The efficacy and feasibility of an immersive virtual reality game to train spatial attention orientation after stroke: a stage 1 pre-registered report

Hanne Huygelier [1], Vero Vanden Abeele [2], Raymond van Ee [1,3,4], Céline R. Gillebert [1]

[1] Brain and Cognition, KU Leuven, Leuven, Belgium

[2] e-Media | Human-Computer Interaction, KU Leuven, Leuven, Belgium

[3] Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, The Netherlands

[4] Philips Research, High tech Campus, Eindhoven, The Netherlands

*Corresponding author: celine.gillebert@kuleuven.be, Department of Brain and Cognition, Tiensestraat 102 box 3711, 3000 Leuven, Belgium, +32 16 37 27 04

Abstract

Introduction. Hemispatial neglect is a disabling post-stroke attention deficit for which there is no effective cognitive intervention. Immersive virtual reality (IVR) may increase treatment engagement and efficacy, as it allows to train cognition in a 3D, dynamic, rich and multisensory environment. We designed an IVR multisensory patient-tailored rehabilitation game to rebalance spatial attention in hemispatial neglect patients. The current study aims to evaluate the efficacy and feasibility of this new rehabilitation approach.

Method. A longitudinal cross-over placebo-controlled study design will be used to compare the effect of 10 hours of an active and 10 hours of a placebo IVR gamified intervention. We will recruit a minimum of 8 and maximum of 24 stroke patients with hemispatial neglect for the left side of space in two rehabilitation centres. The primary outcome is response times on invalid-cued left-sided compared to right-sided targets on a Posner cueing task. The secondary outcomes include performance on a computerized cancellation task, pen-and-paper administered cancellation and line bisection task, performance on the sustained attention to response task and symptoms in daily life evaluated with the Catherina Bergego Scale. To evaluate feasibility, we will systematically document the number of successfully recruited patients, cybersickness symptoms, their experience with the rehabilitation game and study whether patient's experience predicts drop-out from the trial.

Discussion. Our study will assess whether our IVR multisensory patient-tailored rehabilitation game can promote recovery of hemispatial neglect and whether it is feasible in stroke patients.

Keywords: hemispatial neglect, attention, immersive, virtual reality, rehabilitation

1 INTRODUCTION

Hemispatial neglect is a post-stroke attention deficit characterized by a difficulty in responding to events on the contralesional side of space, which significantly restricts patients' functionality and is a negative prognostic factor for stroke rehabilitation outcome (Buxbaum et al., 2004). Hemispatial neglect encompasses spatial and non-spatial deficits (Corbetta & Shulman, 2011; Husain & Rorden, 2003; Robertson, 2001) and is a heterogenous syndrome. Patients can show symptoms in different spatial reference frames, distances to the observer, and different sensory domains (Milner & Harvey, 1994; Van der Stoep et al., 2013). Neglect can affect orientation to visual information, visually imagined information (Buxbaum et al., 2004) or can affect the motor domain (Laplane & Degos, 1983). Although hemispatial neglect is a complex syndrome occurring after both left- and right-hemispheric stroke (Demeyere & Gillebert, 2019; Ten Brink, Verwer, et al., 2017), we will focus on hemispatial neglect that affects processing of visual information in *peri-personal space* following right-hemispheric stroke.

1.1 COGNITIVE REHABILITATION, NON-INVASIVE BRAIN STIMULATION AND PHARMACOLOGICAL INTERVENTIONS

It is estimated that 40% of neglect patients do not experience complete recovery (Demeyere, Gillebert, et al., 2015; Demeyere & Gillebert, 2017; Nijboer et al., 2013). Thus, there is a great need for effective interventions for neglect (Kerkhoff & Schenk, 2012; Luauté et al., 2006; Van Vleet & DeGutis, 2013). Given that there is no strong evidence yet for the effect of cognitive rehabilitation at the group level (Bowen et al., 2013), researchers explored the potential of non-invasive brain stimulation (Cazzoli et al., 2012; Salazar et al., 2018; van Lieshout et al., 2019), pharmacological interventions (Luvizutto et al., 2015; van der Kemp et al., 2017) or re-evaluated classic rehabilitation methods such as prism adaptation (Ten Brink, Visser-Meily, et al., 2017). However, clinical trials have not yielded sufficiently strong evidence for the therapeutic effects or feasibility of these interventions for them to be used as a standard treatment in clinical practice. Harvey (2019) highlighted a salient issue in neglect rehabilitation research, stating that it was uncommon to report the number of referred relative to successfully recruited patients in clinical trials evaluating non-invasive brain stimulation. These data are indeed crucial to evaluate the clinical contribution of therapies, as they provide insight into the percentage of patients that can be treated with the therapy. Thus, it is not surprising that there is little consensus among clinicians about the preferred treatment for hemispatial neglect (Chen et al., 2018).

1.2 VIRTUAL REALITY AS AN INTERESTING AVENUE FOR NEGLECT REHABILITATION

Interestingly, when clinicians were asked about their preferred rehabilitation method in a scenario not hampered by practical constraints, many clinicians expressed a preference for virtual reality rehabilitation (Chen et al., 2018; Kolodziej & Gillebert, 2018). Indeed, immersive virtual reality (IVR), using head-mounted displays (HMD) offers several opportunities for neglect rehabilitation. The immersive nature of IVR may increase treatment engagement (Tieri et al., 2018). Moreover, the HMD offers excellent control over stimulus presentation (Foerster et al., 2016), allows to train spatial orientation in a 3D, dynamic environment (Rizzo et al., 2004) and allows to correct for compensatory head movements. IVR can indeed provide a safe and positive experience for older adults (Huygelier, Schraepen, Ee, et al., 2019). Combining the strengths of IVR with gaming features that may further enhance treatment engagement (Burke et al., 2009), may result in effective and feasible neglect rehabilitation. Therefore, we developed an IVR game to rebalance spatial attention in neglect patients (Huygelier, Schraepen, et al., 2020).

1.3 POTENTIAL LIMITATIONS FOR THE USE OF VIRTUAL REALITY IN REHABILITATION

Although IVR is promising, the clinical utility of IVR remains uncertain to date. For instance, IVR has been notorious for inducing *cybersickness*. However, the newest generation of IVR meets the technological standards necessary to effectively mitigate cybersickness (Kourtesis et al., 2019) and several studies reported little cybersickness using the latest generation of IVR technology in various older populations (Appel et al., 2020; Huygelier, Schraepen, Ee, et al., 2019; Huygelier, Schraepen, et al., 2020; Plechatá et al., 2019). On the other hand, cybersickness has a negative association with sense of presence in the virtual environment (Weech et al., 2019) and may also depend on the design of the IVR application (Davis et al., 2015; Porcino et al., 2017; Stanney & Hash, 1998) as well as on characteristics of the end-users (Arns & Cerney, 2005). It is thus important to monitor cybersickness for each specific IVR application and for each end-user group.

Furthermore, cybersickness has mostly been studied using questionnaires administered after the IVR experience. As symptoms such as fatigue may already be present before using IVR, this procedure cannot clarify whether cybersickness resulted from the IVR experience itself. Indeed, a previous study showed a consistent decline in cybersickness after using an IVR application in six stroke patients (Huygelier, Schraepen, et al., 2020).

Finally, the clinical utility of IVR depends on how IVR is experienced by end-users (Huygelier, Schraepen, Ee, et al., 2019). Nevertheless, only a few studies investigated IVR user experience in older adults and

stroke patients (Dermody et al., 2020; Huygelier, Schraepen, Ee, et al., 2019; Huygelier, Schraepen, et al., 2020; Tuena et al., 2020). One study reported a good usability of a VR assessment using shutter glasses in stroke patients with hemispatial neglect (Fordell et al., 2011). Another study reported a good perceived usability for an exergame using a computer monitor in hemispatial neglect patients (Tobler-Ammann et al., 2017). However, these studies do not clarify whether head-mounted IVR will be positively experienced by neglect patients.

1.4 Design of clinical trials

Many clinical trials on neglect rehabilitation have reported weak or moderate quality evidence in favour of a therapeutic effect (Bowen et al., 2013; Luauté et al., 2006; Salazar et al., 2018; van der Kemp et al., 2017). Indeed, designing a sound clinical trial in stroke rehabilitation poses some challenges.

A first challenge lies in measuring changes in neglect symptoms. It is common to use clinical pen-and-paper assessments to measure effects of rehabilitation, but research has shown large variation in scores on these assessments from test to retest, potentially obscuring rehabilitation effects (Bailey et al., 2004; Machner et al., 2012). Moreover, performance on clinical pen-and-paper tests is often summarized in a way that does not differentiate non-spatial from spatial errors, making it unclear which behavioural aspects may be affected by treatment (Huygelier, Moore, et al., 2020).

A second challenge is to determine the therapy dose prior to conducting a clinical trial. When obtaining non-significant results, researchers often conclude that their therapy dose may have been too low to obtain a therapeutic effect (e.g., Sturm et al., 2013). A possible solution for this salient issue may be to measure symptoms multiple times. Such a longitudinal design can inform whether more therapy hours may have resulted in better treatment effects.

Another challenge is to determine the baseline condition. In neglect rehabilitation an experimental treatment is often compared to usual care (e.g., physical, occupational therapy), rather than to a placebo treatment. However, what exactly constitutes “usual care” is typically underreported and heterogeneous across hospital sites (Negrini et al., 2020). Moreover, clinical trials have often underreported information relevant to assess the feasibility of therapies (Harvey, 2019). These aspects make it difficult to generalize findings from clinical trials to new clinical contexts.

1.5 THE CURRENT STUDY

In the current study we aim to evaluate the *efficacy* and the *feasibility* of a new IVR rehabilitation game for hemispatial neglect.

More specifically, our primary objective is to compare the effect of an active and placebo IVR rehabilitation game on neglect symptoms assessed outside the IVR environment. In the active IVR intervention, multisensory stimulation is more frequently presented in the contralesional than ipsilesional visual field, while in the placebo IVR intervention multisensory stimulation is presented in the central visual field. We will assess neglect symptoms using a Posner task outside the IVR environment, given that pen-and-paper neglect tasks often lack test-retest stability (Bailey et al., 2004; Machner et al., 2012). We hypothesize that the difference in response times to left- versus right-sided targets on the Posner task will decrease more as a function of the active than placebo intervention. In addition, clinical pen-and-paper tasks, daily life functioning and a computerized cancellation task will be used as secondary outcomes. Here again, we hypothesize that the difference in performance between the left- and right visual field will decrease more as a function of the active than placebo intervention. The current study aims to address the impact of our intervention at the cognitive function and activities level and not at the participation level.

Second, we will assess the relationship between therapy dose and symptom recovery. To this end, the Posner task will be administered after 4, 8 and 10 hours of therapy. We hypothesize that there will be a larger effect of intervention hours on neglect symptoms in a non-VR environment in the active than placebo condition.

Third, we will evaluate whether our IVR rehabilitation game can impact non-spatial attention. We hypothesize that patients will improve more in non-spatial attention as a result of the active than the placebo IVR intervention.

Fourth, we will evaluate training effects inside the IVR environment, as patients may show training effects that are only present in the IVR environment. Neglect symptoms will therefore be assessed inside the IVR environment using a visual discrimination task and head orientation. We hypothesize that patients will improve within and outside the IVR environment, but that the improvement in the active IVR condition will be larger within than outside the IVR environment.

Finally, we will evaluate several aspects of the feasibility of our IVR rehabilitation. More specifically, we will report the number of successfully recruited relative to referred patients and report the reasons for missing data and drop-out. We will also assess the impact of a first experience with the IVR rehabilitation

game on user experience and cybersickness. In line with an earlier study (Huygelier, Schraepen, et al., 2020), we expect a positive user experience and less cybersickness after than before the IVR experience. Last, we will monitor cybersickness and user experience throughout the whole intervention period and assess their relationship with the likelihood of drop-out. Here, we hypothesize that cybersickness and user experience will remain stable across sessions and are not related to drop-out.

2 METHOD

2.1 PATIENT RECRUITMENT

Patients will be recruited from two rehabilitation centres in Flanders (Rehabilitation hospital RevArte, University Hospital Leuven campus Pellenberg). Patients residing in the rehabilitation units older than 18 years (no maximum age) with a stroke confirmed by a radiologist and who or their legal representative can provide informed consent can be included for screening. Patients will be excluded after screening when they do not show signs of hemispatial neglect for the left side of space, do not have a right-hemispheric stroke, are left-handed, when the expected discharge is in a period shorter than 10 weeks, they had/have a severe comorbid psychiatric disorder, premorbid diagnosis of a neurodegenerative disease, medical safety contra-indications for IVR (e.g., medical electric implants based on EU safety guidelines, trepanation, history of epileptic seizures), a severe visual or auditory impairment that cannot be corrected while using the IVR system or a severe motor impairment that precludes them from using the IVR system. If patients do not meet any of these exclusion criteria, they will be invited to participate in the clinical trial. Patients or their legal representative will provide written informed consent and all study procedures were approved by the ethical committee of the UZ Leuven/KU Leuven (S61410) and are in accordance with the Helsinki declaration. Our study protocol was preregistered at clinicaltrials.gov (NCT03458611).

2.2 IVR REHABILITATION AND IVR ASSESSMENT

An IVR application was developed in Unity3D for Oculus Rift CV1 (*Oculus Rift | Oculus*, n.d.). Responses will be registered using the right Oculus Touch Controller and head movements are logged. Detailed information and pilot data of the VR game is reported in (Huygelier, Schraepen, et al., 2020) and the game design is illustrated in a video (available online at https://doi.org/10.6084/m9.figshare.6194591.v2). The design of the VR game was iteratively optimized based on pilot studies with stroke patients and neurologically healthy individuals (Huygelier, Schraepen, et al., 2020).

Patients visit a vegetable garden, lake or forest and perform good deeds for their neighbours or friends in the game world (Figure 1). For instance, patients are instructed to catch ladybugs in the neighbour's vegetable garden. To finish each level, they have to do two variations of a visual discrimination task. In Task 1, an audio-visual looming (i.e., grows in size and sound intensity) semi-transparent disk is presented in 50% of the trials for 3s (Figure 1B). Afterwards, a target is presented in front of the white disk for 3s or until patients make a response. In Task 2, the white disk moves from the centre towards the left or right side of the visual field and once the target appears, the cue disappears, and the target moves towards the ground (Figure 1C). On each trial, feedback is presented to indicate whether the patient's response was correct (i.e., green checkmark), incorrect (i.e., red cross) or whether the patient did not make a response (i.e., blue exclamation mark and sound). On 25% of correct trials, patients receive a score. The score is scaled as a function of patient's performance. For instance, if they accurately discriminate only 25% of targets, they receive 4 points; if they accurately discriminate 100% of targets, they receive 1 point. Each rehabilitation level is completed once patients have obtained sufficient points.

In the *active IVR condition,* target locations will be presented more often in the contralesional than ipsilesional visual field, using a patient-tailored design. To tailor the ratio of contra- to ipsilesional stimulation to each individual patient, we first assess patient's spatial attention distribution in the IVR environment. In this *IVR assessment*, patients perform the visual discrimination task in each of the scenes (i.e., vegetable garden, lake, forest). Target locations are uniformly distributed within an area subtending 30° in the left to the right side of the visual field and 5° in the upper and lower visual field. Targets are presented for 3s or until a response is made. There are no rewards and the assessment is finished when a total of 225 trials (i.e., 75 trials per scene) are completed. Before each scene, patients perform 10 practice trials. A model is estimated on their responses and this model is mirrored along the x-axis to obtain a biased target probability distribution (Figure 1A).

This probability distribution is used to sample target locations, resulting in more contralesional than ipsilesional targets. For instance, a patient who participated in a pilot study with our game detected 32% of left-sided (< 2°), 56% of right-sided (> 2°) and 60% of targets in the centre of the visual field ([-2, 2°]) during the IVR assessment. During the active IVR rehabilitation, 61% of targets were located on the left (< 2°), 32% were located on the right (> 2°) and 7% in the centre of the visual field ([-2, 2°]). The target locations will appear at a maximum of 30° in the left and right side of the visual field and 5° in the upper and lower visual field.

In the *placebo* IVR *condition*, target locations will be sampled from a uniform distribution centred on 0° of the visual field with a horizontal angle of 1.5° in the left and right visual field and a vertical visual angle of 5° in the upper and lower visual field. All other game aspects will be identical between the two conditions.

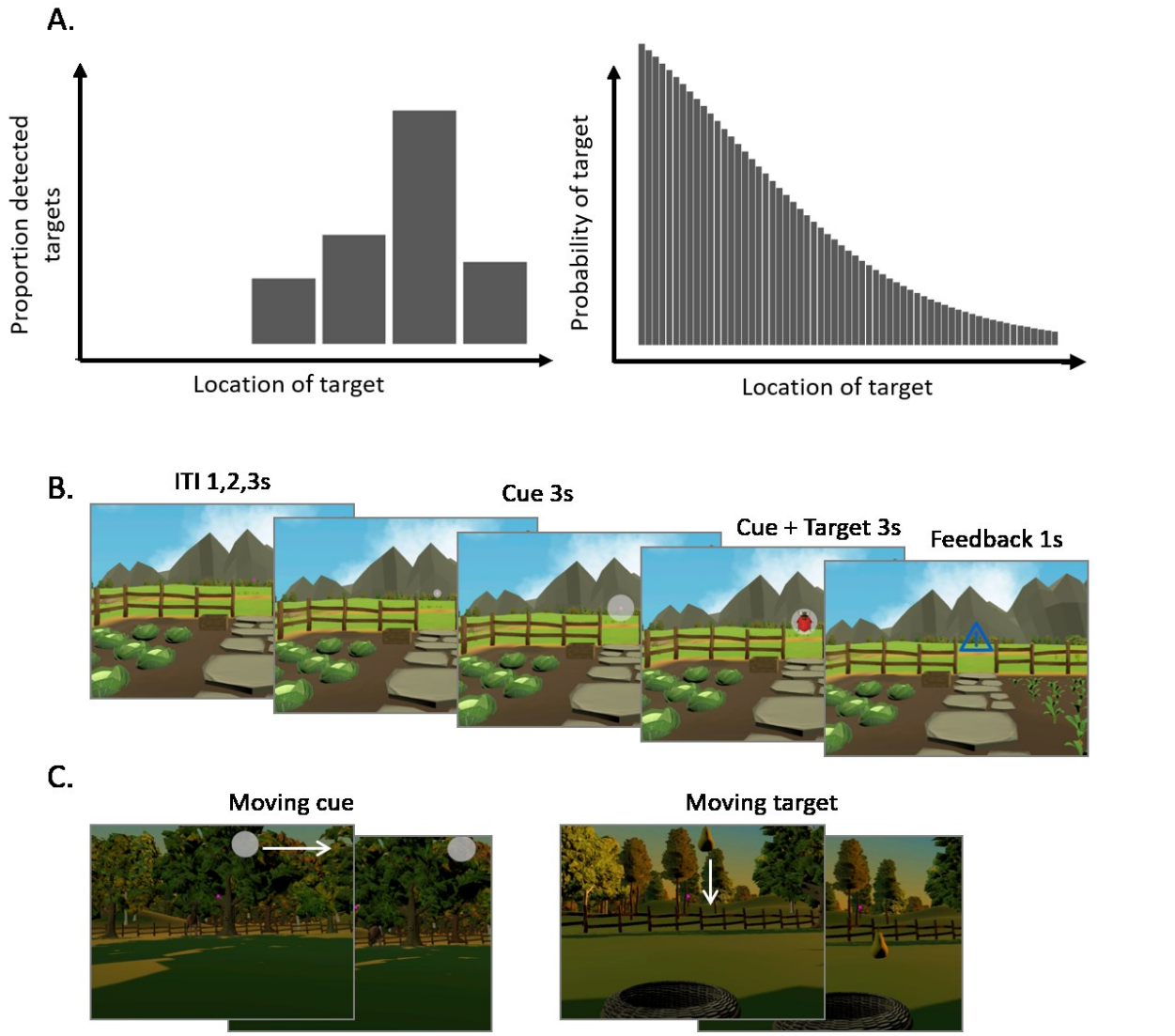

Figure 1. Illustration of patient-tailored design (A) and design of the tasks (B, C). If patients detect less targets for the left visual field during the assessment (panel A, left figure), than the probability that a target appears at those locations will be higher during rehabilitation (panel A, right figure). In Task 1 (panel B), a cue was presented for 3s in 50% of trials. Then, the target and cue were presented for 3s or until a response was provided. Afterwards, feedback was presented. In Task 2 (panel B), the cue moved towards a location after being presented in the centre of the visual field for 3 seconds and the target was presented without the cue and moved downward. Figure licensed under CC BY 4.0 by the authors. Retrieved from https://doi.org/10.6084/m9.figshare.12187710.v1.

2.3 MATERIALS FOR SCREENING, PRIMARY AND SECONDARY OUTCOMES

2.3.1 Semi-structured interview

A clinical psychologist will evaluate the eligibility of patients to participate in the study (e.g., history of epilepsy, pacemaker, cochlear implant), collect basic demographic information (e.g., date of birth, gender, handedness, date of stroke) and obtain information about patients' medical history.

2.3.2 Questionnaires

To evaluate the feasibility of the treatment we will use several questionnaires. *Cybersickness* will be measured with the Simulator Sickness Questionnaire (SSQ)[41] that was translated to Dutch by our research team. Each of the 16 SSQ items will be rated on a scale with four levels representing no, mild, moderate or severe discomfort. The *User Experience scale* consists of 23 items rated on a 5-point Likert scale going from totally disagree (1) to totally agree (5) with 3 as a neutral midpoint. Participants will answer questions about the usability of the touch controllers, their sense of presence and their intrinsic motivation to play the IVR game. The motivation items are based on the intrinsic motivation inventory (McAuley et al., 1989) and the spatial presence items are translations of the International Test Commission Sense of Presence Inventory items (Lessiter et al., 2001). At the end of each game session we will ask participants whether they experienced any physical discomforts and which discomforts they experienced. We will also ask them to rate their general game experience that day on a 5-point Likert scale.

To characterize patients' mood and fatigue at the start of the study we will administer two questionnaires in an interview format. We will administer the *Dutch Hospital Anxiety Depression Scale*. We will use the recommended cut-off score (≥ 8), which corresponds to an 80% sensitivity and specificity in detecting depression and anxiety (Bjelland et al., 2002). In addition, we will administer the Dutch *Fatigue Severity Scale*, which has good reliability and validity in stroke patients (Nadarajah et al., 2017).

2.3.3 Neuropsychological pen-and-paper assessment

To screen for general cognitive impairments, the *Dutch Oxford Cognitive Screen (OCS-NL)* will be administered (Demeyere, Riddoch, et al., 2015). Age-adjusted norms will be used to interpret test scores (Huygelier, Schraepen, Demeyere, et al., 2019).

Several pen-and-paper cancellation and a bisection task that are routinely used in clinical practice by clinical neuropsychologists and occupational therapists will be used in screening or as secondary outcomes (Checketts et al., 2020; Evald et al., 2020). All tasks are administered on A4 paper in landscape orientation.

The *OCS-NL hearts cancellation task* is a cancellation task with 50 full-outlined hearts and 150 hearts with a gap on the left or right side. The two parallel versions of the OCS-NL hearts cancellation task will be administered intermittently across the assessment sessions (i.e., A-B-A-B-A). The *Random Shape Cancellation task* consists of 360 shapes (i.e., 60 targets, 300 distractors) randomly placed in an array of 24 cm by 19 cm. The *Star Cancellation test* (BIT, Halligan, Cockburn, & Wilson, 1991) will be administered following the test manual instructions. To interpret whether there is a significant difference between right and left cancellations, we will use a Bayesian contingency table test, similar to the approach of the Frequentist z-test of proportions (Huygelier, Moore, et al., 2020). We will also administer the BIT *figure copy task* (Halligan, Cockburn, & Wilson, 1991).

The *McIntosh line bisection task* will be administered (McIntosh, 2017; McIntosh et al., 2005). There are 4 line conditions (i.e., condition A: line from -4 cm to 4cm, condition B: line from -8 to 4 cm, condition C: line from -4 to 8 and condition D: line from -8 to 8). Each line condition is presented 8 times on the page in a randomized order. The page is placed with the middle aligned to the patient's body midline. The patient is instructed to mark the middle of each line and tap the table in between each response. Performance is summarized using the *endpoint weighting bias* (EWB). The cut-off scores based on healthy controls are equal to -0.125 for right-sided neglect and 0.075 for left-sided neglect (McIntosh et al., 2017). Two parallel versions will be administered intermittently.

Finally, we will use the *Catherina Bergego scale* (CBS) (Azouvi et al., 2003; Ten Brink et al., 2013), which is a systematic observation scale frequently used by several health disciplines in clinical practice to measure how hemispatial neglect affects activities of daily living (Checketts et al., 2020). In addition, patients will be asked to rate themselves on the scale items and the difference between their own rating and the rating by the examiner will be used as a measure of anosognosia (Grattan et al., 2018).

2.3.4 Computerized assessment

Neglect symptoms will also be measured using several computerized tasks. All computer tasks will be administered on an LCD monitor with a resolution of 1920 by 1080 pixels. Patients will be seated approximately 70cm from the monitor. All code is written in Python 2.7 using Psychopy 1.90.3 (Peirce, 2007). Responses will be registered with a standard keyboard or computer mouse.

2.3.4.1 Posner task

A Posner paradigm is used to measure the primary outcome. We chose exogenous cueing, as a meta-analysis indicated that neglect patients showed more pronounced spatial attention orientation differences

in exogenous than endogenous cueing paradigms (Losier & Klein, 2001). Three squares with a size of 1.5°, 2 located at 7° to the left and right of the fixation cross and 1 in the center of the screen are presented. A cue (i.e., color change of a square) is presented for 100ms. Subsequently, a target is presented 150ms or 1100ms after cue onset for 100ms, in the left or right square (size of 1.4°) (Figure 2A). Cues and targets appear on the left or right side of the screen with equal probability. The cue can be valid (i.e., same side as target) in 40% of trials, invalid (i.e., opposite to target side) in 40% of trials or not followed by a target in 20% of trials (i.e., catch trials). Patients have to respond as quickly as possible when they see the target in a 4s time limit. They have to press the space bar with their right hand. After a response or when the time limit passes, the fixation cross dissapears for 1s to indicate the end of a trial. Patients will perform 10 practice trials in which feedback (i.e., “correct” or “incorrect”) is shown for 1s. There will be 400 experimental trials that are presented in 4 blocks of 100 trials. The order of the trials is randomized.

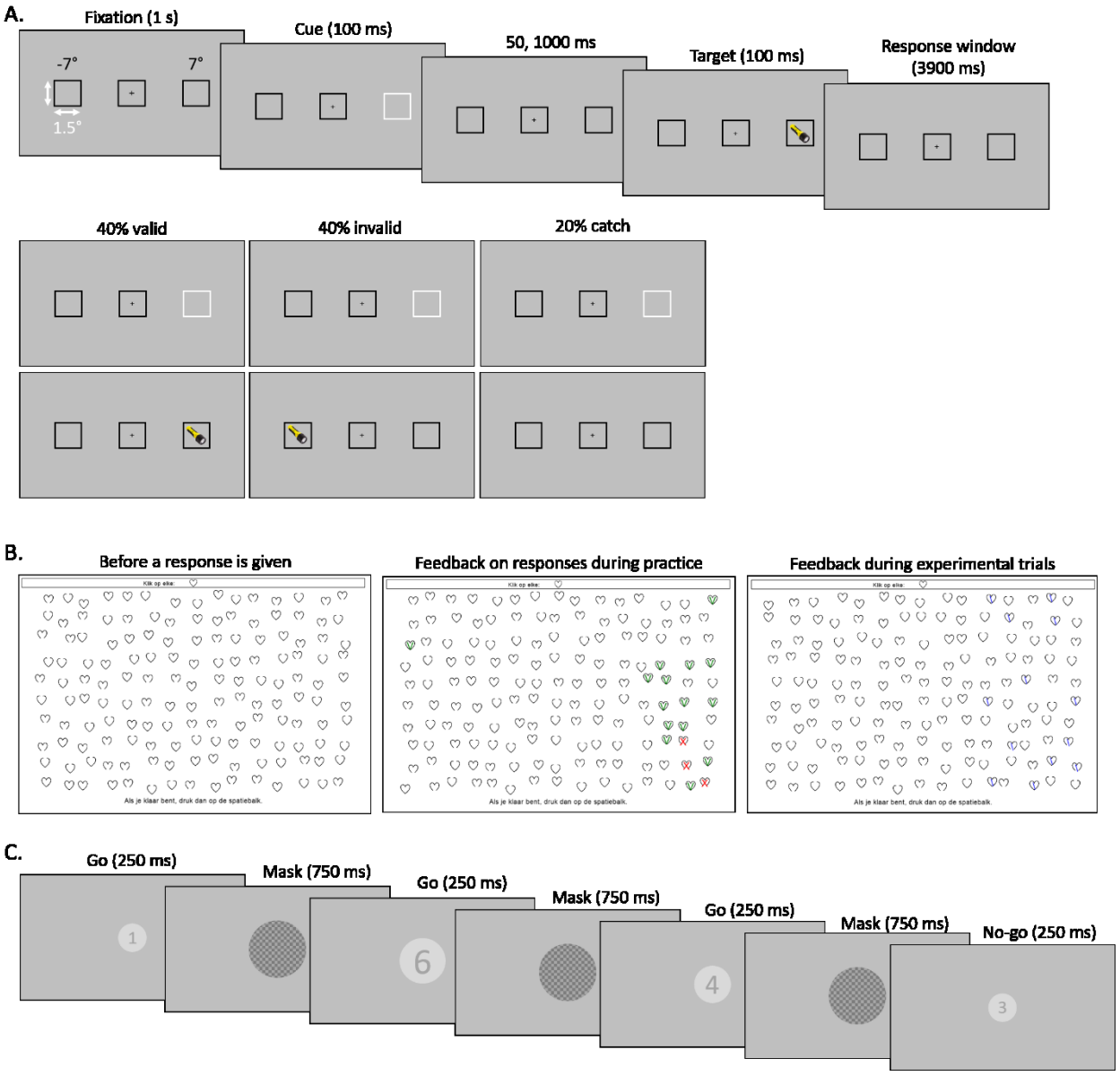

Figure 2. Procedure of Posner task and conditions (A), examples of cancellation arrays of the computerized cancellation task (B) and procedure of the sustained attention to response task (C).

2.3.4.2 Computerized cancellation task

A computerized cancellation task is administered as one of the secondary outcomes. Targets (i.e., full-outlined line drawings) and distractors (i.e., line drawings with upper or lower gap) will be presented in a grid with a width of 28cm and height of 18cm (Figure 2B). At a distance of 70cm, this corresponds to stimuli being placed within a horizontal angle of 11° to the left and right side and a vertical angle of 7.3° to the upper and lower side of the visual field. The grid is divided into 15 equally spaced columns and 10 equally spaced rows. Stimuli are located in the grid at 15 horizontal locations of which their position relative to

the centre of the screen is: -13.01, -11.2, -9.33, -7.47, -5.6, -3.73, -1.87, 0, 1.87, 3.73, 5.6, 7.47, 9.33, 11.2 and 13.01cm. Stimuli are located at 10 vertical locations: -8.1, -6.3, -4.5, -2.7, -0.9, 0.9, 2.7, 4.5, 6.3 and 8.1 cm. The stimuli have a size of 0.9 by 0.9cm. A random amount of jitter is added to make the search array disorganized with a maximum displacement of 0.45cm.

For each trial, 50 targets and 100 distractors will be presented. The target stimuli are spread randomly across the grid. For each set of 3 trials, the target is presented once in each cell. Each trial is presented for 4 minutes or until the patient indicates that he has finished the task by pressing the space bar. The patient is instructed to click on the targets using the left mouse button with their right hand. Once a target has been clicked, a blue line appears on the target. A total of 12 trials are presented. One practice trial is presented in which 50 targets and 100 distractors are shown and feedback is provided. When the patient clicks on a target stimulus, a green "V" sign appears and a 400Hz tone is presented for 150ms. When the patient clicks on a distractor stimulus, a red "X" sign appears and a 200Hz tone is presented for 150ms.

2.3.4.3 Sustained Attention to Response Task

To measure non-spatial attention, we will use the sustained attention to response task (Robertson et al., 1997). Numbers from 1 to 9 will be presented in a randomized order at a frequency of 1 Hz for 250ms with 5 different sizes (Figure 2C). Patients are instructed to press the space bar every time a number appeared and to withhold their response when number '3' appeared. They have to do this task for a duration of 3.75 minutes, completing a total of 225 trials (i.e., 25 no-go trials and 200 go-trials). In addition, 27 practice trials will be presented in which a short beep indicates errors.

2.4 Design

A placebo-controlled longitudinal study will be used. Longitudinal designs are increasingly recognized as a more powerful approach relative to pre-post designs (National Research Council (US) Panel on Handling Missing Data in Clinical Trials, 2010), since they allow to measure symptom trajectories over time. Given the lack of test-retest stability in commonly used measures of neglect (Bailey et al., 2004), this is especially important, as it may provide more reliable assessment of the symptom evolution. In addition, a longitudinal design allows to evaluate the dose-response relation.

We will administer a placebo and an active version of the IVR game to each patient. The order of the placebo and active IVR game will be counterbalanced between subjects, creating two treatment groups (Figure 4A). We will use a minimization algorithm to allocate patients to a treatment group. The first patient is allocated to group A or B using a random number generator. Then, patients will be assigned to

group A or B in a way that minimizes the difference between groups in time since stroke[1]. This minimization approach has been shown to be more effective to balance prognostic factors between treatment groups than stratified randomization for samples sizes of less than 100 patients (Kernan, Viscoli, Makuch, Brass, & Horwitz, 1999). As both groups only differ in the order of treatments, it is important to match the groups on patient characteristics that may influence recovery speed such as time since stroke. Previous research has indeed established a fast recovery of neglect symptoms during the first weeks, and a more stable recovery later on (Nijboer et al., 2013). Matching the two groups on potential moderators of response to treatment (e.g., age, anosognosia) is less important, as we will compare the active and placebo intervention conditions within patients and not between treatment groups.

Patients will not be explicitly informed about the placebo and active intervention. However, complete blinding of patients cannot be guaranteed, as patients may notice a difference between the interventions. For this reason, we will ask patients whether they have noticed a difference at the end of the follow-up session. The clinician who administers the pen-and-paper assessment and completes the behavioural observation scale will be blinded to the treatment group. Blinding of the clinician will be checked by letting them guess the treatment group the patient was allocated to after the follow-up session. Important to note is that the intervention will be added onto the care as usual (e.g., physical, occupational therapy). Given that care as usual can vary between rehabilitation centres and patients (Negrini et al., 2020), we will document the types of therapy and number of therapy hours on a daily basis for each patient.

2.5 PROCEDURE

First, adult stroke patients will be invited to take part in a cognitive screening. When patients provide informed consent, three screening sessions will be administered on consecutive days (Figure 3A). Based on this screening, the eligibility of patients to take part in the clinical trial will be evaluated and patient characteristics that may moderate treatment effects will be recorded. If the proportion of cancelled targets on the left side of the computerized cancellation task is statistically significantly lower, according to a Bayesian contingency table test, than the proportion cancelled targets on the right side, patients will

[1] We will calculate the hypothetical average time since stroke for each treatment group if the new patient would be included in either of the two groups. Then, we allocate the patient to that group that minimizes the difference. For instance, if patients in group A have an average time since stroke of 60 days and group B consists of patients with an average time since stroke of 30 days, a new patient with a time since stroke of 50 days is best allocated to treatment group B.

be considered as left-sided neglect patients. The eligible patients will be invited to take part in the clinical trial.

Patients will be trained to use the IVR game. We will administer a simulator sickness questionnaire (SSQ) before and after IVR exposure and administer a User Experience scale to gauge patient's experience with the game (Figure 3B). In the same week, patients will also complete an IVR assessment, which will be used to tailor the game to each patient. Then, patients will complete 10 1-hour game sessions in the active condition and 10 1-hour game sessions in the placebo condition.

In addition, patients participate in eight weekly assessment sessions to measure the primary outcome (i.e., performance on a Posner task), and four secondary assessment sessions in which a battery of attention tests are administered. After completing the active and placebo intervention, there is a one-week follow-up assessment (Figure 3A). There are 8 parallel versions of the Posner task that only differ in the target shape (Figure 4A). There are five parallel versions of the computerized cancellation task that differ in the line drawings (Figure 4B). The order of the parallel versions will be randomized between participants and matched between treatment group A and B. The IVR assessment is re-administered when patients switch intervention condition and after completing both intervention conditions (Figure 3A).

This study protocol involves a 1-hour daily session. If a session cannot be completed due to practical constraints (e.g., technical problems), we will shift all sessions by 1 day. Through this procedure we aim to minimize missing data.

Details of the data analysis are reported in Appendix A.

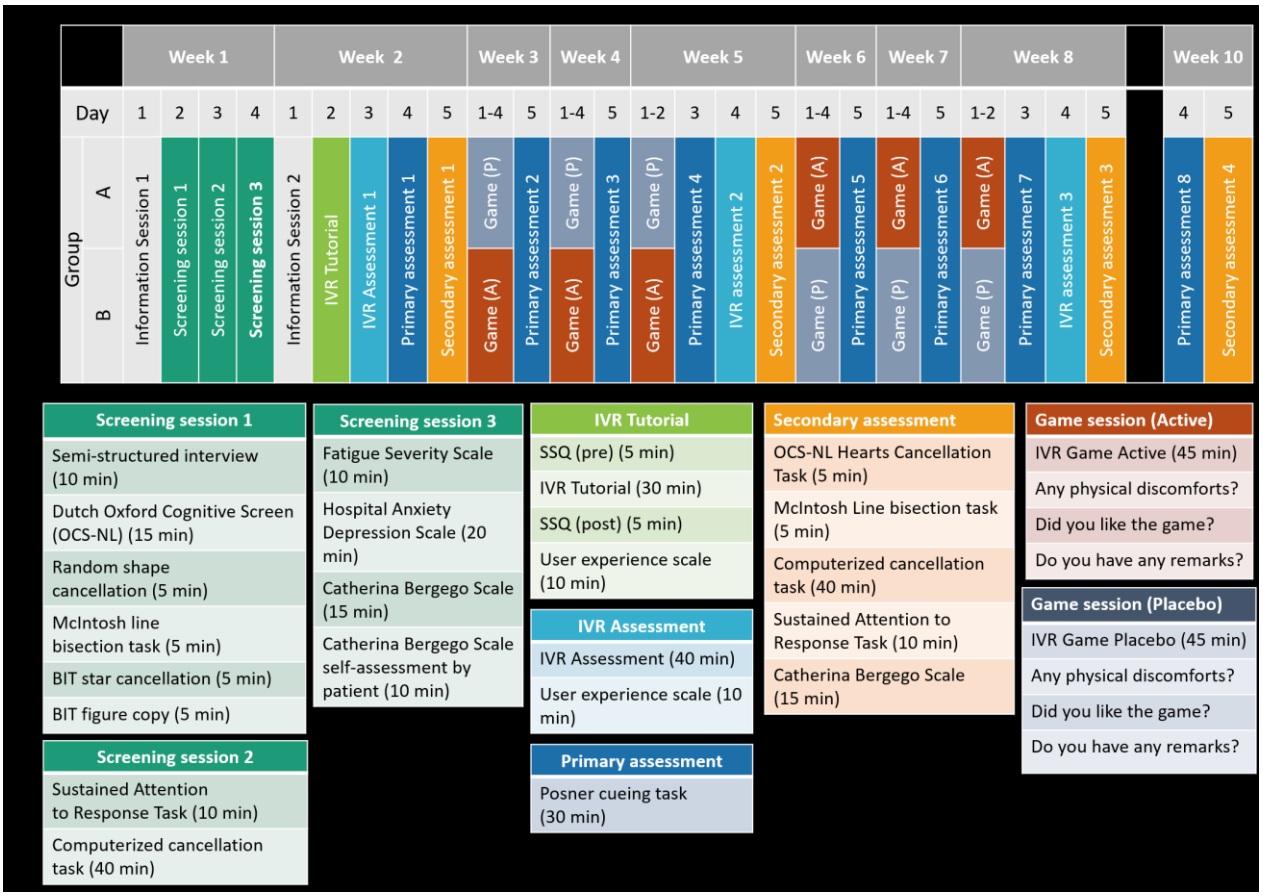

Figure 3. Trial Flowchart (A) and instruments per session (B). SSQ = simulator sickness questionnaire.

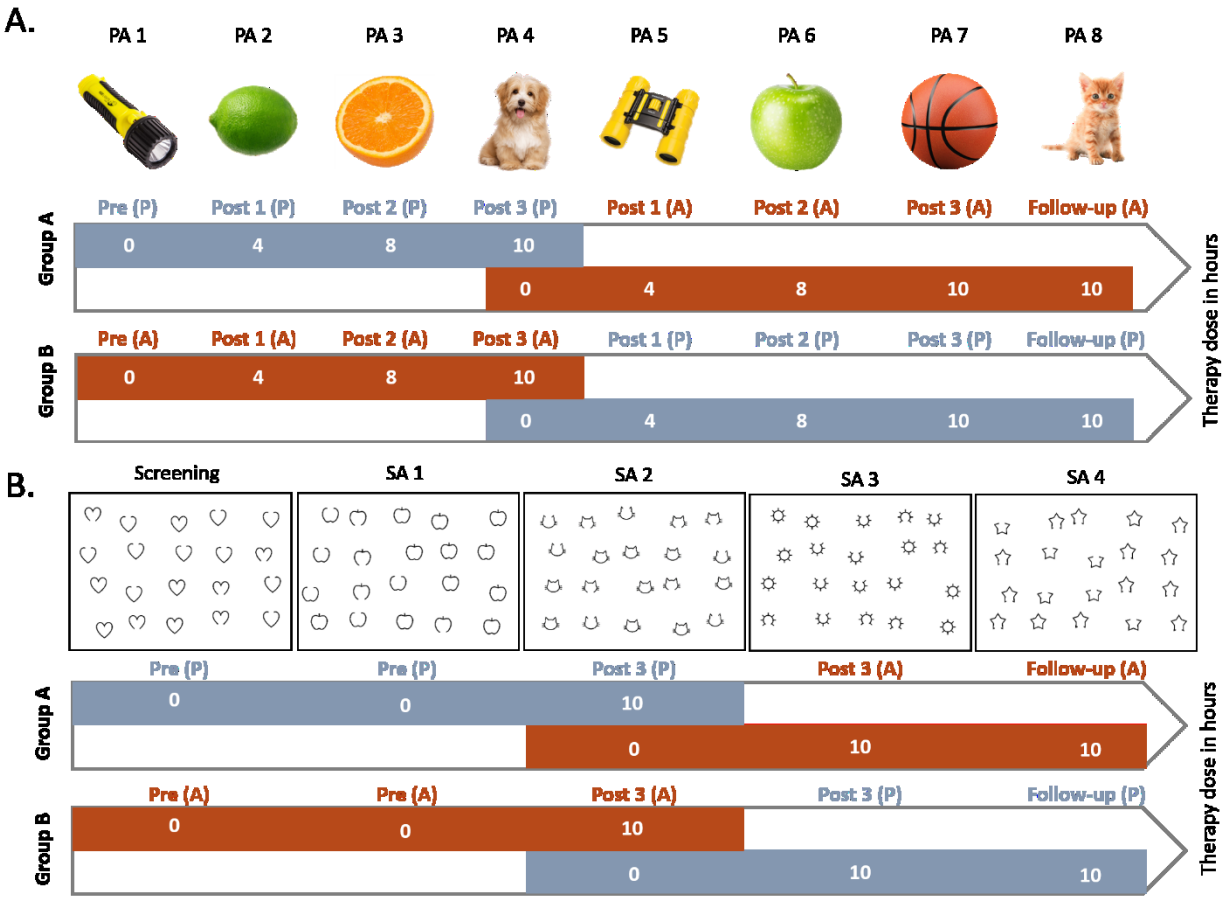

Figure 4. Parallel versions of the Posner task (A) and computerized cancellation task (B). An example of one sequence of the parallel versions is visualized. The sequence of versions will be randomized across participants and matched between group A and B. P = placebo intervention, A = active intervention. PA = primary assessment, SA = secondary assessment.

Acknowledgements

This work was funded by a research grant of the Flemish Fund for Scientific Research (FWO) awarded to H.H. (1171717N, 1171719N), R.v.E. and V.V.A. (G078915N) and C.R.G. (G072517N).

3 REFERENCES

Andersen, S. W., & Millen, B. A. (2013). On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharmaceutical Statistics*, *12*(1), 7–16. https://doi.org/10.1002/pst.1548

Appel, L., Appel, E., Bogler, O., Wiseman, M., Cohen, L., Ein, N., Abrams, H. B., & Campos, J. L. (2020). Older Adults With Cognitive and/or Physical Impairments Can Benefit From Immersive Virtual Reality Experiences: A Feasibility Study. *Frontiers in Medicine*, *6*. https://doi.org/10.3389/fmed.2019.00329

Arns, L. L., & Cerney, M. M. (2005). The relationship between age and incidence of cybersickness among immersive environment users. *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, 267–268. https://doi.org/10.1109/VR.2005.1492788

Azouvi, P., Olivier, S., de Montety, G., Samuel, C., Louis-Dreyfus, A., & Tesio, L. (2003). Behavioral assessment of unilateral neglect: Study of the psychometric properties of the Catherine Bergego Scale. *Archives of Physical Medicine and Rehabilitation*, *84*(1), 51–57. https://doi.org/10.1053/apmr.2003.50062

Bailey, M. J., Riddoch, M. J., & Crome, P. (2004). Test–retest stability of three tests for unilateral visual neglect in patients with stroke: Star Cancellation, Line Bisection, and the Baking Tray Task. *Neuropsychological Rehabilitation*, *14*(4), 403–419. https://doi.org/10.1080/09602010343000282

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69–77. https://doi.org/10.1016/S0022-3999(01)00296-3

Bowen, A., Hazelton, C., Pollock, A., & Lincoln, N. B. (2013). Cognitive rehabilitation for spatial neglect following stroke. In The Cochrane Collaboration (Ed.), *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. http://doi.wiley.com/10.1002/14651858.CD003586.pub3

Burke, J. W., McNeill, M. D. J., Charles, D. K., Morrow, P. J., Crosbie, J. H., & McDonough, S. M. (2009). Optimising Engagement for Stroke Rehabilitation Using Serious Games. *Vis. Comput.*, *25*(12), 1085–1099. https://doi.org/10.1007/s00371-009-0387-4

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Buxbaum, L. J., Ferraro, M. K., Veramonti, T., Farne, A., Whyte, J., Ladavas, E., Frassinetti, F., & Coslett, H. B. (2004). Hemispatial neglect Subtypes, neuroanatomy, and disability. *Neurology*, *62*(5), 749–756. https://doi.org/10.1212/01.WNL.0000113730.73031.F4

Cazzoli, D., Müri, R. M., Schumacher, R., von Arx, S., Chaves, S., Gutbrod, K., Bohlhalter, S., Bauer, D., Vanbellingen, T., Bertschi, M., Kipfer, S., Rosenthal, C. R., Kennard, C., Bassetti, C. L., & Nyffeler, T. (2012). Theta burst stimulation reduces disability during the activities of daily living in spatial neglect. *Brain: A Journal of Neurology*, *135*, 3426–3439. https://doi.org/10.1093/brain/aws182

Checketts, M., Mancuso, M., Fordell, H., Chen, P., Hreha, K., Eskes, G. A., Vuilleumier, P., Vail, A., & Bowen, A. (2020). Current clinical practice in the screening and diagnosis of spatial neglect post-stroke: Findings from a multidisciplinary international survey. *Neuropsychological Rehabilitation*, *0*(0), 1–32. https://doi.org/10.1080/09602011.2020.1782946

Chen, P., Pitteri, M., Gillen, G., & Ayyala, H. (2018). Ask the experts how to treat individuals with spatial neglect: A survey study. *Disability and Rehabilitation*, *40*(22), 2677–2691. https://doi.org/10.1080/09638288.2017.1347720

Corbetta, M., & Shulman, G. L. (2011). Spatial neglect and attention networks. *Annual Review of Neuroscience*, *34*, 569–599. https://doi.org/10.1146/annurev-neuro-061010-113731

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). *Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters*. *167*, 12.

de Haan, B., Clas, P., Juenger, H., Wilke, M., & Karnath, H.-O. (2015). Fast semi-automated lesion demarcation in stroke. *NeuroImage : Clinical*, *9*, 69–74. https://doi.org/10.1016/j.nicl.2015.06.013

Demeyere, N., & Gillebert, C. (2017). Dissociating ego- and allocentric neglect after stroke: Prevalence, laterality and outcome predictors. *BioRxiv*, 200386. https://doi.org/10.1101/200386

Demeyere, N., Gillebert, C., Loftus, L., & Humphreys, G. (2015). Egocentric and allocentric neglect after right and left hemisphere lesions in a large scale neglect study of acute stroke patients: Prevalence and recovery. *Journal of Vision*, *15*(12), 179. https://doi.org/10.1167/15.12.179

Demeyere, N., & Gillebert, C. R. (2019). Ego- and allocentric visuospatial neglect: Dissociations, prevalence, and laterality in acute stroke. *Neuropsychology*, *33*(4), 490–498. https://doi.org/10.1037/neu0000527

Demeyere, N., Riddoch, M. J., Slavkova, E. D., Bickerton, W., & Humphreys, G. W. (2015). The Oxford Cognitive Screen (ocs): Validation of a Stroke-specific Short Cognitive Screening Tool. *Psychological Assessment*, *27*(3), 883–894. https://doi.org/10.1037/pas0000082

Dermody, G., Whitehead, L., Wilson, G., & Glass, C. (2020). The Role of Virtual Reality in Improving Health Outcomes for Community-Dwelling Older Adults: Systematic Review. *Journal of Medical Internet Research*, *22*(6), e17331. https://doi.org/10.2196/17331

Evald, L., Wilms, I., & Nordfang, M. (2020). Assessment of spatial neglect in clinical practice: A nationwide survey. *Neuropsychological Rehabilitation*, *0*(0), 1–16. https://doi.org/10.1080/09602011.2020.1778490

Foerster, R. M., Poth, C. H., Behler, C., Botsch, M., & Schneider, W. X. (2016). Using the virtual reality device Oculus Rift for neuropsychological assessment of visual processing capabilities. *Scientific Reports*, *6*(37016), 1–10. https://doi.org/10.1038/srep37016

Fordell, H., Bodin, K., Bucht, G., & Malm, J. (2011). A virtual reality test battery for assessment and screening of spatial neglect. *Acta Neurologica Scandinavica*, *123*(3), 167–174. https://doi.org/10.1111/j.1600-0404.2010.01390.x

Gelman, A., Meng, X.-L., & Stern, H. (1996). POSTERIOR PREDICTIVE ASSESSMENT OF MODEL FITNESS VIA REALIZED DISCREPANCIES. *Statistica Sinica*, *6*(4), 733–760.

Grattan, E. S., Skidmore, E. R., & Woodbury, M. L. (2018). Examining Anosognosia of Neglect. *OTJR : Occupation, Participation and Health*, *38*(2), 113–120. https://doi.org/10.1177/1539449217747586

Halligan, P. W., Cockburn, J., & Wilson, B. A. (1991). The behavioural assessment of visual neglect. *Neuropsychological Rehabilitation*, *1*(1), 5–32. https://doi.org/10.1080/09602019108401377

Harvey, M. (2019). *Neglect rehabilitation: Strength and challenges of treatments using non-invasive brain stimulation*. 7th scientific meeting of the Federation of the European Societies of Neuropsychology, Milan, Italy. https://www.fesn2019.it/relations/Harvey_Monika_FESN_2019.pdf

Husain, M., & Rorden, C. (2003). Non-spatially lateralized mechanisms in hemispatial neglect. *Nature Reviews Neuroscience*, *4*(1), 26–36. https://doi.org/10.1038/nrn1005

Huygelier, H., Moore, M. J., Demeyere, N., & Gillebert, C. R. (2020). Non-spatial impairments affect false positive neglect diagnosis based on cancellation tasks. *Journal of the International Neuropsychological Society*, *26*, 668–678. https://doi.org/10.1017/S1355617720000041

Huygelier, H., Schraepen, B., Demeyere, N., & Gillebert, C. R. (2019). The Dutch version of the Oxford Cognitive Screen (OCS-NL): Normative data and their association with age and socio-economic status. *Aging, Neuropsychology, and Cognition*, 1–22. https://doi.org/10.1080/13825585.2019.1680598

Huygelier, H., Schraepen, B., Ee, R. van, Abeele, V. V., & Gillebert, C. R. (2019). Acceptance of immersive head-mounted virtual reality in older adults. *Scientific Reports*, *9*(1), 4519. https://doi.org/10.1038/s41598-019-41200-6

Huygelier, H., Schraepen, B., Lafosse, C., Vaes, N., Schillebeeckx, F., Michiels, K., Note, E., Abeele, V. V., Ee, R. van, & Gillebert, C. R. (2020). An immersive virtual reality game to train spatial attention orientation after stroke: A feasibility study. *Applied Neuropsychology: Adult*, *0*(0), 1–21. https://doi.org/10.1080/23279095.2020.1821030

Kass, R. E., & Raftery, Adrian. E. (1995). Bayes Factors. *Journal of the Americal Statistical Association*, *90*(430), 773–795.

Kerkhoff, G., & Schenk, T. (2012). Rehabilitation of neglect: An update. *Neuropsychologia*, *50*(6), 1072–1079. https://doi.org/10.1016/j.neuropsychologia.2012.01.024

Kolodziej, W., & Gillebert, C. R. (2018). *Diagnostiek en revalidatie van hemispatiaal neglect in Vlaanderen en Nederland*. KU Leuven.

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Technological Competence Is a Pre-condition for Effective Implementation of Virtual Reality Head Mounted Displays in Human

Neuroscience: A Technological Review and Meta-Analysis. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00342

Laplane, D., & Degos, J. D. (1983). Motor neglect. *Journal of Neurology, Neurosurgery, and Psychiatry*, *46*(2), 152–158.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, *10*(3), 282–297. https://doi.org/10.1162/105474601300343612

Losier, B. J. W., & Klein, R. M. (2001). A review of the evidence for a disengage deficit following parietal lobe damage. *Neuroscience & Biobehavioral Reviews*, *25*(1), 1–13. https://doi.org/10.1016/S0149-7634(00)00046-4

Luauté, J., Halligan, P., Rode, G., Rossetti, Y., & Boisson, D. (2006). Visuo-spatial neglect: A systematic review of current interventions and their effectiveness. *Neuroscience & Biobehavioral Reviews*, *30*(7), 961–982. https://doi.org/10.1016/j.neubiorev.2006.03.001

Luvizutto, G. J., Bazan, R., Braga, G. P., Resende, L. A. de L., Bazan, S. G. Z., & Dib, R. E. (2015). Pharmacological interventions for unilateral spatial neglect after stroke. *Cochrane Database of Systematic Reviews*, *11*. https://doi.org/10.1002/14651858.CD010882.pub2

Machner, B., Mah, Y.-H., Gorgoraptis, N., & Husain, M. (2012). How reliable is repeated testing for hemispatial neglect? Implications for clinical follow-up and treatment trials. *J Neurol Neurosurg Psychiatry*, *83*(10), 1032–1034. https://doi.org/10.1136/jnnp-2012-303296

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58. https://doi.org/10.1080/02701367.1989.10607413

McIntosh, R. D. (2017). *The end of the line: Antagonistic attentional weightings in unilateral neglect* [Article in Press]. Scopus. https://doi.org/10.1016/j.cortex.2017.07.011

McIntosh, R. D., Ietswaart, M., & Milner, A. D. (2017). Weight and see: Line bisection in neglect reliably measures the allocation of attention, but not the perception of length. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2017.09.014

McIntosh, R. D., Schindler, I., Birchall, D., & Milner, A. D. (2005). Weights and measures: A new look at bisection behaviour in neglect. *Cognitive Brain Research*, *25*(3), 833–850. https://doi.org/10.1016/j.cogbrainres.2005.09.008

Milner, A. D., & Harvey, M. (1994). Toward a taxonomy of spatial neglect. *Neuropsychological Rehabilitation*, *4*(2), 177–181.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*(mar23 1), c869–c869. https://doi.org/10.1136/bmj.c869

Nadarajah, M., Mazlan, M., & Goh, H. (2017). Test-retest reliability, internal consistency and concurrent validity of Fatigue Severity Scale in measuring post-stroke fatigue. *European Journal of Physical and Rehabilitation Medicine*, *53*(5), 7.

National Research Council (US) Panel on Handling Missing Data in Clinical Trials. (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK209904/

Negrini, S., Arienti, C., & Kiekens, C. (2020). Usual care: The big but unmanaged problem of rehabilitation evidence. *The Lancet*, *395*(10221), 337. https://doi.org/10.1016/S0140-6736(19)32553-X

Nijboer, T. C., Kollen, B. J., & Kwakkel, G. (2013). Time course of visuospatial neglect early after stroke: A longitudinal cohort study. *Cortex*, *49*(8), 2021–2027.

Oculus Rift | Oculus. (n.d.). Retrieved February 27, 2020, from https://www.oculus.com/rift/#oui-csl-rift-games=mages-tale

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Plechatá, A., Sahula, V., Fayette, D., & Fajnerová, I. (2019). Age-Related Differences With Immersive and Non-immersive Virtual Reality in Memory Assessment. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01330

Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., & Valente, L. (2017). Minimizing cyber sickness in head mounted display systems: Design guidelines and applications. *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*, 1–6. https://doi.org/10.1109/SeGAH.2017.7939283

Rizzo, A. A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation*, *14*(1–2), 207–239.

Robertson, I. H. (2001). Do we need the "lateral" in unilateral neglect? Spatially nonselective attention deficits in unilateral neglect and their implications for rehabilitation. *Neuroimage*, *14*(1), S85–S90.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). `Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization. *Neuroimage*, *61*(4), 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Salazar, A. P. S., Vaz, P. G., Marchese, R. R., Stein, C., Pinto, C., & Pagnussat, A. S. (2018). Noninvasive Brain Stimulation Improves Hemispatial Neglect After Stroke: A Systematic Review and Meta-Analysis. *Archives of Physical Medicine and Rehabilitation*, *99*(2), 355-366.e1. https://doi.org/10.1016/j.apmr.2017.07.009

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Stanney, K. M., & Hash, P. (1998). Locus of User-Initiated Control in Virtual Environments: Influences on Cybersickness. *Presence: Teleoperators and Virtual Environments*, *7*(5), 447–459. https://doi.org/10.1162/105474698565848

Sturm, W., Thimm, M., Binkofski, F., Horoufchin, H., Fink, G. R., Küst, J., Karbe, H., & Willmes, K. (2013). Combined Space and Alertness Related Therapy of Visual Hemineglect: Effect of Therapy Frequency. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00373

Ten Brink, A. F., Verwer, J. H., Biesbroek, J. M., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2017). Differences between left- and right-sided neglect revisited: A large cohort study across multiple domains. *Journal of Clinical and Experimental Neuropsychology*, *39*(7), 707–723. https://doi.org/10.1080/13803395.2016.1262333

Ten Brink, A. F., Visser-Meily, J. M. A., Schut, M. J., Kouwenhoven, M., Eijsackers, A. L. H., & Nijboer, T. C. W. (2017). Prism Adaptation in Rehabilitation? No Additional Effects of Prism Adaptation on Neglect Recovery in the Subacute Phase Poststroke: A Randomized Controlled Trial. *Neurorehabilitation and Neural Repair*, *31*(12), 1017–1028. https://doi.org/10.1177/1545968317744277

Ten Brink, A., Nijboer, T., Van Beekum, L., Van Dijk, J., Peeters, R., Post, M., & Visser, J. (2013). De Nederlandse Catherine Bergego schaal: Een bruikbaar en valide instrument in de CVA-zorg. *Wetenschappelijk Tijdschrift Voor Ergotherapie*, *6*(3), 27–36.

Tieri, G., Morone, G., Paolucci, S., & Iosa, M. (2018). Virtual reality in cognitive and motor rehabilitation: Facts, fiction and fallacies. *Expert Review of Medical Devices*, *15*(2), 107–117. Scopus. https://doi.org/10.1080/17434440.2018.1425613

Tobler-Ammann, B. C., Surer, E., Knols, R. H., Borghese, N. A., & Bruin, E. D. de. (2017). User Perspectives on Exergames Designed to Explore the Hemineglected Space for Stroke Patients With Visuospatial Neglect: Usability Study. *JMIR Serious Games*, *5*(3), e18. https://doi.org/10.2196/games.8013

Tuena, C., Pedroli, E., Trimarchi, P. D., Gallucci, A., Chiappini, M., Goulene, K., Gaggioli, A., Riva, G., Lattanzio, F., Giunco, F., & Stramba-Badiale, M. (2020). Usability Issues of Clinical and Research Applications of Virtual Reality in Older People: A Systematic Review. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00093

Van den Noortgate, W., & Onghena, P. (2003). Combining Single-Case Experimental Data Using Hierarchical Linear Models. *School Psychology Quarterly Fall 2003*, *18*(3), 325–346. https://doi.org/10.1521/scpq.18.3.325.22577

van der Kemp, J., Dorresteijn, M., Ten Brink, A. F., Nijboer, T. C. W., & Visser-Meily, J. M. A. (2017). Pharmacological Treatment of Visuospatial Neglect: A Systematic Review. *Journal of Stroke and Cerebrovascular Diseases*, *26*(4), 686–700. https://doi.org/10.1016/j.jstrokecerebrovasdis.2017.02.012

Van der Stoep, N., Visser-Meily, J. M. A., Kappelle, L. J., de Kort, P. L. M., Huisman, K. D., Eijsackers, A. L. H., Kouwenhoven, M., Van der Stigchel, S., & Nijboer, T. C. W. (2013). Exploring near and far regions of space: Distance-specific visuospatial neglect after stroke. *Journal of Clinical and Experimental Neuropsychology*, *35*(8), 799–811. https://doi.org/10.1080/13803395.2013.824555

van Lieshout, E. C. C., van Hooijdonk, R. F., Dijkhuizen, R. M., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2019). The Effect of Noninvasive Brain Stimulation on Poststroke Cognitive Function: A Systematic Review. *Neurorehabilitation and Neural Repair*, *33*(5), 355–374. https://doi.org/10.1177/1545968319834900

Van Vleet, T. M., & DeGutis, J. M. (2013). The nonspatial side of spatial neglect and related approaches to treatment. *Progress in Brain Research*, *207*, 327–349. https://doi.org/10.1016/B978-0-444-63327-9.00012-6

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00158

4 APPENDIX A: DATA-ANALYSIS

4.1 DESCRIPTION OF PATIENT SAMPLE

We will report the mean, median, standard deviation and range of a patient's age in years, years of formal education, and time since stroke (i.e., days between hospital admission and first screening session). We will report the number of men and women. We will also report the results from the measure of anosognosia, post-stroke depression and post-stroke fatigue. We will report on which subtests patients performed lower than age-adjusted cut-offs on the OCS-NL and how patients performed on the neglect tasks at screening. In addition, we will report the number of patients with ischemic versus haemorrhagic stroke and visualize the lesion locations. Lesions will be delineated on clinical computer tomography (CT) or Fluid-attenuated inversion recovery (FLAIR) or T2-weighted magnetic resonance imaging (MRI) scans using the Matlab Clusterize toolbox (de Haan et al., 2015) or manually using MRIcron. Scans will be converted from native to MNI space using age-specific CT and MRI templates of the Matlab SPM clinical toolbox (Rorden et al., 2012). Moreover, we will test whether there was a difference between the time since stroke between treatment groups A and B using a Bayesian paired t-test. These data will be reported for the per-protocol and intention-to-treat samples.

4.2 QUALITY MONITORING OF CLINICAL TRIAL PROCEDURE AND MISSING DATA

We will report the number of protocol deviations, type of deviation and reason for deviations (e.g., technical problems, patients not available). In addition, to evaluate the extent to which patients and assessors were blinded, we will report the percentage of accurate guesses of the intervention conditions. Furthermore, we will report the number of hours and types of daily therapy (e.g., occupational therapy, physical rehabilitation) that patients received in the rehabilitation unit. Last, we will report the ratio of contra- and ipsilesional stimulation in the IVR game and the total number of hours that patients used the IVR game in the placebo and active IVR conditions. Any technical problems with the IVR game will also be reported.

Our efficacy analyses will be performed on the *per-protocol sample* (i.e., patients who complete at least 80% sessions), because it is crucial that there is no difference in the number of completed assessments between the placebo and active intervention for our within-subject comparison. To assess the extent to which the per-protocol sample represents the *intention-to-treat sample*, we will report on the prevalence of early drop-out and the missing data pattern. In addition, we will report the reasons for early drop-out.

4.3 PLANNED ANALYSES

4.3.1 What are the effects of an active IVR intervention compared to a placebo IVR intervention and its relationship with therapy dose on neglect symptoms outside the IVR environment?

Our *primary outcome measure* are the response times on invalid-cued targets for the shortest SOA on the Posner task. We will test whether there is a stronger decrease in the difference between left- and right-sided performance on the Posner task for the active compared to the placebo intervention. Moreover, we will test whether performance improves or remains stable for the left and right visual field. In addition, we will evaluate the relation between therapy dose and the change in the left- and right-sided performance. To this end, we will estimate the primary outcome as a function of the main effects of therapy dose (i.e., 4, 8 and 10 hours of intervention), type of intervention (i.e., placebo and active) and visual field (i.e., left and right side), their pairwise interactions and three-way interaction.

If there is no evidence in favour of a group treatment effect for our primary outcome, we will explore whether there is evidence in favour of between-subject differences in the treatment effect. To inform future research, we will then explore whether between-subject differences in the treatment effect may be related to anosognosia, post-stroke depression, post-stroke fatigue and lesion neuro-anatomy.

We will additionally evaluate the treatment effect on our *secondary outcomes*: the proportion cancelled targets on the computerized cancellation task, proportion cancelled targets on the OCS-NL hearts cancellation task, CBS score and EWB score on the McIntosh line bisection task. For the secondary outcomes, we will test the main effect of the test moment (i.e., pre- or post-treatment), type of intervention and the effect of visual field (i.e., left versus right), their pairwise interactions and three-way interaction.

4.3.2 Does non-spatial attention outside the IVR environment recover due to IVR spatial attention training?

Additionally, we will evaluate whether potential improvements as a result of the IVR training are specific to spatial attention or generalize to non-spatial attention. To this end, we will evaluate whether patients improve on the sustained attention to response task more during the active than placebo intervention. To evaluate this hypothesis, we will estimate performance on the sustained attention to response task as a function of the test moment and intervention type and their interaction.

4.3.3 Are training effects inside the IVR environment larger than outside the IVR environment?

If patients do not improve on any of our outcomes obtained outside the IVR environment, it is important to establish whether patients improved inside the IVR environment. The latter can clarify whether training effects in the IVR environment may not have transferred. To test this, we will compare the effect of our intervention between two within-subject conditions: the IVR assessment and the computerized cancellation task. We selected these two outcome measures, as performance on these two tasks are measured on a similar scale (i.e., proportion detected targets). To test our hypothesis, we will estimate the proportion detected targets as a function of the main effect of test moment, intervention, visual field (i.e., left versus right) and task and include their interactions.

Additionally, to assess the extent to which patients improve inside the IVR environment, we will evaluate whether patient's make more head movements towards the left visual field in the IVR assessment after the IVR training compared to before. We hypothesize a stronger increase in left-oriented head movements as a result of the active than the placebo IVR training. To test this hypothesis, we will estimate horizontal head orientation (i.e., negative values indicate left-sided and positive values right-sided orientations) as a function of the interaction of intervention condition and test moment.

4.3.4 Is IVR rehabilitation feasible?

We will evaluate whether self-reported cybersickness and user experience predict the number of sessions completed by patients during the clinical trial. We will also compare cybersickness before and after IVR exposure and report descriptive statistics of the User Experience scale and of the user evaluations at the end of each game session. Moreover, we will document the number of referred and successfully recruited patients and reasons for excluding patients, following the Consolidated Standards of Reporting Trials guidelines (Moher et al., 2010). These feasibility results will be reported for the intention-to-treat sample.

4.4 General modelling approach

We will analyse the data with Bayesian mixed models using the R brms package (Bürkner, 2017), since mixed models can accurately model time-unstructured data (Andersen & Millen, 2013; Van den Noortgate & Onghena, 2003). Moreover, mixed models can clarify whether there were significant between-subject differences in treatment effects. To estimate the effect of the treatment on our secondary outcomes, we will use a multivariate regression model. The latter allows to statistically test for differences in effects between the secondary outcome measures. If the multivariate regression model does not fit well, we will use separate regression models for each outcome. We will use a lognormal or shifted lognormal

distribution to estimate response times, which are typically skewed, and logistic regression to estimate proportion correct. We will evaluate model fit using posterior predictive checks (Gelman et al., 1996). We prefer a Bayesian approach, because it allows to quantify the strength of evidence in favour of the null hypothesis (Wagenmakers, 2007).

In addition, we will calculate the Bayes Factor (BF) for contrasts using the paired t-tests of the Bayes Factor package (Rouder et al., 2009). We will interpret the BF according to the rule of Kass and Raftery (1995). A BF_{10} larger than 3 suggests substantial evidence and larger than 10 suggests strong evidence in favour of the alternative model. BF_{10} smaller than 1/3 represents substantial and smaller than 1/10 represents strong evidence for the null model. BF_{10} in between 1/3 and 3 represent inconclusive evidence.

4.5 SAMPLE SIZE DETERMINATION

Efficient data collection in neglect rehabilitation research is important, as it is very difficult to recruit sufficient patients (Harvey, 2019). For this reason, we will use a sequential Bayes Factor design (Schönbrodt et al., 2017), in which we sequentially calculate the BF_{10} for our *main contrast of interest* to determine when to stop data collection. We predict a smaller difference between left- and right-sided response times to invalid-cued targets for the shortest stimulus onset asynchrony (SOA) in the Posner task (*R-L score*) after than before active treatment and we expect no difference in the R-L score after than before placebo treatment. Thus, our *main contrast of interest* is the pre-post active intervention difference in the R-L score compared to the pre-post placebo intervention difference in the R-L score. We will evaluate our main contrast of interest after 8, 16, 20, and 24 patients who completed at least 80% of the trial sessions. If the BF_{10} exceeds a threshold of 10 or 0.1 before reaching a sample size of 24 patients, we will stop data collection. Otherwise, we stop data collection when reaching a sample size of 24 patients.

To evaluate the probability of obtaining inconclusive or misleading evidence, we ran simulations following the principles of Schönbrodt and Wagenmakers (2018). First, we simulated response times on invalid-cued left and right targets for the 150ms SOA on the Posner task. To obtain realistic estimates of the averages, between-subject, within-subject variances and associations between the variables, we based our simulated data on Posner data of 5 stroke patients (Table A1). We sampled effect sizes from a normal distribution (M=0.2, SD=0.10). The treatment effect was scaled for each case using the within-subject variance in response times. Thus, patients with a higher pre-treatment mean response time are expected to have a larger decrease in response times. This corresponded to a reduction of left-sided response times of 96ms due to active treatment at the group level (Cohen's d = 0.28).

We estimated the BF_{10} using a Bayesian paired t-test of the Bayes Factor package (Rouder et al., 2009) for our main contrast of interest for 50.000 samples under the alternative hypothesis and 5000 samples under the null hypothesis, each consisting of 24 cases. These simulations revealed that 66% of trials under the alternative hypothesis resulted in a BF_{10} larger than 10 and 78.7% trials resulted in a BF_{10} larger than 3. A total of 55% trials reached a BF_{10} larger than 10 at 8 patients, 0.1% at 16, 0.02% at 20 and 11% at 24 patients. A total of 18.7% trials resulted in inconclusive evidence at 24 participants. None of the trials resulted in strong evidence and 2.6% trials resulted in moderate-strength evidence in favour of the null hypothesis. If the null hypothesis was true, none of the BF_{10} exceeded the 0.10 threshold, while 1% of trials resulted in a BF_{10} exceeding 10 (i.e., false positive result) before reaching 24 participants. A total of 65% trials reached a BF_{10} smaller than 1/3 at 24 participants, 1.4% resulted in a BF_{10} larger than 3 (i.e., false positive result) and 33% trials reached an inconclusive BF at 24 participants. We evaluated whether a sample size of 30 patients would produce better results, but a total of 27% of trials still reached an inconclusive BF at 30 patients.

Thus, the probability of a false negative result is 2.6% and the probability of a false positive result is 2.4%. In the Frequentist framework a power of 90% would correspond to a 10% probability of a false negative result. Thus, in comparison, our study has a low probability of producing false results.

Table A1. Characteristics and performance on the Posner task of 5 pilot participants.

Patient	Patient characteristics				Response times M (SD)				Proportion detected M (SD)			
	Age (y)	Gender	TSS	Stroke side	L I	L V	R I	R V	L I	L V	R I	R V
P1*	62	M		RH	1.87 (1.28)	1.85 (1.34)	0.61 (0.17)	0.59 (0.36)	.20 (.41)	.33 (.47)	1.0 (.0)	1.0 (.0)
P2	48	M	48	RH	0.44 (0.03)	0.43 (0.06)	0.39 (0.06)	0.38 (0.03)	1.0 (.0)	1.0 (.0)	1.0 (.0)	1.0 (.0)
P3	58	F	91	RH	0.92 (0.48)	0.77 (0.37)	0.76 (0.38)	0.62 (0.17)	1.0 (.0)	1.0 (.0)	.98 (.16)	1.0 (.0)
P4	65	F	129	RH	1.18 (0.67)	0.81 (0.39)	0.73 (0.47)	0.54 (0.14)	.98 (.16)	.95 (.22)	1.0 (.0)	1.0 (.0)
P5	44	M	113	RH	0.70 (0.33)	0.66 (0.25)	0.52 (0.15)	0.59 (0.24)	.80 (.41)	.73 (.45)	.76 (.42)	.95 (.22)

Note. y = years, TSS = time since stroke in days, M = mean, SD = standard deviation, RH = right-hemispheric, LH = left-hemispheric, L I = left invalid, L V = left valid, R I = right invalid, R V = right valid. * The CT or MRI scan or radiologist report were not available to determine the date of stroke.