STATISTICAL ANALYSIS PLAN

Final

A randomised clinical trial
of a new binocular treatment for amblyopia

NIHI

STATISTICAL ANALYSIS PLAN APPROVAL SHEET

Study: Binocular treatment for amblyopia using videogame (BRAVO)

Title: A randomised clinical trial of a new binocular treatment for amblyopia

Authors: Yannan Jiang and Varsha Parag

Version: Final

Version date: 8th December 2016

Who else was involved in the discussion of the analysis:

Benjamin Thompson, Cindy Guo, Joanna Black, Tina Gao

The undersigned have reviewed this document and find it to be consistent with the requirements of the protocol as it applies to their respective areas. The author/reviewer also finds this document to be in compliance with ICH-E9 as well as NIHI SOP BS-0004.

_____   8/12/2016
Varsha Parag   Date
Study Statistician

_____   8/12/2016.
Benjamin Thompson   Date
Principal Investigator

Preface

The purpose of this Statistical Analysis Plan (SAP) is to provide a detailed statement of the intended statistical analyses that will be performed in the analysis of data from the BRAVO study. This document is intended to be stand-alone from the protocol and adhere to the main points in the analysis summary specified in the protocol. However it is envisaged that the SAP can undergo revision outside of the protocol. It is not anticipated that revisions to the SAP that are in the spirit of the specified protocol analysis would require review by an ethics committee.

It is assumed that the study documentation provided by Data Management follows standard operating procedures. These procedures describe the process for setting up and maintaining study documentation, recording decisions affecting data handling, as well as methods of data capture and the algorithms to ensure accurate data are collected and maintained. It is assumed that the SAP will use data from a locked database that has been verified against the clinical record and is a true record of the data collected from the participant. Any data that are missing will be flagged and assumed to be undiscoverable.

The SAP will also outline the proposed layout of tables/figures that will be presented.

Scope

Please note that the scope of this SAP is intended to cover ONLY those main analyses described in the protocol. For additional research questions not detailed within this SAP and/or questions requiring further exploratory analyses please refer to other separate SAPs.

To keep within timeframe of issue of Biostatistics report, input into required analysis needs to be pre-specified in this analysis plan. Requests for additional analyses must be minimal after sign-off of this document. This enables all pre-programming programs to be performed ahead of database lock.

TABLE OF CONTENTS

1. INTRODUCTION

Amblyopia is a common neurodevelopmental disorder of vision that is characterised by visual impairment in one eye and compromised binocular visual function. Existing evidence-based treatments for children include patching the non-amblyopic eye to encourage use of the amblyopic eye. Currently there are no widely accepted treatments available for adults with amblyopia. The aim of this trial is to assess the efficacy of a new binocular, videogame based treatment for amblyopia in older children and adults.

2. STUDY OBJECTIVES

The primary aim of this trial is to investigate whether the active treatment leads to an improvement in visual acuity of the amblyopic eye by comparing logMAR visual acuity scores obtained six weeks after randomization to baseline scores.

This trial will also assess whether the active treatment improves stereopsis and reduces interocular suppression. Furthermore, compliance and participant acceptability of this home-based treatment and any improvement in quality of life will also be addressed.

3. STUDY DESIGN

Please refer to the study protocol for a full description of the study design [1]. BRAVO is a two-arm, double-blind, placebo-controlled, randomised clinical trial to assess the effectiveness of a novel video-game based binocular treatment for amblyopia. Participants will be randomised to receive 6 weeks of active or placebo home-based binocular treatment.

A total of 108 participants will be recruited from Auckland, New Zealand; Melbourne, Australia; Montreal and Waterloo, Canada and Hong Kong. Three age groups are targeted: 7-12 years, 13-17 years and >17 years.

Participants will be assessed at baseline, and at 3, 6, 12 and 24 weeks post-randomisation.

3.1. Inclusion criteria

- ≥7 years of age.
- Amblyopia associated with the presence or history of strabismus, anisometropia or both (mixed mechanism).
- Unilateral amblyopia, defined as best-corrected amblyopic eye visual acuity (VA) of 0.30-1.00 logMAR inclusive, fellow eye VA ≤0.10 logMAR and an interocular VA difference ≥ 0.20 log units. VA will be measured using the electronic Early Treatment for Diabetic Retinopathy Study (E-ETDRS) protocol presented on an Electronic Visual Acuity (EVA) testing system.
- Strabismic amblyopia, defined as amblyopia in the presence of heterotropia at distance and/or near fixation, a history of strabismus surgery, or resolution of strabismus following hyperopic spectacle correction.
- Anisometropic amblyopia, defined as amblyopia in the presence of a spherical equivalent difference ≥ 0.50D between the eyes, or a difference of astigmatism in any meridian ≥ 1.50D and no strabismus.
- Mixed mechanism amblyopia, defined as amblyopia in the presence of both strabismus and anisometropia.

- An ability to align a dichoptically presented nonius cross within the screen area of an iPod touch device. This criterion ensures that successful play of the treatment game is possible.
- Willing and able to provide written informed consent for participation in the study.
- Stable visual acuity with full optical correction.

3.2. Exclusion criteria

- Myopia of spherical equivalent > -6.00D in either eye with spectacle correction
- Alternating strabismus under normal binocular viewing conditions at either near or distance
- Presence of amblyopia that is not due to strabismus and/or anisometropia, eg deprivational amblyopia
- Previous intraocular surgery
- Any co-existent ocular pathology
- Any known neurological anomalies.

3.3. Study intervention

For this trial, participants will be required to wear appropriate vision correction for both eyes using spectacles or contact lenses before and during treatment. Participants who have had full optical correction for more than 16 weeks will be randomized immediately after baseline examination. Participants who have not worn optical correction full time for 16 weeks will need a refractive adaptation period after baseline examination and before randomisation, as there can be some visual improvement in the amblyopic eye from full correction of refractive error alone, particularly in children. If new spectacles or contact lenses are required according to refractive tests at baseline, these will be provided by BRAVO. This procedure of refractive adaptation will exclude the possibility that the final visual acuity change is partially or purely due to adaptation to new refractive correction. Participants will be randomized if visual acuity in the amblyopic eye is stable over two measurements made at least four weeks apart during this refractive adaptation period. Participants will not be eligible if their visual acuity remains unstable over 16 weeks of refractive adaptation or if their amblyopic eye visual acuity improves to the point that they no longer meet the eligibility criteria.

Eligible participants will be randomized to receive *six weeks* home-based use of the iPod-based amblyopia treatment or an iPod-based placebo that will run identical software to the active treatment device, except that both eyes will see the same images with no contrast offset. All participants will be instructed to play the game on their allocated device for 1-2 hours every day while wearing their appropriate refractive correction and anaglyphic glasses, and this will be done by a study member not involved in data collection. For the active treatment game, Tetris blocks will be presented to the amblyopic eye at 100% contrast and to the non-amblyopic eye at a contrast determined by a random dot kinematogram measure of suppression and an assessment of the participant's ability to play the game by an unblinded assessor. The contrast presented to the fellow eye will be incremented using an automatic algorithm based on successful game play. If the trial shows that the treatment is effective, all members of the control group will be provided with six weeks of treatment free of charge at the end of the study.

3.4. Randomisation

Eligible participants will be allocated randomly in a 1:1 ratio by computer to either active or placebo treatment group, using minimisation stratified by three age groups (7-12 years, 13-17 years and >17 years).

3.5. Sample size

Recruiting 36 participants (18 per arm) in each of the three age groups (7-12 years, 13-17 years and >17 years), will provide 90% power at 5% level of significance (two-sided) to detect a minimal clinically important difference of 0.2 LogMAR for change in distance visual acuity at six weeks between the two treatment arms, assuming a standard deviation of 0.17 LogMAR[2] and 10% loss to follow-up rate. The study will therefore aim to recruit a total of 108 patients (54 per arm).

This sample size also allows for the consistency of effects for pre-specified subgroups, namely severity of amblyopia, and presence or absence of strabismus, to be assessed.

4. STUDY OUTCOMES

4.1. Primary outcome

The primary outcome measure is the change in distance visual acuity in the amblyopic eye, from baseline to six weeks post-randomisation.

4.2. Secondary outcomes

The secondary outcomes are (post-randomisation):

- Change from baseline in distance visual acuity of the amblyopic eye, fellow eye and both eyes at 3, 6, 12 and 24 weeks.

- Change from baseline in near visual acuity of the amblyopic eye, fellow eye and both eyes at 3, 6, 12 and 24 weeks.

- Change from baseline in stereopsis at 3, 6, 12 and 24 weeks measured using Randot Preschool Stereotest, Titmus Stereo Test and iPod-based Stereo test.

- Change from baseline in angle of strabismus (where applicable), measured using the simultaneous prism cover tests at 3, 6, 12 and 24 weeks.

- Change from baseline in interocular suppression at 3, 6, 12 and 24 weeks measured using an iPod-based version of the dichoptic motion coherence test and near and distance Worth 4-dot test.

- Change from baseline in interocular contrast within the Tetris game at 3 and 6 weeks based on the log file extracted from the iPod.

- Treatment compliance at 3 and 6 weeks based on the log file extracted from the iPod (or a participant diary if the log file is unavailable).

- Treatment acceptability at 3 and 6 weeks measured using a modified version of the Amblyopia Treatment Index questionnaire for children and adults.

- Change from baseline in quality of life, measured in adult participants (>17 years of age) at 24 weeks using the World Health Organization Quality of Life.

- Safety: Information regarding adverse events, device events and whether they are related to the active treatment will be collected throughout the trial period.

5. DATA SOURCES

Data collected from the following BRAVO case report forms and questionnaires which will be extracted from the NIHI Oracle database into SAS for the analyses:

- Screening logs
- Form A: Registration and baseline
- Form C: Follow-up (3, 6, 12 and 24 weeks)
- Form D: Device record (baseline, 3, 6 weeks; any device replacement)
- Form M: Concomitant medications (baseline; any change in medications)
- Form Q: WHOQOL – BREF (baseline, 24 weeks)
- Form R: Refraction (0, 4, 8, 12 and 16 weeks)
- Form T: Treatment acceptability questionnaire (3 and 6 weeks)
- Form V: Protocol violation form
- Form X: Adverse events

5.1. Log files extracted from the Tetris game iPods

Currently each participant has a separate log file containing a row of data for each day they used the game. The study manager will combine the following data from the log files for all participants into one excel file so that there is one row per participant:

- Contrast at baseline, 3 and 6 weeks (for intervention group only)

- Total duration participant spent on playing the game at 3 and 6 weeks

Note if the log file is not available due to identified device or technical issues, participants' diaries (form D tables 1 and 3) will be used to obtain the above data.

6. VARIABLE DEFINITIONS

The following definitions relate more specifically to the derived variables and non-standard definitions required for analysis of the data that are specific to the analyses referred to throughout this SAP. Please refer to the study protocol for a complete list of abbreviations set out for this study.

6.1. Baseline data

The baseline data that will be used in the analyses and all results tables will be obtained from Form A (if the participant does not need to go through refractive adaption) or the last Form R (if the refractive adaptation is required for the participant).

6.2. Age groups

The age groups (7-12 years, 13-17 years and >17 years) that will be used in the analyses and all results tables will be based on age at randomisation [Note this was used to stratify the randomisation by age groups].

However the results for participant accountability and the initial part of the consort diagram involving pre-randomisation counts (in sections 9.1 and 9.2) will be presented by age groups based on age at registration. If the age group counts are slightly different pre and post randomisation in the consort diagram then a footnote will be added to it indicating the difference and the reason for it.

6.3. Visual acuity

Amblyopia impairs a broad range of visual functions in the affected eye including visual acuity[3], contrast sensitivity[4] and motion perception[5, 6]. As visual acuity is the gold standard measure of visual function in clinical settings, a widely-accepted clinical definition of amblyopia is a difference in best corrected visual acuity between the two eyes of 0.2 log units in the absence of any ocular or optic nerve pathology[2].

The following visual acuity (VA) measures are collected separately for OD (right eye), OS (left eye), and OU (both eyes):

- Distance VA: Measured using the highly standardised E-ETDRS protocol adopted by the Amblyopia Treatment Study group, and recorded using the threshold score in Form A (Q7.04 – Q7.06), Form R (Q5.04 – Q5.06) and Form C (Q3.04 – Q3.06). The threshold score will be converted to the standard LogMAR unit using the conversions shown in Appendix I (note lower scores indicate better vision).

- Near VA: Measured using the Lighthouse ETDRS near visual acuity chart, and recorded using standard LogMAR unit in Form A (Q7.07 – Q7.09), Form R (Q5.07 – Q5.09) and Form C (Q3.07 – Q3.09).

All VA results will be presented as amblyopic eye, fellow eye and both eyes (and not as right eye and left eye). Hence the following new variables will be created for the above distance and near VA measures:

- If the right eye is amblyopic (where form A Q13.01=right) then amblyopic eye VA is equal to OD (right eye) measure and fellow eye VA is equal to OS (left eye) measure.

- Otherwise If the left eye is amblyopic (where form A Q13.01=left) then amblyopic eye VA is equal to OS (left eye) measure and fellow eye VA is equal to OD (right eye) measure.

6.4. Refraction

The cycloplegic refraction measures are collected at baseline in form A Q11.01-11.02, and each are measured in 3 parts. Each of these refraction measures will be converted using the formula:

$$\text{Mean Spherical Equivalent} = \text{Sphere} + (\text{Cylinder})/2$$

$$= \text{Part 1} + (\text{Part 2})/2$$

The astigmatism of the amblyopic eye is recorded at baseline in part 2 and 3 of form A Q11.01-11.02.

These results will be presented as amblyopic eye and fellow eye (and not as right eye and left eye) using form A Q13.01.

6.5. Stereopsis

Amblyopia is commonly associated with impaired stereoscopic depth perception under ordinary (binocular) viewing conditions[7]. Stereopsis will be measured using three tests, including two clinically well-established tests (Titmus Stereo Test and Randot Preschool Test) and one newly developed iPod-based Stereo test. The Titmus Stereo Test and Randot Preschool test have a limited number of possible results (as the tests present a certain number of stereoscopic depths) vs the iPod based stereo test which gives a more refined level of stereopsis (with an associated standard deviation).

Results of the stereo tests will be recorded using the unit of seconds of arc in Form A (Q10.02, Q10.04 and Q10.05), Form R (Q8.02, Q8.04 and Q8.05) and Form C (Q6.02, Q6.04 and Q6.05). A lower measurement (seconds of arc) represents a better result.

Many participants will have no measurable stereopsis which will be recorded as zero in the form. For the Titmus Stereo Test and iPod-based Stereo test the zero values will not be included in the analyses, but a count of the number of zero values will be presented separately.

The following two methods will be used to convert the Randot Preschool test measure:

1) Using log scale:
   - Non-zero values converted to log values as follows: 40" (1.60), 60" (1.78), 100" (2.00), 200" (2.30), 400" (2.60), 800" (2.90)
   - Zero values set to a value of 1600" (3.20). Sensitivity analyses will also be conducted where zero values set to a value of 1000" (3.00) and 10,000 (4.00) instead.

2) Using a Binocular Function score combining the Randot Preschool test and Worth 4-Dot (from Form A (Q9.15), Form R (Q7.15) or Form C (Q5.15)) results:
   - Non-zero values converted to Log10
   - Zero values set to:
     o log(10,000)=4 if the Worth 4-Dot result is 4 or 5 dots
     o log(100,000)=5 if the Worth 4-Dot result is 2 or 3 dots

6.6. Angle of strabismus

Strabismus is a cause and a type of amblyopia, which can be with or without anisometropia. The types of amblyopia that are included in this trial involve amblyopia associated with unilateral strabismus, anisometropia or both (mixed mechanism).

The angle of strabismus will be measured using simultaneous prism cover tests at near and at distance[8], and recorded in the unit of prism dioptres in Form A (Q8.02 for near and Q8.04 for distance), Form R (Q6.02 for near and Q6.04 for distance) and Form C (Q4.02 for near and Q4.04 for distance) at scheduled visits (3, 6, 12 and 24 weeks).

For participants where the cause of amblyopia is strabismus only (form A Q13.03 equal to Yes) or mixed (strabismus and anisometropia, form A Q13.04 equal to Yes) the angle or amount of strabismus is collected in the following 5 variables for the eye indicated in Q8.01 or Q8.03 (right eye or left eye, which corresponds to the amblyopic eye defined in form A Q13.01):

- horizontally eso
- horizontally exo
- vertically hyper
- vertically hypo
- rotation of the eye termed a cyclo deviation

These results will be presented separately for each of these 5 variables. Lower values indicate a better outcome [Note participants recorded as 'No strabismus' will not be included].

Change from baseline to 6 weeks in the angle of strabismus will only be calculated for participants who are strabismus only or mixed causes of amblyopia at baseline

(Yes to either form A Q13.03 or Q13.04). Counts of the following will also be presented:

- The frequency of enlarged angle (increased value for any of form C Q4.02 and Q4.04 greater than or equal to 6 prism dioptres [9] for any direction of deviation) at 6 week follow-up compared to the baseline value (form A Q8.02 and Q8.04) for strabismus and mixed causes of amblyopia (Yes to either Q13.03 or Q13.04)

- The frequency of decreased angle (decreased value for any of form C Q4.02 and Q4.04 greater than or equal to 6 prism dioptres for any direction) at 6 week follow-up compared to the baseline value (form A Q8.02 and Q8.04) for strabismus and mixed causes of amblyopia (Yes to either Q13.03 or Q13.04)

- The frequency of new strabismus (Right/left eye for form C Q4.01 or Q4.03 and value other than '0' for any of form C Q4.02 and Q4.04) for participants who are Anisometropia only (Yes to form A Q13.05)

6.7. Interocular contrast

For the active game, the contrast presented to the amblyopic eye will be 100% and the contrast presented to the fellow eye will be adjusted automatically through the training period (6 weeks). For the intervention group the contrast within the Tetris game at baseline, 3 and 6 weeks will be obtained in a dataset put together by the study manager (see section 5.1).

6.8. Interocular suppression

Previous studies have demonstrated that stronger interocular suppression is associated with poorer vision in adults and children with amblyopia[10-12], and also in amblyopic primate animal models[13]. The interocular suppression will be measured by an iPod-based version of the dichoptic motion coherence test (binocular and contrast thresholds) and Worth 4-dot test at near and distance.

The results of five trials from the iPod-based suppression test binocular and contrast thresholds (the contrast presented to the amblyopic eye was 80% for contrast thresholds, but see next paragraph) will be recorded as a percentage in Form A (Q9.04-Q9.13), Form R (Q7.04-Q7.13) and Form C (Q5.04-Q5.13). An average percentage of the five trials for the binocular and contrast thresholds will be calculated. For binocular thresholds, lower values indicate a better outcome. For contrast thresholds, higher values indicate a better outcome. Where the average binocular threshold is >85%, the corresponding contrast thresholds at that visit will not be used because these indicate poor performance on the global motion task and the test results are not reliable.

Some of the participants in the Waterloo centre had the contrast presented to the amblyopic eye at 100% instead of 80%. A list of these participants and the visits affected will be provided to the study statistician. The main analyses will be conducted using the actual values provided, and additional sensitivity analyses will also be conducted using the ratio value calculated as the average percentage of the five contrast threshold trials divided by the 80% or 100% contrast as appropriate.

Worth 4-dot tests at near and distance will be recorded in Form A (Q9.14 and Q9.15), Form R (Q7.14 and Q7.15) and Form C (Q5.14 and Q5.15). The Worth 4-dot test presents separate red/green lights to the right and left eyes. The possible responses are 4 lights (normal vision), 2 lights (representing left eye suppression) or 3 lights (representing right eye suppression) or 5 lights (representing double vision).

6.9. Treatment compliance and dose response

Treatment compliance and dose response will be based on the total time that participants spent on playing the game at 3 and 6 weeks which will be obtained in a dataset put together by the study manager from the log files extracted from the iPod (see section 5.1). Prescribed treatment dose is 1-2 hours per day (i.e. 21 hours at 3 weeks visit and 42 hours at 6 weeks visit).

- A participant is considered to be compliant to the study protocol if he/she has received more than 25% of the prescribed dose (i.e. > 5.25 hours at 3 weeks visit AND > 10.5 hours at 6 weeks visit).

- Time spent playing the active or placebo game at or less than 25% of the prescribed dose is defined as poor treatment compliance

For treatment dose response the total time that participants who are able to play the game spent playing the game will be used as both a continuous variable and categorical variable (quintiles will be used based on the data). Note a maximum score of less than 1000 indicates that a participant may not have been able to play the game, so their data will be excluded from these analyses.

6.10. Treatment acceptability

Treatment acceptability will be recorded in form T and measured in parents of participants ≤17 years of age (Q2.01-2.20) and participants >17 years of age (Q3.01-3.19) at 3 weeks and 6 weeks follow-up visits using a modified Amblyopia Treatment Index (ATI) parental questionnaire which has been validated in previous studies [14, 15]. The parent questionnaire consists of 20 Likert-type items (19 items for adult questionnaire) with 5 response choices coded from 1 ("strongly agree") to 5 ("strongly disagree"). All items (except for those reverse scored which are indicated in the table below) will be recoded to have values 5 ("strongly agree") to 1 ("strongly disagree"), so that higher scores indicate more (adverse) impact. A sixth choice of "not applicable" was offered for questions 7 and 8 in the parent questionnaire and for questions 6 and 7 in the adult questionnaire.

Questionnaires with 3 or more missing or not-applicable responses will be excluded from the analyses. For the remaining questionnaires, the missing or not-applicable responses will be imputed using the average score for all completed items. Three subscales (adverse effects, treatment compliance, and social stigma) will be calculated by taking the mean of the items for the subscale listed in the table below. These subscales will be calculated separately for the parent and adult questionnaires.

Table for Amblyopia Treatment Index subscales and corresponding items

Subscale	Parent questionnaire items	Adult questionnaire items
Items composing adverse effects of treatment subscale	2 Worry that child on treatment may miss out on fun activities	2
	3 Treatment affects child's learning	3
	4 Treatment makes it hard for child to play outside	4
	9 Difficult for my child to draw, color, or write	8
	10 Worry that child on treatment will become injured	9
	11* Child can see well on treatment	10*
	15 Child clumsy on treatment	14
	18 Treatment makes it difficult for child to play	17
Items composing treatment compliance	1* Child does not seem to mind treatment	1*
	5 Trouble applying treatment to child	5

subscale	6 Treatment is a source of tension or conflict with child	-
	7 Treatment is a source of tension or conflict with another family member	6
	8 Treatment is a source of tension or conflict with teacher or babysitter	7
	12 Child complains when it is time for treatment	12
	14 Worry child not getting enough treatment	13
Items composing social stigma subscale	13 Treatment makes my child's eye or eyelids red or irritated	11
	16 Other children stare at child	15
	20 Worry that child feels different	19
Items not used in any subscale**	17*	16*
	19	18

* Items are reverse scored
** Note these items are not loaded strongly on any single subscale and are not listed

Because the treatment acceptability questionnaire was modified from a questionnaire intended to assess occlusion treatment, there is some ambiguity regarding questions describing tasks that can be performed concurrently with patching but cannot be done concurrently with playing the videogame. These are Form T Q2.04, 2.09 and 2.18 for children and Q3.04, 3.08, 3.17 for adults, relating to "outdoor activities", "near vision activities like drawing or writing" and "playing with toys/daily activities such as cooking, driving or typing" respectively. The main analyses for the adverse effects subscale will be calculated including these questions, and sensitivity analyses will also be conducted excluding these questions to assess if these make a difference.

6.11. Quality of life

Quality of life will be measured in adult participants (aged >17 years old) using the World Health Organization Quality of Life (WHOQOL)-BREF questionnaire, at baseline and 24 weeks using Form Q. This questionnaire contains 26 items where the first two items are about general quality of life and health, and the remainder of the items are used to calculate the following four domain scores: physical health, psychological, social relationships, and environment. Each of the domain scores has range 0 to 100, where higher scores indicate better quality of life.

Full instructions on how these domains are calculated can be found in the WHOQOL-BREF Instructions Manual, in brief:

- Reverse the scores for items 3, 4 and 26 which are negatively phrased items (note this has already been done in the coding)

- The following mean score of items within each domain is used to calculate the domain score:

  1. Physical health domain: Q3, Q4, Q10, Q15, Q16, Q17, Q18

  2. Psychological domain: Q5, Q6, Q7, Q11, Q19, Q26

  3. Social relationships domain: Q20, Q21, Q22

  4. Environment domain: Q8, Q9, Q12, Q13, Q14, Q23, Q24, Q25

- Mean scores are then multiplied by 4 in order to make domain scores comparable with the scores used in the WHOQOL-100 (not used in analyses). This transforms the domain scores to a range of 4-20.

- The domain scores are then transformed again to have range 0-100 using: (domain score – minimal possible score) / range of possible score, i.e. (domain score – 4) / 16.

- When more than 20% of the data is missing from the questionnaire, the score will not be calculated and will be treated as missing. Also if more than two items are missing from a domain, the domain score will not be calculated (with the exception of social relationships domain, where the domain will only be calculated if ≤1 item is missing). When an item is missing, the mean of other items in the domain will be substituted.

6.12. Adverse events

All adverse events and device-related incidents occurring during the trial that are observed by study personnel or reported by the participant will be recorded in form X, whether or not attributed to trial treatment. The following information will be recorded: description, date of onset and end date, severity, assessment of relatedness to trial medication, other suspect drug or device and action taken. Follow-up information should be provided as necessary. The severity of events will be assessed on the following scale: 1 = mild, 2 = moderate, 3 = severe.

Adverse events considered related to the trial treatment as judged by a qualified investigator will be followed either until resolution, or the event is considered stable. It will be left to the judgement of the Steering Group to decide whether or not an adverse event is of sufficient severity to require discontinuing the participant from the study. A participant may also voluntarily withdraw from participating in the study due to what he or she perceives as an intolerable adverse effect. If either of these occurs, the participant must undergo follow-up visits for trial assessment and be given appropriate care under medical supervision until symptoms cease, or the condition becomes stable.

7. BIOSTATISTICS QUALITY ASSURANCE

Well in advance of study data-lock, programming across all analyses will commence with dummy data. This will allow sufficient time to turn around the 'real' analyses as quickly as possible. Early programming will also allow for a preliminary run at merging of the various datasets into the correct data structure required for analysis. Specifically, the datasets in section 5 will be taken before data-lock and passed through the necessary biostatistical processes (which will include checking that all the datasets have been imported correctly into SAS and are all able to be correctly merged together).

8. ANALYSIS POPULATIONS

8.1. Intention to Treat

Treatment evaluations will be primarily performed on the principle of 'Intention To Treat' (ITT). The ITT population will consist of all randomised participants regardless of whether they actually satisfied the entry criteria, the treatment actually received and subsequent withdrawal or deviation from the protocol.

8.2. Per Protocol

A per protocol analysis will also be performed on the primary outcome in order to check the robustness of the results. Those randomised participants who have no major protocol violations will be included in this subset for analysis. Relevant protocol violations may include errors in treatment assignment, the use of excluded medication, poor compliance, loss to follow-up and missing data.

The use of the per protocol set may maximise the opportunity for a new treatment to show additional efficacy in the analysis, and most closely reflects the scientific model underlying the protocol. However, the corresponding test of the hypothesis and estimate of the treatment effect may or may not be conservative depending on the trial; the bias, which may be severe, arises from the fact that adherence to the study protocol may be related to treatment and outcome.

For this trial, possible conditions of major protocol violation include:

1. Poor treatment compliance (see section 6.9)
2. Loss to follow-up and missing outcome
3. Failure to wear red/green glasses (as instructed) whilst playing the game during the 6 week treatment period (Form D, Q4.01): where either the 3 or 6 week visit is 'No' for this question.
4. Failure to wear glasses/contacts whilst playing the game during the 6 week treatment period (Form C, Q2.03): where either the 3 or 6 week visit is 'No' for this question.
5. Poor compliance with glasses/contacts-wearing (Form C, Q2.01)
6. Follow-up visits fall out of time window: out of ± 7 days for 3 and 6 weeks follow-up visits, ± 21 days for 12 weeks follow-up visit, and ± 28 days for 24 week follow-up visit.

The management committee will conduct a blinded review of all participants with protocol violations before final data analysis, and make decision on the list to be excluded from the per protocol set.

9. STATISTICAL ANALYSIS

All statistical analyses will be performed using SAS version 9.4. All statistical tests will be two-tailed and at 5% significance level throughout the analyses and all treatment evaluations will be performed on the principle of 'intention to treat' unless otherwise specified. No adjustments for multiplicity are planned for any of the outcomes.

All results will be presented overall and by the three age groups (based on age at randomisation); 7-12 years, 13-17 years and >17 years. Note if insufficient participants are randomised to any of the age groups then the regression analyses will not be conducted by the age group, and instead the age group will be combined with one of the other age groups.

Summaries of continuous variables which are normally distributed will be presented as means and standard deviations or medians and inter-quartiles for skewed data, while categorical variables will be presented as frequencies and percentages. Continuous variables will be compared with t-tests or Mann-Whitney tests and categorical data with chi-squared tests as appropriate.

The study statistician will conduct all the analyses. A separate EXCEL file containing the SAP table's specific to this study (Appendix II) will be provided based on all the analyses stated below.

9.1. CONSORT statement

All participants who were invited to participate in this study should be accounted for and a CONSORT statement prepared. The reasons for non-participation will be discussed in relation to the external validity of the study and the pattern of protocol violations considered as potential sources of bias. Reasons for early withdrawal will be listed for all participants that prematurely discontinued intervention or the study. The number of participants that were screened but not randomised will also be presented from the Screening Log.

[Note the results by age groups in the initial part of the consort diagram involving pre-randomisation counts will be based on age at registration. If the age group counts are slightly different pre and post randomisation then a footnote will be added to the consort diagram indicating the difference and the reason for it.]

9.2. Participant accountability

Tables describing participant accountability will be produced. The number of participants who were screened, registered, fulfilled eligibility criterion, together with reasons for exclusion will be summarised in table 1a (note results by age groups for this will be based on age at registration). The status of all randomised participants at each visit, and the number of major protocol violations will also be summarised in table 1a.

Table 1b will contain a summary of sections 1 and 2 in form R.

9.3. Baseline characteristics

Demographics and baseline characteristics collected on all randomised participants will be summarised by treatment groups in table 1c. Since any differences between the groups at baseline could only have occurred by chance, no formal significance testing will be conducted.

9.4. Concomitant medications

A line listing of all concomitant medications recorded in form M at baseline and for the duration of the study will be included in table 2a.

9.5. Primary outcomes analysis

The change from baseline to 6 weeks in distance visual acuity of the amblyopic eye will be assessed using linear regression (ANCOVA), and adjust for baseline outcome value and age groups in the model. The main analyses for the primary outcome will be ITT, where the Last Value Carried Forward (LVCF) approach will be used to replace any missing outcomes at 6 weeks. Model-adjusted treatment difference will be reported, with associated 95% confidence interval and p-value. Results will be presented in table 3a.

The following sensitivity analyses will be conducted to test the robustness of the results:

- Per protocol analysis (excluding major protocol violations described in section 8.2)
- Complete case analysis excluding any missing outcomes at 6 weeks
- Replacing any missing outcomes at 6 weeks using multiple imputations method (instead of LVCF)
- Also adjusting the model for treatment dose using the total time that participants spent on playing the game as both a continuous variable and categorical (where quintiles will be used based on the data)

9.6. Secondary outcomes analysis

Descriptive summary statistics for each follow-up visit and treatment group will be presented for all secondary outcomes in tables 2b and 2c. Participants with missing outcome data will be excluded from these analyses.

Repeated measures analyses will be conducted on continuous outcomes measured at more than one follow-up visit (such as visual acuity, stereopsis, interocular

suppression, angle of strabismus) using the random mixed models. The interaction effect between treatment and visit will be tested to assess the size of treatment effects over time. An overall intervention effect will be estimated if there is no significant interaction, i.e. when there is no time effect in the difference between two groups. Missing data will be dealt in the mixed model (as part of its functionality), assuming they are missing at random. However if all follow-up data is missing then the participant will be excluded from the analyses. Continuous outcomes measured at only one follow-up visit (such as quality of life) will be analysed using linear regression (ANCOVA). For all continuous outcomes the baseline outcome value and age groups (stratification factor) will be adjusted for in the regression models. Model-adjusted treatment difference will be reported, with associated 95% confidence interval and p-value. Note the distribution of the continuous outcomes will be first assessed for normality and skewed data will be subjected to an appropriate transformation before analysis. Non-parametric analysis (Mann-Whitney tests) will be used if data is skewed and cannot be transformed to be normally distributed. Results will be presented in table 3b.

Generalised linear regression models will be applied to categorical outcomes as appropriate, and age groups will be adjusted for in the model.

Treatment compliance and acceptability will be summarised using descriptive statistics at the 3 and 6 week follow-up visits by treatment group.

9.7. Safety

All adverse events (serious and non-serious) recorded in Form X for the duration of the study will be summarised. Table 4a will summarise the adverse event counts by categories collected in form X (type, severity and relationship to study treatment). Line listings of all adverse events will be included in table 4b.

9.8. Subgroup analysis

Note sufficient study power was planned on the primary outcome to conduct analyses separately for each age group (7-12 years, 13-17 years and >17 years), if 36 or more participants were recruited in each age group.

Tests for heterogeneity will be conducted to evaluate the consistency of treatment effects between the age groups for the primary outcome. This will be done by including age group along with its interaction with treatment in the regression model. A test of whether the treatment effect differs across the levels of the age subgroup will be constructed by assessing the significance of this interaction term.

In addition the following subgroups of interest will also be assessed on the primary outcome, if a sufficient number of participants are recruited in each subgroup:

- Distance visual acuity of the amblyopic eye at baseline (pre-randomisation) (Form A: Q7.04 or Q7.05 or last Form R: Q5.04 or Q5.05)

  - Moderate amblyopia (EVA scores of 50-70)

  - Severe amblyopia (EVA scores <50)

- Causes of amblyopia [16]

  - Strabismic amblyopia (Form A Q13.03 equal to Yes)

  - Anisometropic amblyopia (Form A Q13.05 equal to Yes)

  - Combined mechanism amblyopia (Form A Q13.04 equal to Yes)

  [Note strabismic and combined mechanism amblyopia will be collapsed if there is insufficient data to look at these individually].

- Stereopsis at baseline based on Randot Preschool Test results: 0, greater than 0 (Form A: Q10.04 or last Form R: Q8.04)
- Prior occlusion treatment for amblyopia (Q3.08 or Q3.12 equal to Yes)

10. DATA SAFETY AND MONITORING

Ellenburg et al. (2002) provide guidelines for deciding whether or not a data safety monitoring committee (DSMC) needs to be established for a trial. They propose that if two or more of the following criteria are met then a DSMC is required. 1) The trial is intended to provide definitive information about the effectiveness and/or safety of a medical intervention. 2) There is prior data to suggest that the intervention being studied has the potential to induce potentially unacceptable toxicity. 3) The trial is evaluating mortality or another major endpoint such that inferiority of one treatment arm has safety as well as effectiveness implications. 4) It would be ethically important for the trial to stop early if the primary question addressed has been definitively answered, even if secondary questions or complete safety information were not fully addressed. This trial meets only point 1 indicating that an independent DSMC does not need to be established for the trial, which is in concordance with the independent decision made by the HRC committee. Data safety and regular monitoring will be performed by the Steering Group Committee. Compliance with the protocol and accuracy in relation to source documents will be evaluated.

An independent monitor will be involved in BRAVO at each study site. The independent monitor will check the existence and correct date for all signed consent forms. The monitor will sample over 10% of all randomized participant to check correct data collection and data entry for the key points of this study, including visual acuity and active/placebo treatment placement. The monitor will confirm collected data on hard copies of CRFs with the source data (saved EVA files on the laptop for visual acuity and licence of the video-game program for the treatment) and check the data entry from the hard copy to the electronic database.

11. PEER REVIEW

The statistical analysis and report for the primary outcome will be peer reviewed by an independent member of the NIHI Biostatistics team (not involved in the study).

12. BUDGET

Level of involvement and engagement of statistical analyses as well as details prepared in the statistical report will depend on the study budget for Biostatistics team. Please note that for any additional analysis that is not specified in this statistical analysis plan or beyond the given budget, additional allowance will be required for further analyses in consultation with the principal investigator and/or NIHI operations manager.

13. DISSEMINATION

At the end of study (after data lock), all planned statistical analysis will be carried out. A statistical analysis report will be produced to summarise the main findings. ANNEX tables will be attached to present all detailed results.

The study statistician will therefore take full responsibility on this report and attached tables. The study statistician does not take any responsibility for any modifications made in the interpretation, conclusions, or tables/figures beyond those presented in statistical analysis report.

14. REFERENCES

1. Guo CX, Babu RJ, Black JM, Bobier WR, Lam CS, Dai S, Gao TY, Hess RF, Jenkins M, Jiang Y, Kowal L, Parag V, South J, Staffieri SE, Walker N, Wadham A, Thompson B; BRAVO study team. Binocular treatment of amblyopia using videogames (BRAVO): study protocol for a randomised controlled trial. Trials, 2016, Oct 18; 17(1):504.

2. To L, Thompson B, Blum JR, Maehara G, Hess RF, Cooperstock JR. A game platform for treatment of amblyopia. IEEE Transactions on neural systems and rehabilitation engineering, 2011, 19:280-289.

3. Holmes, J.M. and M.P. Clarke, Amblyopia. The Lancet, 2006. 367(9519): p. 1343-1351.

4. Hess, R.F. and E.R. Howell, The threshold contrast sensitivity function in strabismic amblyopia: evidence for a two type classification. Vision Res, 1977. 17(595414): p. 1049-1055.

5. Simmers, A.J., et al., Deficits to global motion processing in human amblyopia. Vision research, 2003. 43(6): p. 729-738.

6. Thompson, B., et al., Abnormal cortical processing of pattern motion in amblyopia: evidence from fMRI. Neuroimage, 2012. 60(22285220): p. 1307-1315.

7. McKee, S.P., D.M. Levi, and J.A. Movshon, The pattern of visual deficits in amblyopia. Journal of Vision, 2003. 3(5).

8. Scheiman, M.M., et al., Randomized trial of treatment of amblyopia in children aged 7 to 17 years. Archives of ophthalmology, 2005. 123(4): p. 437-447.

9. Pediatric Eye Disease Investigator, G., Interobserver reliability of the prism and alternate cover test in children with esotropia. Arch Ophthalmol, 2009. 127(1): p. 59-65.

10. Black, J.M., et al., The measurement and treatment of suppression in amblyopia. Journal of visualized experiments: JoVE, 2012(70).

11. Li, J., et al., How Best to Assess Suppression in Patients with High Anisometropia. Optometry & Vision Science, 2013. 90(2): p. e47-e52.

12. Li, J., et al., The role of suppression in amblyopia. Investigative ophthalmology & visual science, 2011. 52(7): p. 4169-4176.

13. Bi, H., et al., Neuronal responses in visual area V2 (V2) of macaque monkeys with strabismic amblyopia. Cereb Cortex, 2011. 21(21263036): p. 2033-2045.

14. Cole, S.R., et al., The amblyopia treatment index. Journal of American Association for Pediatric Ophthalmology and Strabismus, 2001. 5(4): p. 250-254.

15. Holmes, J.M., et al., Further validation of the Amblyopia Treatment Index parental questionnaire. Journal of American Association for Pediatric Ophthalmology and Strabismus, 2008. 12(6): p. 581-584.

16. Birch, E.E., et al., Binocular iPad treatment for amblyopia in preschool children. Journal of American Association for Pediatric Ophthalmology and Strabismus, 2015. 19(1): p. 6-11.

APPENDIX I: EVA THRESHOLD SCORE AND LOGMAR CONVERSIONS

Threshold score	LogMAR		Threshold score	LogMAR		Threshold score	LogMAR
0	1.7		34	1.02		68	0.34
1	1.68		35	1		69	0.32
2	1.66		36	0.98		70	0.3
3	1.64		37	0.96		71	0.28
4	1.62		38	0.94		72	0.26
5	1.6		39	0.92		73	0.24
6	1.58		40	0.9		74	0.22
7	1.56		41	0.88		75	0.2
8	1.54		42	0.86		76	0.18
9	1.52		43	0.84		77	0.16
10	1.5		44	0.82		78	0.14
11	1.48		45	0.8		79	0.12
12	1.46		46	0.78		80	0.1
13	1.44		47	0.76		81	0.08
14	1.42		48	0.74		82	0.06
15	1.4		49	0.72		83	0.04
16	1.38		50	0.7		84	0.02
17	1.36		51	0.68		85	0
18	1.34		52	0.66		86	-0.02
19	1.32		53	0.64		87	-0.04
20	1.3		54	0.62		88	-0.06
21	1.28		55	0.6		89	-0.08
22	1.26		56	0.58		90	-0.1
23	1.24		57	0.56		91	-0.12
24	1.22		58	0.54		92	-0.14
25	1.2		59	0.52		93	-0.16
26	1.18		60	0.5		94	-0.18
27	1.16		61	0.48		95	-0.2
28	1.14		62	0.46		96	-0.22
29	1.12		63	0.44		97	-0.24
30	1.1		64	0.42		98	-0.26
31	1.08		65	0.4		99	-0.28
32	1.06		66	0.38		100	-0.3
33	1.04		67	0.36			

APPENDIX II: LIST OF TABLE TEMPLATES

Section 1. Randomisation and baseline information
Table 1a. Participant accountability
Table 1b. Refraction summary
Table 1c. Baseline variables

Section 2. Descriptive summary of follow-up data
Table 2a. Line listing of all concomitant medications
Table 2b. Continuous follow-up variables
Table 2c. Categorical follow-up variables

Section 3. Analysis
Table 3a. Primary outcome analysis
Table 3b. Secondary outcome analysis

Section 4. Adverse events
Table 4a. Adverse event counts by categories
Table 4b. Line listing of all adverse events