Statistical Analysis Plan for TRIDENT: TRIal Designs for DElivery of Novel Therapies for Neurodegeneration

ISRCTN No:	52651778	Version Number:	1.0 (10 March 2023)

Final Plan

Based on protocol version: 4.1 (21 Nov 2019)

SAP Revision History

Protocol version	Updated SAP version no.	Section number changed	Description and reason for change	Date changed

ROLES AND RESPONSIBILITIES

Trial Statistician: Ms Sharon Ayayo			
Role: Statistics			
Date:	10/03/2023	Signature:	Email
Senior Statistician: Dr Philip Pallmann			
Role:			
Date:		Signature:	
Chief Investigators: Prof Anne Rosser, Prof William Gray			
Role:			
Date:		Signature:	
Other non-signatory contributor to the SAP: Dr Cheney Drew			
Role:			

1. INTRODUCTION

This Statistical Analysis Plan provides guidelines for the final presentation and analysis for the TRIDENT trial (Drew et al. 2021). This plan, along with all other documents relating to the analysis of this trial, will be stored in the Statistics Master File electronically and/or in hard signed copy formats.

2. BACKGROUND

2.1 RATIONALE AND RESEARCH QUESTION

Neurodegenerative conditions are a common cause of dementia and disability and represent a huge societal burden. Currently most are untreatable, but we are on the cusp of a new era of potential disease-modifying therapies for neurodegeneration, with many of the most promising requiring direct delivery into the central nervous system (CNS).

Cell replacement therapy (CRT) may provide a way of treating the many neurodegenerative conditions for which the underlying cellular pathology is currently uncertain. It can be considered in any condition in which there is a relatively focal loss of cells in the CNS. Neurodegenerative diseases comprise a large number of relatively rare genetic conditions that together constitute a substantial disease burden, and indeed even "common" conditions, such as Parkinson's (PD), are now known to be heterogeneous in terms of genetic origin and cellular pathogenesis. This presents a problem for developing therapies dependent on targeting specific cellular pathways, but CRT has the potential to circumvent this situation by directly treating the cell loss.

Huntington's (HD) is a powerful paradigm for understanding and treating neurodegeneration. It is the most common monogenetic neurodegenerative condition of the CNS. It is an autosomal dominant disease with full penetrance; thus it is possible to make a firm diagnosis in life, even prior to symptom onset, which provides substantial power for clinical studies. There is relatively focal loss of a specific neuronal cell type, striatal medium spiny neurons (MSNs), which makes it suitable for CRT. The vision is that the principles underlying effective cell therapy in HD will be applicable to other neurodegenerative conditions. Previous work has shown that CRT in HD is safe and potentially efficacious, but to date only a relatively small number of cells have been transplanted due to prior concerns regarding overgrowth of the graft. This trial intends to transplant a much higher number of cells as it is hoped increased cell number would provide greater efficacy.

For HD patients, previous trials typically transplanted 2-10 million cells per striatum, with one trial transplanting up to 20 million cells in a few patients. Surgical studies used 1-3 burr holes and a varied number of trajectories (3-10) and 5-6 deposits per trajectory.

To be ethically acceptable trials of cell therapy must attempt to deliver the safest cells using the safest device. The current gold standard in HD is human foetal cell transplantation against which all future cell therapies will be judged. However, graft success depends on both graft delivery and the subsequent ability of the cells to survive and integrate. Data from structural and functional imaging to measure the health of acute or very early cell grafts is either limited or non-existent. We therefore cannot easily dissociate the independent but serially linked effects of the efficacy of the delivery device and the characteristics of the cells themselves on graft survival and functional outcome. This means there can be but scant data on the best device for cell delivery apart from safety data. Currently the number of approved delivery devices is extremely limited and we are independently developing theoretically superior devices which are at pre- and peri-clinical stages of development. To our knowledge there is currently only one commercially available device CE marked for the delivery of cells to human brain, manufactured by Elekta. Whilst we work towards optimising this aspect of neural cell transplantation for HD and other neurodegenerative diseases, for this trial we will therefore use this CE marked device for cell transplantation in conjunction with an in house

manufactured inner cannula. We will use this cannula combination to investigate the safety of transplanting higher numbers of cells than have been used previously and delineate optimal trial processes for future studies.

2.2 OBJECTIVES

The primary objective of this study is to evaluate the safety of transplantation surgery using increased numbers of human foetal ganglionic eminence cells for the treatment of patients with HD.

The secondary objectives of this study are to:

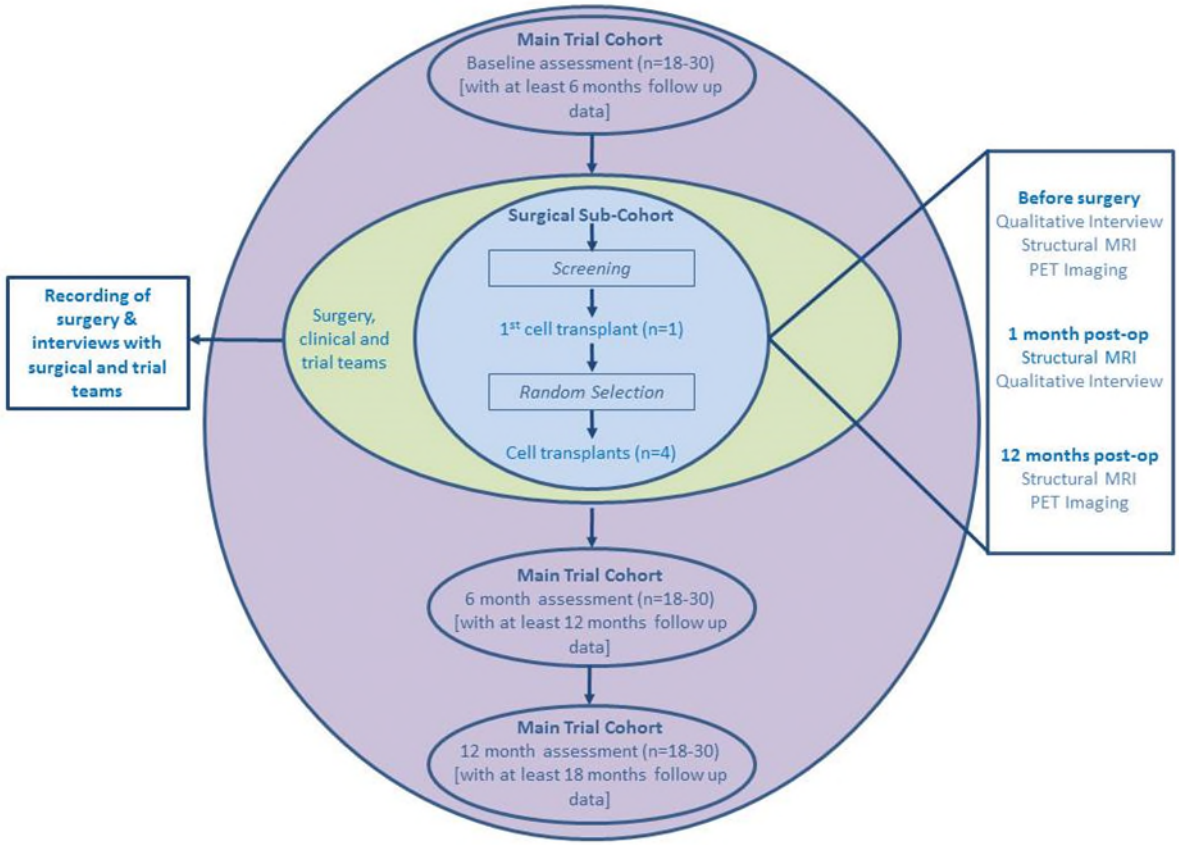
- define a framework for assessing the fidelity of cell transplantation devices and procedures;
- explore effect estimates to inform sample size calculations for future trials;
- evaluate feasibility of health economic evaluation for future trials;
- explore attitudes and understanding, feasibility and acceptability of this process in HD patients and their supporters/carers, trial deliverers, and health professionals;
- capture the social experience of patients and family members/carers over the entire lifecycle of the cell transplantation process, including the time period before, during and after the event;
- identify the support needs of patients undergoing neural transplantation and their family members/carers;
- explore the expectations, attitudes and clinical equipoise of health professionals engaged in the activity of neural transplantation towards transplantation process and trial processes (e.g. randomisation).

3. STUDY MATERIALS

3.1 TRIAL DESIGN

This is a single-site, open-label (only outcome assessors are blinded), phase 1 feasibility/safety trial within a cohort (TWiC). The design is illustrated in Figure 1.

Figure 1: Trial schema.



3.2 RANDOMISATION

While not a randomised trial, potentially eligible participants in the TRIDENT observational cohort who have been assessed for surgical suitability (the surgically suitable cohort) will be approached to undergo surgery using a random process. A set of computer-generated random numbers will be used to order the surgically suitable cohort, and this will be used to inform the selection of participants to be invited to undergo surgery. This will provide a direct assessment of the willingness to be randomised to such an intervention. The process is described in full in the Randomisation Protocol held securely within the Statistics Master File.

3.3 SAMPLE SIZE

Up to five participants will receive the neural transplantation. This sample size was decided upon as it restricts the number of participants being asked to take part in a highly novel, high-risk trial, consistent with the standard approach for phase 1 trials, whilst allowing for a number of trial processes (such as randomisation, surgical procedure and process evaluation) to be evaluated. Additionally, resource available dictated that it was only possible to test the intervention in a small number of participants as they will be required to be followed up for life. The participants undergoing transplantation will be identified from the trial observational cohort of approximately 18-30 participants. Thus, the remaining 13-25 will act as control participants.

3.4 FRAMEWORK

This is an early-phase feasibility/safety study.

3.5 INTERIM ANALYSES

The primary safety outcomes measured at 4 weeks post-surgery and other safety-related data such as MRI scans will be reviewed by the Trial Steering Committee (TSC) after every transplant before proceeding to the next one, in a multi-disciplinary meeting as described in section 14.6 of the Protocol. If the TSC members are satisfied with the safety outcomes, they may also approve a stepwise increase in the number of cells transplanted up to a total of 22 million cells, with a maximum increase of 2.5 million cells per successive transplant, and later expansion to bilateral transplants.

3.5.1 PLANNED SAMPLE SIZE ADJUSTMENT

Not applicable.

3.5.2 STOPPING RULES

There are no formal statistical criteria for stopping the study early, but the TSC can terminate the study if they are concerned about the safety or ethics of continuing in the light of safety data accrued from previous transplants and/or information external to the study e.g. about advances in cell transplantation devices.

3.6 TIMING OF FINAL ANALYSIS

All outcomes will be analysed collectively once all participants have completed the trial.

3.7 TIMING OF OUTCOME ASSESSMENT

A schedule of trial procedures, including the timing of outcome assessments, is provided in Table 1.

Table 1: Schedule of trial procedures and outcome assessments.

Main Trial Cohort

	Enrollment	Month of Trial Participation											
	Up to -6 months	0	1	2	3	4	5	6	7	8	9	10	11	12
Discusion of trial	x													
Informed consent	x													
Random allocation to comparator cohort	x													
CAPIT-HD2 assessment battery		x												x
Truncated CAPIT-HD2 assesment battery								x						
EMG Recording		x												x
SF-12 Quality of Life		x												x
Client Service Receipt Inventory		x												x

Surgical Sub-Cohort

	Enrollment and pre-op assessment					Month of post surgical follow up																			
	-6 to -4	-3	-2	-1	0	1		2		3		4		5		6		7		8		9		10		11		12
Number of outpatient and research visits pre w	1	1	1	1		2	1	1	1		1		1		1		1					1					1	
Discussion of trial	x																											
Informed consent	x																											
Invitation to pre-surgical assessment	x																											
Blood Tests: FBC, U's &E's, LFTs, CRP	x				x	x	x	x	x		x		x		x		x					x					x	
Viral serology	x				x																							
Drug levels					x	x	x	x	x		x		x		x		x					x					x	
Pregnancy Test (Serum HCG)	x																											
Urine pregnancy test			x		x				x			x		x		x						x					x	
ECG	x																											
CAPIT-HD2 assessment battery				x																								x
Truncated CAPIT-HD2 assesment battery																x												
EMG Recording				x																								x
SF-12 Quality of Life				x																								x
Client Service Receipt Inventory				x																								x
CT Scan					x	x																						
3T MRI Scan		x						x																				x
1.5T MRI Scan				x																								
PET Scan			x																									x
Qualitative Interview				x				x																				
Transplant surgery					x																							
Adverse event monitoring					x	x	x	x	x		x		x		x		x					x					x	
Immunosupressive drug compliance					x	x	x	x	x		x		x		x		x					x					x	

Denotes inpatient stay

4. STATISTICAL PRINCIPLES

4.1 LEVELS OF CONFIDENCE AND P-VALUES

All confidence intervals presented will be 95% and two-sided. As this is a feasibility study no statistical hypothesis tests will be performed, therefore no significance thresholds for p-values need to be specified.

4.1.1 ADJUSTMENT FOR MULTIPLICITY

Not applicable.

4.2 ADHERENCE AND PROTOCOL DEVIATIONS

4.2.1 DEFINITION AND ASSESSMENT OF ADHERENCE

For participants undergoing surgery, compliance with their immunosuppressant medication regimen will be checked at every follow-up visit through pill counts and blood level monitoring as described in the Protocol.

4.2.2 PRESENTATION OF ADHERENCE

Participants' pill counts and blood level data will be tabulated per follow-up visit and/or displayed graphically.

4.2.3 DEFINITION OF PROTOCOL DEVIATION

Any protocol deviations will be classified as major or minor.

4.2.4 PRESENTATION OF PROTOCOL DEVIATIONS

Protocol deviations will be summarised descriptively.

4.3 ANALYSIS POPULATION

All participants with data will be included in relevant analyses (complete cases).

5. STUDY POPULATION

5.1 SCREENING DATA

Owing to the multi-stage consent model of the study, screening data will be reported as numbers and percentages (as appropriate) of participants screened for potential eligibility 1) to take part in the study, 2) to be included in the surgical sub-cohort and undergo pre-surgical assessments, and 3) to undergo surgery.

5.2 ELIGIBILITY

The numbers and percentages of participants considered eligible and approached at each stage of consent (as defined in 5.1 above) will be reported. The numbers and percentages of participants falling into each exclusion criterion will be reported, again by stage of consent. The number of ineligible participants recruited or included, if any, will be reported as well, with reasons for ineligibility.

5.3 RECRUITMENT

The numbers and percentages of participants consented and recruited or included at each stage of consent (as defined in 5.1 above) will be reported.

In addition to the summaries provided in 5.1 and 5.2, a CONSORT flow diagram will summarise the numbers 1) screened, 2) considered eligible and approached, 3) consented and recruited or included, by stage of consent, 4) completing each assessment visit.

5.4 WITHDRAWAL/FOLLOW UP

5.4.1 LEVEL OF WITHDRAWAL

The numbers and percentages of participants falling into each of the following categories will be reported:

- withdrawal from trial observational cohort
- withdrawal from trial intervention (i.e. prior to surgery)
- withdrawal from qualitative interviews
- withdrawal from video recording of surgery
- withdrawal from imaging assessments
- withdrawal of consent to all of the above
- withdrawal from use of data already collected

5.4.2 TIMING OF WITHDRAWAL

The numbers and percentages of participants withdrawing or lost at each stage or assessment time point will be reported, by sub-cohort.

If a participant decides to withdraw from the study after undergoing transplantation, they will be required to adhere to the follow-up schedule pertaining to the assessment of the immunosuppression medication on a clinical basis.

Any participant who withdraws or is withdrawn prior to surgery will be replaced by another eligible participant from the non-transplanted trial observational cohort.

5.4.3 REASONS FOR WITHDRAWAL

Reasons for withdrawal will be reported descriptively (if known).

Participants may be withdrawn prior to surgery if:

- they develop an acute medical condition that precludes their inclusion in the trial;
- they are not suitable for transplant surgery as deemed by neurosurgical opinion;
- a female participant is found to be pregnant.

5.4.4 PRESENTATION OF WITHDRAWAL/LOSS TO FOLLOW-UP

Participants in the observational cohort will be considered as lost to follow-up if they fail to attend the 12-month follow up assessment within +4 weeks of the planned assessment date. Appointments may be re-scheduled a maximum of 3 times within the given time frame.

Due to the nature of mandatory follow-up of participants who have undergone transplantation surgery, we do not anticipate that transplantation participants will be lost to

follow-up during the trial period. It is possible that participants may move outside of Wales and fall under another specialist HD service within the UK; should this occur, participants will be invited to return to the research site to undergo follow-up assessments. If this is not possible, attempts will be made to obtain follow-up data from the participant via their new clinical service.

Numbers and reasons (if known) for withdrawal, loss to follow-up and/or exclusion from analysis at each stage, by sub-cohort, will be presented in a CONSORT diagram.

5.5 BASELINE PARTICIPANT CHARACTERISTICS

5.5.1 LIST OF BASELINE DATA

In addition to reporting on the outcomes listed in Table 1 that are collected at baseline (e.g. CAPIT-HD2 assessments, SF-12), participants will be described with respect to age, gender, ethnicity and CAG repeat length.

5.5.2 DESCRIPTIVE STATISTICS

Categorical data will be summarised by number and percentage, overall and by sub-cohort (those selected for the surgical sub-cohort vs. those always remaining in the observational cohort). Continuous data will be summarised by mean and standard deviation (SD), or median and interquartile range (IQR) if notably skewed, minimum and maximum, overall and by sub-cohort.

6. ANALYSIS

6.1 OUTCOME DEFINITIONS

6.1.1 PRIMARY OUTCOME(S)

The primary outcome measure of the trial will be safety at 4 weeks after surgery as defined by the lack of incidence of:

- significant additional, permanent neurological deficits;
- a clinically significant intra-cranial haemorrhage;
- clinically significant intra-cranial infection

as assessed by the TSC.

6.1.2 TIMING, UNITS AND DERIVATION OF PRIMARY

This information is provided under 6.1.1. above.

6.1.3 LIST OF SECONDARY OUTCOMES

The secondary outcome measures will be:

- feasibility and acceptability of clinical trial processes as determined by recruitment, retention, participant and carer experiences;
- fidelity of neurosurgery defined by evaluation of successful delivery of cells and accurate neurosurgical graft placement assessed by MRI and PET scanning;

- long term (12 months) safety of transplantation, defined by growth profile of graft as assessed by MRI and PET, and absence of development of clinically significant inflammatory or immune reaction as assessed by the clinician and TSC;
- documentation of research, treatment and immunosuppression costs to aid full health economic analysis in future trials;
- development of fidelity markers through analysis of video data capture of the surgery and graft survival over 1 year as determined by structural MRI/PET;
- outcomes in the motor, cognitive, psychiatric and functional domain that are part of the CAPIT-HD2 battery of assessments (Table 2).

Table 2: Schedule of trial procedures and outcome assessments (CAPIT-HD2 battery).

Name	Description	Time to complete	Time point
Motor Domain			
UHDRS TMS *	The unified Huntington's disesase rating scale total motor assessment is the gold standard measure for assessing motor severity in people with HD	95 minutes	Baseline and 12 months (* 6 months in truncated battery)
Q-Motor speeded and metronome tapping	The participant is required to tap their index finger on a force transducer according to cues. The duration and variability of finger taps are recorded		
Q-Motor grasping and lifting	The participant is required to grip and lift a device fitted with a force transducer and hold it stable. Grip force, 3D position and orientation of the object are recorded		
Q-Motor dynamic cue and force matching and reproduction tasks	Participants are required to complete a series of tests where they generate force on a transducer with their index finger. They will be asked to; match force patterns for which they have previously receivedvisual feedback, match a sinusoidal pattern, generate increasing and decreasing force patterns with and without visual feedback. Deviations from target forces and patterns are recorded.		
Q-trail	Participants are required to make a trail between specific numbers and/or letters using a stylus on a force transducer. Total distance travelled, total time used, precision of target identification (inclduing total errors) and path precision are all recorded.		
Q-eye	Participant are required to look at visual stimuli on a projected screen whilst their head is stabilised using a brow bar and chin rest. Eye movements (saccades, smooth pursuit and optokinetic nystagmus) in response to the stimuli will be recorded.		
Cognitive domain			
Mattis Dementia Rating Scale	The MATTIS is a well-documented global measure of cognitive status, especially sensitive in sub cortical affections	60 minutes	Baseline and 12 months (* 6 months in truncated battery)
Hopkins Verbal learning Test	The HVLT is composed of 12 items, orgnaised into 3 sematic categories and presented over three consectuive learning trials.		
Controlled Oral word association tests	The participant is asked to name as many words (excluding porpoer nouns)beginning with a specific given letter		
Category fluency	Participants are asked to name as many things in one category as possible in a given time (usually 60 seconds)		
Stroop test*	Participants are presented with a series of words pertaining to colours and are asked to read them out loud. Initially the words are coloured to correpsond to the word (i.e red, green). Participants are asked to repeat the task with words written in contrasting colours, but they ahve to say the name of the colour the word is written in (i.e green, blue, red).		
Symbol digit modalities test*	Participants are presented with a series of symbols and a code assigning a number (1-9) to each symbol. They have 90 seconds to write the corresponding number for the symbol for as many symbols as possilble.		
Relationship and Life events	The relationship questionnaire is composed of 48 items. For each question, six possible responses are proposed: 'absolutely true', 'true', 'mostly true', 'mostly false', 'false', 'absolutely false'. Life events assessment will be performed by using the Holmes & Rahe's scale. Patients are asked to tally 43 life events which allow us to provide a score of Events during the last year.		
Psychiatric domain			
Problem behaviours assessment	This is a semi structured clinical interview measuring the presence, severity and frequency of 11 key behavioural symptoms. Detailed severity scoring criteria are provided for each item	60 minutes	Baseline and 12 months (* 6 months in truncated battery)
Apathy scale, Irritability scale	Standardised questionanire to assess apathy and irratibilty		
Frontal Systems Behvioural Scale	This is a brief, participant completed behavior rating scale with demonstrated validity for the assessment of behavior disturbances associated with damage to the frontal-subcortical brain circuits		
Maze task	Participants are asked to make decisions when offered a choice between objects (decision making under limited choice) and when there is no list of options to select from (decision making under unlimited choice). Subjects are told to make the decision as quickly as they can. The decision outcome is then recorded.		
Persistence Test	This is intended to assess loss of motivation (an aspect of apathy). Participants are informed that they must race their icon against an opponent's icon. They are also informed that their icon is fitted with a speed boost that the computer will activate at a random point in the race. Latency to quitting/completion is measured		
Functional Domain			
UHDRS Total Functional Capacity (TFC)	UHDRS TFC scale assesses how people with HD manage their work, finances, daily living, domestic chores and their care arrangements	25 minutes	Baseline and 12 months
C3T	The Clinch Token Transfer Test (C3T) is a dual-task assessment of bilateral, upper motor function that consists of three-coin transfer tasks which increase in difficulty (baseline simple, baseline complex and a dual task). The time taken to pick up and transfer the coins from dominant to non-dominant hand and place into a purpose developed box is recorded. The addition of cognitive load increases the task complexity.		

6.1.4 ORDER OF TESTING

Not applicable.

6.1.5 TIMING, UNITS AND DERIVATION OF SECONDARIES

The recruitment rate is defined as the number of participants consented to take part in the study divided by the total number of patients approached, screened and determined to be eligible.

The retention rate is defined as the number of participants completing at least one task or questionnaire at the 12 month assessment divided by the total number of participants consented.

The complete CAPIT-HD2 battery will be assessed at baseline and 12 months (+/- 4 weeks), and a truncated battery (as indicated in Table 2) at 6 months (+/- 2 weeks) for participants who have undergone surgery only.

Motor domain

Unified Huntington's Disease Rating Scale (UHDRS) Total Motor Score (TMS): This is a measure of motor severity in HD. It consists of 31 items, each rated on a five-point Likert scale (from 0=best to 4=worst outcome) that can be added up to obtain a total score (Huntington Study Group 1996, Reilmann & Schubert 2017) between 0 and 124. Higher scores indicate more severe symptoms.

Q-Motor speeded and metronome tapping: Participants are asked to tap their index finger on a force transducer according to cues. The duration and variability of finger taps are recorded.

Q-Motor grasping and lifting: Participants are asked to grip and lift a device fitted with a force transducer and hold it stable. Grip force, 3D position and orientation of the object are recorded.

Q-Motor dynamic cue and force matching and reproduction tasks: Participants are asked to complete a series of tests where they generate force on a transducer with their index finger. They are asked to match force patterns for which they have previously received visual feedback, match a sinusoidal pattern, generate increasing and decreasing force patterns with and without visual feedback. Deviations from target forces and patterns are recorded.

Q-Trail: Participants are asked to make a trail between specific numbers and/or letters using a stylus on a force transducer. Total distance travelled, total time used, precision of target identification (including total errors) and path precision are recorded.

Q-Eye: Participants are asked to look at visual stimuli on a projected screen whilst their head is stabilised using a brow bar and chin rest. Eye movements (saccades, smooth pursuit and optokinetic nystagmus) in response to the stimuli are recorded.

Cognitive domain

Mattis Dementia Rating Scale (MDRS-2): This is a global measure of cognitive status that is widely used for assessing dementia. It consists of five subscale scores (attention, initiation and perseveration, construction, conceptualisation, memory) that can be added up to obtain a total score (Mattis 1988, Marson et al. 1997).Lower scores indicate worse cognitive status.

Hopkins Verbal Learning Test (HVLT-R): This is a test of verbal learning and memory that consists of three consecutive trials, each of which is a list-learning and free-recall task comprising 12 items (four from each of three semantic categories) followed by yes/no recognition (Brandt 1991). There are six equivalent alternate forms to circumvent learning effects due to item familiarity in serial testing. For the immediate and postponed free-recall tasks numbers of correct answers, perseverations and intrusions are recorded; for the recognition task numbers of correct answers, related and unrelated false-positives and total errors are recorded. Higher numbers of correct words indicate better verbal learning and memory.

Controlled Oral Word Association Tests (COWAT): This is a test of verbal fluency where participants are asked to name as many words (excluding proper nouns) beginning with a specific given letter as possible in 60 seconds (Bechtoldt et al. 1962, Ruff et al. 1996). This is repeated three times with three different letters. Total numbers of correct words per 15-second interval (0-15, >15-30, >30-45, >45-60 seconds) are recorded for each letter. Higher numbers of correct words indicate better verbal fluency.

Category fluency test (CFT): This is a test of category fluency where participants are asked to name as many words fitting a given semantic category (e.g. animals) as possible in 120 seconds (Butters et al. 1987, Ho et al. 2002). Total numbers of correct words in 1 and 2 minutes, respectively, as well as intrusions (i.e. incorrect words) and perseverations (i.e. repeated words) are recorded. Higher numbers of correct words indicate better category fluency.

Stroop test: This is an assessment of cognitive ability that requires participants to read out loud lists of colour words (Stroop 1935, Scarpina & Tagini 2017). In the first trial the words match the colour they are printed in (e.g. “red” printed in red). In the second trial there is a mismatch (e.g. “red” printed in green) and participants are asked to say the name of the colour the word is written in. Total numbers of correct answers, errors and self-corrected errors are recorded for each attempt. Higher numbers of correct answers and lower numbers of errors, respectively, indicate better cognitive performance.

Symbol Digit Modalities Test (SDMT): This is a measure of cognitive impairment where participants are presented with a series of symbols and a code assigning a single-digit number to each symbol. They are asked to match as many symbols with their corresponding numbers as possible in 90 seconds (Smith 2007). Total numbers of correct matches and errors are recorded. Higher numbers of correct matches and lower numbers of errors, respectively, indicate better cognitive performance.

Relationship questionnaire: This is a measure of the subjective quality of everyday social relationships. It is composed of 49 items assessed on a six-point Likert scale (from -3=“absolutely false” to +3=“absolutely true”) from which a negative score (sum of all negatives), a positive score (sum of all positives) and a total score (sum of all negatives and positives) between -147 and 147 flipping the scales for items 1, 3, 4, 7, 8,11, 13, 15, 16, 17, 19, 20, 22, 23, 25, 26, 29, 30, 32, 34, 35, 36, 37, 38, 40, 42, 45, 46, 48, 49 will be calculated. Higher scores indicate better/healthier/more secure relationships. Zero imputation will be done for missing items if ≥ 50% of the total questions are complete.

Life events questionnaire: This is a measure of stress based on various life events from the Social Readjustment Rating Scale (Holmes & Rahe 1967). It consists of 43 items, each associated with a score between 11 and 100 (depending on severity) if the event occurred in the past year and 0 otherwise, which will be added up to obtain a total score. Higher scores indicate more stress.

Psychiatric domain

Problem Behaviors Assessment for HD, short form (PBA-s): This is the short version of a semi-structured interview to assess neuropsychiatric symptoms relevant to HD (Callaghan et al. 2015). It consists of 11 items, each rated for the period of four weeks preceding the interview on a five-point Likert scale separately for severity (from 0="absent" to 4="severe") and frequency (from 0="never/almost never" to 4="daily/almost daily for most/all of the day"). The severity and frequency ratings are multiplied to yield an overall score of 0, 1, 2, 3, 4, 6, 8, 9, 12 or 16 for each symptom and then summed to obtain a total score ranging from 0 to 176. Patients who have been interviewed previously are additionally rated for the worst level of severity since the previous interview. Total sum of severity (and worst severity), total sum of frequency ranging from 0 to 44 will also be calculated. Higher scores indicate more severe and/or frequent symptoms. Mean imputation will be done for missing values if ≥50% of the questions are completed.

Apathy Evaluation Scale (AES-S): This is a self-reported measure of apathy consisting of 18 items (Marin et al. 1991), each rated on a four-point Likert scale (from 0="not at all" to 3="a lot"). A total score between 0 and 54 is obtained by adding up all item scores (flipping the scales for items 6, 10 and 11). Higher scores indicate less apathy.

Irritability Scale (IS): This is a measure of irritability consisting of 14 items (Chatterjee et al. 2005), each rated on a four-point Likert scale (from 0="not at all" to 3="a lot"). A total score between 0 and 42 is obtained by adding up all item scores (flipping the scales for items 3, 5, 9, 11 and 13). Higher scores indicate more irritability.

Frontal Systems Behaviour Scale, completed by participants (FrSBe-P): This is a self-report measure for assessing behavior disturbances linked to frontal lobe pathology (Grace & Malloy 2001, Stout et al. 2003). It consists of three subscale scores (apathy, disinhibition, executive dysfunction) that can be added up to obtain a total score. It comprises a total of 24 items, each rated on a five-point Likert scale for frequency (from 0="almost never" to 4="almost always") and distress (from 0="not at all distressing" to 4="extremely distressing or very severe") which will be multiplied per item to obtain an overall score which can take on values 0, 1, 2, 3, 4, 6, 8, 9, 10, 12, 16, these will then be added up to obtain an overall score ranging from 0 to 384. Total sum of frequency and total sum of distress ranging from 0 to 96 will also be calculated. If the behaviour is not applicable to the person, the frequency is 'almost never'. Higher scores indicate more severe and/or more frequent behaviour disturbances. Mean imputation will be done for missing values if ≥50% of the questions are completed.

Maze task: Participants are asked to make decisions when offered a choice between objects (decision making under limited choice) and when there is no list of options to select from (decision making under unlimited choice). Subjects are told to make the decision as quickly as

they can. The decision outcome is then recorded. The reaction time is measured in milliseconds.

Persistence test: This is intended to assess loss of motivation (an aspect of apathy). Participants are informed that they must race their icon against an opponent's icon. They are also informed that their icon is fitted with a speed boost that the computer will activate at a random point in the race. The latency to quitting or completion is measured in seconds.

Functional domain

UHDRS Total Functional Capacity (TFC): This is a measure of functional capacity that consists of five items, each rated on a three- or four-point Likert scale (from 0=worst to 2 or 3=best outcome) that can be added up to obtain a total score between 0 and 13 (Huntington Study Group 1996). Higher scores indicate higher functional capacity.

UHDRS HD Functional Capacity Scale (HDFCS): This is another measure of functional capacity that consists of 25 binary items, each of which can be answered yes or no (Huntington Study Group 1996). They can be added up to obtain a total score between 0 and 25. Higher scores indicate higher functional capacity.

UHDRS Independence Scale (IS): This is a measure of independence rated on a scale from 0 to 100. Higher scores indicate greater independence.

Clinch Token Transfer Test (C3T): This is a functional upper limb assessment that consists of three token transfer tasks (Clinch et al. 2018). It requires participants to pick up each of eight coins with their non-dominant hand, transfer them to their dominant hand, and release them into a moneybox in order of size (Baseline Transfer), in order of value (Complex Transfer), or in order of size whilst reciting the alphabet (Dual Transfer). The time to perform the transfer tasks is measured in seconds, and the accuracy (accounting for value/size errors, transfer errors and dropped tokens) is recorded. Time and accuracy are combined into a total score (number of tokens transferred, divided by time, multiplied by accuracy) for each of the tasks. For the Dual Transfer task, an alphabet rate (number of correct letters recited per second) is calculated and compared against baseline performance prior to the task. Higher total scores and alphabet rates indicate better performance.

12-item Short Form survey (SF-12): This is a general health questionnaire constructed using questions drawn from each of the 8 dimensions of the Medical Outcome Study (MOS) 36 item short form survey (SF-36). It is used to measure observable and tangible limitations due to poor health and/or bodily pain in physical, social and role activities. Two summary scores are reported from the SF-12 – a mental component score (MCS-12) and a physical component score (PCS-12). The items can be added up to obtain a total score between 0 and 100. Higher scores indicate better physical and mental health functioning. A score of 50 or less on the PCS-12 is recommended as cut-off to determine a physical condition and a score of 42 or less on the MCS-12 may indicate 'clinical depression'.

6.2 ANALYSIS METHODS

6.2.1 LIST OF METHODS AND PRESENTATION

The analysis of all secondary outcomes will be primarily descriptive.

- Continuous variables will be summarised as means and SDs, or medians and IQRs if notably skewed, per time point (baseline, 6 months, 12 months) and separated by sub-cohort (those selected for the surgical sub-cohort vs. those always remaining in the observational cohort).
- Categorical variables will be summarised as frequencies and percentages, per time point and by sub-cohort.
- Variables like PBA-s and FrSBe-P scores will be summarised as medians and IQRs, per time point and by sub-cohort. These variables are ordinal but not interval scaled (e.g. a severity score of 4 indicates greater severity than a score of 2 but not necessarily twice the severity, and the interval between scores of 2 and 3 is not necessarily the same as between 3 and 4), skewed (e.g. many participants will likely have a score of 0) and discrete (e.g. symptom scores of 0, 1, 2, 3, 4, 6, 8, 9, 12 and 16 are possible but 5, 7, 10, 11, 13, 14 and 15 are not).
- The patients who were in the surgical cohort have multiple scores since they had more than one pre-screening visits. The mean score across all the visits will be used in the analysis for these patients.

95% CIs will be calculated for differences of sub-cohort means (or medians), but no formal statistical hypothesis testing will be performed.

The data will also be presented graphically, for example using jittered dot plots and/or box or violin plots (Weissgerber et al. 2019) per time point (baseline, 6 months, 12 months) and separated by sub-cohort.

If data quality allows, we will explore the use of mixed-effects models for repeated measures (Detry & Ma 2016) to model trajectories of outcome variables across the different time points, with sub-cohort as a fixed and participant as a random effect variable.

Motor domain

UHDRS TMS: The total score will be summarised as mean and SD per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort means.

Motor score: A total motor score (obtained by combining motor speeded and metronome tapping, motor grasping and lifting and motor dynamic cue and force matching reproduction tasks) will be summarized as mean and SD per sub-cohort and overall. A 95% CI will be calculated for the difference between sub-cohort means.

Q-Trail: Total distance travelled, total time used, precision of target identification (including total errors) and path precision will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

Q-Eye: Self-paced saccades and optokinetic nystagmus (average of the 3 attempts) will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means. Optokinetic nystagmus will be presented for right and left eyes separately.

Cognitive domain

MDRS: Each subcategory (attention, initiation and perseveration, construction, conceptualisation and memory) together with the total score will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

HVLT: Each subcategory (immediate free recall, postponed free recall and recognition) will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

COWAT: Sums will be calculated per letter across all 15-second intervals, per 15-second interval across all letters, and a grand total. The subscores of total correct letters within 0-15 seconds, 16-30 seconds, 31-45 seconds and 46-60 seconds together with the grand total (3 minutes) across all letters will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

CFT: Numbers of words in 1 minute and 2 minutes, intrusions and perseverations will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

Stroop: Total correct, total errors and total self-corrected errors for each attempt will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

SDMT: Total numbers of correct matches and total errors will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

Relationship questionnaire: The total score will be summarised as mean and SD per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort means.

Life events questionnaire: The total score will be summarised as mean and SD per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort means.

Psychiatric domain

PBA-s: The overall score will be summarised as median and IQR per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort medians.

AES: The total scores will be summarised as mean and SD per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort means.

IS: The total score will be summarised as mean and SD per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort means.

FrSBe-P: The total sum of frequency, distress and frequency*distress total score will be summarised as median and IQR per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort medians.

Maze task: The reaction times will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

Persistence test: The latencies will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

Functional domain

UHDRS TFC: The total score will be summarised as mean and SD per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort means.

UHDRS HDFCS: The total score will be summarised as mean and SD per sub-cohort. A 95% CI will be calculated for the difference between sub-cohort means.

UHDRS IS: Participants' independence will be summarised as frequency and percentage per sub-cohort and also by mean and SD with a 95% CI for the difference between sub-cohort means.

C3T: Time taken for baseline transfer task and the total score, time taken for complex transfer task and the total score, time taken for dual transfer task and the total score will be summarised as mean and SD per sub-cohort. 95% CIs will be calculated for the differences between sub-cohort means.

6.2.2 COVARIATE ADJUSTMENT

Not applicable.

6.2.3 ASSUMPTION CHECKING

No formal checks of distributional assumptions will be performed as sample sizes will be too small for any such checks to be meaningful.

6.2.4 ALTERNATIVE METHODS IF DISTRIBUTIONAL ASSUMPTIONS NOT MET

Not applicable.

6.2.5 SENSITIVITY ANALYSES

No sensitivity analyses will be performed.

6.2.6 SUBGROUP ANALYSES

No formal subgroup analyses will be performed.

6.3 MISSING DATA

Number and percentage of missing values will be reported for each variable. Imputation of missing values will not be performed.

6.4 ADDITIONAL ANALYSES

The group of participants who do not receive the intervention will be further divided into those who were not selected for surgery and those who were initially selected and approached but did not receive the neural transplantation (e.g. refusal, ineligibility for surgery).

Exploratory evaluations will be carried out to explore plausible trial designs for subsequent (i.e. larger) randomised controlled trials evaluating the efficacy of neural transplantation in this population. Design considerations and parameters of interest will include, but are not limited to:

- prevalence of HD and projected recruitment rates;
- current and projected timelines for clinical procedures prior to surgery (including arrival distribution of foetal cells);
- number of available surgeons;
- participant retention (i.e. availability of outcome data);
- quality of intervention receipt (i.e. the extent to which the transplantation was successful and the immunosuppression regimen was adhered to);
- approaches to minimise the required sample size, including:
    - use of repeated measures outcomes;
    - within-patient designs (e.g. individual stepped-wedge designs, multiple baseline design, etc.);
    - time-matched controls;
    - response-adaptive designs.

6.5 HARMS

Adverse events will be summarised descriptively, by sub-cohort, including information on severity, causality and expectedness.

6.6 STATISTICAL SOFTWARE

The analysis will be carried out using Stata (version 17 or higher) and/or R (version 4.0 or higher). Graphs will be generated using ggplot2 (Wickham 2016). Other packages such as SAS may be used if necessary.

7. REFERENCES

7.1 NON-STANDARD STATISTICAL METHODS

Bechtoldt HP, Benton AL, Fogel ML (1962) An application of factor analysis in neuropsychology. Psychological Record, 12, 147-156.

Brandt J (1991) The Hopkins Verbal Learning Test: development of a new memory test with six equivalent forms. Clinical Neuropsychologist, 5, 125-142.

Butters N, Granholm E, Salmon DP, et al. (1987) Episodic and semantic memory: a comparison of amnesic and demented patients. Journal of Clinical and Experimental Neuropsychology, 9, 479-497.

Callaghan J, Stopford C, Arran N, et al. (2015) Reliability and factor structure of the Short Problem Behaviors Assessment for Huntington's disease (PBA-s) in the TRACK-HD and REGISTRY studies. Journal of Neuropsychiatry and Clinical Neurosciences, 27, 59-64.

Chatterjee A, Anderson KE, Moskowitz CB, et al. (2005) A comparison of self-report and caregiver assessment of depression, apathy, and irritability in Huntington's disease. Journal of Neuropsychiatry and Clinical Neurosciences, 17, 378-383.

Clinch SP, Busse M, Lelos MJ, Rosser AE (2018) Rethinking functional outcome measures: the development of a novel upper limb token transfer test to assess basal ganglia dysfunction. Frontiers in Neuroscience, 12, 366.

Detry MA, Ma Y (2016) Analyzing repeated measurements using mixed models. JAMA, 315, 407-408.

Drew CJG, Sharouf F, Randell E, et al. (2021) Protocol for an open label: phase I trial within a cohort of foetal cell transplants in people with Huntington's disease. Brain Communications, 3, fcaa230.

Grace J, Malloy PF (2001) Frontal Systems Behavior Scale: professional manual. Psychological Assessment Resources, Lutz, FL.

Ho AK, Sahakian BJ, Robbins TW, et al. (2002) Verbal fluency in Huntington's disease: a longitudinal analysis of phonemic and semantic clustering and switching. Neuropsychologia, 40, 1277-1284.

Holmes TH, Rahe RH (1967) The Social Readjustment Rating Scale. Journal of Psychosomatic Research, 11, 213-218.

Huntington Study Group (1996) Unified Huntington's Disease Rating Scale: reliability and consistency. Movement Disorders, 11, 136-142.

Marin RS, Biedrzycki RC, Firinciogullari S (1991) Reliability and validity of the Apathy Evaluation Scale. Psychiatry Research, 38, 143-162.

Marson DC, Dymek MP, Duke LW, Harrell LE (1997) Subscale validity of the Mattis Dementia Rating Scale. Archives of Clinical Neuropsychology, 12, 269-275.

Mattis S (1988) Dementia Rating Scale professional manual. Psychological Assessment Resources, Odessa, FL.

Reilmann R, Schubert R (2017) Motor outcome measures in Huntington's disease clinical trials. In: Feigin AS, Anderson KE (editors) Handbook of Clinical Neurology, volume 144: Huntington Disease. Elsevier, Amsterdam, The Netherlands.

Ruff RM, Light RH, Parker SB, Levin HS (1996) Benton Controlled Oral Word Association Test: reliability and updated norms. Archives of Clinical Neuropsychology, 11, 329-338.

Scarpina F, Tagini S (2017) The Stroop color and word test. Frontiers in Psychology, 8, 557.

Smith A (2007) Symbol Digits Modalities Test: manual. Western Psychological Services, Torrance, CA.

Stout JC, Ready RE, Grace J, Malloy PF, Paulsen JS (2003) Factor analysis of the Frontal Systems Behavior Scale (FrSBe). Assessment, 10, 79-85.

Stroop JR (1935) Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Weissgerber TL, Winham SJ, Heinzen EP, et al. (2019) Reveal, don't conceal: transforming data visualization to improve transparency. Circulation, 140, 1506-1518.

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer, New York, NY. https://ggplot2.tidyverse.org

Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996 Mar;34(3):220-33. doi: 10.1097/00005650-199603000-00003. PMID: 8628042.

Woodgate S, Morgan-Jones P, Clinch S, Drew C, Playle R, Bennasar M, Hicks Y, Holt C, Bachoud-Lévi AC, Massart R, Craufurd D, Kirby N, Hamana K, Schubert R, Reilmann R, Rosser A, Busse M. Objectively characterizing Huntington's disease using a novel upper limb dexterity test. J Neurol. 2021 Jul;268(7):2550-2559. doi: 10.1007/s00415-020-10375-8. Epub 2021 Feb 8. PMID: 33555419; PMCID: PMC7868671.

7.2 DATA MANAGEMENT PLAN

S:\Centre for Trials Research\Research\Mixed Studies\Cell transplantation studies\TRIDENT\1. e-TMF\Section 8.0 Data Management\8.1 Data Management\8.1.1 Data Management Plan

7.3 TRIAL MASTER FILE AND STATISTICAL MASTER FILE

S:\ Centre for Trials Research\Research\Mixed Studies\Cell transplantation studies\TRIDENT\1. e-TMF

S:\ Centre for Trials Research\Research\Mixed Studies\Cell transplantation studies\TRIDENT\1. e-TMF\Section 8.0 Data Management\8.5 Statistics

7.4 OTHER SOPS OR GUIDANCE DOCUMENTS

SOP/008/1 – Design and implementation of a randomisation strategy template

SOP/008/2 – Statistical analysis plan

SOP/008/3 – Sample size calculations

SOP/008/4 – Statistical reporting

SOP/008/5 – Statistical analysis quality assurance

SAP/ISAP DEVIATION LOG

Document number:		**Document version:**	
Reason for deviation:			

8. APPENDICES