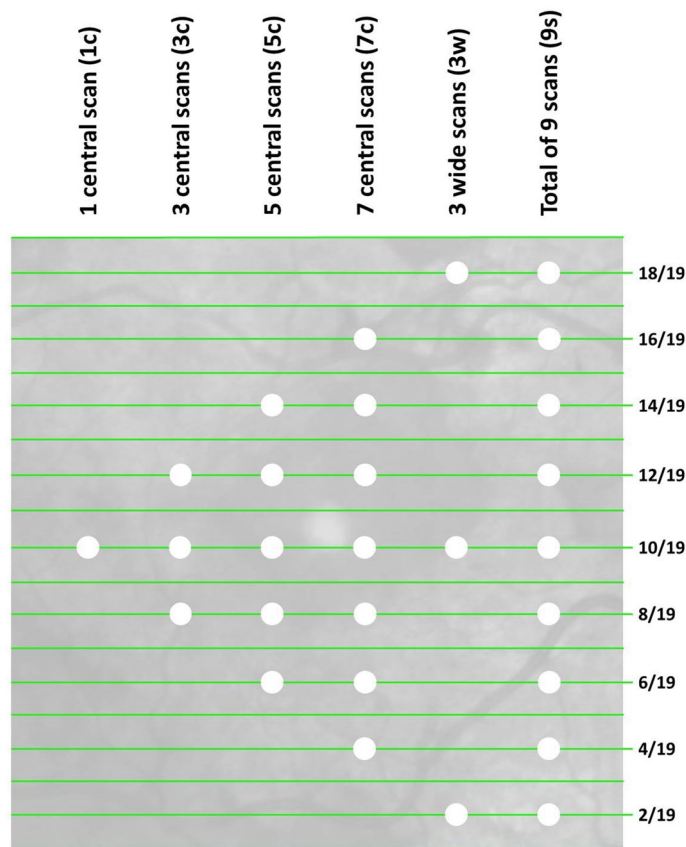
Figure 1. Illustration of an infrared image linked to an OCT volume scan that consists of 19 B-scans. The white dots indicate which scans (green lines) have been included in the different sub selections of B-scans (columns).

Scan sub selection	Median VI difference from reference standard (interquartile range)	Limits of agreement compared to reference standard
1 central scan (1c)	0.006 (0.009)	[−0.039;0.037]
3 central scans (3c)	0.005 (0.011)	[−0.033;0.032]
5 central scans (5c)	0.004 (0.009)	[−0.028;0.026]
7 central scans (7c)	0.001 (0.004)	[−0.009;0.009]
3 wide scans (3w)	0.003 (0.005)	[−0.014;0.013]

Table 1. Deviation between the reference standard (VI means of 9 B-scans) and average VI values from the sub selections of B-scans and respective limits of agreement. VI, vitreous intensity.

trial endpoint. Thus, we assessed the minimum required number of B-scans to reliably quantify vitreous inflammation in this study.

Results

Current image data of 49 eyes of 49 patients (31 females, 18 males) examined at a tertiary referral centre were included. Uveitis was classified as intermediate in 8 eyes, posterior in 33 eyes and panuveitis in 8 eyes. Mean age at examination was 70 ± 12 years; mean logMAR BCVA at examination was 0.5 ± 0.3 and 44 eyes were pseudophakic.

The mean distance between two B-scans was 243 ± 8 μm (Fig. 1, individual B-scans are represented by green lines). Across all B-scans, the median VI was 0.029 (interquartile range: 0.032), ranging from 0.0026 to 0.394. The mean VI per eye did not differ significantly between phakic and pseudophakic eyes ($P = 0.919$). The intra-class correlation coefficient of the VI values across the 9 B-scans was 0.923 (95% confidence interval 0.886 – 0.952), indicating high agreement between VI values.

Smaller median differences indicate less variation from the chosen reference standard (i.e. the mean VI value from 9 B-scans, Fig. 1). Table 1 shows that the median differences between the reference standard and the average values of scan selections (1 central scan, 1c; 3 central scans, 3c; 5 central scans, 5c; 7 central scans, 7c and 3 widely distributed scans, 3w) were noticeably different. For instance, the difference from the total VI average in

200

Scan sub selection	AUC [95% CI]
1 central scan (1c)	0.862 [0.753; 0.971]
3 central scans (3c)	0.870 [0.767; 0.974]
5 central scans (5c)	0.929 [0.855; 1.0]
7 central scans (7c)	0.985 [0.954; 1.0]
3 wide scans (3w)	0.965 [0.912:.1.0]

Table 2. Area under the curve values for the detection of values larger than the dataset's median, per scan sub selection. AUC, area under the curve; CI, confidence interval.
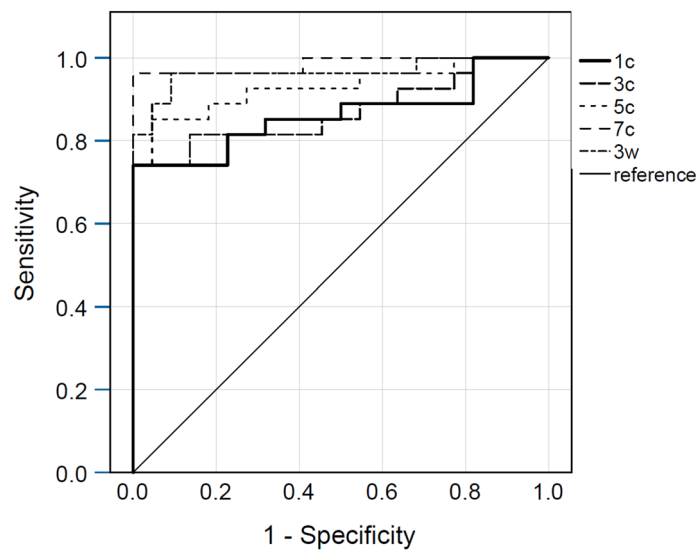
Figure 2. Receiver operating curves of the different scan sub selections. The state variable was the dataset's median. 1c, 1 central scan; 3c, 3 central scans; 5c, 5 central scans; 7c, 7 central scans; 3w, 3 wide scans.

9 B-scans was smaller in the sub selection 3w compared to 5c (Holm-Bonferroni adjusted $P = 0.022$). It was also smaller in the sub selection 7c compared to 3w (Holm-Bonferroni adjusted $P = 0.003$).

Linear regression analysis revealed no significant associations between the two axes on Bland–Altman plots with each other, comparing the reference standard with the VI scores from individual scan selections (comparators: 1c, p = 0.907; 3c, p = 0.120; 5c, p = 0.172; 7c, p = 0.604; 3w, p = 0.243).

All area under the curve (AUC) values from ROC analysis were > 0.8 (Table 2), indicating high sensitivity and specificity for the detection of values larger than the dataset's median, i.e. detection of eyes with statistically "higher inflammation" in contrast to eyes with "lower inflammation". The scan sub selections 7c and 3w achieved the highest sensitivity and specificity values (Fig. 2). We achieved similar results using a comparison with the dataset's upper and lower quartile as state variables (data not shown).

Discussion

The results indicate that Vitreous/RPE-relative intensity is consistent across OCT scans in individuals with uveitis. Averaging the VI of several B-scans within one OCT volume scan further improved measurement reliability because it reduces the influence of local structural alterations. VI calculation from three widely distributed B-scans (average distance 1944 µm) achieves comparable results with VI calculation from nine equally distributed B-Scans (average distance 486 µm) and allows for sufficient discrimination of different levels of inflammation.

Our results indicate that less dense scan patterns compare well to more dense scans in the quantification of VI. As the biomarker itself as well as the measurement of vitreous inflammation using OCT are relatively novel, no data comparing different scan densities for this purpose are available so far. However, different OCT scan patterns of the macula have been compared in the literature with respect to detection of retinal biomarkers such as the presence of intraretinal fluid and quantification of retinal layer thicknesses. Sayanagi and colleagues did not find a significant difference in retinal thickness measurements of patients with macular diseases between a dense OCT volume scan and a radial scan consisting of 6 lines[14]. Other groups confirmed that retinal thickness can be measured almost as reliably with low B-scan density scans compared to high density scans in individuals with defined macular diseases such as diabetic macular oedema[15,16], age-related macular degeneration[17] and retinal vein occlusion[18]. Different studies showed the detection of fluid to be almost as sensitive in scan patterns

only 25–50% as dense as the respective reference standard in age-related macular degeneration, diabetic macular oedema, retinal vein occlusion and other retinal diseases[17–23].

Our main result that a pattern with a smaller number of OCT B-scans-based measurements is similarly sensitive as a denser reference standard is thus consistent with findings reported in the literature. The scan sub selection "3w" including a central B-scan and two peripheral B-scans was superior to a single central scan, 3 central scans and 5 central scans. Reliability can be increased with 7 B-scans, i.e. the difference from the reference standard was significantly lower, but the relevance of this small decrease in mean difference is unclear and needs to be considered against the increased workload.

Of note is that the minimum scan density that can be recommended for the measurement of VI is lower than the one recommended for use in retinal diseases to assess retinal thickness or presence of macular oedema reported in the literature. The minimum number of B-scans required for these purposes varies between five and 32 scans[17–23]. The diffuse nature of the signal in the vitreous cavity in inflammatory diseases in contrast to clearly locatable pathologies in only a small part of the retina in retinal diseases might be a potential explanation for this. However, the impact of local heterogeneity in vitreous haze or accumulation of inflammatory cells (e.g. snowballs) on the OCT-based parameter and its changes with eye movement require further investigation.

The strengths of our study include a standardized imaging protocol for all participants, a relatively homogenous sample consisting of only uveitis patients (excluding anterior uveitis) and the use of a previously developed algorithm which is already clinically validated. Limitations include the relatively small sample size with only a limited number of B-scans per subject available, the use of only one device (Spectralis, Heidelberg Engineering), the relatively high age of the participants for a uveitis population, and the limited availability of clinical data. In this study we have considered measurements taken at a single visit, and therefore have not considered the stability of the signal over time.

Overall, our study shows that automated determination of VI is reliable across OCT B-scans in uveitis patients. The recommended minimum B-scan density for future research based on this parameter is three horizontal scans: One central scan and two peripheral scans located approximately 2000 μm inferiorly and superiorly from the central B-scan (3w). Measurements were even more stable across scans in a pattern of 7 horizontal scans (7c) but we interpret this improvement as not clinically relevant compared to the recommended pattern. In the future, further correlation of the data with clinical vitreous haze scores and other clinical variables as well as further reliability analysis based on these values is warranted.

Methods

The retrospective study took place at the department of ophthalmology of the University of Bonn, Germany. The institutional Ethics Committee (University Hospital Bonn, Germany) approved the use of retrospective data for study purposes and approved that informed consent can be waived due to the use of retrospective data only (no. 103/18). The study adhered to the principles of the declaration of Helsinki. Participants were included if they had a form of Posterior Segment Involving Uveitis (PSIU) i.e. one of Intermediate, posterior or panuveitis as classified according to the SUN criteria[3].

Image data. Macular OCT volume scans were retrospectively collected. OCT is a light-based, non-invasive technique frequently applied in ophthalmology. It is based on local interference between two signals (object signal and reference signal)[24,25]. Using software, B-scans (e.g. axial) are automatically calculated from A-scans. Retinal OCT B- scans show parts of the posterior vitreous cavity, the retinal layers as well as choroidal structures. The volume scans were obtained with the Spectralis SD-OCT (Heidelberg Engineering, Heidelberg, Germany), with a B-scan image resolution of 512×496 pixels and 5 images averaged (automated real-time tracking mode = 5). Inclusion criteria were volume scans consisting of 19 B-scans each and a B-scan size of $20° \times 15°$. Exclusion criteria were insufficient image quality (HEYEX software image quality score < 20 in > 3 B-scans), incomplete scan, fixation errors and a presumed disease aetiology other than uveitis. Besides image data, age, uveitis classification, best-corrected visual acuity (BCVA) and lens status of all included patients were collected.

Image analysis. Every other B-scan was selected from the OCT volume scans, resulting in 9 B-scans per volume scan available for analysis (Fig. 1). As one of the previous VI algorithm validation studies included a reference of 7 B-scans per volume, we used a comparable density as our gold standard[11]. The image data and additional image acquisition parameters were imported into MATLAB, Version R2016a (The MathWorks, Natick, Massachusetts, USA). The VI parameter Vitreous/Retinal Pigment Epithelium (RPE)-relative intensity was automatically calculated per B-scan according to an algorithm that has previously been described and clinically validated[8–12,26]. In summary, pre-processing steps include opening, thresholding and adjustment as outlined by Keane et al.[9]. The posterior part of the vitreous cavity is automatically detected and the OCT sum signal in this area is quantified relative to the RPE signal intensity in order to lower the impact of media opacities on the outcome parameter. The overall vitreous reflectivity is increased in inflammation which has been explained e.g. by inflammatory components and proteins in the vitreous cavity[9,26]. B-scan quality was assessed for all selected B-scans and the distance between B-scans was obtained per individual volume scan.

Statistical analyses. The intra-class correlation coefficient between all VI values per volume scan was calculated. The single VI value of the central B-scan (1c) and averaged VI values of five combinations of B-scans (3, 5 and 7 central scans (3c, 5c, 7c), all 9 scans available for analysis (9 s), 3 widely distributed scans (3w); Fig. 1) were computed for all volume scans included. The averaged VI value of 9 B-scans was used as the standard reference. Mean absolute differences between this reference and a single central B-scan VI as well as the averaged VI values listed above (3c, 5c, 7c, 3w) were calculated. 95% limits of agreement (LoA) were calculated according to

the formula LoA = mean ± 1.96 × standard deviation of the differences between the two measurements. Linear regression analysis was performed based on Bland–Altman plots to identify associations between the above mentioned mean absolute differences (e.g. 9 s-1c, 9 s-3c, 9 s-5c, etc.) and their respective means, excluding four cases that were likely outside of the sensitivity range of our study (mean VI score > 0.1). In addition, we performed receiver operating curve characteristic (ROC) analysis for discrimination of VI values greater or equal and VI values smaller than the median VI value out of all B-scans.

Statistical analyses were performed with SPSS Statistics, version 25 (IBM Corporation, Armonk, New York, USA) and R, version 3.5.0 (R Core Team, Vienna, Austria). Paired samples were compared with the Wilcoxon rank sum test correcting for multiple comparisons using the Holm-Bonferroni method[27]. The level of statistical significance was $P < 0.05$.

References

1. Miserocchi, E., Fogliato, G., Modorati, G. & Bandello, F. Review on the worldwide epidemiology of uveitis. *Eur. J. Ophthalmol.* **23**, 705–717 (2013).
2. Tsirouki, T. *et al.* A focus on the epidemiology of uveitis. *Ocul. Immunol. Inflamm.* **26**, 2–16 (2018).
3. Standardization of Uveitis Nomenclature (SUN) Working Group. Standardization of uveitis nomenclature for reporting clinical data Results of the First International Workshop. *Am. J. Ophthalmol.* **140**, 509–516 (2005).
4. Kempen, J. H., Ganesh, S. K., Sangwan, V. S. & Rathinam, S. R. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. *Am. J. Ophthalmol.* **146**, 813-818.e1 (2008).
5. Davis, J. L. *et al.* Scale for photographic grading of vitreous haze in uveitis. *Am. J. Ophthalmol.* **150**, 637-641.e1 (2010).
6. Tugal-Tutkun, I. & Herbort, C. P. Laser flare photometry: a noninvasive, objective, and quantitative method to measure intraocular inflammation. *Int. Ophthalmol.* **30**, 453–464 (2010).
7. Herbort, C. P., Guex-Crosier, Y., de Ancos, E. & Pittet, N. Use of laser flare photometry to assess and monitor inflammation in uveitis. *Ophthalmology* **104**, 64–71 (1997) (**discussion 71-2**).
8. Keane, P. A. *et al.* Objective measurement of vitreous inflammation using optical coherence tomography. *Ophthalmology* **121**, 1706–1714 (2014).
9. Keane, P. A. *et al.* Automated analysis of vitreous inflammation using spectral-domain optical coherence tomography. *Transl. Vis. Sci. Technol.* **4**, 4 (2015).
10. Sreekantam, S. *et al.* Quantitative analysis of vitreous inflammation using optical coherence tomography in patients receiving sub-Tenon's triamcinolone acetonide for uveitic cystoid macular oedema. *Br. J. Ophthalmol.* **101**, 175–179 (2017).
11. Montesano, G. *et al.* Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis. *Sci. Rep.* **8**, 1–7 (2018).
12. Zarranz-Ventura, J. *et al.* Evaluation of objective vitritis grading method using optical coherence tomography: influence of phakic status and previous vitrectomy. *Am. J. Ophthalmol.* **161**, 172-80.e1–4 (2016).
13. Coric, D. *et al.* Objective quantification of vitreous haze on optical coherence tomography scans: no evidence for relationship between uveitis and inflammation in multiple sclerosis. *Eur. J. Neurol.* **27**, 144-e3 (2020).
14. Sayanagi, K., Sharma, S. & Kaiser, P. K. Comparison of retinal thickness measurements between three-dimensional and radial scans on spectral-domain optical coherence tomography. *Am. J. Ophthalmol.* **148**, 431–438 (2009).
15. Nittala, M. G., Konduru, R., Ruiz-Garcia, H. & Sadda, S. R. Effect of OCT volume scan density on thickness measurements in diabetic macular edema. *Eye (London)* **25**, 1347–1355 (2011).
16. Taban, M., Sharma, S., Williams, D. R., Waheed, N. & Kaiser, P. K. Comparing retinal thickness measurements using automated fast macular thickness map versus six-radial line scans with manual measurements. *Ophthalmology* **116**, 964–970 (2009).
17. Velaga, S. B. *et al.* Impact of optical coherence tomography scanning density on quantitative analyses in neovascular age-related macular degeneration. *Eye (London)* **31**, 53–61 (2017).
18. Rahimy, E., Rayess, N., Maguire, J. I. & Hsu, J. Radial versus raster spectral-domain optical coherence tomography scan patterns for detection of macular pathology. *Am. J. Ophthalmol.* **158**, 345-353.e2 (2014).
19. Fang, P. P. *et al.* Minimal optical coherence tomography B-scan density for reliable detection of intraretinal and subretinal fluid in macular diseases. *Retina* **39**, 150–156 (2019).
20. Baranano, A. E., Keane, P. A., Ruiz-Garcia, H., Walsh, A. C. & Sadda, S. R. Impact of scanning density on spectral domain optical coherence tomography assessments in neovascular age-related macular degeneration. *Acta Ophthalmol.* **90**, e274–e280 (2012).
21. Sadda, S. R., Keane, P. A., Ouyang, Y., Updike, J. F. & Walsh, A. C. Impact of scanning density on measurements from spectral domain optical coherence tomography. *Investig. Ophthalmol. Vis. Sci.* **51**, 1071–1078 (2010).
22. de Niro, J. E. *et al.* Sensitivity of fluid detection in patients with neovascular and using spectral domain optical coherence tomography high-definition line scans. *Retina* **34**, 1163–1166 (2014).
23. Adam, M. K., Rayess, N., Rahimy, E., Maguire, J. I. & Hsu, J. Radial versus raster spectral-domain optical coherence tomography scan patterns for detection of macular fluid in neovascular age-related macular degeneration. *Br. J. Ophthalmol.* **100**, 491–494 (2016).
24. Podoleanu, A. G. Optical coherence tomography. *J. Microsc.* **247**, 209–219 (2012).
25. Huang, D. *et al.* Optical coherence tomography. *Science (New York, N.Y.)* **254**, 1178–1181 (1991).
26. Denniston, A. K., Keane, P. A. & Srivastava, S. K. Biomarkers and surrogate endpoints in uveitis: the impact of quantitative imaging. *Investig. Ophthalmol. Vis. Sci.* **58**, BIO131–BIO140 (2017).
27. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**, 65–70 (1979).

Author contributions

J.H.T. and R.P.F. designed the study, G.O., G.M., X.L., P.A.K. and A.K.D. created the software used in this work, G.O., J.H.T. and M.L. acquired and analysed the data. J.H.T., M.W.M.W. and R.P.F. interpreted the data, J.H.T. and R.P.F. drafted the manuscript, G.O., G.M., M.W.M.W., M.L., X.L., P.A.K., D.P.C. and A.K.D. substantively revised the manuscript. All authors have approved the submitted version and have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of

any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Competing interests

The Department of Ophthalmology, University of Bonn received imaging devices from Heidelberg Engineering, Optos, Carl Zeiss Meditec, and CenterVue. JHT: No additional conflicts of interest. GO: None. GM: None. MWMW: Travel grant/reimbursement from DigiSight Technologies, Heine Optotechnik GmbH, ASKIN & CO GmbH and Berlin-Chemie AG, grants from Heine Optotechnik and Berlin-Chemie AG, consultant for Heine Optotechnik GmbH, imaging devices from DigiSight Technologies, D-EYE Srl and Heine Optotechnik GmbH, free analysis from Eyenuk Inc. ML: No additional conflicts of interest. XL: None. PAK: Grants from Allergan, Bayer, Carl Zeiss Meditec, Haag-Streit, Heidelberg Engineering, Novartis, Topcon, consultant for Deepmind. AKD: None. DPC: Speaker fees from Allergan, Bayer, Santen; unrestricted funding from Allergan, Santen; consultancy with Centervue—all outside the remit of the submitted work. RPF: Consultant for Bayer, Novartis, Santen, Opthea, Novelion, Santhera, Inositec, Alimera and RetinaImplant; honoraria from Bayer, Ellex, Alimera.

Additional information

Correspondence and requests for materials should be addressed to J.H.T. or R.P.F.

Reprints and permissions information is available at www.nature.com/reprints.

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

Article

Merging Information From Infrared and Autofluorescence Fundus Images for Monitoring of Chorioretinal Atrophic Lesions

Giovanni Ometto[1,2], Giovanni Montesano[1,2], Saman Sadeghi Afgeh[3], Georgios Lazaridis[2,4], Xiaoxuan Liu[5−7], Pearse A. Keane[2,7,8], David P. Crabb[1], and Alastair K. Denniston[5−8]

[1] Division of Optometry and Visual Sciences, School of Health Sciences, City, University of London, London, UK
[2] Moorfields Eye Hospital NHS Foundation Trust, London, UK
[3] Data Science Institute, City, University of London, London, UK
[4] Centre for Medical Image Computing, University College London, London, UK
[5] Department of Ophthalmology, Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[6] Academic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, UK
[7] Health Data Research UK, London, UK
[8] NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, UK

Purpose: To develop a method for automated detection and progression analysis of chorioretinal atrophic lesions using the combined information of standard infrared (IR) and autofluorescence (AF) fundus images.

Methods: Eighteen eyes (from 16 subjects) with punctate inner choroidopathy were analyzed. Macular IR and blue AF images were acquired in all eyes with a Spectralis HRA+OCT device (Heidelberg Engineering, Heidelberg, Germany). Two clinical experts manually segmented chorioretinal lesions on the AF image. AF images were aligned to the corresponding IR. Two random forest models were trained to classify pixels of lesions, one based on the AF image only, the other based on the aligned IR-AF. The models were validated using a leave-one-out cross-validation and were tested against the manual segmentation to compare their performance. A time series from one eye was identified and used to evaluate the method based on the IR-AF in a case study.

Results: The method based on the AF images correctly classified 95% of the pixels (i.e., in vs. out of the lesion) with a Dice's coefficient of 0.80. The method based on the combined IR-AF correctly classified 96% of the pixels with a Dice's coefficient of 0.84.

Conclusions: The automated segmentation of chorioretinal lesions using IR and AF shows closer alignment to manual segmentation than the same method based on AF only. Merging information from multimodal images improves the automatic and objective segmentation of chorioretinal lesions even when based on a small dataset.

Translational Relevance: Merged information from multimodal images improves segmentation performance of chorioretinal lesions.

Introduction

Punctuate inner choroidopathy (PIC) is a rare condition that was first recognized by Watzke in 1984 as a group of the White Dot Syndromes.[1] The disease is an inflammatory choroiditis that does not affect the anterior chamber or vitreous cavity. It generally affects eyes of young and myopic women and its cause is unknown.[2] It is characterized by the

appearance of multifocal, well-circumscribed, small lesions that resolve in a few weeks, leaving atrophic spots with variable pigmentation. These episodes are symptomatic, with patients reporting blurred central vision, flashes of light, and paracentral scotomas. Symptoms can disappear with the resolution of the lesion, but about 40% of the patients experience more severe visual loss with the development of choroidal neovascularization (CNV).[3–5]

The detection and monitoring of PIC are assisted by a number of imaging techniques including optical coherence tomography (OCT) and fundus autofluorescence (AF).[6,7] Whereas OCT provides good three-dimensional views of the evolution of individual inflammatory lesions (and detection of CNV), the AF provides the best overview of the number of atrophic PIC lesions, their size and the total area of the macula that has been affected. Specialists therefore routinely use the evaluation of hypoautofluorescent areas on AF images to estimate disease progression over time and to assess efficacy of treatment; additionally hyperautofluorescent areas may indicate new disease activity although this is a less consistent phenomenon.[8] Visual assessment is, however, significantly hampered by its subjective nature (based on direct visual comparison of scans between visits) and is not supported by any numerical information that could be used to provide objective indices to support treatment decisions or progression monitoring. Estimates of lesions in AF can be improved by semiautomatic segmentation, such as the one provided by the Region Finder software (Heidelberg Engineering, Heidelberg, Germany), a region-based segmentation that often requires manual correction.[9] There is also considerable interest around the automated segmentation of geographic atrophic lesions in patients with age-related macular degeneration.[10–19] However, to the best of our knowledge, there are no published algorithms for the automated segmentation of chorioretinal lesions in PIC and similar uveitic syndromes, and clinicians are therefore currently dependent on subjective, qualitative assessment to detect change between visits and inform treatment decisions. This is likely due to the scarcity of large datasets, rarity of the disease and the morphological complexity of these lesions that make the training of such algorithms a challenging task. Such complexity would also translate into a taxing endeavor for clinicians, required to manually correct segmentations of multiple sparse lesions.

Chorioretinal atrophic lesions are also visible in images acquired in infrared (IR) and color modalities. In these images, lesions appear as sharply demarcated regions of absent retinal pigment epithelium through which the choroid or sclera is visible. Although their appearance in these images might not be as contrasted as in AF, they provide complementary information that can improve the performance of algorithms for automatic segmentation.

In this work, we present a proof of concept for a machine learning algorithm, which combines the information of IR and AF images to produce an automatic segmentation of PIC atrophic lesions. We demonstrate that it is feasible to develop and test these methods using a small dataset. We compare the results of the proposed method with those from another algorithm, based on the same model but trained on the AF only. Finally, we present a case study to explore the potential benefits of the technique for the monitoring of progression.

Methods

Dataset

All patients attending the specialist PIC clinic at University Hospitals Birmingham NHS Foundation Trust, UK, have a standardized set of scans (the "Birmingham PIC Protocol") which comprises: 30° OCT of macula ("fast macula") and retinal nerve fiber layer (RNFL); 30° Bluepeak AutoFluorescence (AF); 30° multicolor (three laser reflectance), 55° wide-field posterior pole OCT, AF, and Multicolor, Ultrawide field Optos Color, and AF. This protocol was approved from the NRES East Midlands Ethics Committee (Ref: 14/EM/1163). Written informed consent was gathered from all subjects. This protocol adhered to the tenets of the declaration of Helsinki. The scans of 16 patients (with 18 affected eyes) with PIC had at least one visit with all modalities acquired and were identified for this study. Each of the 18 eyes had a 6 mm × 6 mm macular OCT volume with an associated 768 × 768-pixels IR image and a 768 × 768-pixels AF image. All volumes and AF images were acquired using a Spectralis HRA+OCT device (Heidelberg Engineering, Heidelberg, Germany) using 820 nm and 488 nm wavelengths for IR and AF, respectively. All patients enrolled in this study had "classic" PIC with predominantly central lesions; patients who had multifocal choroiditis without the central lesions or who had progressive subretinal fibrosis were not included in this study. All stages of PIC were eligible for inclusion.

Manual Segmentation

A clinical expert (GM) manually segmented the pixels of chorioretinal lesions on the AF image based on the appearance of the AF image supported by the

INFRARED ALIGNED AF AF AND IR MANUAL SEGMENTATION
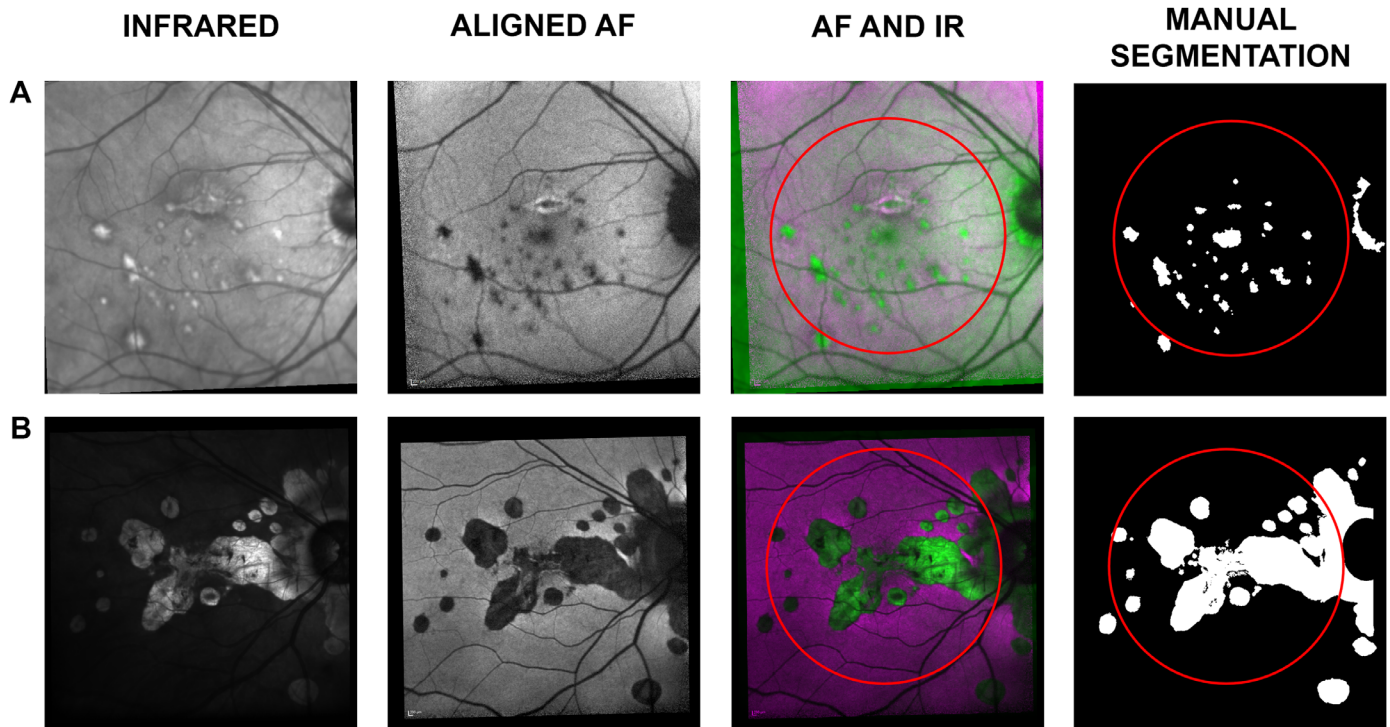
Figure 1. Rows A and B illustrate data from two of the 18 selected eyes. The first column shows the IR image, acquired with the macular OCT scan. The second column shows the aligned AF image. The third column shows a combination of the IR and AF, where the *magenta* represents intensities in the AF higher than in the IR and vice versa in *green*. The fourth column shows the manual segmentation as a binary map of "0" (non-lesion, in *black*) and "1" (lesion, in *white*). The *red circles* in the third and fourth columns show the central 22.5°, the area used for training and testing the automatic classification.

IR and OCT. The segmentation was then reviewed by a second clinical expert (XL). The task was carried out using the ImageSegmenter app available in Matlab R2019a (Mathworks, Natick, MA, USA) with the Image Processing Toolbox. The segmentation produced eighteen 768×768 binary maps, classifying each pixel of the IR image as "0" (nonlesion) or "1" (lesion) (Fig. 1). Areas of peripapillary atrophy were also classified as lesions. This choice was necessary to avoid ambiguities in the segmentation of lesions that merged with these areas (see Fig. 1B).

AF-IR Image Alignment

Matlab image registration function *imregconfig* was used to automatically align normalized AF images to the relative, normalized IR using a 100-step optimization process. The function was set to align mono-modal images, as the information captured by these two modalities was largely non-complementary, for the exception of areas with lesions. The clinical grader (GM) visually inspected the results of the alignment and performed manual matching where the automated algorithm failed. Manual alignment was obtained through the localization of four landmarks on the pair of images. The landmarks were used to calculate the parameters of the local weighted mean transformation, which was then used to align the AF. Control points and transformation parameters were obtained using the Image Processing Toolbox functions *cpselect* and *fitgeotrans*, respectively.

Automatic Segmentation Methods

A machine learning classifier (random forest with 25 trees) was trained to categorize pixels of the images into two classes, "0" (non-lesion) or "1" (lesion), using the manual segmentation as the reference.[20]

Each observation was identified by a pixel location of the IR and aligned AF and was characterized by eight attributes. The latter were obtained with an adaptive histogram equalization of both IR and AF images operated on 4 differently sized, neighboring regions: 15×15, 31×31, 151×151 and 301×301 pixels. This process, equivalent to a pre-processing stage of local-intensity normalization, generated eight equalized images, four from IR and four from AF. Therefore, at each pixel location, these images provided

207

an attribute that incorporated information of neighboring intensities.

Only pixels within the central 22.5° (3/4 of the field of view of the lens) were used in the validation of the model (red circles in Fig. 1). This restriction was introduced to exclude peripheral areas, often noisy due to scarce illumination, and to guarantee that the included macular area had been captured by both IR and AF despite acquisition misalignments.

Finally, a proportion (10%) of randomly selected observations from the "lesion" class was selected and matched by the same number of randomly selected "non-lesions." This produced balanced training-datasets while allowing faster training.

The same random forest used for the classification of AF-IR was trained using the four attributes from the AF only, obtaining a new classification model.

Validation

The random forest classifiers were evaluated using a leave-one-out cross-validation: the model was trained on observations from 17 of the 18 eyes, and was validated on the eye 'left-out', repeating the process for each eye.[21] Results were analyzed to report the percentage of correct classifications, sensitivity and specificity of the model, as well as Dice's coefficient of similarity with the segmentation by the grader.[22]

Case Study

A time series of IR and associated AF images acquired over 42 months from 35 consecutive visits was identified in the database and used for a case study. The 34 IR from follow-up visits were aligned to the IR of the first visit. Then, all AF images were aligned to their associated IR. Resulting pairs of IR-AF were classified by the trained random forest classifier. The generated classification-maps were processed to fill the holes in the segmentation; to remove spurious classifications of individual pixels as lesions with a morphological opening and closing; and to force pixels classified as lesion at a time point to retain the classification for the rest of the time series. Using the pixel-mm^2 conversion provided by the manufacturer (1 pixel = 0.01118 mm^2), the total segmented area at each visit was converted to millimeters squared. The first derivative of the total area was calculated to estimate the expansion speed in mm^2/days.

Results

The 16 patients selected for the study were all female with a mean (range) age of 41 (31, 62) years. All but one patient had both eyes affected by the condition and twelve patients were on systemic immunosuppression.

Automatic alignment was successful in nine out of the 18 AF-IR pairs, with the others requiring manual intervention. Failed alignments were associated with the presence of large lesions or large areas with low illumination in at least one of the two photographs (IR or AF).

The AF-IR model correctly classified 95.9% of the pixels in the dataset with sensitivity and specificity of 0.83 and 0.98 respectively. Dice's coefficient was 0.85, showing a good similarity between the automatic and manual segmentations. The AF model correctly classified 94.6% of the pixels in the dataset with sensitivity and specificity of 0.79 and 0.97, respectively; Dice's coefficient was 0.80. For reference, we trained the same model on IR only. The IR model correctly classified 90.0% of the pixels; sensitivity and specificity were 0.40 and 0.98 respectively; Dice's coefficient was 0.53 showing poorer correlation with the reference segmentation than the other models. Figure 2 shows the results of the automatic segmentation for a randomly selected subset of the dataset. Segmentation results of the whole dataset are available in the Supplementary Figure.

Discussion

This work introduces a novel method for the segmentation of atrophic chorioretinal lesion. We demonstrate how this method could be feasibly used to provide clinicians with real-time objective metrics such as lesion area and growth rate.

PIC was used as a case-example of the wider group of chorioretinal inflammatory diseases (posterior uveitis) because there is a clear "use case" here in that the presence or absence of lesions, and their change over time, directly impacts on treatment decisions.

Our approach to chorioretinal lesion segmentation used a combination of two standard scanning laser ophthalmoscopy (SLO) imaging techniques—IR and AF—both of which can be routinely acquired from the Heidelberg Spectralis system. Our automated segmentation technique shows strong agreement with manual segmentation by a clinical grader while using only 18 images for the training of the algorithm.
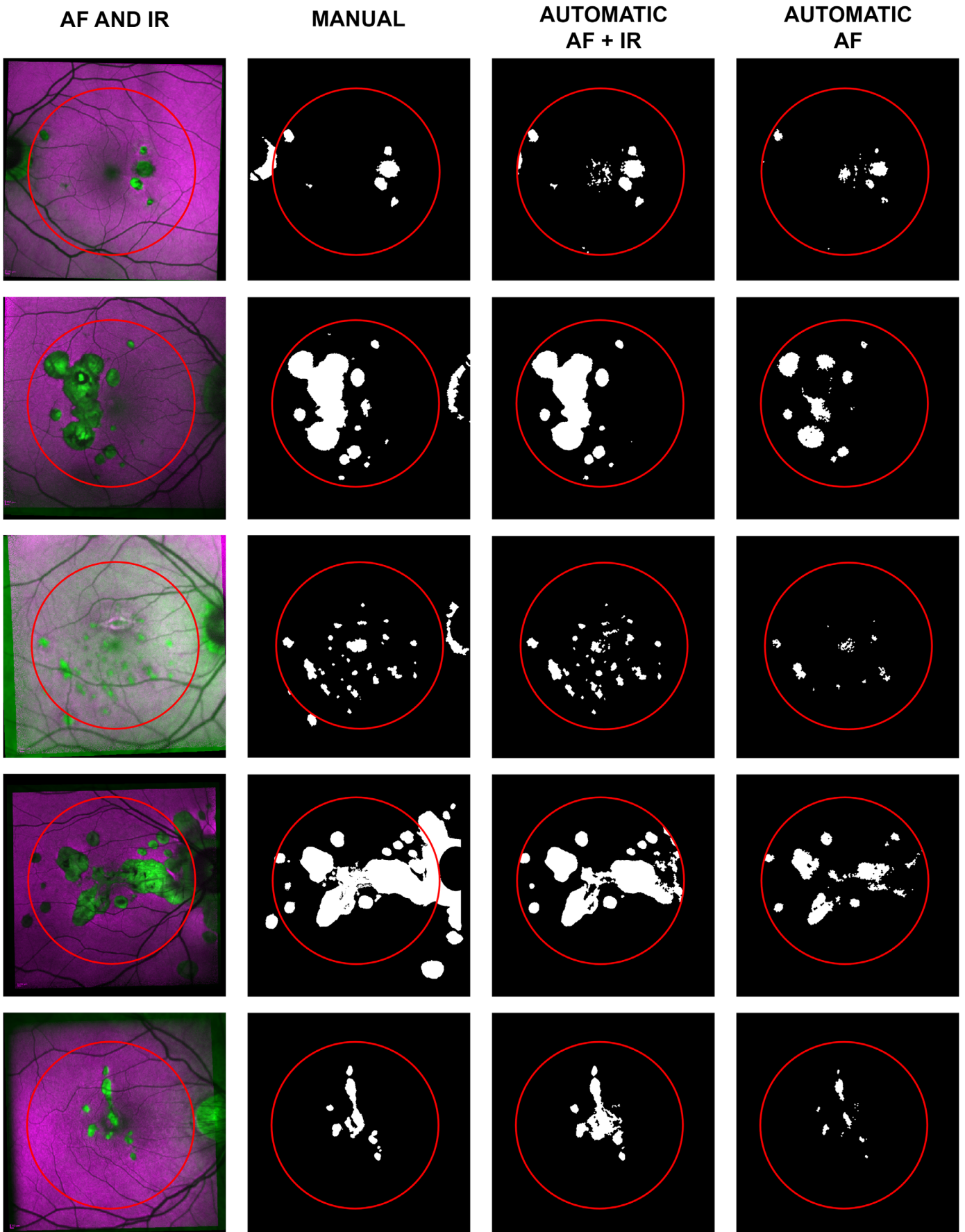
Figure 2. A random subset of five of the 18 selected eyes. The first column shows a combination of the IR and AF. The second column shows the manual segmentation as a binary map of "0" (non-lesion, in *black*) and "1" (lesion, in *white*). The third column shows the automatic segmentation based on IR and AF for the central 22.5°, delimited by a *red circle*. The fourth column shows the results of the same classification model trained on AF only.

The model based on the IR only performed poorly compared to the other two (AF only and IR-AF) and this result is consistent with previous literature.[8,9] This report demonstrates new knowledge because merging the information from multimodal images proved to be effective, outperforming the classification model based on AF only. Percentage of correctly classified pixels and specificity do not highlight major differences in performance due to the much higher number of pixels from non-lesions. However, higher sensitivity and higher Dice's coefficient achieved by the IR-AF based model reflects a significant improvement in the segmentation, also clearly visible inspecting the segmentation results (Fig. 2 and Supplementary Figure). In fact, a challenge in lesion segmentation with traditional single modality techniques, is delineating between a pathological feature and a normal structure such as fovea, optic disc, and retinal vessels that could be wrongly classified as lesions. This problem requires particular attention and can be time-consuming when the task is performed semiautomatically with the aid of the Heidelberg Engineering Region Finder software.[23] In this task, our proposed method outperforms the same algorithm trained on AF only, in part because the combination of IR and AF reinforces features of normal structures and differentiation of abnormalities: retinal features like the fovea and vessels (all darker in AF images and possibly confused with atrophic lesions) can therefore be ignored, and pathological features can be highlighted. In particular, better identification of lesions in the foveal region is of extreme importance due to their sight threatening implications. The better performance of the model based on multimodal images can be explained in part by its ability to exploit the most informative features provided by each acquisition modality, such as the generally sharper features of IR images and the intensities of lesions represented in AF images.

Although the proposed method still requires some modest manual intervention for the alignment, this allows for the tracking of areas that become atrophic during follow-ups. Thanks to the alignment, newly developed atrophic areas can be directly identified and highlighted for each visit (Fig. 3). This tracking not only increases the level of detail in the monitoring of the condition but can also represent an important quantification tool for outcomes for prospective research studies.

We have previously reported the use of commercial OCT segmentation software (HEYEX; Heidelberg Engineering) to identify new inflammatory PIC lesions on the OCT volume scans, using the heat-map function including the ability to generate heat-maps of change from a baseline scan.[7] We see these techniques as complementary, since they provide information about different aspects of the disease process at different stages in the pathway. The heat-map technique (and indeed direct careful perusal of the volume scans) will identify inflammatory PIC lesions (and PIC-associated CNV) from a very early stage. At these early stages, the lesions may not be detected by our IR-AF technique if they have not caused sufficient disruption to the RPE to be seen as a hypoautofluorescent signal. The IR-AF technique detects these lesions slightly later in their development i.e. once they have caused loss of the RPE. Yet the technique segments the lesions themselves rather than the retinal layers and this represents a significant advantage. In contrast, the heat-map technique is primarily qualitative as it does not provide a direct measurement of individual lesions, total lesion area or change in lesion burden.

Although, to the best of our knowledge, ours is the first work to combine IR and AF images for semantic segmentation of PIC, some have used AF images for semantic segmentation of Geographic Atrophy (GA). These contributions can be roughly divided into three categories. First, region-based approaches, level-set methods and other computer vision techniques;[10,15,16] second, heuristic methods and handcrafted features that are then input to machine learning classifiers;[14,17] third, supervised Deep Learning methods, which do not require handcrafted features, but automatically learn useful features directly from the input data.[18,19] Works in the first category typically achieved lower accuracy than methods using supervised machine learning, because they tend to generalize less well to unseen cases. However, these do not rely on manually segmented labels for model training and can therefore be applied successfully to situations where a large labeled dataset is difficult or impossible to obtain, as for example in the case of rare pathologies. Acquiring training labels is typically an expensive step in
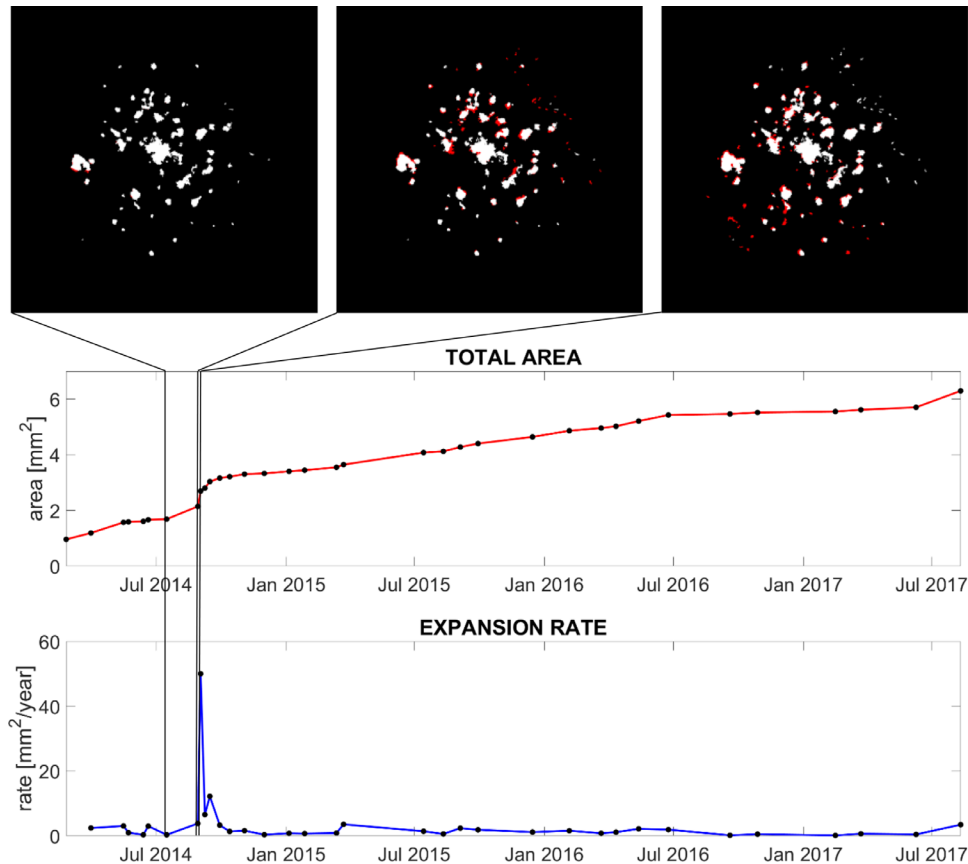
Figure 3. Case study of chorioretinal lesions development in the left eye of a young woman with PIC. The *red line* shows the total area of PIC atrophic lesions measured from the segmentation of the time series. The *blue line* shows the first derivative of the total area, or expansion rate. The three images on top of the plots show the segmentation of three acquisitions taken just before the peak in the expansion rate. *Black* represents nonlesions; *white* represents lesions already segmented in the previous visit; *red* highlights newly segmented lesions.

any machine-learning pipeline, and data scarcity led models in the third category to poor generalization. Models in the second category can represent a compromise between those in the other two, attempting to strike a balance between data efficiency and generalization. These models are well suited in clinical contexts where labelled data is scarce. Our results suggest that the combination of information from different image modalities can generate a new class of handcrafted features, which could help improve the performance of this category of models.

Our methods and the study used to evaluate them have some limitations.

Although the proposed method was able to train on a small dataset, calculated performance metrics can be affected by the small number of images available.

Our estimate for measured sensitivity (0.83) indicates a limitation of our automated segmentation technique. This suggests that our technique was unable to identify all pixels that were classified as abnormal by the manual approach. This may be due

to the variable reflectivity of larger lesions in the IR, which could have confused the classification model, resulting in the underestimation of some atrophic areas.

One limitation of blue AF images is the masking from the macular pigment near the fovea (see for example the misclassification of the foveal lesions in Fig. 2). The recent introduction of green AF (514 nm wavelength) could overcome this issue and provide better segmentation results[24,25] using essentially the same methodology but simply pairing the IR with green (rather than blue) AF.

It should be recognized that the areas provided with our technique are estimates based on the manufacturer-provided conversion factor of 1 pixel = 0.01118 mm^2. This is an average conversion factor, because the actual area that pixels equate to will vary slightly between patients. Although this means that there is some uncertainty in these estimates of area if comparing between patients, these conversion factors will remain constant within the same patient, and therefore the primary

objective of monitoring disease within the same patient will likely be unaffected.

In terms of implementation, the main limitation of the proposed method is the need for aligned images. The automatic alignment could only complete the task successfully on nine of the 18 eyes evaluated, and this is therefore an area that requires further development. Unless a more robust algorithm is available, manual alignment may be required in a significant number of cases. It should be recognized however that the cases selected in this series are likely to have been particularly challenging given the advanced stage of the condition for many of the selected eyes, and that alignment performance across an unselected population would be expected to be better than this. Additionally, although automatic alignment is preferred, manual alignment is not onerous because it is facilitated with the selection of only four control points per image (eight points per alignment) making the process fast enough even for a clinical setting. We suspect that most clinicians would find this a reasonable investment of time in order to gain better objective quantitative metrics of chorioretinal lesions that the technique provides.

Our case study illustrates the clinical application of the technique as applied to a patient in their 30s with PIC over a 3.5-year time period, which we have visualized in both static and dynamic graphical forms (Fig. 3 and Supplementary Video). This example shows how the quantification of the chorioretinal lesions enabled by our technique could bring greater precision to monitoring progression including a sharp increase in the expansion rate, which might have been missed by simple visual inspection. In PIC and other forms of sight-threatening posterior uveitis, treatment decisions depend on the evaluation of these chorioretinal inflammatory lesions.

In addition to its value to routine clinical practice, our approach may provide a sensitive, reliable, objective measure for clinical trials which include patients with PIC or other posterior uveitis syndromes. We think this is particularly noteworthy. Currently, most such trials "lump together" all forms of "posterior segment involving uveitis" (PSIU). One way of dealing with the wide variation in which these forms of uveitis may demonstrate disease activity is to include "new or active chorioretinal lesion" as part of a composite endpoint of "active disease" or "treatment failure." This therefore reduces a complex disease process (chorioretinitis) to a binary variable based on a subjective evaluation. The technique we have described here would provide objectivity and enable a more nuanced approach to evaluating impact of any intervention in these conditions.

No test, whether diagnostic or monitoring, should be considered in isolation, but rather within the context of the care pathway it supports. Future work should include the evaluation of this method within its testing pathway, with consideration of the actual consequences of the provision of this test data to clinicians, the treatment decisions made and the short and long-term consequences of those decisions.

Automated segmentation of chorioretinal lesions using multimodal images shows closer alignment to traditional manual segmentation than segmentation based on AF only, as indicated by a high Dice's coefficient. The proposed technique provides an automatic and objective segmentation of chorioretinal lesions that could offer much-needed quantitative measurements in clinical practice as demonstrated by its performance in lesion detection, automated area estimation and progression tracking in the sight-threatening posterior uveitis syndrome, PIC.

Acknowledgments

References

1. Watzke RC, Packer AJ, Folk JC, Benson WE, Burgess D, Ober RR. Punctate inner choroidopathy. *Am J Ophthalmol*. 1984;98:572–584.
2. Amer R, Lois N. Punctate inner choroidopathy. *Surv Ophthalmol*. 2011;56:36–53.
3. Essex RW, Wong J, Fraser-Bell S, et al. Punctate inner choroidopathy: clinical features and outcomes. *Arch Ophthalmol*. 2010;128:982–987.
4. Zarranz-Ventura J, Sim DA, Keane PA, et al. Characterization of punctate inner choroidopathy using enhanced depth imaging optical coherence tomography. *Ophthalmology*. 2014;121:1790–1797.

5. Sa-Cardoso M, Dias-Santos A, Nogueira N, Nascimento H, Belfort-Mattos R. Punctate Inner Choroidopathy. *Case Rep Ophthalmol Med*. 2015;2015:371817.

6. Ahnood D, Madhusudhan S, Tsaloumas MD, Waheed NK, Keane PA, Denniston AK. Punctate inner choroidopathy: a review. *Surv Ophthalmol*. 2017;62:113–126.

7. Madhusudhan S, Keane PA, Denniston AK. Adjunctive use of systematic retinal thickness map analysis to monitor disease activity in punctate inner choroidopathy. *J Ophthalmic Inflamm Infect*. 2016;6:9.

8. Yung M, Klufas MA, Sarraf D. Clinical applications of fundus autofluorescence in retinal disease. *Int J Retina Vitreous*. 2016;2:12.

9. Hua R, Liu L, Chen L. Evaluation of the progression rate of atrophy lesions in punctate inner choroidopathy (PIC) based on autofluorescence analysis. *Photodiagnosis Photodyn Ther*. 2014;11:565–569.

10. Hu Z, Medioni GG, Hernandez M, Hariri A, Wu X, Sadda SR. Segmentation of the geographic atrophy in spectral-domain optical coherence tomography and fundus autofluorescence images. *Invest Ophthalmol Vis Sci*. 2013;54:8375–8383.

11. Liefers B, Colijn JM, González-Gonzalo C, et al. A deep learning model for segmentation of geographic atrophy to study its long-term natural history. *Ophthalmology*. 2020;127:1086–1096.

12. Yehoshua Z, Rosenfeld PJ, Gregori G, et al. Progression of geographic atrophy in age-related macular degeneration imaged with spectral domain optical coherence tomography. *Ophthalmology*. 2011;118:679–686.

13. Holz FG, Bindewald-Wittich A, Fleckenstein M, et al. Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am J Ophthalmol*. 2007;143:463–472.e2.

14. Feeny AK, Tadarati M, Freund DE, Bressler NM, Burlina P. Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images. *Comput Biol Med*. 2015;65:124–136.

15. Deckert A, Schmitz-Valckenberg S, Jorzik J, Bindewald A, Holz F, Mansmann U. Automated analysis of digital fundus autofluorescence images of geographic atrophy in advanced age-related macular degeneration using confocal scanning laser ophthalmoscopy (cSLO). *BMC Ophthalmol*. 2005;5:8.

16. Schmitz-Valckenberg S, Brinkmann CK, Alten F, et al. Semiautomated image processing method for identification and quantification of geographic atrophy in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:7640–7646.

17. Hu Z, Medioni GG, Hernandez M, Sadda SR. Automated segmentation of geographic atrophy in fundus autofluorescence images using supervised pixel classification. *J Med Imaging*. 2015;2:014501.

18. Hu Z, Wang Z, Sadda SR, editors. Automated segmentation of geographic atrophy using deep convolutional neural networks. *Medical Imaging 2018: Computer-Aided Diagnosis*; 2018: International Society for Optics and Photonics.

19. Wang Z, Sadda SR, Hu Z, editors. Deep learning for automated screening and semantic segmentation of age-related and juvenile atrophic macular degeneration. *Medical Imaging 2019: Computer-Aided Diagnosis*; 2019: International Society for Optics and Photonics.

20. Ho TK, editor. Proceedings of the 3rd International Conference on Document Analysis and Recognition. Random decision forests. IEEE; 1995:278–282.

21. Friedman J, Hastie T, Tibshirani R. *The elements of statistical learning: Springer series in statistics*. New York: Springer; 2001.

22. Faes L, Liu X, Wagner SK, et al. A clinician's guide to artificial intelligence: how to critically appraise machine learning studies. *Transl Vis Sci Technol*. 2020;9:7.

23. RegionFinder—The Perfect Use: Heidelberg Engineering; 2011. Available from: https://www.heidelbergengineering.com/media/e-learning/Totara/Dateien/pdf-tutorials/93410-001_SPECTRALIS_Regionfinder_EN.pdf.

24. Wolf-Schnurrbusch UE, Wittwer VV, Ghanem R, et al. Blue-light versus green-light autofluorescence: lesion size of areas of geographic atrophy. *Invest Ophthalmol Vis Sci*. 2011;52:9497–9502.

25. Pfau M, Goerdt L, Schmitz-Valckenberg S, et al. Green-light autofluorescence versus combined blue-light autofluorescence and near-infrared reflectance imaging in geographic atrophy secondary to age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2017;58:BIO121–BIO30.

List of publications during doctoral studies unrelated to the thesis

List of references provided only

1. Khan SM and Liu X, et al. (2020). A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *The Lancet Digital Health*.

2. Cruz Rivera S, et al. (2020). SPIRIT-AI and CONSORT-AI Working Group; SPIRIT-AI and CONSORT-AI Steering Group; SPIRIT-AI and CONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 26(9):1351-1363.

3. Liu X, et al. (2020) SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 26(9):1364-1374.

4. Rivera SC, et al. (2020) SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ. 370:m3210.

5. Liu X, et al. (2020) SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ. 370:m3164.

6. Liu X, et al. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health*.

7. Cruz Rivera S, et al. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *The Lancet Digital Health*.

8. Wicks P et al. (2020). Going on up to the SPIRIT in AI: will new reporting guidelines for clinical trials of AI interventions improve their rigour? BMC Med. 18(1):272.

9. Faes L, et al. (2020). A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. Transl Vis Sci Technol. 9(2):7.

10. Wagner SK, et al. (2020) Insights into Systemic Disease through Retinal Imaging-Based Oculomics. Transl Vis Sci Technol.9(2):6.

11. Hui BTK, et al. (2020). Bilateral acute anterior uveitis and iris atrophy caused by moxifloxacin. BMJ Case Re. 13(6):e233528.

12. Korot E, et al. (2020). Will AI Replace Ophthalmologists? Transl Vis Sci Technol. 9(2):2.

13. Akbarali S*, et al*. (2020). Imaging based uveitis surveillance in juvenile idiopathic arthritis: feasibility, acceptability and diagnostic performance. *Arthritis Rheumatol*.

14. Sounderajah V, et al. (2020). Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. Nat Med. 26(6):807-808.
15. Liu X, et al. (2019). Optical coherence tomography (OCT) in unconscious and systemically unwell patients using a mobile OCT device: a pilot study. BMJ Open. 9(11):e030882.

16. Liu X and Faes L, et al. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*. 1(6):e271-e297.

17. Faes L and Wagner SK, et al. (2019). Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *The Lancet Digital Health*.

18. Liu X, et al; CONSORT/SPIRIT-AI Extension Group. (2019). Extension of the CONSORT and SPIRIT statements. Lancet. 2019 Oct 5;394(10205):1225.

19. Liu X and Kelly SR, et al. (2019). Evaluating the Impact of Uveitis on Visual Field Progression Using Large-Scale Real-World Data. Am J Ophthalmol. 207:144-150.