Clinical evaluation of a non-immersive virtual environment in stroke rehabilitation
Judi Edmans, John Gladman, Dave Hilton, Marion Walker, Alan Sunderland, Sue Cobb, Tony Pridmore
and Shirley Thomas
Clin Rehabil 2009; 23; 106
DOI: 10.1177/0269215508095875

The online version of this article can be found at:
http://cre.sagepub.com/cgi/content/abstract/23/2/106

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Clinical Rehabilitation can be found at:

Email Alerts: http://cre.sagepub.com/cgi/alerts

Subscriptions: http://cre.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.co.uk/journalsPermissions.nav

Citations http://cre.sagepub.com/cgi/content/refs/23/2/106

Downloaded from http://cre.sagepub.com at UNIV OF NOTTINGHAM on April 26, 2010
Clinical evaluation of a non-immersive virtual environment in stroke rehabilitation

Judi Edmans, John Gladman Division of Rehabilitation & Ageing, Dave Hilton VIRART, Marion Walker Division of Rehabilitation & Ageing, Alan Sunderland School of Psychology, Sue Cobb VIRART, Tony Pridmore School of Computer Science and Shirley Thomas Division of Rehabilitation & Ageing, University of Nottingham

Received 25th January 2008; returned for revisions 2nd March 2008; revised manuscript accepted 16th June 2008.

Background and purpose: We describe our attempts to evaluate the effectiveness of a virtual environment developed to rehabilitate stroke patients in the task of making a hot drink.

Methods: Single case studies were performed in 13/138 (9%) stroke patients undergoing rehabilitation in a UK stroke unit. Participants in AB/BA (n = 5) and ABA (n = 2) design studies received 5 one-hour sessions of attention control training (A phase) and 5 one-hour sessions of virtual environment training (B phase). An AB design with random duration of A and B phases (minimum duration of A and B phases 3 and 5 days respectively, with total duration of 3 weeks) was used in 6 participants.

Results: Visual inspection of scores across all cases showed a trend towards improvement over time in both real and virtual hot drink making ability in both control and intervention phases. There was no significant difference (Wilcoxon, \( p > 0.05 \)) in the improvements in real and virtual hot drink making ability during all control and intervention phases in the 13 cases. Ceiling effects limited the evaluation of effectiveness in 5 of the 8 cases in which daily performance measures were used.

Conclusions: Few people in this setting were suitable for this intervention. The case studies showed no evidence of a strong effect of this intervention, but we had great difficulty in performing single case studies. We conclude that more testing and development of this system is required before it is subjected to rigorous testing of clinical effectiveness.

Introduction

Virtual reality is a computer-generated technology that allows the user to interact with a simulated three-dimensional environment (the virtual environment). Virtual environments are used either to replicate real world settings and objects or visualize imaginary or abstract concepts. The concept of virtual reality was developed in the early 1960s but it has only been since the 1990s that the technology was sufficiently powerful and affordable to be considered for use in rehabilitation, and it remains in the early stages of development in
this role. The popular view of virtual reality reflects the immersive form of it, where the user is enveloped within the virtual environment, for example, using a head-mounted display. However, in practice, virtual environments are usually experienced in two dimensions on an ordinary computer screen: this is referred to as non-immersive virtual reality. Early experience in clinical settings with immersive virtual reality noted considerable problems with dizziness and nausea, and so attention in the rehabilitation field has tended to focus upon non-immersive virtual reality.

The reasons to consider the use of any form of virtual reality technology in rehabilitation include the following:

- Potentially dangerous activities can be practised in safety in a virtual environment.
- Virtual reality systems can potentially be used by patients without direct support from a therapist. This could enable the rehabilitation task to be repeated, which could lead to greater efficacy.
- Virtual reality systems can be designed to deliver the principles of neuro-rehabilitation. For example, in virtual reality, the rehabilitation task can be made progressively harder as the patient improves, it can be deliberately constrained to prevent compensatory actions by the patient, or it could be constrained only to permit errorless learning.
- Virtual systems are often enjoyable and this may motivate the patient towards their rehabilitation.

Two steps are necessary before virtual reality-based systems for rehabilitation can be used in clinical practice. First, a useable system has to be developed. Second, its effectiveness in clinical practice has to be established. The standard research methodology to develop a useable system is user-centred design. This is an iterative, developmental process whereby a concept and then successive prototypes of the system are discussed with, shown to, and used by all those who will be involved in the eventual product. The point at which the developmental process is halted often depends upon when it is judged that a product suitable for market or application has been designed. However, systems used in clinical settings need to undergo clinical effectiveness testing before they can be widely adopted. In this paper, we describe our experience of the early clinical evaluation of a non-immersive virtual reality approach to stroke rehabilitation and reflect upon our findings and the methodological issues involved in moving from the evaluation of a product during its development to the evaluation of its clinical effectiveness.

Development of the virtual reality rehabilitation system

Our project aimed to develop and evaluate a virtual environment for the rehabilitation of making a hot drink after stroke. User-centred design was used to identify the hot drink-making task, by consulting patients and therapists. Early versions of the system were developed, using user-centred design, with particular focus upon the interface between the patient and the system. We found that stroke patients did not want to be asked to use a system that they thought would require pre-existing computer expertise, as many of them were not computer literate. Furthermore, rehabilitation therapists wished to use naturalistic interfaces with the virtual environment, such as using real hot drink-making objects, so that not only the cognitive aspects of task performance could be trained but also the physical aspects. This led us to explore the use of machine vision so that the movement of real objects could be ‘seen’ by the system and so control the making of a virtual hot drink in the virtual environment. We have reported on these developments elsewhere.

Not only did we need to explore the use of different interfaces with the virtual environment, at the same time we had to develop the virtual environment itself, and ensure that it would support the task of making a virtual hot drink. A virtual environment was constructed using a standard software programme (Virtools™), where the necessary objects (cup, spoon, kettle, etc.) were created with properties necessary for their use in the virtual task (for example, the lid of the instant coffee jar needed to be removed before the spoon could be inserted). The use of these objects was defined and constrained by a task model, representing how
The task model was constructed so that subtasks could be done in any order that was practical and safe. The virtual environment was programmed to replicate how an occupational therapist would give feedback to the patient in terms of whether each subtask had been performed correctly or not, whether it has already been done or whether another subtask needed to be done first. An audible instruction could be given if the patient did not complete a subtask despite these prompts, and the subtask was demonstrated if the patient was still unable to complete it. This virtual environment therefore allowed for repetition of the task, it imposed a degree of constraint to task performance and encouraged errorless learning. Since it could be timed and scored, patients could be encouraged to perform the task more quickly and more accurately over time. At this stage of development, we felt that the virtual environment was suitable to support the rehabilitation of the task of making a hot drink.

The first working version of this system did not use a naturalistic interface, because these had not been sufficiently developed, but used a simple touch-sensitive computer screen with which the user interacted using a special pointer. At this stage, we had developed a system that was theoretically capable of supporting only the rehabilitation of the cognitive aspects of making a hot drink (for example, object recognition, sequencing and attention). Even though we had not developed a full system with a naturalistic interface, it was plausible that this version with the touch-sensitive screen interface might have some value in clinical rehabilitation. Our first clinical tests of this system established that performance on this virtual task by stroke patients undergoing rehabilitation was similar to making a real drink in that it was influenced by similar impairments and patient factors. However, making a virtual hot drink was harder than making a hot drink in a rehabilitation kitchen, and it was evident from the comparison of video-recordings that people with stroke made different cognitive mistakes when making a real and virtual hot drink. This indicated that the cognitive functions tested and trained in the virtual task may be different from those tested and trained in a real kitchen.

One option for the team would have been to continue to develop the virtual environment further before further clinical testing – either to develop the naturalistic interface or to alter the task model, or both. However, we decided to undertake clinical testing of the system with the touch-sensitive screen at this stage. Our reasons for doing so were that it was possible that our simple system even at this stage could be clinically effective, but also that it was important to establish how clinically feasible it was to use this system as a treatment in a clinical setting, and also to gather data that would inform later clinical evaluations. We reflect on these reasons in the Discussion in the light of our experience.

**Method**

We initially planned to undertake a pilot single blind randomized controlled trial (RCT) of the touch-sensitive screen version of the virtual environment. However, we found the recruitment rate too low to allow an adequate sample size to be obtained. Accordingly the research plan was amended and single case studies were undertaken to optimize the information available from the small number of available patients. We reflect upon these methodological decisions in the Discussion. In this paper, we report the results of the first participants enrolled into the RCT as single cases, and present their outcomes as if they were AB or BA case studies (where A and B represent control and treatment phases). These are described as ‘Design I’ studies. Once we stopped recruiting to the RCT, we chose to employ aspects of single case design in the subsequent case studies. Two research designs were attempted, Designs II and III. Table 1 summarizes the three designs used.

**Design I (AB/BA)**

Data from the few cases recruited and randomized into the RCT are presented here as AB or BA case studies with random allocation to AB or BA and fixed duration of phases.
Participants received five 1-hour sessions of virtual environment training over five week days and five 1-hour sessions of attention control training over five week days.

**Design II (ABA)**

ABA studies were conducted once recruitment to the RCT was halted. These had a fixed duration of phases. The control phases (A) were for five week days, during which participants received five 1-hour sessions of attention control training. The intervention phases (B) were also for five week days, during which participants received up to 5 hours of virtual environment training.

**Design III (rAB)**

To improve the scientific quality of the studies and take account of changes associated with passage of time as opposed to the intervention, AB studies with random duration of the control (A) and intervention (B) phases were undertaken. The minimum duration of the A and B phases were three and five days respectively, and the total duration of the studies was three weeks.

**Participant identification and recruitment**

Participants were identified from patients undergoing rehabilitation on the stroke unit at the Queens Medical Centre, Nottingham, UK. Inclusion criteria were having a goal of returning home and being unable to make a real hot drink. Exclusion criteria were dementia, major psychiatric illness, epilepsy triggered by screen images, no upper limb function, enrolment in other studies, or the presence of clinically significant co-pathology that made intervention unfeasible as judged by the research therapist. Written consent was obtained. The identification and recruitment process was the same for all three study designs.
Intervention and attention control treatment phases

The experimental treatment phases involved the patient being encouraged to use the virtual environment to make a virtual hot drink. In all studies, intervention was provided by a research therapist who was a senior stroke occupational therapist with experience both in stroke rehabilitation and in the use of the virtual environment (JE). Interventions were given in the usual hospital intervention areas and the research therapist worked closely with the multidisciplinary team for each patient. During the attention control phases, treatments were individually chosen for each participant to represent a period of therapist–patient interaction that was meaningful to the participant, and which lasted a similar amount of time to the intervention treatments, but which did not involve making a hot drink. These treatments included arm therapy, playing games, personal ADL therapy and the provision of stroke-related information.

Measurement of outcome

In all designs, participants were assessed at baseline and at the end of each intervention phase on hot drink-making performance in the real and virtual hot drink-making environment. The real world assessment required the same sub-tasks to be performed as in the virtual environment, and in both a score between 0 and 100 was given depending upon the number of tasks completed and the degree of assistance required. Performance in the real world was scored by a trained, research psychologist’s (ST) assessment, masked to treatment phase, of a video-recording of the participant’s performance. A score for the performance in the virtual environment was generated by the system’s software by comparing the actions taken with the built-in task model.

Video-recordings of the real task performance performed at baseline and at the end of each intervention phase also permitted the identification of any errors made, as a measure of cognitive function. These errors were classified using a schedule based on published literature and clinical experience comprising problems with initiation, attention, neglect, addition, sequence omission, perseveration, selection, object use, problem-solving, dexterity, quantity and spatial awareness (Appendix 1). Errors were determined at baseline by the research therapist and at the end of each intervention phase by a trained, research psychologist both of whom assessed the video-recordings masked to treatment phase. High interreliability for 25/27 stages of the real world performance score and the classification of the 12 errors had been established on a sample of 20 videos (kappa 0.6–1.0).

In designs II and III, a daily measure of performance on one aspect of real drink-making suitable for each patient was identified with the aim of providing a time-series measure sensitive to error variation within intervention phases. Daily video-recordings of the aspect of real hot drink-making were made by the research therapist, and assessed by a trained research psychologist masked to treatment phase.

The Barthel ADL Index was recorded at baseline by the research therapist, and on completion of all studies (but not between phases) by a research psychologist. This was to help describe the clinical characteristics of the population studies and to provide a measure of the trend towards recovery over the time-course of each study.

Analysis

The hot drink-making scores (real world, virtual and any subscores) were plotted graphically and visual inspection was used in each case. Data from all cases were summed and the improvements in real and virtual hot drink-making scores between phases, and the Barthel scores at baseline and study completion were compared (Wilcoxon paired test).

Ethics

Approval for the study and alterations was granted from the Nottingham Local Research Ethics Committee and the R&D Department at Queens Medical Centre, Nottingham, UK.

Results

Of 138 patients assessed over 16 months, only 16 were recruited, of whom three withdrew during the study. Two participants withdrew due to discharge.
planning complications and one withdrew due to changes in his clinical condition, but none withdrew due to problems related to the virtual environment. Reasons for the 122 who were not included are given in Table 2.

Results were available from five participants using design I, two using design II and six using design III. The characteristics of the 13 participants (9% of the total 138 patients) are shown in Table 3.

**Design I (AB/BA)**

Figure 1 shows real and virtual hot drink-making scores for the five cases in which design I was used. Visual inspection across all cases suggests a trend towards improvement over time in both real and virtual hot drink-making, with no particular relationship to the intervention phase. Between these five participants, 23 errors were made during real world performance of making a hot drink at baseline: they made 9 errors at the end of the control phases and 11 at the end of the intervention phases.

**Design II (ABA)**

Figure 2 shows real and virtual hot drink-making scores for the two cases in which design II was used. Visual inspection across both cases suggests a general trend towards improvement over time in both real and virtual hot drink-making, with no particular relationship to the intervention phase.

The daily assessment measure in case 6 (participant was given a jug of water, a cordless kettle with kettle base and a power socket and asked to boil a kettle of water, all items were placed in the same position as in the real hot drink-making assessment) appeared to improve during the treatment phase but went back to baseline at the end of that phase, and is most likely to represent variation in performance rather than a treatment effect. The daily assessment measure in case 7 (the same as for case 6) appeared to improve during the first control phase to reach the maximum, making it impossible for any effect to be seen during the treatment phase.

Between these two participants, 12 errors were made during real world performance of making a hot drink at baseline: they made 7 errors at the end of the first control phase (A), 5 at the end of the intervention phase (B) and none at the end of the second control phase (A).

**Design III (rAB)**

Figure 3 shows real and virtual hot drink-making scores for the six cases in which design III was used. Visual inspection across all cases suggests a general trend towards improvement over time in both real and virtual hot drink-making, with no particular relationship to the intervention phase.
Figure 1  Studies using design I.

Figure 2  Studies using design II.
Ceiling effects were seen on the daily assessment measure during the control phases in four of the six cases (8, 9, 11 and 13), making it impossible to assess improvement on these measures during the treatment phase. There was no improvement in the daily assessment measure during the treatment phase in the remaining two cases.

Between these six participants, 29 errors were made during real world performance of making a hot drink at baseline: they made 14 errors at the end of the control phase and 9 at the end of the intervention phase.

**Combined analyses**

Confirming the visual interpretation of the results, the changes in real and virtual hot drink-making scores for all 13 cases were not significantly different between the control and intervention phases: the median improvements in the real scores for the control and intervention phases were 9.78 and 0.60, respectively (Wilcoxon $z = -1.64$, $P > 0.05$); the median improvements in the virtual scores in the control and intervention phases were 4.77 and 7.90 respectively (Wilcoxon $z = -0.80$, $P > 0.05$).

The median Barthel score increased significantly from 7.45 (interquartile range (IQR) 4–10) at baseline to 10.85 (IQR 6.5–10) at the end of the studies (Wilcoxon paired test $z = 2.74$, $P < 0.01$).

**Discussion**

When tested in a clinical setting and under ordinary clinical conditions, we found that there was
limited scope to use this system, as only 9% of those undergoing stroke rehabilitation could be recruited. Even in these highly selected patients, there was no evidence that use of the virtual environment had any noticeable clinical effect on hot drink-making ability when compared with attention control training, but there may be several reasons for this as discussed below.

An important methodological observation from this research is that we were unable to recruit to a RCT. In hindsight, it may have been wiser to have chosen to undertake case studies in the first instance. It may, however, have been harder to have obtained research funding if we had made such plans.

Another important observation is that we did not find it easy to conduct convincing single case experiments. The cases we reports as design I, which were the first five cases entered into the RCT, did not have frequent daily measurements to enable a more detailed time series analysis. Accordingly it was not possible to detect any small or modest effects of the intervention over and above the general trend towards improvement over time. Even though designs II and III had daily assessments intended to permit a closer examination of the relationship between clinical changes during the control and intervention phases, our studies were limited by ceiling effects: participants tended to reach the maximum on these measures during the control phases thereby making it impossible to use these measures to detect further improvement.

Although our initial research plan seemed justified at the time, with hindsight we may have been better at this stage simply to have used unstructured case studies, aiming for the therapist to try out using the virtual environment in patients in a variety of settings (not just the stroke unit), for her to select the duration and frequency of treatment according to clinical preferences (rather than stick to an arbitrary pre-set protocol), and to make a comprehensive assessment of health changes during treatment as well as to elicit patient views. It may even have been better to have tested the system in other groups of people who have difficulty with making a hot drink, such as those with early dementia or learning disabilities. If we had done this, we may have had a better understanding of who potentially benefited from the use of this system, and in what way. This would have enabled us either to have designed better experiments, or to have undertaken yet more development work. In short, one reason why we did not see any convincing effect of the use of the virtual reality system upon outcome is that we were trying to test it using quantitative research designs before we had understood or developed the intervention sufficiently. We had attempted rigorous evaluation of clinical effectiveness before we had conducted sufficient user testing.

These methodological problems reduce what can be concluded from our work with regards to the virtual reality system under test. Nevertheless, we can conclude that a system such as ours may have only a small part to play in stroke rehabilitation given the small number of people who could use and benefit from them. Further technical development may enable the system to be more widely used and accepted, so that fewer people would be unable or unwilling to use it. The largest single group of people who were excluded from this study, thereby limiting the potential scope of our system, was those who were not going home and who would therefore not be required to make a hot drink. Another large group that was unsuitable was those who were already able to make a hot drink. One way to deal with this would be to change our development strategy dramatically, and to focus upon more general cognitive and motor functions that might be more widely applicable, rather than to concentrate upon one specific task. Yet another approach would be to develop a suite of such devices – for example, not only systems for making a hot drink, but also for managing money, driving, making a meal, and so on.

In our studies, we found that those who appeared suitable for using this treatment, by and large, did just as well without using it. The system was not associated with outstanding improvements in task performance or a reduction in the number of cognitive errors observed during task performance. There are no grounds for introducing this sort of system into clinical practice at this stage. Indeed, this sort of system might need to be tested in different patient groups or settings before its potential is evident.

In conclusion, we found that this early attempt to undertake rigorous clinical evaluation of an emerging technology for stroke rehabilitation...
was largely unsuccessful, and we recommend greater use of unstructured case studies before attempting to use formal research designs including single case studies. Our experience also provides an instance that suggests that a great deal of development will be required before new technologies such as virtual reality can be usefully applied in clinical practice.

**Clinical message**

- New rehabilitation technologies should be well developed and user tested before rigorous evaluation of clinical effectiveness.

**Acknowledgements**

We thank staff and patients on the Queen’s Medical Centre stroke unit for accommodating this study. We are grateful to Professor Derick Wade for constructive comments on an earlier version of this paper.

**Author contributions**

Judi Edmans was the research therapist, and contributed to the design of the protocol, conducted the data collection and analysis of results, and prepared the manuscript. John Gladman was the Principal Investigator, and was involved in the design of the protocol, the management of the project, the analysis of the data and writing the manuscript. Dave Hilton developed and maintained the virtual environment for the project. Marion Walker contributed to the design of the protocol and the preparation of the manuscript. Alan Sunderland contributed to the design of the protocol, analysis and preparation of the manuscript. Sue Cobb supervised the development of the virtual environment. Tony Pridmore supervised the development of the virtual environment. Shirley Thomas was the research psychologist who acted as the independent assessor, conducted all the outcome measures and contributed to the preparation of the manuscript.

**Competing interests**

There were no competing interests.

**Funding support**

The Stroke Association funded the project.

**References**

Appendix 1 – Classification and description of errors in task performance

- Initiation: Does not automatically begin the task or a stage
- Attention: Does not attend to an individual event (e.g. pays no attention to the fact the kettle has boiled)
- Neglect: Cannot find an object or does not respond to a visual or auditory cue to the affected side (e.g. unable to locate teapot positioned on his/her affected side)
- Addition: Adds an abnormal action (e.g. rips a teabag open and pours loose tea into the teapot)
- Sequence omission: Performs an action at the wrong time within the activity, which is unsafe (e.g. switches electrical power on with an empty kettle) or omits a stage (e.g. fails to put any coffee in the mug)
- Perseveration: Continually repeats a stage or action (e.g. repeatedly pours the milk into the cup)
- Selection: Does not select the correct object to accomplish a stage (e.g. stirs the tea with a finger or pours milk into the teapot)
- Object use: Does not use object appropriately (e.g. uses the kettle as a teapot)
- Problem solving: Gives unmistakable signs of not knowing what to do (e.g. continues to place the cup near the spout of kettle without picking up the kettle or looks hesitantly at the objects, picking them up, turning them over, putting them down and trying with another object)
- Dexterity: Fumbles when attempting to use objects (e.g. spills coffee when spooning)
- Quantity: Misjudges the amount of something (e.g. fills the cup with more milk than tea)
- Spatial awareness: Misjudges the location of objects (e.g. misses the cup and pours the tea onto the table)