Statistical modelling of Upper Limb Functional Ability using Motion Data: Validity Study

J. Serradilla* J. Q. Shi*§ Y. Cheng* G. Morgan†§ C. Lambden‡ and J. A. Eyre‡ §
*School of Mathematics & Statistics, Newcastle University, UK
† School of Computing Science, Newcastle University, UK
‡Institute of Neurosciences, Newcastle University, UK
§Email: *jian.shi@ncl.ac.uk* or *graham.morgan@ncl.ac.uk* or *janet.eyre@ncl.ac.uk*

***Abstract*—The objective of this study is to estimate the cross-sectional and longitudinal validity (sensitivity to change) of a novel algorithm as a new measure to assess upper-limb functional ability in stroke survivors. This algorithm models functional ability by mapping an array of kinematic variables extracted from the analysis of movements made by patients while playing bespoke, professionally-written action video games to the CAHAI-10 (Chedoke Arm and Hand Activity Inventory). A second aim of the research is to determine how the output from the model compares with existing measures of functional ability to distinguish change in patients in the acute/chronic stages of their recovery.**

Keywords—–health, upper-limb functional ability, validity, sensitivity to change, stroke, video games, ROC analysis-health, upper-limb functional ability, validity, sensitivity to change, stroke, video games, ROC analysise

I. Introduction

With ever improving medical care the chances of surviving stroke is more common now than at any point in the past. At current rates of improvement in care it is expected survivors will increase beyond 70 million by 2030. Unfortunately, on surviving stroke approximately 80% of patients endure hemiparesis [17], [9]; partial or complete paralysis due to brain injure. The road to recovery for such patients requires rehabilitation therapy. However, the lack of resources (therapists, money, clinics) hindered further by geographic dispersion of patients results in a practical inability to monitor and tailor appropriate therapies per-patient. This brings a pressing priority to seek a practical, cost efficient, delivery of rehabilitative care to stroke patients.

One possible avenue for cost efficient delivery of rehabilitative care is computer enabled serious gaming. Research has shown that it has become possible to encode rehabilitative strategies into a serious game (e.g., [16], [14], [7]). As games themselves become more physically interactive there is a real possibility of utilising off-the-shelf gaming devices for rehabilitation (e.g., Kinect, PSMove, Wii Motes) (e.g., [15], [11]). Most promising is the recent results from clinical studies that assess the suitability of serious gaming [10]. Such research is important as it clearly identifies the benefits of serious gaming as successfully motivating patients to participate in a positive way to attain their rehabilitative goals. Cumulatively, the literature clearly identifies serious gaming as promoting physical activity in patients (not limited to stroke) beyond that achievable without gaming involved.

Rehabilitative literature within the serious gaming community has so far concentrated on the delivery of rehabilitative strategies embodied within video games. The reason why video games have proved so popular for rehabilitation is that motor learning underpins recovery after stroke. The rehabilitative strategies encoded within a video game encourage movement that in turn promotes the relearning of movements or an ability to compensate for the lack of movement. Such motor learning forms the basis for assessing patient progress: motor learning in normal subjects and functional neuroplasticity leading to post-stroke motor recovery have been shown to share the same molecular and genetic substrates and brain networks [13]. The alternative to motor learning is to observe behavioural outcomes, a much more time consuming task and bespoke for each patients. Therefore, motor learning provides an ideal marker of the biological system underpinning rehabilitation. Measuring motor learning is a time consuming task that requires clinical intervention (therapist). An ideal scenario would be for the video game itself to provide a clinical assessment of motor learning.

An important function of video games used in rehabilitation is to motivate via reward. To achieve this the patient is monitored and rewarded on their ability to move in relation to the games goals. One may envisage that the capturing of such data will translate into a suitable assessment strategy for motor learning. Unfortunately, this is not the case as such data is substantial and carries a large degree of variance. In the literature only broad assumptions can be made regarding the degree of movement and directional activity of patients from in-game data. This is OK for scoring and encouraging participation, but not suitable for replacing clinical assessment for directing an intervention strategy. Therefore, serious games are still benchmarked against a therapists scores for motor learning. A popular scoring technique is the Chedoke Arm and Hand Activity Inventory (CAHAI) for upper limb. This in itself is a time consuming and costly task, requiring therapist and patient to be co-located and may take up to 3 hours to complete per-patient.

Although a complete serious game rehabilitative solution would require in-game decision-making regarding intervention (i.e., exercise schedule based on in-game analysis), an important first step is to create a statistical model that can successfully score a patient using only in-game data. This will require the video game to at least be as competent as a therapist, or future interventions directed by the game itself may be more harmful than beneficial. In this paper we present a statistical model that uses only in-game data that can achieve at least the fidelity of therapist directed CAHAI assessment to determine change in stroke patients. This study is part of a wide reaching (commercial cloud based delivery of rehabilitative care) research project undertaken within the NHS in the UK to deliver upper limb rehabilitative intervention via serious gaming to stroke patients. Data has been derived from the commercial video Circus Challenge [19] and a statistical model created specifically based on the output data produced by a prototype commercial off-the-shelf game input device from Sixense http://sixense.com/hardware/wireless. The data was gathered from over 100 patients over the period of a year. During the study, therapists also carried out CAHAI assessments. These therapist-based assessments were used to benchmark the statistical model. We show for the first time that a serious game is not limited to encouraging patient recovery, but can cost efficiently monitor patient change to a clinical standard in the home.

II. Methods

Ethical approval was obtained from the National Research Ethics Committee and all work undertaken was in accordance with the Declaration of Helsinki. Written, informed consent from all the subjects was obtained.

A. Participants

A cohort of 26 of stroke survivors without significant cognitive or visual impairment were recruited for the study. Full patient characteristics are provided in Table I. Further inclusion criteria required that participants were able to grasp game controllers with their paretic hand and move their affected limb against gravity. Patients had a wide range of upper limb function as reflected in their Chedoke Arm and Hand Activity Inventory (CAHAI) scores. None had previously played video games but all participated in a home-based rehabilitation programme using the Circus Challenge video games over a 3 month period. The games can be played either standing or sitting down.

	Characteristic	Value
1	Genger (M/F)	18/8
2	Stratification (chronic/acute)	18/8
3	Age*	
4	+ Chronic	58 (43-78)
5	+ Acute	50 (33-77)
6	Weeks since stroke*	
7	+ Chronic	59 (36-414)
8	+ Acute	4 (1-6)
9	CAHAI-10*	
10	+ Chronic	42 (13-70)
11	+ Acute	49 (14-65)

TABLE I: Note: * Median (Range)

B. Measures

Chedoke Arm and Hand Activity Inventory.: The CAHAI-13 is a measure of upper limb functional ability with 13 items which are assessed using a 7-point quantitative scale[1]. Shortened versions of the CAHAI with 7,8 and 9 items (CAHAI-7, CAHAI-8 and CAHAI-9 respectively) also exist and have been proven to maintain the same degree of validity as the full version of the measure [3]. A shortened version with the first 10 items of the CAHAI-13 is used in this study which we will refer to as CAHAI-10. The maximum and minimum scores of this subtest are 70 and 10 respectively.

Circus Challenge Assessment Game (CCAG).: Circus Challenge (CC) is a new stroke rehabilitation tool based on 10 computer-based video games. Control of the video games is achieved via 100 separate upper limb actions based on identified patterns of co-ordinated bimanual movements, which together form the functional basis for activities of daily living [12].

The upper-limb functional ability measure is derived from a shortened version which we refer to as the Circus Challenge Assessment Game. This version comprises 40 representative actions, ranging from the simplest mirrored movements where the same movement is performed simultaneously by each upper limb, to co-ordinated movements where each arm and hand performed different movements in a coordinated manner. Actions in the CCAG are presented in order of increasing difficulty and the data generated from measuring the arm and hand movements whilst patients performed these actions are used to derive an algorithm measuring upper-limb functional ability [19]. The maximum and minimum scores of this subtest are similar to those of the CAHAI-10.

C. Design

Data is collected using a longitudinal study design. Patients' assessments were made at baseline and then weekly for the first four weeks followed by an assessment every other week for a further eight weeks; eight assessments are made in total. Some patients are still being followed in the study and hence the number of observations per subject varies.

Research assessments were carried out in the patient's own home. At each occasion, an occupational therapist trained in the administration and scoring of the CAHAI undertook a blinded clinical assessment of upper limb function using the CAHAI-10 as a measure. Patients were also introduced to Circus Challenge at baseline and asked to play the game in their home each day for approximately thirty minutes. From occasion 2 through to occasion 8, patients were also asked to play the CHAG, once they had familiarized themselves with Circus Challenge.

Individuals were stratified, a-priori, into two groups according to the amount of change in their upper limb functional ability expected throughout the duration of the study: Group 1, the acute group, consisted of participants who enrolled into the study within three months of their first ever stroke; and Group 2, the chronic group, was formed by participants who were 6 months of more postonset of stroke, Table I.

III. Data Analysis

Prior to showing the comparative study between the CCAG and the CAHAI-10, we first check the validity of the latter measure.

A. CAHAI-10 validity

Shortened CAHAI versions with 7,8 and 9 items have been studied elsewhere [3] and found to have similar validity and sensitivity to change as the original CAHAI-13. This study does not collect data for the full CAHAI but all the above mentioned shortened versions of the CAHAI contain all the items measured by the CAHAI-10 version. We, arbitrarily, choose the CAHAI-9 as a reference for comparison with the CAHAI-10 in a convergent construct validation process.

Between-subjects correlation between CAHAI-9 and CAHAI-10 raw scores is 0.99 ($p < .001$). Likewise, within-subjects correlation between CAHAI-9 and CAHAI-10 raw scores is 0.98 ($p < .001$). These results are similar to those reported for other shortened versions of the CAHAI and demonstrate that the CAHAI-10 is an appropriate measure to account for upper-limb functional ability.

B. CCAG model development from motion data

Collecting movement data using the CCAG varies depending on the familiarity of the player but, on average, it does not require more than 20 minutes; likewise, the assessment game has the added advantage that it does not require of the physical presence of an occupational therapist and it lends itself to be monitored remotely.

Each time the CCAG is played position and orientation data for each of the 40 movements which comprise the game are stored in a computer. That raw data is then processed and up to 320 kinematic variables are derived from it. Those variables are related to underlying features which define how well the movement is performed; namely, speed, fluency, synchrony and accuracy.

The full inner-workings of the model building and selection process is explained elsewhere [19]. In brief, however, our approach is to use those movement covariates to build a regression model for the validated CAHAI-10 scores. After a variable selection process, twelve variables are selected. Furthermore, to account for patient heterogeneity we resort to a linear mixed model [4] with random effects. The model is linear with respect to the selected movement covariates; and the random effects can be thought of as a tool which allows for patients to have differing upper-limb functional ability (i.e random intercept) and for acute patients to experience different recovery rates (i.e. random slope). But more importantly, the random effects permit us to predict the recovery curve for future patients (not in the study) by taking into account patient-to-patient and within-patient variability estimated from subjects in the study.

The proportion of the total variability in the response variable (CAHAI-10) accounted for by this model is 91%. A check of the model performance can be seen in Figure 1, where the model fitted values are plotted against the clinically assessed CAHAI-10 for the chronic and acute group. Figure 2 (chronic group) and Figure 3 (acute group) allow for a closer inspection of model performance by graphing the observed and fitted values longitudinally for each of the patients.

C. Sensitivity to change

No gold standard exists as a measure of upper-limbs functional ability. We therefore resort to a *convergent construct validation process* [20, p.257] whereby the correlation between the proposed measure (CCAG) and an *accepted reference measure* (CAHAI-10) is calculated. Validity will be confirmed if the new measure is shown to be correlated with the reference measure.

1) Cross-sectional validity: The interest when assessing cross-sectional validity is in understanding the pattern of variation across patients, i.e. are subjects with high values of the reference measure also likely to have high values of the new measure and vice-versa? In answering this question, we resort to a correlation coefficient between the subject means for each of the two measures weighted by the number of observations for the subject [6]. We will refer to this measure as a *between-subjects correlation*.

The between-subjects correlation (i.e. the correlation between the subject means weighted by the number of observations per subject) between the CAHAI-10 and the CCAG model output scores is 1.00 ($p < .001$). This is a good indication that the CCAG model output can be used as a surrogate measure to determine upper-limb functional ability.

2) Longitudinal validity: The situation is different when investigating longitudinal validity, which is the main interest in the study of sensitivity to change. In this case, we are interested in how the measures vary across time: is an increase/decrease of the reference measure within a subject associated with an increase/decrease of the new measure? To assess this question, we consider the following model

$$y_{ij} = \alpha_i + \beta z_{ij} + \varepsilon_{ij} \tag{1}$$

where y_{ij}, z_{ij} are the CAHAI-10 and CCAG model scores respectively for patient i at occasion j. The coefficient β is the parameter of interest, indicating weather two measures change in the same way simultaneously. Patients are treated as categorical variables with α_i being an intercept which is different for each patient and being used to remove the variability between subjects; finally, ε_{ij} is an error term.

The analysis of variance table for this model is provided in Table II. Assessment of longitudinal validity is given by the middle row in the table. The effect of CCAG scores on CAHAI-10 can be assessed via a hypothesis test $H_0 : \beta = 0$. The p-value is calculated from the F-test in Table II which is less than 0.001 indicating a significant effect. We further look at the test $H_0 : \beta = 1$ and calculate the p-value from a likelihood ratio test in Eq. (1). The p-vlue is 0.43, indicating that we accept the hypothesis $H_0 : \beta = 1$ and the two measures do change in the same way.

Following discussion given in [5], sensitivity of change can also be assessed by a so called *within-subjects correlation coefficient* which is the proportion of the variation caused by the covariate CCAG scores in Eq. (1) after the within patient variation removed. The within-subject correlation coefficient (r) is calculated from Table II

$$\mathrm{r} = \sqrt{\frac{1114.51}{1114.51 + 1651.42}} = 0.63\,,$$

for which the p-value testing $\mathrm{H_o} : r = 0$ is also given from the F-test in Table II ($p < .001$). This hypothesis is equivalent at testing $H_0 : \beta = 0$.

Breaking down the within-subjects correlation in groups, we have obtained that $r = 0.34$ ($p = .002$) for chronic group and $r = 0.80$ ($p < .001$) for acute group. It makes sense

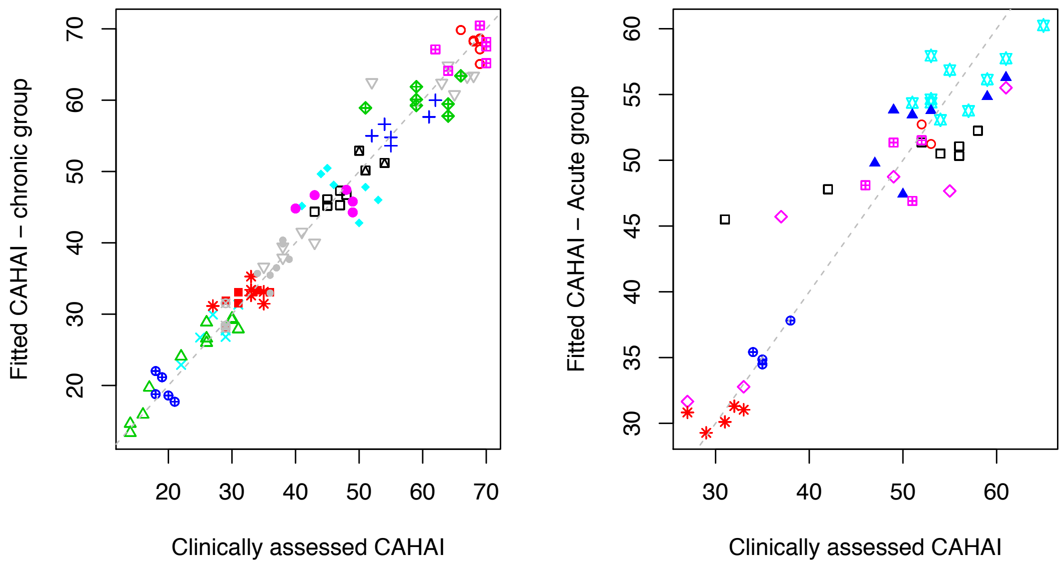

Fig. 1: Fitted CAHAI using random-effects model vs clinically assessed CAHAI. Panels split patients into chronic and acute; symbols and colours are used to differentiate subjects.

intuitively that r for the acute is higher than the chronic as it is debatable whether chronic patients show any improvement at all in the study. The longitudinal plots presented in Figures 2 and 3 confirm the findings.

The result presented in this section shows that the within-subject correlation between both measures is significant, indicating that the CCAG model scores are able to pick up longitudinal changes within patients particularly for acute patients.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio	p-value
Patients	25	29153.90	1166.16	83.33	–
z_{ij}	1	1114.51	1114.51	79.64	0.0000
Residuals	118	1651.42	14.00		

TABLE II: Analysis of variance for the model in Eq. (1).

3) ROC analysis: We also interested in answering the research question as to which of the two measures, the CAHAI-10 or the CCAG, is more sensitive to change. A well accepted approach to compare the performance of different measures for classification purposes is to measure the area under the curve (AUC) in an ROC (Receiver Operating Characteristic) plot. The higher the area, the better the classification power of the measure.

The ROC plots for each of the two measures have been built as follows: (1)a priori patients are stratified into two groups, chronic and acute, based on the expected amount of change they are likely to undergo during the study; (2) for each patient, the amount of change in the corresponding measure is determined by subtracting the value of the measure at the last occasion from its baseline; (3) the performance of the measure to classify patients is assessed over the range of all possible values in the amount of change observed. The result has been graphed in Figure 4.

In comparing the area under the curves for each measure we have taken into account that the curves are correlated (i.e. both curves derive from multiple measurements on the same sample) and use the method by Delong et al. [8] as implemented by Robin et al. [18]. The null hypothesis for the comparison is that there is no difference in classification performance between the two methods against a two-side alternative hypothesis. The conclusion of this study is that there is no evidence to conclude one of the methods is superior to the other $(p = .63)$.

IV. Conclusions

The main aim of this article is to assess the cross-sectional and longitudinal validity of a model derived from video game data (Circus Challenge) as a measure of upper limb functional ability. In that endeavour, as a side-product, we have first checked that the CAHAI-10 was a valid measure. This approach deserves a brief explanation. Versions of the CAHAI with 13, 9, 8 and 7 items had all been previously validated [2], [3]. The CAHAI-10 version we have used as a reference contain all the items in the CAHAI-9 plus the additional activity *zip up the zipper*. The more important reason as to why we resorted to this new version is rooted in the design of Circus Challenge as a therapeutic tool for stroke survivors: there are several movements embedded throughout Circus Challenge mimicking activities of daily living and which are of a very similar nature to this additional task. The conclusion of the validation process is that the CAHAI-10 maintain the same degree of validity as other shortened

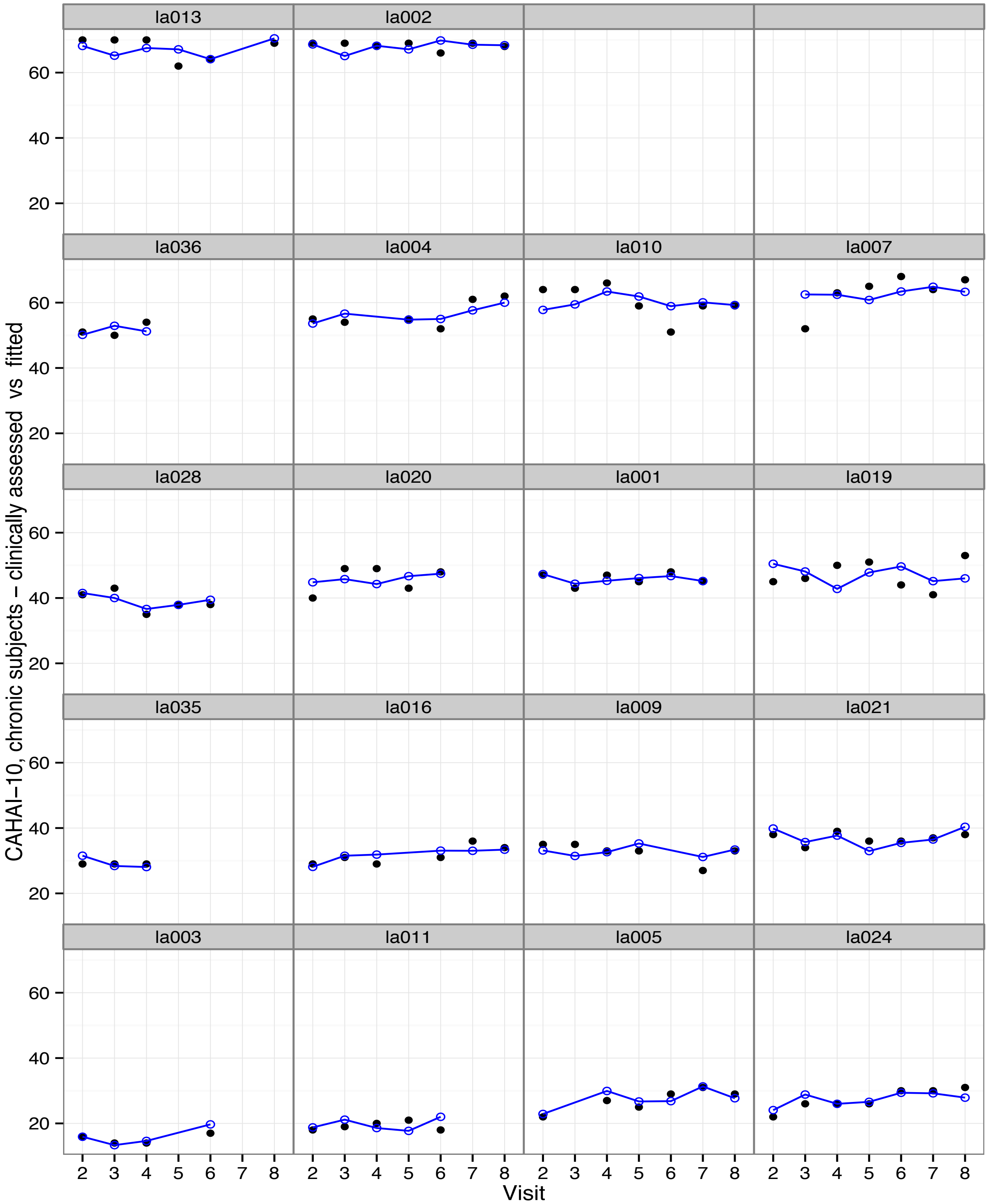

Fig. 2: Chronic subjects clinically assessed CAHAI (solid dots) vs fitted CAHAI for random-effects (solid blue line) model. Panels represent patients and are ordered, from bottom left to top right, by increasing mean CAHAI level.

versions which in line with previous findings by Barreca et al. cited above.

We have adopted a convergent construct validation approach using the CAHAI-10 scores as reference to show that the Circus Challenge Assessment Game is a valid (cross-sectionally and longitudinally) tool to make inferences about upper limb functional ability. Its classification performance using ROC curve analysis has been found to be no different from that of the CAHAI-10 which brings about the question as

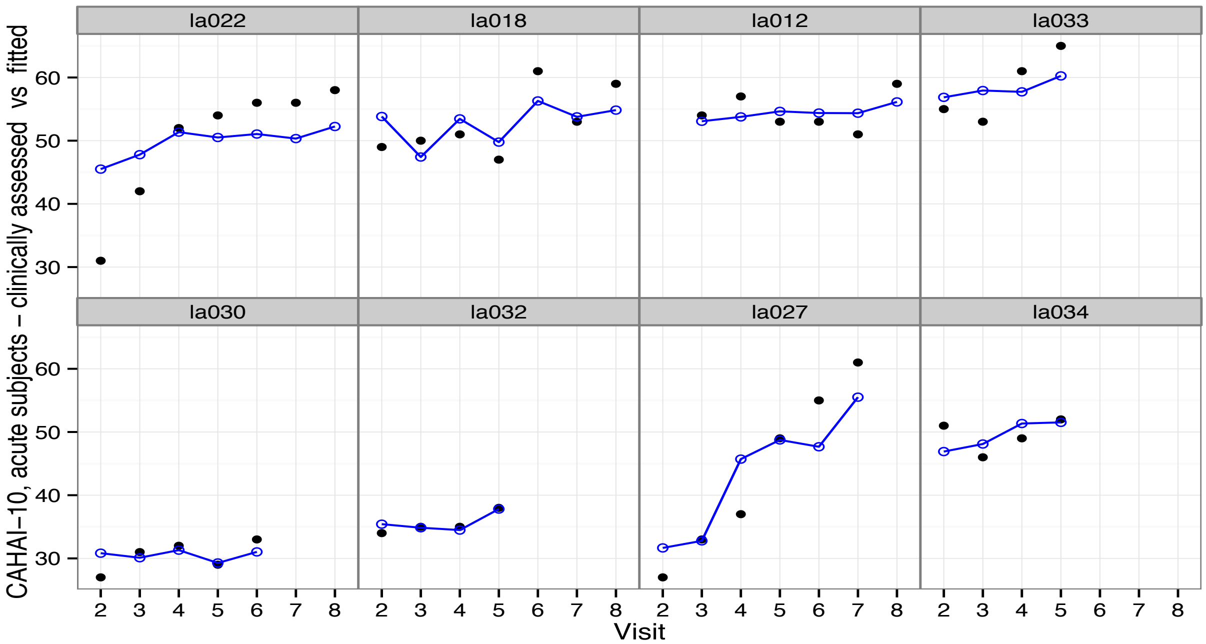

Fig. 3: Acute subjects clinically assessed CAHAI (solid dots) vs fitted CAHAI for random-effects (solid blue line) model. Panels represent patients and are ordered, from bottom left to top right, by increasing mean CAHAI level.

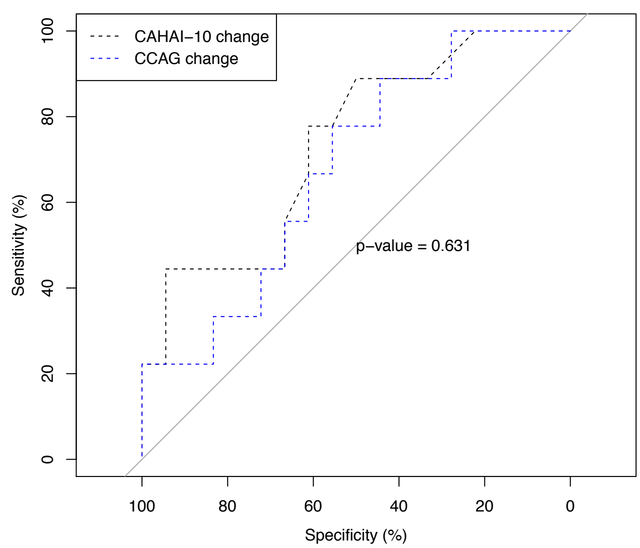

Fig. 4: ROC curves for stroke patient classification into chronic/acute groups using the CAHAI-10 and the CCAG measures. There is no significance difference in any of the measures performance ($p = .63$).

to why a new measure is needed. One of the main advantages of the game as a rehabilitation tool is that it is designed to be played at the patients' own home. Likewise, the model allows making inferences about upper-limb functional ability remotely and with limited therapists intervention. This should only be seen in the context of reaching out to an ever-increasing population of stroke survivors: with the same amount of resources, remote monitoring allows therapists to effectively follow up many more patients.

Acknowledgements.: Funding from the Health Innovation Challenge Fund (Department of Health and Wellcome Trust) and technical support from Limbs Alive Ltd (www.limbsalive.com), making Circus Challenge available for the research, are gratefully acknowledged.

Disclaimer.: This publication presents independent research commissioned by the Health Innovation Challenge Fund (Grant number HICF 1010 020), a parallel funding partnership between the Wellcome Trust and the Department of Health. The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health.

References

[1] Barreca, S., Gowland, C., Stratford, P., Huijbregts, M., Griffiths, J., Torresin, W., Dunkley, M., Miller, P., Masters, L.: Development of the Chedoke Arm and Hand Activity Inventory: Theoretical constructs, item generation, and selection. Top Stroke Rehabil 11, 31–42 (2004)

[2] Barreca, S., Stratford, P., Lambert, C., Masters, L., Streiner, D.: Test-Retest Reliability, Validity, and Sensitivity of the Chedoke Arm and Hand Activity Inventory: a New Measure of Upper-Limb Function for Survivors of Stroke. Arch Phys Med Rehabil 86, 1616–1622 (2005)

[3] Barreca, S., Stratford, P., Masters, L., Lambert, C., Griffiths, J., McBay, C.: Validation of Three Shortened Versions of the Chedoke Arm and Hand Activity Inventory. Physiother. Can. 58, 1–9 (2006)

[4] Bates, D.: lme4: Mixed-Effects Modeling with R (2010)

[5] Bland, J.M., Altman, D.G.: Statistics Notes: Calculating Correlation Coefficients with Repeated Observations: Part 1 - Correlation within Subjects. BMJ 310, 446 (1995)

[6] Bland, J.M., Altman, D.G.: Statistics Notes: Calculating Correlation Coefficients with Repeated Observations: Part 2 - Correlation between Subjects. BMJ 310, 633 (1995)

[7] Burke, J.W., A, B: Optimising engagement for stroke rehabilitation using serious games. The Visual Computer 25.12, 1085–1099 (2009)

[8] DeLong, E., DeLong, D., Clarke-Pearson, D.: Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 44, 837–845 (1988)

[9] for Disease Control, C., (CDC), P.: Prevalence of stroke–United States, 2006-2010. MMWR Morb Mortal Wkly Rep 61, 379–82 (2012)

[10] Fritz, S.L.a.A., B: Active video-gaming effects on balance and mobility in individuals with chronic stroke: A randomized controlled trial. Topics in stroke rehabilitation 20(3), 218–225 (2013)

[11] Herz, N.B., A, B: Nintendo wii rehabilitation (wii-hab) provides benefits in parkinsonś disease. Parkinsonism & Related Disorders (2013)

[12] Kimmerle, M., Mainwaring, L.: The functional repertoire of the hand and its application to assessment. Am J Occup Ther 57, 489–498 (2003)

[13] Krakauer, J.: Motor learning: Its relevance to stroke recovery and neuroehabilitation. Curr Opin Neurol 19, 84–90 (2006)

[14] Lange, B., A, B: Development of an interactive game-based rehabilitation tool for dynamic balance training. Topics in Stroke Rehabilitation 17(5), 345–352

[15] Lange, B., A, B: Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE (2011)

[16] Lange, B., Flynn, S., Rizzo, A.: Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation. Physical Therapy Reviews 14.5, 355–363 (2009)

[17] Party, I.W.: National clinical guideline for stroke. London: Royal College of Physicians, 4^{th} edn. (2012)

[18] Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Mller, M.: pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. BMC Bioinformatics 12, 77 (2011)

[19] Shi, J., Cheng, Y., Serradilla, J., Morgan, G., Lambden, C., Ford, G., Price, C., Rodgers, H., Cassidy, T., Rochester, L., Eyre, J.: Evaluating Functional Ability of Upper Limbs after Stroke Using Video Game Data. In: Imamura, K., Usui, S., Shirao, T., Kasamatsu, T., Schwabe, L., Zhong, N. (eds.) International Conference on Brain and Health Informatics. Lecture Notes in Artificial Intelligence, vol. 8211, pp. 181–192. Springer (2013)

[20] Streiner, D., Norman, G.: Health measurement scales: a practical guide to their development and use (4th edition). Oxford University Press, Oxford, UK (2008)