Check for updates

RESEARCH ARTICLE

# REVISED The STArgardt Remofuscin Treatment Trial (STARTT): design and baseline characteristics of enrolled Stargardt patients

[version 3; peer review: 2 approved]

Patty P.A. Dhooge[1,2], Philipp T. Möller[3,4], Camiel J.F. Boon[5,6], Andrew J. Lotery[7], Philipp Herrmann[3,8], Maurizio Battaglia Parodi[9], Wolfgang Klein[10], Mario G. Fsadni[10,11], Thomas H. Wheeler-Schilling[12], Oliver Jungmann[13], Hans Müller[13], Frank G. Holz[3,4], Steffen Schmitz-Valckenberg[3,4,14], Tobias M. Peters[12], Katarina Stingl[12,15], Carel B. Hoyng[1,2], Soraprazan Consortium

[1]Department of Ophthalmology, Radboud University Nijmegen Medical Centre, Nijmegen, 6500HB, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, 6525EN, The Netherlands
[3]Department of Ophthalmology, University of Bonn, Bonn, 53127, Germany
[4]GRADE reading center, Bonn, 53127, Germany
[5]Department of Ophthalmology, Leiden University Medical Center, Leiden, 2333ZA, The Netherlands
[6]Department of Ophthalmology, Amsterdam University Medical Center, Amsterdam, 1081HV, The Netherlands
[7]Faculty of Medicine, University of Southampton, Southampton, SO17 1BJ, UK
[8]Center for Rare Diseases Bonn (ZSEB), University of Bonn, Bonn, 53113, Germany
[9]Department of Ophthalmology, Ospedale San Raffaele, Milan, 20132, Italy
[10]Katairo GmbH, Kusterdingen, 72127, Germany
[11]International Pharm-Med Ltd, Hemel Hempstead, HP1 1LD, UK
[12]University Eye Hospital, Center for Ophthalmology, Institute for Ophthalmic Research, University of Tübingen, Tübingen, 72076, Germany
[13]Smerud Medical Research Germany GmbH, Mannheim, D-68161, Germany
[14]John A. Moran Eye Center, University of Utah, Salt Lake City, Utah, 84132, USA
[15]Center for Rare Eye Diseases, University of Tübingen, Tübingen, 72076, Germany

**v3** **First published:** 19 Aug 2021, **1**:96
https://doi.org/10.12688/openreseurope.13872.1
**Second version:** 22 Jul 2022, **1**:96
https://doi.org/10.12688/openreseurope.13872.2
**Latest published:** 27 Sep 2022, **1**:96
https://doi.org/10.12688/openreseurope.13872.3

## Abstract

**Background:** This report describes the study design and baseline characteristics of patients with Stargardt disease (STGD1) enrolled in the STArgardt Remofuscin Treatment Trial (STARTT).
**Methods:** In total, 87 patients with genetically confirmed STGD1 were randomized in a double-masked, placebo-controlled proof of concept trial to evaluate the safety and efficacy of 20 milligram oral remofuscin for 24 months. The primary outcome measure is change in mean quantitative autofluorescence value of an 8-segment ring centred on the fovea ($qAF_8$). Secondary efficacy variables are best corrected visual

**Open Peer Review**

**Approval Status** ✓ ✓

| | 1 | 2 |
|---|---|---|
| **version 3** (revision) 27 Sep 2022 | | ✓ view |
| **version 2** (revision) 22 Jul 2022 | | ? view |
| **version 1** 19 Aug 2021 | ✓ | ? |

acuity (BCVA), low-luminance visual acuity (LLVA), mesopic microperimetry (mMP), spectral domain optical coherence tomography (SD-OCT), reading speed on Radner reading charts, and patient-reported visual function as assessed by the National Eye Institute Visual Functioning Questionnaire 25 (NEI VFQ-25) and Functional Reading Independence (FRI) Index.
**Results:** Mean age of participants was 35±11 years with 49 (56%) female. Median $qAF_8$ value was 438 Units (range 210-729). Median BCVA and LLVA in decimal units were 0.50 (range 0.13-0.80) and 0.20 (range 0.06-0.63), respectively. The median of the mean retinal sensitivity with mMP was 20.4 dB (range 0.0-28.8). SD-OCT showed median central subfield retinal thickness of 142 µm (range 72-265) and median macular volume of 1.65 $mm^3$ (range 1.13-2.19). Compared to persons without vision impairment, both reading performance and patient-reported visual function were significantly lower ($p<0.001$, one sample t-test). Mean reading speed was 108±39 words/minute with logRAD-score of 0.45±0.28. Mean VFQ-25 composite score was 72±13. Mean FRI Index score 2.8±0.6.
**Conclusions:** This trial design may serve as reference for future clinical trials as it explores the utility of $qAF_8$ as primary outcome measure. The baseline data represent the largest, multi-national, STGD1 cohort to date that underwent standardized qAF imaging, reading speed assessment and vision-related quality of life measures which all contribute to the characterization of STGD1.
**EudraCT registration:** 2018-001496-20 (09/05/2019)

**Keywords**
Stargardt disease, STGD1, ABCA4, Remofuscin, Soraprazan, qAF, quantitative autofluorescence

This article is included in the Horizon 2020 gateway.

This article is included in the Clinical Medicine gateway.

This article is included in the Medical Physics collection.

1 2
view view

1. **Xiangrong Kong**, Johns Hopkins University, Baltimore, USA
2. **Victor Chong**, Janssen R&D LLC, Raritan, USA

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Disease Management and Treatment collection.

**Corresponding author:** Carel B. Hoyng (carel.hoyng@radboudumc.nl)

**Author roles: Dhooge PPA**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Möller PT**: Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Boon CJF**: Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Lotery AJ**: Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Herrmann P**: Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Battaglia Parodi M**: Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Klein W**: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Fsadni MG**: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Wheeler-Schilling TH**: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Jungmann O**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Müller H**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Holz FG**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Schmitz-Valckenberg S**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Peters TM**: Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Stingl K**: Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Hoyng CB**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing;

**Competing interests:** Wolfgang Klein: financial support, personal financial interest, and consultant for Katairo GmbH. Mario G Fsadni: paid consultant to Katairo GmbH, outsourced from International Pharm-Med Ltd in United Kingdom. Oliver Jungmann: commercial relationships via employment Smerud Medical Research Hans Müller: commercial relationships via employment Smerud Medical Research Philipp T. Möller, Frank G. Holz and Steffen Schmitz-Valckenberg: contracted research for reading center services by Katairo GmbH All other authors: no relevant competing interests disclosed.

**Grant information:** This research was financially supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement No. 779317 (Soraprazan - a new regenerative therapy for Stargardt's disease [Soraprazan]).
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2022 Dhooge PPA *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Dhooge PPA, Möller PT, Boon CJF *et al.* **The STArgardt Remofuscin Treatment Trial (STARTT): design and baseline characteristics of enrolled Stargardt patients [version 3; peer review: 2 approved]** Open Research Europe 2022, **1**:96 https://doi.org/10.12688/openreseurope.13872.3

**First published:** 19 Aug 2021, **1**:96 https://doi.org/10.12688/openreseurope.13872.1

**REVISED** **Amendments from Version 2**

We have changes our discussion section

Section added: The sample size was calculated based on the baseline $qAF_8$ score of 484±109 units (mean± standard deviation (SD) by the single center, single operator study by Burke *et al.* 2014[12] . In our multi-center study, actual baseline values were slightly more variable, namely 438 (210–729, 133) median (range, IQR). We therefore chose to elongate the trial period from 12 to 24 months, Opposed to the cohort of Burke *et al.* 2014[12], for the STARTT, we specifically selected patients with preserved retinal function because they are likely to show disease progression without effective treatment. We therefore included patients with a BCVA between 0.20 and 0.80, because better baseline BCVA was associated with a greater yearly rate of decline[25,26] ..... With these baseline values, we should be able to provide the first proof of effectiveness of remofuscin within 24 months if treatment works.

Section added: The current trial design was therefore designed to be able to align qAF with functional endpoints including microperimetry, BCVA and reading speed. Nonetheless, as both qAF and visual function are highly variable. It might be hard to demonstrate a structure-function correlation in such a small sample size and future studies need to be undertaken.

The wording of conclusion has been weakened: In conclusion, the STARTT design demonstrates a unique approach to explore the utility of $qAF_8$ as primary outcome measure while simultaneously evaluating the safety and efficacy of a new treatment agent for STGD1.

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

Stargardt disease (STGD1) is an inherited retinal disease caused by mutations in the *ABCA4* gene which results in progressive vision loss[1,2]. In the absence of a functional ABCA4 protein, bisretinoid products are formed and deposited in postmitotic retinal pigment epithelium (RPE) cells as lipofuscin[3]. The lipofuscin build-up in the RPE of patients with STGD1 cannot yet be reversed and results in degeneration of RPE cells and, subsequently, corresponding photoreceptors leaving residual atrophy. This leads to worsening of visual acuity and could progress to blindness[4].

Currently, there is no treatment available which can remove the accumulated toxic lipofuscin to allow RPE cell regeneration. However, the ability of soraprazan to remove lipofuscin from RPE cells has been demonstrated in both primates and ABCA4-/- knock out mice[5–7]. Soraprazan was originally designed to treat gastro-oesophageal reflux disease (GERD). Its safety and tolerability following oral administration were already demonstrated in various phase I and phase II clinical trials[8,9]. To evaluate the safety and efficacy of oral soraprazan in patients with STGD1, we designed the Soraprazan Project as part of the European Union's Horizon 2020 research and innovation program (grant agreement No 779 317). Although the international non-proprietary name of the active ingredient remains soraprazan, the investigational drug is renamed remofuscin®. The European Commission has granted an orphan designation (EU/3/13/1208) to use soraprazan, renamed remofuscin, for the treatment of STGD1. The STArgardt Remofuscin Treatment Trial (STARTT) is a phase II, prospective, multicentre, randomized, double-masked, placebo-controlled proof of concept trial. The current paper describes the study design of the STARTT and the baseline characteristics of enrolled patients.

The primary objective of the STARTT is to evaluate the efficacy of remofuscin in reducing the amount of lipofuscin in RPE cells of subjects with STGD1 by assessing the change in quantitative fundus autofluorescence (qAF) levels from baseline compared to placebo after treatment with remofuscin for up to 24 months. qAF acts as a marker of lipofuscin levels in the retina and allows for quantification of RPE changes at distinct retinal locations[10]. With qAF being available since 2011, only limited data is available on qAF levels in STGD1 patients. In a single centre cohort of 77 STGD1 patients, 76,6% had higher qAF levels as compared to age-related controls[11]. qAF levels were up to 8-fold higher in a different single centre STGD1 cohort of 42 patients[12]. In addition, qAF was able to discriminate STGD1 from other retinal diseases that appear similar on clinical examination[13–15]. With 87 patients enrolled in the multicentre STARTT, this study represents the largest international STGD1 cohort to date that include repetitive, prospective standardized qAF imaging.

It is important to include outcomes parameters that are relevant from the patients' perspective. Therefore, secondary efficacy parameters of the STARTT include best corrected visual acuity (BCVA), low luminance visual acuity (LLVA), mesopic microperimetry (mMP) and reading speed as functional measures. To our knowledge, Radner reading charts[16] have not been previously used to assess reading acuity and speed in patients with STGD1. In addition, we include patient-reported outcome measures regarding visual function and quality of life as assessed by the National Eye Institute Visual Functioning Questionnaire 25 (NEI VFQ-25) and Functional Reading Independence (FRI) Index[17,18]. Knowledge on the patient-reported vison-related quality of life of STGD1 patients is still limited. The baseline data presented in the current paper will therefore aid in characterizing the burden of STGD1 and provide outcome measures and possible indicators of cost effectiveness for all future clinical trials.

## Methods

STARTT is a randomized, double-masked, placebo-controlled study to evaluate safety and efficacy of oral remofuscin in subjects with STGD1 (EudraCT No. 2018-001496-20). An international consortium was formed consisting of six European investigator sites: two centres in the Netherlands (Radboud University Medical Centre Nijmegen, Nijmegen and Leiden University Medical Centre, Leiden), two centres in Germany (University Eye Hospital Tübingen, Tübingen and Department of Ophthalmology, University of Bonn, Bonn), one centre in the United Kingdom (University of Southampton, Southampton) and one centre in Italy (Ospedale San Raffaele, Milano), a Contract Research Organisation (CRO) (Smerud Medical Research), and a start-up company as a sponsor (Katairo). The study is conducted according to the principles of the Declaration of Helsinki (Version 2013) and is being managed and

conducted according to the latest International Conference on Harmonisation (ICH) guidelines for Good Clinical Practice (GCP). For all sites, the study protocol was approved by the Institutional Review Boards (CCMO, project number NL68179.091.18; Commissie Mensgebonden Onderzoek Arnhem-Nijmegen, project number 2018-4988; BfArM, project number 4043121; Ethik-Kommission an der Medizinischen Fakultät der Eberhard-Karls-Universität and am Universitätsklinikum Tübingen, project number 843/2018AMG1; MHRA, project number 50612/001/001-0001; North-West Haydock Research Ethics Committee, project number 18/NW/0845; AIFA, project number 2018-001496-20; Comitato Etico Ospedale San Raffaele Milano, project number 2018-001496-20). The study was registered at www.clinicaltrialsregister.eu (EudraCT No. 2018-001496-20, registered May 9th, 2019). Potential subjects were identified and recruited from the investigator site's database and all patients gave written informed consent before enrolment. Study staff were certified before patients were recruited. All study site personnel as well as the CRO personnel involved in the monitoring or conduct of the study were blinded to the individual subject treatment assignments. A data safety and monitoring board (DSMB) composed of a physician, an ophthalmologist not involved in the study conduct and a statistician reviews the blinded data on a regular basis in order to assure independent assessment and ensure safety of the subjects. Currently, the study is ongoing. Last patient last visit is planned for September 2022.

## Eligibility criteria

A complete list of inclusion and exclusion criteria is given in Table 1. The primary variable of interest is the change in mean qAF value of an 8-segment circular ring centred on the

**Table 1. Inclusion and exclusion criteria for STArgardt Remofuscin Treatment Trial.**

| | Criteria |
|---|---|
| Inclusion | – Male or female of any ethnicity and ≥ 18 years old<br>– Clinical diagnosis of typical autosomal recessive Stargardt macular dystrophy (STGD1)<br>– Genetic report indicating at least two *ABCA4* mutations (one with confirmed pathogenesis by a certified lab, one reported previously)<br>– Onset of STGD1 disease before the age of 45 years<br>– Elevated $qAF_8$ in at least one eye at screening (value ≥ 300 units)<br>– Visual acuity of the study eye: BCVA 0.2-0.8 (decimal unit)<br>– Study eye must have clear ocular media and adequate pupillary dilation, including no allergy to dilating eye drops and with sufficient fixation to permit good quality retinal imaging.<br>– Female subjects of childbearing potential and male subjects participating in the study who are sexually active must use acceptable contraception from screening and until one month after intake of the last IMP dose. Male and female subjects documented<br>– as being of non-child bearing potential (e.g. infertile, surgically sterile, postmenopausal) are exempt from the contraceptive requirements.<br>– Willing to avoid excessive exposure to sun light (e.g. by using a hat, ultraviolet absorbing sunglasses and sunscreen with a minimum SPF of 30) |
| Exclusion | – Intolerance to acid pump antagonists<br>– Hypersensitivity to Soraprazan or to any of the excipients<br>– Intake of prohibited medications/supplements (supplements containing vitamin A or beta-carotene, medications to treat any liver disease, or oral retinoid medications) within 28 days prior to screening and throughout the study<br>– Intake of other drugs with a pH dependent absorption, e.g. ketoconazole<br>– Breastfeeding or positive urine pregnancy test at screening or visit 2 (first intake of IMP).<br>– At screening, clinically significant abnormal haematology or biochemistry findings or levels >1.5 x upper limit of normal of aspartate aminotransferase (AST), alanine aminotransferase (ALT), and/or total bilirubin<br>– Acute or unstable severe disease or history of disease which in the opinion of the investigator would preclude participation in the study<br>– Active or history of an additional ocular disorder in the primary study eye that, in the opinion of the investigator, may confound the study results. These include, but are not limited to, any reason that might interfere with the imaging techniques used in the study (such as optic media opacity or poor pupil dilatation), inflammatory eye<br>– disease, other retinal disorders besides STGD, confirmed glaucoma or baseline intraocular pressure of ≥25mmHg, optic neuritis, high myopia (>8D spherical equivalent), amblyopia<br>– Intraocular surgery or injections in the primary study eye within 180 days of the screening visit<br>– Clinically significant abnormal electrocardiogram, or a corrected QT interval (QTc) of ≥450ms in males or ≥470ms in females<br>– Participation in any other investigational clinical trial within 28 days of the screening visit |

BCVA, best corrected visual acuity; IMP, investigational medicinal product; SPF, sun protection factor; STGD1, Stargardt disease; $qAF_8$, mean quantitative autofluorescence value of an 8-segment circular ring centred on the fovea

fovea ($qAF_8$) from baseline to the key time point at Month 12 (Visit 10), and Month 24 (Visit 16). To be included, the $qAF_8$ value must be ≥300 units at screening in the study eye. If the quality of the qAF image at screening is not deemed acceptable by the central reading centre, the qAF measurement is repeated at an additional visit (Visit 1B). Further, the study eye requires a BCVA between 0.20 and 0.80 (decimal units) at screening. If only one eye meets the above-mentioned criteria, this eye will be defined as the primary study eye. In case both eyes meet the criteria, the eye with the higher BCVA score will be defined as the primary study eye.

### Study treatment and randomization

In the STARTT, the investigational medicinal product (IMP) consists of two 10 milligram tablets, containing either remofuscin or placebo. Participants will have to take two IMP tablets orally per day in the late evening for up to 24 months. In the previous phase I and II trials using soraprazan to treat GERD, a daily dose of 20 milligram per day was considered safe[8,9], and was used in STARTT. An oral dose of 6 mg/kg/day administered to monkeys for 1-year resulted in a full removal of lipofuscin without any ocular adverse effects. This dose in monkeys is approximately 3-fold higher than humans dosed at 20 mg/day. As a full effect on monkey lipofuscin removal was observed at 6 mg/kg/day after 12 months, it is assumed that a daily dose of 20 mg/day will also result in lipofuscin reduction in Stargardt disease patients. For placebo tablets, remofuscin is replaced by the same amount of cellulose. In order to minimize any symptoms related to the expected secondary pharmacodynamics effects on stomach acid reduction, dosing of remofuscin is planned before hours of sleep. The treatment is randomized in a 2:1 ratio (remofuscin:placebo).

### Visit procedures

The study consists of 17 standardized study visits spread over a review period of 24 months. Procedures took place within the investigator sites. Patients were screened between June 2019 and August 2020. The last patient visit is planned for September 2022. The study flow-chart is set out in Table 2 and Table 3. Comparison will be made between remofuscin and placebo treatment arms at month 12 (primary timepoint) and month 24 for the following functional and structural outcome measures: (1) $qAF_8$ value; (2) BCVA taken in decimal as measured by Early Treatment Diabetic Retinopathy Study (ETDRS) charts; (3) BCVA taken in decimal under low luminance conditions by placing a 2.0-log-unit neutral density filter in front of the study eye and having the subject read a normally illuminated ETDRS chart (LLVA); (4) Low Luminance Deficit (LLD) calculated as the difference between BCVA and LLVA; (5) binocular and monocular maximum reading speed and critical print size as assessed by Radner reading charts[16]; (6) macular functional response as assessed by mMP; (7) patient-reported visual function as assessed by the NEI VFQ-25[17] and FRI Index[18]; (8) central subfield retinal thickness (CSRT) and macular volume as assessed by SD-OCT; (9) macular atrophy as assessed by fundus autofluorescence (FAF); (10) changes on colour fundus photography (FP). In addition, the centre in Tübingen will assess changes in (11) pupillographic campimetry and (12) adaptive optics.

Safety values will include (1) evaluation of clinical and laboratory adverse events; (2) blood haematology and clinical chemistry; (3) vital signs including blood pressure and heart rate after at least 5 minutes of rest in supine; (4) 12 lead Electrocardiogram; (5) physical examinations; (6) body mass index (BMI); (7) number of patients with an increase in Lens Opacities Classification System grading by ≥2 grades.

### Retinal imaging

Retinal image acquisition and data transmitting are performed according to the standardized operating procedures from the Grading of Digital fundus Examination (GRADE) Reading Centre, Department of Ophthalmology, University of Bonn, Germany. This reading centre performs the standardized scientific analysis of the study endpoints involving qAF, FAF, mMP, SD-OCT, and FP. All equipment has been approved by the central reading centre and each photographer was certified prior to obtaining retinal images for the study.

***Microperimetry***. To assess retinal sensitivity and fixation, mMP is obtained using the CenterVue Macular Integrity Assessment (MAIA), Padova, Italy. With this device, fundus-controlled perimetry was performed with automated real-time fundus tracking. After at least 5 minutes of dark adaptation, retinal sensitivity was obtained with a custom-made test pattern using 54 Goldmann III stimuli of 200ms with an initial test brightness of 2.6±0.5 asb and a background luminance of 1.27cd/m2. The test pattern was automatically placed on the fovea and could be manually corrected by the instructor if needed. We use a custom-made test to ensure that microperimetry testing areas fall within the qAF grid as described by Delori. In this way we can correlate the qAF changes to the microperimetry changes in selected areas.

To reduce learning effects, a training exam containing 8 stimuli was conducted for both eyes before the custom-made test was performed and the study eye was always examined last. mMP always preceded qAF imaging, dilated ophthalmoscopy, and FP, as these procedures may temporarily affect retinal sensitivity. From the second visit onwards, the mesopic assessment of the screening visit serves as a reference and follow-up mode was used to automatically place all follow-up scans in the same location as the baseline scan. If Fixation Losses are above 30% or if the 95% Bivariate Contour Ellipse Area is above 50°, the test is deemed inaccurate and was repeated at least once. The mean MP sensitivity level was calculated per eye by averaging the point-wise retinal sensitivity in decibel.

***Quantitative fundus autofluorescence and fundus autofluorescence***. qAF and FAF images were obtained using a Spectralis device (Heidelberg Engineering, Heidelberg, Germany) with a qAF imaging mode including installation of the qAF reference standard and an upgrade to HEYEX software version 5.6 or higher. In order to reduce photopigment absorption, the retina was exposed to the blue excitation mode (488nm) for at least 20 seconds. For FAF, one image centred on the macula was obtained per eye (field of view 30° × 30°, 768 × 768 pixels, high-speed mode, 30 single frames using automated real-time

**Table 2. Study flow-chart: screening to month 12.**

| Study Period | Screening | | Randomisation | Treatment | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Clinic Visit (V) Number** | **V1** | **V1b (if required)[6]** | **V2** | **V3** | **V4** | **V5** | **V6** | **V7** | **V8** | **V9** | **V10** |
| **Study Weeks (Months)** | **-4 to -1** | **between V1 and V2** | **0** | **4 (1)** | **8 (2)** | **12 (3)** | **16 (4)** | **24 (6)** | **32 (8)** | **40 (10)** | **52 (12)** |
| **Study Day** | **-28 to -7** | | **0** | **28 (±3)** | **56 (±3)** | **84 (±3)** | **112 (±3)** | **168 (±3)** | **224 (±3)** | **280 (±3)** | **364 (±3)** |
| Informed Consent | X | | | | | | | | | | |
| Inclusion/Exclusion Criteria | X | | X | | | | | | | | |
| Review Medical history[1] (incl. Stargardt disease history) for any ongoing diseases | X | | | | | | | | | | |
| Confirmation of ABCA4 mutations | X | | | | | | | | | | |
| Check for concomitant medication | X | | X | X | X | X | X | X | X | X | X |
| Bodyweight and Height[2] | X | | | | | | | | | | X |
| Physical examination | X | | | | | | | | | | X |
| Vital signs | X | | X | X | X | X | X | X | X | X | X |
| Eye examination | X | | X | | X | | X | X | X | X | X |
| 12-lead electrocardiogram | X | | X | X | X | X | X | X | X | X | X |
| Blood sample for haematology and clinical chemistry[3] | X | | X | X | X | X | X | X | X | X | X |
| Urine pregnancy test (if applicable)[4] | X | | X | X | X | X | X | X | X | X | X |
| BEst corrected visual acuity (BCVA) | X | | X | X | X | X | | X | | X | X |
| Low luminance visual acuity (LLVA), low luminance deficit (LLD) | | | X | | | | | X | | | X |
| Reading Speed | | | X | | | | | | | | X |
| NEI VFQ25 and FRI Index | | | X | | | | | X | | | X |
| Retinal sensitivity (microperimetry) | BE[5] | | BE | BE | BE | BE | | BE | | BE | BE |
| Quantitative auto-fluorescence (qAF) | BE | BE | BE | BE | BE | BE | BE | BE | BE | BE | BE |
| SD-OCT | BE | | BE | BE | BE | BE | | BE | | BE | BE |
| Colour Fundus Photography | BE | | BE | BE | BE | BE | | BE | | BE | BE |
| Adaptive Optics[7] | | | X | | | | | X | | | X |
| Pupillographic Campimetry[7] | | | X | | | | | X | | | X |
| Definition of the "primary study eye" | | | X | | | | | | | | |
| Adverse events | X | | X | X | X | X | X | X | X | X | X |
| Randomisation | | | X | | | | | | | | |
| IMP dispensation | | | X | X | X | X | X | X | X | X | |
| IMP return/accountability | | | | X | X | X | X | X | X | X | X |

[1]ongoing diseases or stopped within 12 months prior to screening

[2]height is only required at screening

[3]central laboratory

[4]urine stick test before IMP intake at the site

[5]BE = both eyes (examination of both eyes)

[6]V1b is required if the quality of the qAF image taken at V1 is not acceptable according to the central reading centre. In these cases, qAF should be repeated before V2

[7]only at the clinical trial site in Tübingen

**Table 3. Study flow-chart: month 14 to month 24, including early termination (ET) visit and follow-up visit.**

| Study Period | Treatment | | | | | | Follow up |
|---|---|---|---|---|---|---|---|
| **Clinic Visit (V) Number** | **V11** | **V12** | **V13** | **V14** | **V15** | **V16 (or ET)[1]** | **VS** |
| **Study Weeks (Months)** | **60 (14)** | **68 (16)** | **76 (18)** | **84 (20)** | **92 (22)** | **104 (24)** | **Last Visit (LV) +1** |
| **Study Day** | **420 (±3)** | **476 (±3)** | **532 (±3)** | **588 (±3)** | **644 (±3)** | **728 (±3)** | **LV +28 (±3)** |
| Check for concomitant medication | X | X | X | X | X | X | X |
| Bodyweight | | | | | | | |
| Physical examination | | | | | | X | |
| Vital signs | X | X | X | X | X | X | |
| Eye examination | X | X | X | X | X | X | |
| 12-lead electrocardiogram | X | X | X | X | X | X | |
| Blood sample for haematology and clinical chemistry[2] | X | X | X | X | X | X | X |
| Urine pregnancy test (if applicable)[3] | X | X | X | X | X | X | X |
| BEst corrected visual acuity (BCVA) | X | X | X | X | X | X | X |
| Low luminance visual acuity (LLVA), low luminance deficit (LLD) | | | | | | X | |
| Reading Speed | | | | | | X | |
| NEI VFQ25 and FRI Index | | | | | | X | |
| Retinal sensitivity (microperimetry) | BE[4] | BE | BE | BE | BE | BE | BE |
| Quantitative auto-fluorescence (qAF) | BE | BE | BE | BE | BE | BE | BE |
| SD-OCT | BE | BE | BE | BE | BE | BE | BE |
| Colour Fundus Photography | BE | BE | BE | BE | BE | BE | BE |
| Adaptive Optics[5] | | | | | | | |
| Pupillographic Campimetry[5] | | | | | | | |
| Adverse events | X | X | X | X | X | X | X |
| IMP dispensation | X | X | X | X | X | | |
| IMP return/accountability | X | X | X | X | X | X | |

[1]ET = Early Termination, V16 will be performed

[2]central laboratory

[3]urine stick test before IMP intake at the site

[4]BE = both eyes (examination of both eyes)

[5]only at the clinical trial site in Tübingen

(ART) modus). From the FAF images, if present, the macular atrophy, interpreted as complete loss of autofluorescence (analogous to definite decreased autofluorescence as described by Kuehlewein *et al.*[19]), was calculated using the Heidelberg RegionFinder software version 2.6.4.0.

qAF imaging was performed according to the method developed by Delori *et al.*[10] Two qAF movies for both eyes were taken, each with a series of 12 frames per movie. Quality was evaluated so that every movie has at least 9 frames that are of equal brightness without shadowing or flickering. With the frames of good quality, a mean color-coded image was computed without normalization. The color-coded qAF images were adjusted for corneal curvature measurement performed at the screening visit and adjusted for patient's age. The Delori pattern was placed, centred on the fovea and its border moved in direction of the optic nerve heads temporal border. Individual segmentation was performed to exclude atrophic areas and

vessels and the mean grey level was calculated for each segment of the Delori pattern. $qAF_8$ value was calculated by taking the mean of the grey levels in the 8 middle segments of the Delori pattern, as previously described by Delori *et al.*[10]

***Spectral domain optical coherence tomography***. SD-OCT images were obtained with either the Spectralis OCT or Spectralis HRA+OCT device (Heidelberg Engineering, Heidelberg, Germany) using the Spectralis Software Version 6.9a or newer. Imaging consisted of different scan patterns per eye: (1) central line scan (30°, centred on the fovea, single B-scans, 30 frames for ART modus); (2) central volume scan (30° x 30°, centred on the fovea, 31 B-scans, 4 frames for ART modus); (3) enhanced-depth imaging (EDI) volume scan (30° x 25°, centred on the fovea, 241 B-scans, 9 frames for ART modus). The AutoRescan$^{tm}$ tool was used to automatically place all follow-up scans in the same location as the baseline scan. Central subfield retinal thickness (CSRT), and macular volume (centre 3 x 3 mm area) were calculated by manual segmentation.

***Fundus photography.*** FP were acquired using a standard digital fundus camera (field 30–40°, minimum resolution of 1536 x 1536 pixels), centred on (1) optic disc; (2) fovea; (3) temporal part of macula.

### Sample size and baseline statistics

The sample size was calculated on the basis of the primary hypothesis that $qAF_8$ score will be 50 units lower with active treatment compared to placebo after 12 months of treatment. Baseline $qAF_8$ score was expected to be 484±109 units (mean± standard deviation (SD)) based on Burke *et al*. 2014[12]. With an assumed correlation of 0.75 between the $qAF_8$ score values at baseline and after 12 months of treatment, the corresponding changes from baseline will have a standard deviation (SD) of 77 units. With 58 subjects on active treatment and 29 subjects on placebo, computer simulations show that analysis of covariance (ANCOVA) will have at least 80% power to detect a statistically significant treatment difference when using a 5% level of significance.

In this manuscript we report the descriptive statistics at baseline including mean ± SD, median, range and interquartile range for continuous variables and proportions for categorical variables. For Radner and NEI-VFQ 25, comparisons with previously published norm scores were assessed by a one sample t-test[16,17]. A p-value of <0.05 was considered significant. Statistical analysis was performed using the SPSS statistics package for Windows; version 22 (SPSS IBM, New York, USA).

## Results

A total of 112 patients were screened of whom 87 patients fitted the inclusion criteria and were enrolled in the STARTT between June 2019 and September 2020[20]; 24 patients by the Radboud University Medical Centre Nijmegen, 12 patients by Leiden University Medical Centre, 23 patients by University Eye Hospital Tübingen, 8 patients by University of Bonn, 9 patients by University of Southampton and 11 patients by Ospedale San Raffaele. We included 49 females (56%) and 38 males (44%) with a mean age of 35 ± 11 years. Patients identified themselves with the following ethnicities: White (82 patients, 94%), Asian (3 patients, 4%), and Hispanic (2 patients, 2%). Age at onset of STGD1 was between 15 and 44 years with a median of 28 years.

The baseline study eye characteristics of all enrolled patients are summarized in Table 4. To provide a complete overview of $qAF_8$ values, we used the screening values instead of the baseline values in 15 cases because of insufficient grading quality of the image (7 times), use of wrong settings (2 times) and incomplete data (6 times). At least one eye of all 87 enrolled patients had a $qAF_8$ value ≥300 units and a BCVA between 0.20 and 0.80 (decimal units) at screening (visit 1). For 5 patients, $qAF_8$ values for both eyes dropped below 300 units at the baseline visit (visit 2). Accordingly, baseline $qAF_8$ values

**Table 4. Baseline study eye characteristics of patients enrolled in STArgardt Remofuscin Treatment Trial.**

| Study eye characteristics | Median (range, IQR) |
|---|---|
| **$qAF_8$** | 438 (210–729, 133) |
| **BCVA in decimals** | 0.50 (0.13–0.80, 0.38) |
| **LLVA in decimals** | 0.20 (0.06–0.63, 0.16) |
| **LLD in decimals** | 0.18 (–0.05–0.56, 0.25) |
| **Mean retinal sensitivity in decibels (available for 81 eyes)** | 20.4 (0.0–28.8, 9.05) |
| **Atrophic area in mm² (present in 42/83 gradable eyes)** | 0.97 (0.05–21.3, 4.58) |
| **Central subfield retinal thickness in µm (available for 82 eyes)** | 142 (72–265, 47.3) |
| **Macular Volume in mm³ (available for 82 eyes)** | 1.65 (1.13–2.19, 0.28) |

BCVA, best corrected visual acuity; IQR, interquartile range; LLVA, BCVA under low luminance conditions; LLD, Low Luminance Deficit; $qAF_8$, mean quantitative autofluorescence value of an 8-segment circular ring centred on the fovea.

of the study eye were between 210 and 729 units with a median value of 438 units. For 4 patients, BCVA dropped below 0.20 at visit 2, resulting in baseline BCVA values between 0.13 and 0.80 with a median acuity of 0.50. If retinal imaging had poor quality, grading was not performed. Good quality mMP, FAF and SD-OCT images were available for 81, 83 and 82 eyes respectively. RPE atrophy was present in 42 of 83 gradable eyes (51%).

The patient-reported visual function was assessed with the NEI VFQ-25 in all 87 enrolled patients and the FRI in 86 patients, as 1 patient did not complete the FRI questionnaire. The results of the NEI VFQ-25 are reported in Table 5. Scores range from 0 to 100, where 100 indicates no disability[17]. The overall NEI VFQ-25 composite score, averaging the vision-targeted subscale scores while excluding general health and the driving question, was 72±13. As compared to a reference group without eye diseases[17], STGD1 patients report a significantly lower visual function for almost all aspects (p<0.001, one sample t-test), except for general health (p=0.88, one sample t-test), colour vision (p=0.061, one sample t-test) and ocular pain (p=0.127, one sample t-test). The results of the FRI questionnaire are summarized in Table 6. FRI scores range from 1 to 4, with higher scores indicating greater independence. FRI level score is an ordinal classification using 4 functional levels: level 1, unable to do; level 2, help some or most of the time; level 3, moderately independent; level 4, totally independent[18].

Besides the patient reported FRI, we measured reading speed directly using Radner reading charts. An overview of the results is presented in Table 7. Unfortunately, the reading task was not performed correctly in 20 patients, excluding these results from the analysis. As there are age-related changes in baseline reading acuity and speed, we compared our cohort to a reference group without eye diseases and with the same mean age of 35 years[16]. Reading performance of our STGD1 cohort was significantly lower for all test values as compared to the reference group (p<0.001, one sample t-test).

## Discussion

The current paper describes the study design and baseline characteristics of the STARTT which evaluates the safety and efficacy of oral remofuscin in subjects with STGD1. In preclinical trials, remofuscin removed lipofuscin from RPE cells[5–7]. To measure the efficacy of remofuscin, $qAF_8$, a direct marker of lipofuscin levels in the retina[10], was chosen as primary outcome measure for the STARTT. The STARTT study design is the first to explore the utility of $qAF_8$ as clinical endpoint. $qAF_8$ is considered to be a biomarker intended to detect both STGD1 activity and the pharmacologic activity of remofuscin.

$qAF_8$ has high potential to become a valid and sensitive endpoint for clinical trials aiming to reduce or limit lipofuscin levels. Currently, atrophy area measured using FAF is the

**Table 5. Results of the National Eye Institute Visual Function Questionnaire (NEI VFQ-25) at baseline.**

| Subscale | STGD1 patients (n=87), mean (SD) | Reference group without eye diseases[1], mean (SD) | Pairwise comparisons*, test value; p-value |
|---|---|---|---|
| General health | 69 (20) | 69 (24) | t=-0.148 ; p=0.88 |
| General vision | 58 (16) | 83 (15) | t=-14.72; p<0.001 |
| Near vision | 59 (19) | 92 (13) | t=-16.035; p<0.001 |
| Distance vision | 65 (18) | 93 (11) | t=-14.549; p<0.001 |
| Peripheral vision | 81 (21) | 97 (10) | t=-7.258; p<0.001 |
| Colour vision | 95 (14) | 98 (8) | t=-1.897; p=0.061 |
| Ocular pain | 87 (16) | 90 (15) | t=-1.541; p=0.127 |
| Vision specific | | | |
| Role difficulties | 63 (28) | 93 (13) | t=-10.012; p<0.001 |
| Dependency | 75 (24) | 99 (6) | t=-9.528; p<0.001 |
| Social functioning | 75 (20) | 99 (3) | t=-10.924; p<0.001 |
| Mental health | 59 (21) | 99 (3) | t=-18.048; p<0.001 |
| Driving (n=43) | 47 (16) | 92 (12) | t=-18.171; p<0.001 |

SD, standard deviation; STGD, Stargardt disease.

*as assessed by one sample t-test

[1]Mangione CM, Lee PP, Gutierrez PR, Spritzer K, Berry S, Hays RD. Development of the 25-item National Eye Institute Visual Function Questionnaire. Arch Ophthalmol 2001;119(7):1050–8

**Table 6. FRI index scores at baseline.**

| FRI Index | STGD1 patients (n=86) |
|---|---|
| FRI item-level score, mean (SD) | |
| Read written print such as books, magazines, newspapers | 2.5 (1.0) |
| Read to pay bills or write a check | 2.8 (0.9) |
| Read in order to take medicine | 2.5 (1.1) |
| Read labels such as price tags, clothing labels, or food labels | 2.7 (0.9) |
| Read numbers on a phone, answering machine, or caller-ID | 3.3 (0.6) |
| Read words or numbers on screen while watching television | 2.6 (1.1) |
| Read when using a computer | 3.0 (0.6) |
| Overall FRI Index score, mean (SD)<br>FRI Level, n (%) | 2.8 (0.6) |
| FRI Level 1 (unable to do) | 1 (1.2%) |
| FRI Level 2 (help some or most of the time) | 19 (21.1%) |
| FRI Level 3 (moderately independent) | 51 (59.3%) |
| FRI Level 4 (totally independent) | 15 (17.4%) |

One patient did not complete the FRI Index questionnaire.

FRI, Functional Reading Independence; SD, standard deviation; STGD1, Stargardt disease.

*one sample t-test, #Overall Mann-Whitney U test

**Table 7. Results of Radner reading speed at baseline.**

| Binocular Radner Reading Chart Results | STGD1 patients (n=67), mean (SD | Reference group without eye diseases and age 35-39[1], mean (SD) | Pairwise comparisons*, test value; p-value |
|---|---|---|---|
| Both eyes | | | |
| Maximum reading speed in words/minute | 134 (50) | 235.71 (24.48) | t=-16.682; p<0.001 |
| Mean reading speed in words/minute | 108 (39) | 212.10 (21.92) | t=-21.750; p<0.001 |
| Critical print size in logRAD | 0.90 (0.24) | 0.085 (0.0745) | t=28.321; p<0.001 |
| Reading acuity in logRAD | 0.46 (0.27) | -0.125 (0.044) | t=17.325; p<0.001 |
| logRAD-score | 0.45 (0.28) | -0.1020 (0.0436) | t=16.170; p<0.001 |

LogRAD, logarithm of the reading acuity determination (reading equivalent of logMAR); SD, standard deviation; STGD1, Stargardt disease.

*as assessed by one sample t-test

[1]Radner W, Benesch T. Age-related changes in baseline reading acuity and speed as measured using RADNER Reading Charts in healthy eyes with best corrected ETDRS distance acuity. Br J Ophthalmol 2019;103(10):1518-23

most widely used endpoint in phase II/III clinical trials for STGD1[21]. However, RPE atrophy progression over time is not adequate for use in patients with early disease stages, i.e. prior to the development of RPE atrophy, and assesses the disease when cell death is already present and cannot be prevented anymore. By contrast, $qAF_8$ provides a quantifiable parameter for assessment of disease status, independent of disease stage, because qAF levels are elevated early in the disease course[10,12,22].

The use of $qAF_8$ as primary efficacy endpoint in late-stage clinical trials is only acceptable if $qAF_8$ is proven to reflect clinical benefit. According to the European Medicines Agency (EMA), a clinical endpoint should primarily evaluate how

the patient functions or feels[23]. The current trial design was therefore designed to be able to align qAF with functional endpoints including microperimetry, BCVA and reading speed. Nonetheless, as both qAF and visual function are highly variable. It might be hard to demonstrate a structure-function correlation in such a small sample size and future studies need to be undertaken. The longitudinal $qAF_8$ data from the STARTT will aid in further optimization of $qAF_8$ technique and can be used to evaluate the intra-visit, inter-visit repeatability of qAF in a multi-center setting. In this way, the STARTT was uniquely designed to aid in the validation of $qAF_8$ in order to make $qAF_8$ acceptable to support regulatory decisions in the approval of a new drug[24]. However, in order to truly establish the repeatability of qAF, a separate study should be set out.

The sample size was calculated based on the baseline $qAF_8$ score of 484±109 units (mean± standard deviation (SD) by the single center, single operator study by Burke *et al*. 2014[12]. In our multi-center study, actual baseline values were slightly more variable, namely 438 (210–729, 133) median (range, IQR). We therefore chose to elongate the trial period from 12 to 24 months, Opposed to the cohort of Burke *et al.* 2014[12], for the STARTT, we specifically selected patients with preserved retinal function because they are likely to show disease progression without effective treatment. We therefore included patients with a BCVA between 0.20 and 0.80, because better baseline BCVA was associated with a greater yearly rate of decline[25,26]. The qAF inclusion criterium of 300 units was chosen on the basis of the natural disease course were $qAF_8$ levels tend to initially increase, reach a ceiling level of approximately 800 units and then decline because of the development of RPE atrophy[11,12,15]. The cut-off point of 300 units is well below the ceiling level, allowing us to measure progression, and above the average level of normal eyes[11,12], allowing us to demonstrate a reduction of $qAF_8$ levels if the treatment works. The highest measured qAF value at baseline was 729 units, thus none of the patients are close to the ceiling level. As atrophy was present in only 42 out of 83 gradable eyes (51%), half of the cohort has an early disease stage where $qAF_8$ levels tend to increase. With these baseline values, we should be able to provide the first proof of effectiveness of remofuscin within 24 months if treatment works.

The STARTT represents the largest therapeutic trial to date to assess disease progression with functional measures, including reading speed assessment and patient-reported outcome measures (PRO's) regarding visual function and quality of life. These data are of utmost importance to characterize the burden of STGD1 and provide outcome measures and possible indicators of cost effectiveness for all future clinical trials.

Both reading speed and VA are considered to be strong determinants of quality of life in STGD1 patients[27]. Even though we included patients with preserved visual function, only 17.4% of patients claim to feel entirely independent as measured by the FRI index. However, a reference cohort without eye diseases does not exist yet for the FRI. The actual reading performance of our STGD1 cohort, as measured by Radner, was significantly lower for all test values as compared to a reference group[16] without eye diseases and with the same mean age of 35 years ($p<0.001$, one sample t-test).

With both reading speed and VA being lower than in a population without eye diseases, our patients report a considerable impact of STGD1 on daily function and quality of life as represented by the NEI VFQ-25. Even though general health was equally scored by STGD1 patients and the control group without eye diseases, STGD1 patients indicate significantly lower social functioning and mental health ($p<0,001$). This accords with earlier observations which showed that patients experience difficulties in the social environment (because of, among others, issues with facial recognition), tend to be frustrated and worried, have difficulties discussing their disease and even exhibit depressive symptoms[28–30]. A lower visual acuity, longer disease duration and younger age at onset resulted in a higher impact of STGD1 and a lower vision-specific quality of life[28,31]. The proposed treatment is targeted to remove lipofuscin which gives hope for recovery of RPE cell integrity, thereby restoring some visual function. Thus, if therapy proves to be effective, the quality of life of STGD1 patients might also improve. Because the STARTT includes PRO's, we can calculate the quality-adjusted life years (QALYs) of individuals which is an important outcome measure that not only reflects clinical benefit, but also aids in assessing the value of medical interventions during economic evaluation[32].

In conclusion, the STARTT design demonstrates a unique approach to explore the utility of $qAF_8$ as primary outcome measure while simultaneously evaluating the safety and efficacy of a new treatment agent for STGD1. The current trial design may therefore serve as reference for future clinical trials. For this, we reported a useful combination of multiple imaging methods, functional outcome measures, and PRO's and described the $qAF_8$ implementation in a multicentre setting. The results of reliability and repeatability of $qAF_8$, the longitudinal progression of STGD1 including the $qAF_8$ levels within this cohort and the safety and efficacy of oral remofuscin will be presented in subsequent publications. These publications will aid in detailed characterization about the natural course of STGD1, identify critical efficacy measures necessary to plan future clinical trials, and will of course demonstrate the safety and efficacy of oral remofuscin.

## Data availability

### Underlying data

DANS: The STArgardt Remofuscin Treatment Trial (STARTT): design and baseline characteristics of enrolled Stargardt patients. https://doi.org/10.17026/dans-x53-b696[20].

This project contains the following underlying data:

- STARTT Baseline for repository.dat
- STARTT Baseline for repository.csv
- Codebook STARTT Baseline.pdf

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

## References


1. Blacharski PA: **Fundus flavimaculaus**. Newsome DA, ed. *Retinal Dystrophies and Degenerations*. New York: Raven Press, 1988; 135–59.
2. Allikmets R, Singh N, Sun H, *et al.*: **A photoreceptor cell-specific ATP-binding transporter gene (*ABCR*) is mutated in recessive Stargardt macular dystrophy.** *Nat Genet.* 1997; **15**(3): 236–46.
PubMed Abstract | Publisher Full Text
3. Lenis TL, Hu J, Ng SY, *et al.*: **Expression of ABCA4 in the retinal pigment epithelium and its implications for Stargardt macular degeneration.** *Proc Natl Acad Sci U S A.* 2018; **115**(47): E11120–E11127.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Sparrow JR, Gregory-Roberts E, Yamamoto K, *et al.*: **The bisretinoids of retinal pigment epithelium.** *Prog Retin Eye Res.* 2012; **31**(2): 121–35.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Julien S, Schraermeyer U: **Lipofuscin can be eliminated from the retinal pigment epithelium of monkeys.** *Neurobiol Aging.* 2012; **33**(10): 2390–7.
PubMed Abstract | Publisher Full Text
6. Taubitz T, Peters T, Pöschel S, *et al.*: **Removal of lipofuscin from the RPE of Abca4$^{-/-}$ mice with THPE: quantitative and toxicity studies.** *Invest Ophthalmol Vis Sci.* 2015; **56**(7): 4199–99.
Reference Source
7. Julien-Schraermeyer S, Illing B, Tschulakow A, *et al.*: **Penetration, distribution, and elimination of remofuscin/soraprazan in Stargardt mouse eyes following a single intravitreal injection using pharmacokinetics and transmission electron microscopic autoradiography: Implication for the local treatment of Stargardt's disease and dry age-related macular degeneration.** *Pharmacol Res Perspect.* 2020; **8**(6): e00683.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Devi N, Singh D, Rawal RK, *et al.*: **Medicinal Attributes of Imidazo[1,2-a]pyridine Derivatives: An Update.** *Curr Top Med Chem.* 2016; **16**(26): 2963–94.
PubMed Abstract | Publisher Full Text
9. Björklund H: **Nycomed Annual Report 2008**.
Reference Source
10. Delori F, Greenberg JP, Woods RL, *et al.*: **Quantitative measurements of autofluorescence with the scanning laser ophthalmoscope.** *Invest Ophthalmol Vis Sci.* 2011; **52**(13): 9379–90.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Müller PL, Gliem M, McGuinnes M, *et al.*: **Quantitative Fundus Autofluorescence in *ABCA4*-related retinopathy -Functional Relevance and Genotype-Phenotype Correlation.** *Am J Ophthalmol.* 2020; **222**: 340–350.
PubMed Abstract | Publisher Full Text
12. Burke TR, Duncker T, Woods RL, *et al.*: **Quantitative fundus autofluorescence in recessive Stargardt disease.** *Invest Ophthalmol Vis Sci.* 2014; **55**(5): 2841–52.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Gliem M, Müller PL, Birtel J, *et al.*: **Quantitative fundus autofluorescence and genetic associations in macular, cone, and cone-rod dystrophies.** *Ophthalmol Retina.* 2020; **4**(7): 737–749.
PubMed Abstract | Publisher Full Text
14. Duncker T, Tsang SH, Lee W, *et al.*: **Quantitative Fundus Autofluorescence Distinguishes *ABCA4*-Associated and Non–*ABCA4*-Associated Bull's-Eye Maculopathy.** *Ophthalmology.* 2015; **122**(2): 345–55.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Duncker T, Tsang SH, Woods RL, *et al.*: **Quantitative fundus autofluorescence and optical coherence tomography in *PRPH2/RDS*- and *ABCA4*-associated disease exhibiting phenotypic overlap.** *Invest Ophthalmol Vis Sci.* 2015; **56**(5): 3159–70.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Radner W, Benesch T: **Age-related changes in baseline reading acuity and speed as measured using RADNER Reading Charts in healthy eyes with best corrected ETDRS distance acuity.** *Br J Ophthalmol.* 2019; **103**(10): 1518–23.
PubMed Abstract | Publisher Full Text
17. Mangione CM, Lee PP, Gutierrez PR, *et al.*: **Development of the 25-item National Eye Institute Visual Function Questionnaire.** *Arch Ophthalmol.* 2001; **119**(7): 1050–8.
PubMed Abstract | Publisher Full Text
18. Kimel M, Leidy NK, Tschosik E, *et al.*: **Functional Reading Independence (FRI) Index: A New Patient-Reported Outcome Measure for Patients With Geographic Atrophy.** *Invest Ophthalmol Vis Sci.* 2016; **57**(14): 6298–304.
PubMed Abstract | Publisher Full Text
19. Kuehlewein L, Hariri AH, Ho A, *et al.*: **Comparison of Manual and Semiautomated Fundus Autofluorescence Analysis of Macular Atrophy in Stargardt Disease Phenotype.** *Retina.* 2016; **36**(6): 1216–21.
PubMed Abstract | Publisher Full Text
20. Dhooge PPA, Hoyng CB: **The STArgardt Remofuscin Treatment Trial (STARTT): Design and Baseline Characteristics of Enrolled Stargardt Patients**. *DANS*. 2021.
http://www.doi.org/10.17026/dans-x53-b696
21. Csaky KG, Richman EA, Ferris FL 3rd: **Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium.** *Invest Ophthalmol Vis Sci.* 2008; **49**(2): 479–89.
PubMed Abstract | Publisher Full Text
22. Sparrow JR, Duncker T, Schuerch K, *et al.*: **Lessons learned from quantitative fundus autofluorescence.** *Prog Retin Eye Res.* 2020; **74**: 100774.
PubMed Abstract | Publisher Full Text | Free Full Text
23. European Medicines Agence: **EU Regulatory Workshop – Ophthalmology – Summary and Report**. 2012.
Reference Source
24. Wickström K, Moseley J: **Biomarkers and Surrogate Endpoints in Drug Development: A European Regulatory View.** *Invest Ophthalmol Vis Sci.* 2017; **58**(6): BIO27–BIO33.
PubMed Abstract | Publisher Full Text
25. Kong X, Strauss RW, Cideciyan AV, *et al.*: **Visual Acuity Change over 12 Months in the Prospective Progression of Atrophy Secondary to Stargardt Disease (ProgStar) Study: ProgStar Report Number 6.** *Ophthalmology.* 2017; **124**(11): 1640–51.
PubMed Abstract | Publisher Full Text
26. Kong X, Strauss RW, Michaelides M, *et al.*: **Visual Acuity Loss and Associated Risk Factors in the Retrospective Progression of Stargardt Disease Study (ProgStar Report No. 2).** *Ophthalmology.* 2016; **123**(9): 1887–97.
PubMed Abstract | Publisher Full Text
27. Murro V, Sodi A, Giacomelli G, *et al.*: **Reading Ability and Quality of Life in Stargardt Disease.** *Eur J Ophthalmol.* 2017; **27**(6): 740–45.
PubMed Abstract | Publisher Full Text
28. Miedziak AI, Perski T, Andrews PP, *et al.*: **Stargardt's macular dystrophy--a patient's perspective.** *Optometry.* 2000; **71**(3): 165–76.
PubMed Abstract
29. Roborel de Climens A, Tugaut B, Dias Barbosa C, *et al.*: **Living with Stargardt disease: insights from patients and their parents.** *Ophthalmic Genet.* 2021; **42**(2): 150–60.
PubMed Abstract | Publisher Full Text
30. Moschos MM, Nitoda E, Lavaris A: **Estimation of depression prevalence in patients with Stargardt disease using PHQ-9 and Zung scores.** *Eur J Ophthalmol.* 2016; **26**(3): 268–72.
PubMed Abstract | Publisher Full Text
31. Bax NM: **Stargardt disease: clinical, genetic and psychological characteristics**. Radboud University, 2019.
Reference Source
32. National Insititue for Health and Care Excellence: **Judging whether public health interventions offer value for money**. 2013.
Reference Source

# Open Peer Review

## Current Peer Review Status: ✔ ✔

**Version 3**

Reviewer Report 28 September 2022

https://doi.org/10.21956/openreseurope.16362.r30158

© **2022 Chong V.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Victor Chong**
Janssen R&D LLC, Raritan, NJ, USA

No further comments.

***Competing Interests:*** VC is an employee of Janssen R&D LLC. Janssen R&D LLC is in retinal drug development but has no direct competing interests. This review is performed by the reviewer outwith the employment of Janssen and not endorsed by Janssen.

***Reviewer Expertise:*** Clinical trial design and endpoint expert in retinal diseases

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Version 2**

Reviewer Report 22 August 2022

https://doi.org/10.21956/openreseurope.16203.r29743

© **2022 Chong V.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

? **Victor Chong**
[1] Janssen R&D LLC, Raritan, NJ, USA
[2] Janssen R&D LLC, Raritan, NJ, USA

I would like to thank the authors for answering my questions. It was very helpful. Most of the

amendments clarify the reasoning and I hope that would strengthen the paper.

My main reservation remains on qAF. I think we might be talking about something different here. Yes, anything can be used as a primary endpoint, but whether it would be a primary outcome measure for drug approval and whether this study would serve as a reference to achieve that position.

My concerns as follows:

A) qAF:

1. 1. The variability is high in this study. In previously published work, they were mainly in single centre using single machine possibly by single operator. If the study is used as a reference, it would even weaken the case for qAF as the primary outcome measure, as the authors had noted qAF is more variable when the study becomes a multi-centre study as in this study. As they now have the baseline data, would the authors consider a recalculation of their study size? I am worried that the study is no longer power adequately. I understand that amendment at this stage is highly unlikely. But why this is not discussed as a limitation?

2. Functional correlation: The authors cited Muller *et al* 2021. I quoted from the same paper as below.

   "When qAF, microperimetry, and ERG data were plotted together, a relationship unfolded in which high qAF was present in eyes with good retinal function. Although qAF was still in the upper range, retinal function deteriorated, and a significant reduction of qAF was consistently associated with loss of function"

   Muller and colleagues found that high qAF were in eyes with good retinal function. A non-scientific interpretation of the data would be that high qAF is good for you. I do understand that high qAF is probably associated as earlier stage of the disease. As RPE cells die, the qAF reduces, and hence worse function. I understand that. As this study aims to reduce qAF, and that is the dissociation. Hence, my former comment on that there is no association of qAF with function. No suggestion that reduce qAF without RPE cell death is associated with visual function. In general, regulators would probably not accept that as approvable primary endpoint (outcome measure). This study might provide evidence that a reduction of qAF is related to better / improve visulal function as mentioned in the discussion. Nonetheless, as both qAF and visual function are highly variable. It might be hard to demonstrate that in such a small sample size. Might be worth discussing that as a limitation. Nonetheless, I am looking forward to the resutls.

   My suggestion is to weaken the wording of the conclusion. If anything, this trial might make qAF less likely as an approvable endpoint. I would not have thought that was the intention of the authors.

B) FAF:

FAF lesion size was measured, and it is a potentially approvable endpoint. Why it is not included in the abstract and in the results as mentioned in my original comments. This looks odd? As a reviewer, something so obvious should be included and it is not included, I wondered why. In fact,

with this sample size, it is more likely that a reduction of qAF is linked with slowing down of FAF lesion growth. FAF lesion size measurement is much less variable than visual function tests. FAF lesion growth is associated with long term visual loss. Why the omission? Why would the authors want to use something more variable (qAF) than something much easier (FAF)?

***Competing Interests:*** The reviewer is an employee of Janssen R&D LLC. Janssen R&D LLC is in drug development for retinal diseases. This review is the personal opinion of the reviewer, outwith Janssen's employment and this review is not endorsed by Janssen R&D LLC.

***Reviewer Expertise:*** Clinical trial design and endpoint expert in retinal diseases

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 18 Sep 2022

**Patty Dhooge**

*Dear Dr. Chong, We understand your concerns and answer them below.*

A) qAF: 1. The variability is high in this study. In previously published work, they were mainly in single centre using single machine possibly by single operator. If the study is used as a reference, it would even weaken the case for qAF as the primary outcome measure, as the authors had noted qAF is more variable when the study becomes a multi-centre study as in this study. As they now have the baseline data, would the authors consider a recalculation of their study size? I am worried that the study is no longer power adequately. I understand that amendment at this stage is highly unlikely. But why this is not discussed as a limitation?
***Answer:*** *We have changed our discussion section accordingly: The sample size was calculated based on the baseline $qAF_8$ score of 484±109 units (mean± standard deviation (SD) by the single center, single operator study by Burke et al. 2014[12] . In our multi-center study, actual baseline values were slightly more variable, namely 438 (210–729, 133) median (range, IQR). We therefore chose to elongate the trial period from 12 to 24 months, Opposed to the cohort of Burke et al. 2014[12], for the STARTT, we specifically selected patients with preserved retinal function because they are likely to show disease progression without effective treatment. We therefore included patients with a BCVA between 0.20 and 0.80, because better baseline BCVA was associated with a greater yearly rate of decline [25, 26] . The qAF inclusion criterium of 300 units was chosen on the basis of the natural disease course were $qAF_8$ levels tend to initially increase, reach a ceiling level of approximately 800 units and then decline because of the development of RPE atrophy [11, 12, 15] . The cut-off point of 300 units is well below the ceiling level, allowing us to measure progression, and above the average level of normal eyes [11, 12] , allowing us to demonstrate a reduction of $qAF_8$ levels if the treatment works. The highest measured qAF value at baseline was 729 units, thus none of the patients are close to the ceiling level. As atrophy was present in only 42 out of 83 gradable eyes (51%), half of the cohort has an early disease stage where $qAF_8$ levels tend to increase. With these baseline values, we should be able to provide the first proof of effectiveness of remofuscin within 24 months if treatment works.*

Functional correlation: The authors cited Muller *et al* 2021. I quoted from the same paper as below. "When qAF, microperimetry, and ERG data were plotted together, a relationship unfolded in which high qAF was present in eyes with good retinal function. Although qAF was still in the upper range, retinal function deteriorated, and a significant reduction of qAF was consistently associated with loss of function" Muller and colleagues found that high qAF were in eyes with good retinal function. A non-scientific interpretation of the data would be that high qAF is good for you. I do understand that high qAF is probably associated as earlier stage of the disease. As RPE cells die, the qAF reduces, and hence worse function. I understand that. As this study aims to reduce qAF, and that is the dissociation. Hence, my former comment on that there is no association of qAF with function. No suggestion that reduce qAF without RPE cell death is associated with visual function. In general, regulators would probably not accept that as approvable primary endpoint (outcome measure). This study might provide evidence that a reduction of qAF is related to better / improve visulal function as mentioned in the discussion. Nonetheless, as both qAF and visual function are highly variable. It might be hard to demonstrate that in such a small sample size. Might be worth discussing that as a limitation. Nonetheless, I am looking forward to the resutls. My suggestion is to weaken the wording of the conclusion. If anything, this trial might make qAF less likely as an approvable endpoint. I would not have thought that was the intention of the authors.
***Answer:*** *We have changed our discussion and conclusion section*: *The current trial design was therefore designed to be able to align qAF with functional endpoints including microperimetry, BCVA and reading speed. Nonetheless, as both qAF and visual function are highly variable. It might be hard to demonstrate a structure-function correlation in such a small sample size and future studies need to be undertaken. The wording of conclusion has been weakened: In conclusion, the STARTT design demonstrates a unique approach to explore the utility of qAF* $_8$ *as primary outcome measure while simultaneously evaluating the safety and efficacy of a new treatment agent for STGD1.*

B) FAF: FAF lesion size was measured, and it is a potentially approvable endpoint. Why it is not included in the abstract and in the results as mentioned in my original comments. This looks odd? As a reviewer, something so obvious should be included and it is not included, I wondered why. In fact, with this sample size, it is more likely that a reduction of qAF is linked with slowing down of FAF lesion growth. FAF lesion size measurement is much less variable than visual function tests. FAF lesion growth is associated with long term visual loss. Why the omission? Why would the authors want to use something more variable (qAF) than something much easier (FAF)?
***Answer:*** *We did not include FAF lesion size, because most patients do not have definitely decreased fundus autofluorescence on FAF images at baseline as this was not part of the inclusion criteria. If atrophy (ddaf) is present, we will measure it, but we cannot track it in all patients and therefore, FAF lesion aria is not in the endpoints.*

***Competing Interests:*** No competing interests were disclosed.

**Version 1**

Reviewer Report 26 January 2022

https://doi.org/10.21956/openreseurope.14956.r28352

**© 2022 Chong V.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Victor Chong**
[1] Janssen R&D LLC, Raritan, NJ, USA
[2] Janssen R&D LLC, Raritan, NJ, USA
[3] Janssen R&D LLC, Raritan, NJ, USA

The authors presented the STARTT trial design and reported the baseline characteristics. Overall, it was very well written.

My main concerns are as follows:

1. The main conclusion was that this trial design may serve as a reference for future clinical trials. I accept it MAY, but there is so far no evidence that qAF is correlated with visual function nor improvement of qAF is beneficial to the patient. In fact, qAF is not validated as a clinical endpoint in terms of reproducibility and clinical significance, either as a predictor of prognosis or visual function. In general, the primary endpoint of a trial should be a validated one. Hence, it is a bad practice. Nonetheless, I do agree that the information collected in this study might be able to validate qAF for future trials.

2. I am surprised that on one hand FAF was performed, but FAF area of atrophy changes were not included as an endpoint, as that is the usual acceptable clinical endpoint for Stargardt disease by regulatory agencies such as FDA. I understand there were patients included in the study with no atrophy at baseline. However, why the changes of the area of atrophy is not an endpoint?

3. Considering, 5 out of 87 patients (6%), the qAF dropped below the inclusion criteria of 300 units and based on the range, at least one patient dropped to 210 units. The hypothesis is that the reduction of qAF is 50 units. So at least with no treatment, at least one patient can drop close to 100 unit between screening and baseline. Furthermore, it was not reported how many other patients had changed qAF between screening and baseline. Once again highlighted the concern of using qAF as primary endpoint.

4. The choice of dosage is based on the study of the same drug in GERD. I assume that was failed. So is there any evidence other than it is safe that this dose has acceptable PK/PD to the retina / RPE. At least, some comments or some more information about that would be useful. The cited references (8 and 9), I was not able to access that to look for information on dose selection. For oral compounds, incorrect dosage is a common reason that a drug fails to show benefit, and since this dose is already failed in GERD, this should be looked into in more details. The authors might have done and it would be good if they can share that information in the article.

Minor suggestions:

1. Letter counting is more commonly used for BCVA. And if decimal is used, it might be useful to know how part of the line read is included. Also it might be helpful to have Snellen equivalent as decimal BCVA is not commonly used in most parts of the world and often confused with LogMar BCVA.

2. Is there a particular reason to use a 54 custom-made test pattern for Stargardt? Had that been validated?

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and does the work have academic merit?**
Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**
Partly

**Are the conclusions drawn adequately supported by the results?**
No

***Competing Interests:*** I am an employee of Janssen R&D LLC. Although Janssen does not have a molecule in development for Stargardt disease, it might have in the future and might be a potential conflict of interest. The review is the opinion of the reviewer alone and is not endorsed by Janssen.

***Reviewer Expertise:*** Clinical trial design and endpoint expert in retinal diseases

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 19 Jul 2022

**Patty Dhooge**

Dear dr. Chong, Thank you for reviewing our manuscript entitled “The STArgardt Remofuscin Treatment Trial (STARTT): design and baseline characteristics of enrolled Stargardt patients” We feel strengthened by the positive comments and useful suggestions we have received. Please find enclosed our detailed point-by-point response to your review.

- ***The main conclusion was that this trial design may serve as a reference for future***

***clinical trials. I accept it MAY, but there is so far no evidence that qAF is correlated with visual function nor improvement of qAF is beneficial to the patient. In fact, qAF is not validated as a clinical endpoint in terms of reproducibility and clinical significance, either as a predictor of prognosis or visual function. In general, the primary endpoint of a trial should be a validated one. Hence, it is a bad practice. Nonetheless, I do agree that the information collected in this study might be able to validate qAF for future trials.***

The EMA has confirmed during protocol advice in 2014 (EMEA/H/SA/2776/1/2014/PA/SME/III), that qAF analysis can be accepted as a primary endpoint in clinical studies. This has also been confirmed at a meeting with the Medicinal Product Agency, Sweden. Multiple studies report repeatability coefficients of 6-12% for intra-visit repeatability 7-14% for inter-visit repeatability (interval up to 64 days).(Delori, Greenberg et al. 2011, Greenberg, Duncker et al. 2013, Burke, Duncker et al. 2014, Duncker, Greenberg et al. 2014, Duncker, Tsang et al. 2015, Duncker, Tsang et al. 2015, Muller, Gliem et al. 2015, Reiter, Told et al. 2019) Measures of qAF were correlated with retinal function assessed by full-field electroretinography (ERG) and fundus-controlled perimetry.(Müller, Gliem et al. 2021)

We believe that the use of qAF as endpoint in this proof-of-concept study qAF is justified because of its special property that it is an early, direct, and objective marker of lipofuscin levels in the retina. Of course, we will also assess the secondary endpoints in order to analyze the efficacy of Remofuscin.

- ***I am surprised that on one hand FAF was performed, but FAF area of atrophy changes were not included as an endpoint, as that is the usual acceptable clinical endpoint for Stargardt disease by regulatory agencies such as FDA. I understand there were patients included in the study with no atrophy at baseline. However, why the changes of the area of atrophy is not an endpoint?***

Macular atrophy as assessed by fundus autofluorescence (FAF) was included as clinical trial endpoint number 9. See paragraph ‘visit procedures’ and ‘retinal imaging’: “From the FAF images, if present, the macular atrophy, interpreted as complete loss of autofluorescence (analogous to definite decreased autofluorescence as described by Kuehlewein et al. 19 ), was calculated using the Heidelberg RegionFinder software version 2.6.4.0.”

- ***Considering, 5 out of 87 patients (6%), the qAF dropped below the inclusion criteria of 300 units and based on the range, at least one patient dropped to 210 units. The hypothesis is that the reduction of qAF is 50 units. So at least with no treatment, at least one patient can drop close to 100 unit between screening and baseline. Furthermore, it was not reported how many other patients had changed qAF between screening and baseline. Once again highlighted the concern of using qAF as primary endpoint.***

Indeed, the real-life test-retest variability of qAF seems higher than assumed. We will take this increased variability into account when analyzing the data.

- ***The choice of dosage is based on the study of the same drug in GERD. I assume that was failed. So is there any evidence other than it is safe that this dose has acceptable PK/PD to the retina / RPE. At least, some comments or some more information about that would be useful. The cited references (8 and 9), I was not able to access that to look for information on dose selection. For oral compounds, incorrect dosage is a common reason that a drug fails to show benefit, and since this***

***dose is already failed in GERD, this should be looked into in more details. The authors might have done and it would be good if they can share that information in the article.***

An oral dose of 6 mg/kg/day administered to monkeys for 1-year resulted in a full removal of lipofuscin without any ocular adverse effects. This dose in monkeys is approximately 3-fold higher than humans dosed at 20 mg/day. As a full effect on monkey lipofuscin removal was observed at 6 mg/kg/day after 12 months, it is assumed that a daily dose of 20 mg/day will also result in lipofuscin reduction in Stargardt disease patients. We will add this information to our manuscript.

**Minor suggestions:**

- ***Letter counting is more commonly used for BCVA. And if decimal is used, it might be useful to know how part of the line read is included. Also it might be helpful to have Snellen equivalent as decimal BCVA is not commonly used in most parts of the world and often confused with LogMar BCVA.***

We will add Snellen equivalents to the manuscript.

1. ***Is there a particular reason to use a 54 custom-made test pattern for Stargardt? Had that been validated?***

We use a custom-made test to ensure that microperimetry testing areas fall within the grid as described by Delori. In this way, we can correlate the qAF changes to the microperimetry changes in selected areas. We will add this information to the manuscript.

***Competing Interests:*** No competing interests were disclosed.

Reviewer Report 15 December 2021

https://doi.org/10.21956/openreseurope.14956.r28043

**© 2021 Kong X.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Xiangrong Kong**

[1] School of Medicine, Johns Hopkins University, Baltimore, MD, USA
[2] School of Medicine, Johns Hopkins University, Baltimore, MD, USA
[3] School of Medicine, Johns Hopkins University, Baltimore, MD, USA

This manuscript reports the study design and baseline characteristics of the STARTT study, a randomized placebo controlled parallel clinical trial on Stargardt disease. There were 6 clinical sites from Europe. Participants were randomized to the intervention 20mg Soraprazan per day or placebo with a ratio of 2:1, and are followed monthly for 24 months. Strengths of the study and manuscript include the high significance given the unmet need for Stargardt disease management, pre-clinical studies supporting the trial to test repurposing of Soraprazan to Stargardt disease, the known safety profile and potential low cost of Soraprazan, many outcome modalities including 2 instruments measuring patient reported outcomes and multiple visual

function outcomes such as low illuminance acuity and reading speed, and the clear description of the study design in the manuscript. The primary outcome measure was based on the change of qAF level between baseline and 24 months. This is novel and compared to the FAF atrophy area, qAF can be used for earlier stage patients where sizable FAF lesions have yet to form. In addition to evaluating the safety and efficacy of the drug, the longitudinal data collected through the trial will provide natural history information on disease progression measures such as qAF, low luminance VA, reading speed and patient reported outcomes, which in and of itself will contribute to the better understanding of the disease progression as there is very limited literature currently on the longitudinal changes of these measures for Stargardt.

A few comments are listed below.

- Page 4, DSMB preforms reviews of the blinded data regularly. There are two considerations here, first, if the DSMB are all masked from the treatment information in the data, there is the question of how to timely identify safety signals. The drug's safety profile is relatively well known from other conditions, but this trial requires long-term use of the drug. Long-term safety needs to be tracked timely during the trial. It is unclear how effective safety tracking can be done if the DSMB is masked. Second, considering the randomization ratio of 2:1, not equal allocation, it is easy to surmise the group with larger sample size would be the intervention. More details would be helpful in the manuscript regarding how the data are presented to DSMB if they need to be blinded.

- Page 5 about eligibility criteria and study eye selection were described clearly. One question is why only use one study eye from a patient. The intervention tested is a systemic treatment. Also as an inherited disease, it is most often bilateral presentation and it seems the study is measuring both eyes on most imaging/tests. Using data from both eyes if both are eligible will provide more information and will be more statistically powerful. It is unclear why the study design would only consider one study eye for primary efficacy analysis.

- The study flow-chart is very nicely presented. Participants are followed monthly for 2 years. Such follow-up frequency is intensive for patients especially these are patients with visual impairment. Most of the testing/imaging are also done monthly. This can be burdensome for patients. For FAF imaging, there is the theoretical concern that the intense short-wavelength (SW) light excitation during imaging may be detrimental to Stargardt eyes (Cideciyan A.V., *et al.* 2007 *Opt Soc Am A Opt Image Sci Vis.*) and reduced illuminance SW imaging protocol has been used in Stargardt imaging. Was this considered in the imaging protocol for this trial? It would be helpful to mention the justification of monthly visits and imaging. Also, patients of Stargardt must be highly motivated, but missed visits can occur especially with such a frequent schedule. Are there measures to help prevent missed visits?

- The baseline data presented are very informative. They suggest the Radner reading test is not a good test for Stargardt patients, as a nontrivial proportion of participants could not preform it correctly. The comparison of NEI-VFQ response to a reference group is interesting and well thought of. The significantly lower scores on all scales except general health and ocular pain are informative and expected.

- Discussion Page 12 first paragraph on the left, "longitudinal $qAF_8$ data from the STARTT will aid in further optimization of $qAF_8$ technique and can be used to test the reliability and

repeatability of $qAF_{8.}$ " . Two comments can be made here. First, normally when designing a trial, the primary outcome measure chosen should be a measure that's already been shown to be able to be measured reliably. A reliable outcome measure can minimize bias and increase power for detecting treatment efficacy. Second, assessments of the reliability and repeatability of qAF measure require its own study, including repeated gradings of the same qAF image for image grading repeatability and multiple images for the same eye within a short period for imaging technique reliability. The longitudinal data as currently described in the manuscript cannot be used for assessing reliability and repeatability of the measure as there would be true biological change in qAF that is confounded with the measurement's own variability due to image grading and imaging process.

- Discussion Page 12 second paragraph on the left about the natural disease trajectory of qAF is very interesting and informative. So the qAF change in Stargardt is not monotone over time. With the trial sample including half participants who already had FAF atrophy at enrollment, where their qAF changes are expected to fall on the natural disease trajectory during the 2 years of follow-up? If their qAF increase to the plateau and then decrease during the 2 years, how to use the value of the change of qAF between baseline and 24 months visits to infer disease course during the 2 years?

- Discussion Page 12 first paragraph on the right, regarding "Thus, if therapy proves to be effective and visual acuity and reading ability improve, the quality of life of the STGD1 patients would also improve", there are two comments. First, this suggests that the proposed treatment is targeted to improve visual function. It is questionable whether such restoring visual function target can be achieved with the presented mechanism of how Soraprazan may work in Stargardt (removing lipofuscin in RPE). It's worth to note that the therapeutic target of delaying disease progression is not the same as the target of reversing visual function loss or improving visual function. If Soraprazan indeed can remove lipofuscin in RPE, it may delay disease progression, but it is unclear how it may reverse the RPE and photoreceptor loss that has already occurred prior to treatment initiation. Second, the baseline quality of life (QOL) of Stargardt patients is worse compared to normal controls. But to evaluate whether a treatment can improve QoL in the patients, it is important to know whether and how QoL as measured by NEI-VFQ changes over a 2-years period.

**References**

1. Cideciyan AV, Swider M, Aleman TS, Roman MI, et al.: Reduced-illuminance autofluorescence imaging in ABCA4-associated retinal degenerations.*J Opt Soc Am A Opt Image Sci Vis*. 2007; **24** (5): 1457-67 PubMed Abstract | Publisher Full Text

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and does the work have academic merit?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Partly

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Clinical trial design, statistics, Stargardt disease natural history

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 19 Jul 2022

**Patty Dhooge**

Dear dr. Kong, Thank you for reviewing our manuscript entitled "The STArgardt Remofuscin Treatment Trial (STARTT): design and baseline characteristics of enrolled Stargardt patients" We feel strengthened by the positive comments and useful suggestions we have received. Please find enclosed our detailed point-by-point response to your review.

- ***Page 4, DSMB performs reviews of the blinded data regularly. There are two considerations here, first, if the DSMB are all masked from the treatment information in the data, there is the question of how to timely identify safety signals. The drug's safety profile is relatively well known from other conditions, but this trial requires long-term use of the drug. Long-term safety needs to be tracked timely during the trial. It is unclear how effective safety tracking can be done if the DSMB is masked. Second, considering the randomization ratio of 2:1, not equal allocation, it is easy to surmise the group with larger sample size would be the intervention. More details would be helpful in the manuscript regarding how the data are presented to DSMB if they need to be blinded.***

The DSMB receives a sub-set of data relevant for safety assessment. The assessment may include but is not limited to: adverse events, safety laboratory tests (haematology and clinical chemistry), qAF8 level, Lens Opacities Classification System grade. The DSMB is blinded to individual subject treatment assignments for the duration of the entire study, but may be un-blinded in order to facilitate prompt analysis of any safety/tolerance issues that may arise. Also, pharmacovigilance staff may un-blind single cases in relation to SUSAR reporting.

- ***Page 5 about eligibility criteria and study eye selection were described clearly. One question is why only use one study eye from a patient. The intervention tested is a systemic treatment. Also as an inherited disease, it is most often bilateral presentation and it seems the study is measuring both eyes on most imaging/tests. Using data from both eyes if both are eligible will provide more information and will be more statistically powerful. It is unclear why the study design would only consider one study eye for primary efficacy analysis.***

Initially, we will look at data from one study eye as standard statistical methods can be employed when analyses are based on a single eye per individual. If needed, second eye data can be included to confirm results.

- **The study flow-chart is very nicely presented. Participants are followed monthly for 2 years. Such follow-up frequency is intensive for patients especially these are patients with visual impairment. Most of the testing/imaging are also done monthly. This can be burdensome for patients. For FAF imaging, there is the theoretical concern that the intense short-wavelength (SW) light excitation during imaging may be detrimental to Stargardt eyes (Cideciyan A.V., et al. 2007 Opt Soc Am A Opt Image Sci Vis.) and reduced illuminance SW imaging protocol has been used in Stargardt imaging. Was this considered in the imaging protocol for this trial? It would be helpful to mention the justification of monthly visits and imaging. Also, patients of Stargardt must be highly motivated, but missed visits can occur especially with such a frequent schedule. Are there measures to help prevent missed visits?**

Indeed, reduced illuminance SW imaging would be an option if the primary goal is to slow down the progression of RPE atrophy as reduced illuminance SW imaging can well be used to measure areas of definitely decreased autofluorescence (i.e. RPE atrophy). However, STARTT focusses on hyper autofluorescent signals, since the mechanism of action of Remofuscin concerns the removal of lipofuscin. Due to its autofluorescence, the amount of lipofuscin in the RPE cells – and therefore also any signs of removal – can be directly analyzed by quantitative autofluorescence (qAF). qAF is not possible with reduced illuminance SW-AF imaging. Patients are therefore already exposed to more intense light during SW-AF imaging. Making additional FAF images requires minimal extra time. Besides, the maximal retinal irradiance during SW-AF and qAF imaging is approximately 2 mW/cm2 for a 10 degrees by 10 degrees image and is therefore well below international standards.(Bindewald, Schmitz-Valckenberg et al. 2005) However, those thresholds are mainly based on unaffected retinas, while diseased retinas may have lower light-damage thresholds. Indeed, simulations suggest that the RPE of STGD1 patients is generally at increased risk of photo-oxidative stress and that exposure during FAF imaging will amplify this risk.(Teussink, Lambertus et al. 2017) However, this theoretical phototoxicity has never been verified with in-vivo data. Currently, participating in a clinical trial is the only option for STGD1 patients to receive treatment. Therefore, patients are highly motivated. In order to prevent missed visits, planning of visits is discussed and coordinated together with patients, and we offer patients reimbursement of all costs.

- ***The baseline data presented are very informative. They suggest the Radner reading test is not a good test for Stargardt patients, as a nontrivial proportion of participants could not preform it correctly. The comparison of NEI-VFQ response to a reference group is interesting and well thought of. The significantly lower scores on all scales except general health and ocular pain are informative and expected.***

Thank you for your compliment.

- ***Discussion Page 12 first paragraph on the left, "longitudinal qAF8 data from the STARTT will aid in further optimization of qAF8 technique and can be used to test the reliability and repeatability of qAF8. " . Two comments can be made here. First, normally when designing a trial, the primary outcome measure chosen should be a measure that's already been shown to be able to be measured reliably. A reliable outcome measure can minimize bias and increase power for detecting treatment***

***efficacy. Second, assessments of the reliability and repeatability of qAF measure require its own study, including repeated gradings of the same qAF image for image grading repeatability and multiple images for the same eye within a short period for imaging technique reliability. The longitudinal data as currently described in the manuscript cannot be used for assessing reliability and repeatability of the measure as there would be true biological change in qAF that is confounded with the measurement's own variability due to image grading and imaging process.***

The EMA has confirmed during protocol advice in 2014 (EMEA/H/SA/2776/1/2014/PA/SME/III), that qAF analysis can be accepted as a primary endpoint in clinical studies. This has also been confirmed at a meeting with the Medicinal Product Agency, Sweden. Indeed, reliability of qAF has not yet been established in STGD1. Multiple studies report repeatability coefficients of 6-12% for intra-visit repeatability 7-14% for inter-visit repeatability (interval up to 64 days).(Delori, Greenberg et al. 2011, Greenberg, Duncker et al. 2013, Burke, Duncker et al. 2014, Duncker, Greenberg et al. 2014, Duncker, Tsang et al. 2015, Duncker, Tsang et al. 2015, Muller, Gliem et al. 2015, Reiter, Told et al. 2019) Indeed, to establish repeatability of qAF, a separate study should be set out. However, before the initiation of treatment, all patients satisfying the other enrollment criteria of this trial underwent qAF imaging at two visits, with two qAF imaging sets per visit. The interval between both visits was between 6-72 days likely providing a small biological change in qAF during this period. Thus, by using these individual measurements, we are able to evaluate the intra-visit, inter-visit repeatability of qAF in a multi-center setting. We agree that the longitudinal data as currently described in the manuscript cannot be used for assessing reliability and repeatability and we will correct the wording.

- ***Discussion Page 12 second paragraph on the left about the natural disease trajectory of qAF is very interesting and informative. So the qAF change in Stargardt is not monotone over time. With the trial sample including half participants who already had FAF atrophy at enrollment, where their qAF changes are expected to fall on the natural disease trajectory during the 2 years of follow-up? If their qAF increase to the plateau and then decrease during the 2 years, how to use the value of the change of qAF between baseline and 24 months visits to infer disease course during the 2 years?***

Before analysis of the qAF level, atrophic areas were excluded by an individual segmentation based on the qAF add-on tool in the HEYEX software (Heidelberg Engineering, Heidelberg, Germany), and with manual alterations if deemed necessary. We believe that the cut-off point of 300 units is well below the ceiling level, allowing us to measure progression, and allowing us to demonstrate a reduction of $qAF_8$ levels if the treatment works. In addition, we can use the microperimetry analysis to differentiate. If treatment works, retinal sensitivity will remain the same or improve when qAF levels decrease. If treatment does not work, retinal sensitivity will decline when qAF levels decline.

- ***Discussion Page 12 first paragraph on the right, regarding "Thus, if therapy proves to be effective and visual acuity and reading ability improve, the quality of life of the STGD1 patients would also improve", there are two comments. First, this suggests that the proposed treatment is targeted to improve visual function. It is questionable whether such restoring visual function target can be achieved with the presented mechanism of how Soraprazan may work in Stargardt (removing lipofuscin in RPE). It's worth to note that the therapeutic target of delaying disease progression is not the same as the target of reversing visual function loss or***

***improving visual function. If Soraprazan indeed can remove lipofuscin in RPE, it may delay disease progression, but it is unclear how it may reverse the RPE and photoreceptor loss that has already occurred prior to treatment initiation. Second, the baseline quality of life (QOL) of Stargardt patients is worse compared to normal controls. But to evaluate whether a treatment can improve QoL in the patients, it is important to know whether and how QoL as measured by NEI-VFQ changes over a 2-years period.***

Indeed, proposed treatment is targeted to remove lipofuscin which gives hope for recovery of RPE cell integrity, thereby restoring some visual function. However, RPE and photoreceptor loss cannot be reversed. We will change the wording. To the best of our knowledge, studies on NEI-VFQ changes over a 2-year period in STGD1 or IRDs have not been carried out so far. STARTT is a placebo-controlled study. Therefore we can compare NEI-VFQ changes between both arms in order to evaluate if treatment can improve the quality of life in STGD1 patients.

***Competing Interests:*** No competing interests were disclosed.