Research Article

The graded redefined assessment of strength sensibility and prehension version 2 (GV2): Psychometric properties

Sukhvinder Kalsi-Ryan [1,2], Colin Chan[3], Mary Verrier[1,2,4], Armin Curt[5], Michael Fehlings [6,7], Marc Bolliger[5], Inge-Marie Velstra[8], GRASSP Cross Sectional Study Team[1], GRASSP Longitudinal Study Team[1]

[1]KITE, Toronto Rehabilitation Institute-UHN, Toronto, Canada, [2]Departtment of Physical Therapy, University of Toronto, Toronto, Canada, [3]Biomedical Engineering, University of Western Ontario, London, Canada, [4]Rehabilitation Sciences Institute, University of Toronto, Toronto, Canada, [5]Spinal Cord Injury Centre, University Hospital Balgrist, Zurich, Switzerland, [6]Department of Surgery, University of Toronto, Toronto, Canada, [7]Krembil Neuroscience Centre, University Health Network, Toronto, Canada, [8]Clinical Trial Unit, Swiss Paraplegic Center, Nottwil, Switzerland

Context: GRASSP Version 1 (GV1) was developed in 2010, is an upper extremity measure specifically designed to assess recovery after traumatic tetraplegia. A second version was developed to reduce length of the test and refine instructions/standardization. The purpose of this *post hoc* analysis was to calculate psychometric properties of GRASSP Version 2 (GV2).

Design/Setting: A *post-hoc* analysis of datasets for the GRASSP cross-sectional ($n = 72$ chronic,) and longitudinal ($n = 127$ acute) studies was conducted. Reliability, validity and MDD were calculated from the chronic sample and responsiveness was re-calculated from the longitudinal sample. Both studies were observational.

Participants: A chronic sample ($n = 72$) and acute longitudinal sample ($n = 127$) of individuals with traumatic tetraplegia (AIS A to D, NLI C2 to C8) were studied.

Outcome Measures: GV1, the Spinal Cord Independence Measure III (SCIM), International Standards of Neurological Classification of Spinal Cord Injury (ISNCSCI) were administered in both studies at all centers and the Capabilities of the Upper Extremity Questionnaire (CUE-Q) was administered in North American sites only. GRASSP-Palmar Sensation, GRASSP-Prehension Performance subtest items included in GV2 were re-analyzed for reliability; validity, MDD and responsiveness.

Results: Inter-rater and test-retest reliability for all subtests ranged between 0.849–0.971 and 0.950–0.971 respectively. Concurrent validity between domains of GV2 were positively and moderately (0.530–0.830, $P < 0.0001$) correlated to SCIM, SCIM self-care subscore (SS) and CUE-Q. MDD values were 4 and 3 points for sensation and prehension performance (single side). Responsiveness values were .84-.88 for GR-Sens and .93–1.22 for GR-PP respectively.

Conclusions: GV2 retains excellent psychometric properties as does GV1.

Keywords: Psychometric properties, Outcomes, Upper limb, Tetraplegia

Introduction

What is GRASSP? – GRASSP Version 1 (GV1) is an outcome measure designed to quantify upper limb impairment after traumatic tetraplegia.[1,2] The overall objective for the assembly of the GRASSP was to develop a clinical research tool that could capture information about the upper limb impairment from the traumatic cervical (C0-T1) spinal cord injury (SCI) population, obtain integrated sensory and motor impairment data, and define the tetraplegic population in greater detail. The GRASSP was intended to:

(1) be reliable, valid, and responsive (sensitive) to change over time
(2) monitor the extent of spontaneous (natural) recovery; and

Correspondence to: Sukhvinder Kalsi-Ryan, Toronto Rehabilitation Institute, 520 Sutherland Drive, Rehabilitation Engineering Lab-REL, Toronto, Ontario M4G 3V9, Canada. Email: Sukhvinder.kalsi-ryan@uhn.ca

DOI 10.1080/10790268.2019.1616950

(3) be utilized in clinical trials and studies to evaluate the effect of interventions.

The GRASSP is recommended for use in the early acute phases to any point in the post-injury time-course, particularly when a change in neurological status is the construct of interest. The GRASSP is intended for use with traumatic cervical SCI participants that are being followed for recovery of the upper limb. GV1 or GV2 is not designed for the pediatric population or the non-traumatic SCI population. GRASSP Version Myelopathy is a third modification that can be applied in non-traumatic cervical SCI (website).

Origin of GRASSP: GV1 development and copyright was completed in 2009,[3] by the GRASSP Research and Development Group (GRDG). GV1 has been made available for use since 2010. It was originally designed using elements of the Link Hand Function Test,[4] Quadriplegia Hand Assessment Tool[5] and the Sollerman Hand Function Test.[6] The psychometric properties of GV1 are excellent; evidence for reliability, validity (1), responsiveness and minimally detectable difference (MDD) are published.[7,8] Predictive validity of GV1 has also been established for some of the subtests within the GRASSP.

GRASSP within the context of other measures: There are a variety of upper limb measures used in the SCI field.[9] Upper extremity monitoring is important particularly in tetraplegia for two reasons: (1) upper extremity recovery is a priority for people with tetraplegia[10,11] and (2) almost all clinical trials are being conducted on individuals with tetraplegia, therefore, any subtle or significant neurological improvements will be noted in the upper limb.[12] There are a wide variety of well, developed measures that assess different constructs of upper extremity deficit in traumatic SCI. However, GRASSP is sensitive to subtle changes in neurological status as well as function of the hand. GRASSP uniquely measures the body structure and function domain (ICF), whereas many of the other upper limb measures, assess function, independence or activity. Therefore, the GRASSP holds a unique position in the field as it is an important measure in acute clinical trials where any degree of signal is of interest as well as neuro-restorative studies where the intent is to have an impact on the neurological state.

Updated GRASSP Version 2: Updating the GV1 to GRASSP Version 2 (GV2) was planned from the release of GV1. The rationale for the modification/s was to reduce items that were redundant and/or clinically irrelevant in turn reducing the administration time of the test without compromising the sensitivity. The evidence to make changes was not yet available in 2010. Therefore, in 2016/2017, Dr. Velstra conducted a rasch analysis with the European longitudinal dataset originally collected to establish responsiveness and minimally clinical important difference (MCID). GRASSP generates ordinal total scores and the applicability as an interval level measurement required adjustments. The rasch analysis was aimed at examining metric characteristics with the intent, to reveal interval level scales for the GRASSP subtests and reduce subtest items if redundant.[13] Velstra *et al*. found that the GRASSP Strength (GR-Str) could be reduced from 10 to 4 items, the GRASSP Sensation (GR-Sens) from 6 to 3 and the GRASSP Prehension Performance (GR-PP) from 6 to 4 items. In addition to the rasch analysis, elements of experience with the tool were considered. In particular, even though the rasch analysis suggested a reduction in GRASSP-Str items, it was not reduced, as the sensitivity of this subtest has been useful in defining subtle change.[8]

The effort has been made to minimize the effect on existing use of GV1. GRASSP is applied broadly in the SCI research community internationally, and is being used within sponsored (Asterias Therapeutics, Vertex Pharmaceuticals, Stem Cells Inc.), investigator driven (Nogo anti-body Study, Riluzole Study), acute clinical and chronic rehabilitation studies. While it is not typically a primary endpoint in most studies, it is used as an important secondary endpoint and more recently a primary measure for enrollment criteria (NeuroRecovery Technologies, Renetx). The GRDG (see team on website www.grassptest.com) has confidence that the transition for clinicians and researchers is manageable and the mergence of GV1 and GV2 data is as well. Information on the website and this manuscript will provide insight to users on how to implement GV2. In addition, new web-tools will be made available on the GRASSP website www.grassptest.com to facilitate this process. GV2 was launched in November of 2017,[14] however, at the time no new psychometric data were released.

The objectives of this analysis were to calculate the psychometric properties (reliability, validity, responsiveness and MDD) of the GRASSP V2, so that revised psychometric properties are available to the field for clinical and research use. Since the launch of GV2 there has been interest in use of the test, however, uptake is limited as there is no literature available to justify GV2 use. The gap had been a lack of available and documented psychometric properties.

Methods and materials

Modifications to GRASSP: Uptake and interest in the GRASSP measure has been significant. GV1 consists of 5 subtests; GRASSP Strength (GR-Str), Dorsal Sensation (GR-DSens), Palmar Sensation (GR-PSens), Prehension Ability (GR-PA) and Prehension Performance (GR-PP). Cumulative end-user feedback from workshops and surveys, in Europe and Canada has been collected and archived since 2010. The sum mean score of GR-Str (isotonic testing) was compared to the upper extremity motor sum score (UEMS) for EU longitudinal data. Rasch analysis of EU longitudinal data was conducted to determine redundancy of items.[13] Modifications were derived and applied to GV1. Based on end-user feedback: (1) language throughout the manual was refined to clarify subtest/item definitions and instructions to reduce variability of interpretation among assessors. Paired t-tests confirmed no significant difference between sum-scores of GR-Str and UEMS sumscores. Rasch analysis confirmed response thresholds to be disordered, for GR-Sens, and GR-PP permitting a smaller set of items to be integrated. Although, the rasch analysis also identified redundant items in GR-Str, the 10 muscles were retained, based on clinical meaningfulness. Modifications to GRASSP included the following:

1. GR-Str, Isotonic Manual Muscle Testing (MMT) changed to isometric MMT.
2. GR-Sens, Reduction of 6 test locations to 3 per hand (elimination of dorsal sensation).
3. GR-PP, Reduction of 6 items to 4 items.
4. Instruction manual revised for clarity and standardization.

In summary, GRASSP was modified to improve objectivity, reduce assessment time, and improve usability (Fig. 1 shows modifications to GRASSP).[14]

Data and Analysis: Data collected for the original cross sectional and longitudinal studies were analyzed for this study. As stated in the cross sectional (1) and longitudinal study manuscripts,[7,8] ethical approval was attained at all institutions participating in the studies (see Table 1 for sites involved). Data elements from both datasets were selected to match the elements of GV2. Study 1: A cross-sectional multi center trial was conducted to establish the reliability and validity of the GRASSP. *Inclusion/Exclusion:* Individuals with chronic (greater than 6 months after injury) traumatic tetraplegia who were neurologically and medically stable, between the ages of 16 and 65 years and able to provide informed consent were included in the study. Individuals with moderate brain injury who were neurologically unstable or individuals with any other pathology causing upper limb impairment were excluded. Seventy two datasets for inter-rater reliability and validity were collected and 45 datasets for test retest reliability were collected. Seven centers participated in the study, 3 European and 4 North American further

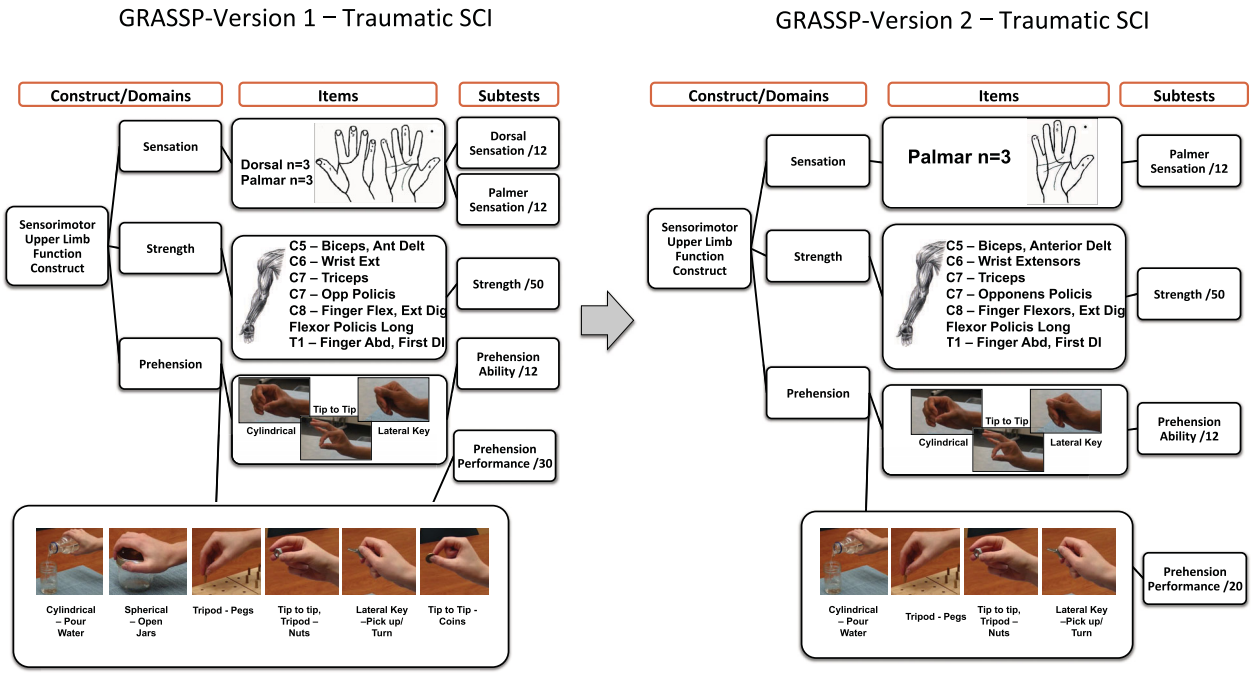

Figure 1 Visual definition of the domains of GRASSP Version 1 and what those domains have been reduced to in GRASSP Version 2.

Table 1 Summary sample characteristics.

Demographics	Chronic Sample	Longitudinal Study
Years of Data Collection	2007–08	2009–14
Analysis for GV2	Inter/test retest reliability, concurrent validity, MDD	Responsiveness
Total datasets used for analysis	72	127
Age (mean ± SD; range)	39.7 ± 10.7; 16–65	49.3 ± 23.8; 18–87
Site of enrollment		
CANADA		
Toronto Rehabilitation Institute (ON, CAN)	15 (21%)	3 (2%)
GF Strong (BC, CAN)	10 (14%)	
Hamilton Health Sciences-2 sites (ON, CAN)		6 (5%)
St. Michael's Hospital (ON, CAN)		6 (5%)
Toronto Western Hospital (ON, CAN)		24 (19%)
USA		
Rehab Institute of Chicago (IL, USA)	10 (14%)	
Thomas Jefferson University (PA, USA)	10 (14%)	
EUROPE		
Klinik Hohe Warte Bayreuth (D, EU)	27 (37%)	25 (20%)
Unfallklinik Murnau (D, EU)		1 (1%)
University Hospital Balgrist (CH, EU)		14 (11%)
Universitätsklinik Heidelberg (D, EU)		9 (6%)
Swiss Paraplegic Center (CH, EU)		25 (20%)
Mechanism of Injury		
Fall (Non-Sports Related)		58 (44%)
Fall (Sports-Related)		27 (20%)
Motor Vehicle Collision		26 (20%)
Collision (Sports-Related)		8 (6%)
Assault		2 (2%)
Other		10 (8%)
AIS (Baseline)		
A	28 (38%)	29 (23%)
B	18 (25%)	17 (13%)
C	14 (19%)	26 (21%)
D	12 (17%)	55 (43%)
Neurological Level (Baseline)		
C1 – C2		18 (15%)
C3		20 (16%)
C4		41 (32%)
C5		29 (23%)
C6	38 (52.5%)	11 (9%)
C7		3 (2%)
C8		2 (1%)
T1		3 (2%)

details of the data collection process are available in the original manuscript.[1] Study 2: A multi-center, observational, longitudinal cohort study was conducted in Canada and Europe, which included 12 sites (Table 1). Enrollment and follow up data collection took place between 2009 and 2013. The studies were conducted separately and the data was merged in 2015. Patients were included if they (1) sustained a traumatic cervical SCI; (2) presented with a minimum UEMS of 1/25 with at least one motor point in the C5 myotome on the right or left side; (3) were graded as ASIA Impairment Scale (AIS) of A, B, C or D; (4) were between the ages of 16 and 85; and (5) were able to provide informed consent. Patients were excluded if (1) they had any additional cause of upper limb neurological impairment and/or (2) a moderate-to-severe brain injury precluding their ability to participate in the study effectively. A total of 127 datasets were collected in this study. Further study design details are available from the published work.[7,8,13] Within, both studies the GV1, International Standards of Neurological Classification of SCI (ISNCSCI), Spinal Cord Independence Measure (SCIM) and Clinician/Participant Questionnaires were administered and the Capabilities of Upper Extremity Questionnaire (CUE-Q) was administered in North America. Note Table 2 defines in detail the n included in each analysis.

A priori we anticipated the following: (1) Inter rater and test retest reliability for subtest scores would be greater than or equal to an intra class correlation coefficient (ICC) value of $r = 0.80$. According to Streiner and

Table 2 Sampling and analysis explanation.

Specific Analysis Done	Sample Source	N available for analysis	Geographical Source
Inter-rater Reliability	Chronic	72	North America and EU
Test Retest Reliability	Chronic	45	North America
Concurrent Validity SCIM/SS	Chronic	72	North America and EU
Concurrent Validity CUE-Q	Chronic	45	North America
MDD	Chronic	72	North America and EU
Responsiveness	Longitudinal	127*	Canada and EU

SCIM-Spinal Cord Independence Measure; SS-Selfcare Subscore; CUE-Q-Capabilities of Upper Extremity Questionnaire; MDD- Minimally Detectable Difference; *varying n based on paired data, see Table 6

Norman, reliability is considered to be good if the ICC is above 0.75.[15,16] Reliability was analyzed by conducting non-parametric intra class correlation coefficients (ICC) on the GV2 GR-Sens and GR-PP subtest total scores. Concurrent validity was analyzed by conducting Pearson correlation coefficients between GV2 subtest scores with SCIM total scores, SCIM Self-care Subscores (SS), and CUE-Q scores. The *a priori* hypothesis was that, concurrent validity would be demonstrated by a moderate association of GV2 subtest scores with the SCIM, SCIM-SS and CUE Q. MDD was analyzed using the ICC values of inter-rater reliability, the Beckerman *et al*. method was applied for this calculation.[17] Responsiveness was calculated using the standardized response mean (SRM) and comparing the values to the previously published SRM values for GV1.[18] Furthermore, the SRM t-statistic was calculated for the GR-Sens and GR-PP subtests for both the GV1 and GV2 and compared to the ISNCSCI Light Touch scores and SCIM-SS scores.

Results

Sample

The data sample is described in Table 1, where the specific analyses conducted on each sample is described. Table 2 simply defines which aspects of the study samples were included in which aspects of the analysis. Tables 3–5 report on the updated values of reliability, MDD, concurrent validity and responsiveness respectively. Shaded in gray are the values re-calculated for the GV2, which can be combined with the previously calculated values for the GR-Str, GR-PA established during the development of GV1.

Table 3 Reliability values of subtest scores within GV2.

Subtest	Inter rater Reliability		Test Retest Reliability	
	ICC	CI	ICC	CI
GR-Palmar Sensation Right	0.84	0.75–0.90	0.95	0.90–0.97
GR-Palmar Sensation Left	0.93	0.89–0.95	0.97	0.94–0.98
GR-Strength Right	0.95	0.93–0.97	0.98	0.98–0.99
GR-Strength Left	0.95	0.92–0.97	0.98	0.96–0.98
GR-Prehension Ability Right	0.95	0.92–0.97	0.98	0.96–0.99
GR-Prehension Ability Left	0.95	0.92–0.97	0.98	0.97–0.99
GR-Prehension Performance Right	0.97	0.95–0.98	0.96	0.92–0.97
GR-Prehension Performance Left	0.96	0.95–0.98	0.97	0.94–0.98

GR-GRASSP; All values with significance level of $P < 0.001$; shaded rows represent newly calculated values for GV2.

Reliability

Inter rater and test retest reliability were calculated for GR-palmar sensation and GR-prehension performance, right and left sides separately. The reliability values re-

Table 4 Minimally detectable difference values for subtest scores within GV2.

	SEM	SRD	# of items	SRD/ items	Change in Scores
GR-Palmar Sensation Right (0–12)	1.41	3.27	3	1.09	4 pts or more
GR-Palmar Sensation Left (0–12)	0.93	2.68	3	0.89	4 pts or more
GR-Strength Right (0–50)	3.34	9.23	10	0.92	5 pts or more
GR-Strength Left (0–50)	3.47	9.59	10	0.95	5 pts or more
GR-Prehension Ability Right (0–12)	0.99	2.76	3	0.92	4 pts or more
GR-Prehension Ability Left (0–12)	0.98	2.76	3	0.92	4 pts or more
GR-Prehension Performance Right (0–20)	1.08	2.89	4	0.75	3 pts or more
GR-Prehension Performance Left (0–20)	1.12	2.94	4	0.74	3 pts or more
GR-Strength (0–100) R + L	5.51	15.20	20	6.71	7 pts or more
GR-Palmar Sensation (0–24) R + L	2.31	4.21	6	0.70	3 pts or more
GR-Prehension Ability (0–24) R + L	1.81	4.90	6	0.81	4 pts or more
GR-Prehension Performance (0–40) R + L	2.48	4.36	8	0.55	3 pts or more

GR-GRASSP; SEM-Standard Error of Measure; SRD-Smallest Real Difference.

Table 5 Concurrent validity values for GV2 subscores.

	SCIM	SCIM-SS	CUE-Q
GRASSP Sensibility	0.53	0.72	0.79
GRASSP Strength	0.59	0.74	0.76
GRASSP Prehension	0.71	0.82	0.83

Pearson correlation coefficient: low concurrence = 0.50–0.59; moderate concurrence = 0.60–0.79; substantial concurrence = 0.80–01.00. All pearson correlation coefficients had significance level of $P < 0.001$.

calculated for the modified subtests actually indicate improved reliability for both inter rater and test retest reliability. See Table 3 for the non-parametric ICC values for the subtests within the GV2, including the confidence interval for each value.

Minimally detectable difference

MDD is calculated with the reliability value, therefore, the re-calculated values for the modified subtests are also decreased from the GV1. Table 4 presents the MDD values for the GR-palmar sensation and GR-prehension performance, right and left sides separately as well as for bilateral scores. See shaded values in Table 4.

Concurrent validity

Concurrent validity reflects the strength of the association between two measures. In the case of GRASSP, which is more of an impairment measure, we anticipate a positive and moderate correlation. The re-calculated values do reflect very similar findings to the original GV1. Table 5 displays all of the concurrent validity values. Right and left data were combined for this analysis and Pearson correlation coefficients were conducted to establish the association between GV2 subtests and the CUE-Q, SCIM and SCIM-SS. It was noted that all associations were positive and of moderate strength with significance (*P*-value) at a level of less than 0.0001.

Responsiveness

Responsiveness for the modified subtests rendered values that show the GV2 to be responsive, however, not as sensitive to the GV1 when considering the GR-PP subtest alone. Table 6 presents the descriptive statistics and responsiveness values for GR-Sens and GR-PP bilateral scores. Four pairs of calculations are presented and compared with the original values and the comparative measures. Baseline to 6 months, baseline to 12 months, 1 month to 6 months and 1 month to 12 months responsiveness values are reported in Table 6.

Discussion

GRASSP V1 and V2 are measures of upper limb impairment specific to traumatic tetraplegia. Both versions capture sensorimotor information that describe the neurological status of the upper limb. GRASSP is currently used in clinical drug trials as it captures subtle differences in sensorimotor function of the upper limb at a time when small changes may be the only detectable benefit of biologics. The GRASSP can be administered as early as a few days post injury, thus, capturing baseline data elements as being feasible. GRASSP is also used as primary endpoint in therapeutic studies that are establishing the benefit of

Table 6 Descriptive statistics and responsiveness values for GV2 new subtests.

Group	GV2 Subtest	N	Baseline	1 month	6 month	12 month
Whole	GR-Sens	71	13.6 (7.7)	14.3 (8.3)	18.1 (6.7)	18.4 (6.4)
AB	GR-Sens	17	11 (8.5)	10.1 (7.9)	14.8 (8.8)	15.4 (8.6)
CD	GR-Sens	47	14.2 (7.2)	16.3 (7.8)	19.1 (5.5)	19.5 (4.9)
Whole	GR-PP	108		16.4 (14.9)	23.5 (14.5)	25.9 (13.2)
AB	GR-PP	33		5.1 (8.7)	13.4 (13.6)	14.7 (12.7)
CD	GR-PP	64		26.8 (14.1)	32.1 (12.3)	31.3 (9.8)
Group	GV2 Subtest	N	Mean Diff	SRM	ES	SE
Whole BL-6mo	GR-Sens	102	3.41 (3.97)	.84	.41	.39
AB BL-6mo	GR-Sens	17	3.89 (4.65)	.86	.45	1.13
CD BL-6mo	GR-Sens	43	3.15 (3.56)	.88	.41	.54
Whole BL-12mo	GR-Sens	60	5.8 (4.5)	1.28	.77	.58
AB BL-12mo	GR-Sens	17	5.65 (5.1)	1.11	.69	1.24
CD BL-12mo	GR-Sens	42	5.8 (4.3)	1.35	.81	.67
Whole 1mo-6mo	GR-PP	99	8.7 (9.4)	.93	.62	.94
AB 1mo-6mo	GR-PP	33	8.5 (9.7)	.88	.97	1.7
CD 1mo-6mo	GR-PP	64	5.76 (8.6)	.67	.41	1.1
Whole 1mo-12mo	GR-PP	89	9.7 (8.6)	1.12	.66	.91
AB 1mo-12mo	GR-PP	31	9.9 (8.1)	1.22	1.11	1.47
CD 1mo-12mo	GR-PP	58	9.6 (8.8)	1.10	.69	1.15

GR-Sens = GRASSP Sensation; GR-PP = GRASSP Prehension Performance; Mean Diff = Mean Difference; SRM = Standardized Response Mean; ES = Effect Size; SE = Standard Error of Mean.

intensive therapies during the post acute and chronic phases of SCI injury. Again, the GRASSP is used in these studies to capture subtle change, but also to capture the changes in neurological state, which is the construct of interest.

In 2009 for the GRASSP to be accepted as the measure of choice for clinical SCI research, psychometric testing with the tetraplegic population was a requirement. The GDRG and partner institutes (Cross sectional and longitudinal study teams) conducted two important studies to capture the data required for reliability, validity, responsiveness, MDD and minimally clinical important difference testing. Because of these studies, the developmental process and robust psychometrics development; GV1 has been used widely within the SCI community; both in clinical drug trials and clinical therapeutic studies and as a clinical outcome measure for day-to-day use in clinical settings. The rasch analysis identified that either the dorsal or palmar sensation test points were redundant. It was decided by the investigative team to remove dorsal locations, as palmar sensation is more reflective of functional outcome. The rasch analysis did identify specifically that the opening of jars and coin tasks were redundant and therefore, those were the tasks removed.[13] Pouring of water incorporates more arm function than some of the other tasks, which is likely why it was not shown to be redundant. Whereas, picking up the coins and dropping them into the slot could be similar to picking up the key and picking up the nuts tasks.

This analysis (study) calculates a new set of psychometric properties for the subtests that have been reduced within GV1. The purpose of conducting these calculations was to provide the SCI research and clinical community the metrics that can support use of GV2. In summary it was determined that the reduction in test items, did not affect the reliability, validity or MDD values, and the responsiveness is only slightly reduced the modified subtests. The GV2 remains as robust as GV1. Therefore, GV2 is as psychometrically sound as GV1.

The modifications made to render the GV2 are summarized in Fig. 1 and in the methods section. The purpose of modifying the measure was to identify any redundant items and reduce the length of the test to improve uptake and feasibility of application. By improving the feasibility of use, uptake can be enhanced. GV2 does take approximately 15 min less time to administer and remains to have robust psychometric properties. Therefore, where researchers have been concerned about burden of outcomes assessments on the patient, GV2 can be used with the same important information captured.

Reliability and validity values remain almost identical to the GV1, with some slight increases in reliability values as well as improved validity. This is likely due to reduction in score ranges, which would typically reduce variability of scores. Therefore, GV2 can be administered repeatedly with reliable and valid results for longitudinal comparison, and by multiple assessors if needed. The MDD values are slightly different, however, remain useful and have been calculated for unilateral and bilateral scores together. These values can be used to determine if the change in score is a true reflection of clinical change. It can be noted that the mean differences in Table 6 reflect true clinical change.

Responsiveness was evaluated by calculating the mean difference between paired data, the SRM, SE and ES. GR-Sens in the GV2 is more sensitive and responsive as a subtest than the light touch of the ISNCSCI and the GV1 Sensation subtest. GR-PP did show to be responsive across timepoints as per the delta of the means, and the SRM values do show the GR-PP subtest to be responsive to change, however, not as sensitive as the GR-PP subtest in the GV1. The SCIM-SS shows to be more responsive by way of SRM values, however, this is mainly due to the larger SD that is inherent in the GR-PP (broader score range). Using a t-statistic of 10.7 for the SCIM-SS and 9.25 for the GR-PP, the responsiveness is much closer. This indicates that the standard deviation values may be reducing the SRM values for GR-PP. However, when reviewing the means over time and the mean differences, the SCIM-SS scores do not change significantly, whereas the GR-PP scores show large increases and are accompanied by larger variances in the sample. GR-PP in GV1 will capture a more sensitive perspective of change in the hand, while GV2 may not be as sensitive. Nonetheless, the GR-PP in GV2 is accompanied by an MDD score that will facilitate the clinician or researcher in understanding which version of the GRASSP to use.

One might ask now when should GV2 be used instead of GV1. Since the use of GRASSP is multi-faceted. The developers suggest where GRASSP is a secondary outcome and there is a general interest in upper extremity status the GV2 can be used. When hand function recovery is the primary interest, then GV1 should be implemented. Furthermore, if one remains unsure about which version to implement, GV1 data elements can be captured and the analysis can be done to reflect GV1 or GV2 or both. This can be done as the

GV2 simply has items eliminated. Currently any investigators that are capturing GV1 data can also conduct a GV2 analysis to determine which version is more suitable for their purposes. However, based on the items reduced the GV1 should be implemented when the hand is the construct of interest and the GV2 should be implemented when the entire upper limb is the construct of interest.

Conclusion

The GRASSP was designed to be a sensitive clinical impairment measure specific to the upper limb for TSCI with a sound theoretical framework and relevant domains. Because of the robust development the GRASSP, it has a unique place in the SCI field today as an important outcome measure that is used in both research and clinical settings. The GRDG has presented a second version, which is a shortened version and shows consistent psychometric properties that place GV2 at the same level as GV1. A third version of the tool called the GRASSP Version Myelopathy (GVM) has also been developed for individuals presenting with non-traumatic SCI. Application of versions may vary based on the detail of measurement desired and whether there is more interest in the hand versus the whole upper limb. The version used should be based on the primary objective/s of the study or clinical setting. GV2 elements can be extracted if GV1 is implemented and enables the selection of the version to take place after data acquisition and/or analysis if appropriate. Ultimately, the GV2 can be implemented in research and clinical settings with the same level of confidence that the GV1 has been implemented with. The developers believe the release of GV2 will be well received by the community as it addresses issues of feasibility of the test and it ensures the field that there is ongoing oversight and development of the GRASSP outcome measure.

Disclaimer statements

Contributors None.

Funding This work has been supported by the Craig Neilsen Foundation, Canadian Institutes of Health Research, Physiotherapy Foundation of Canada, Ontario Neurotrauma Foundation, Rick Hansen Institute and Christopher and Dana Reeve Foundation in North America. European colleagues were supported by IRP and the EMSCI. The author(s) disclose receipt of financial support for the Swiss Paraplegic Center, Nottwil, Switzerland, and the Clinical Research Priority Program NeuroRehab of the University of Zurich, Switzerland.

Disclosure Statement S. Kalsi-Ryan owns the company Neural Outcomes Consulting Inc. that manufactures the GRASSP. The original inventors: S. Kalsi-Ryan, M. Verrier, M. Fehlings, A. Curt receive licensing and sales royalties from the GRASSP commercialization.

ORCID

Sukhvinder Kalsi-Ryan http://orcid.org/0000-0003-2332-5986
Michael Fehlings http://orcid.org/0000-0002-5722-6364

References

1 Kalsi-Ryan S, Beaton D, Curt A, Duff S, Popovic MR, Rudhe C, Fehlings MG, Verrier MC. The graded redefined assessment of strength sensibility and prehension: reliability and validity. J Neurotrauma 2012;29(5):905–14.
2 Kalsi-Ryan S, Curt A, Verrier MC, Fehlings MG. Development of the graded redefined assessment of strength, sensibility and prehension (GRASSP): reviewing measurement specific to the upper limb in tetraplegia. J Neurosurg Spine 2012;17(1 Suppl):65–76.
3 Kalsi-Ryan S, Curt A, Fehlings MG, Verrier MC. Assessment of the hand in tetraplegia using the graded redefined assessment of strength sensibility and prehension (GRASSP): impairment versus function. Top Spinal Cord Inj Rehabil 2009;14(4):34–46.
4 Link C. The Link Hand Function Test for Participants with a Cervical Spinal Cord Injury: An Intra-Rater and Inter-Rater Reliability and Expert Opinion Evaluation Study (2004). Hogeschool van Amsterdam, Institute of Occupational Therapy, The Netherlands; University College-South, School of Occupational Therapy and Physiotherapy, Denmark; Karolinska Institutet Division of Occupational Therapy, Sweden.
5 Kalsi-Ryan S, Beaton D, McIlroy W, *et al.* The development of the quadriplegia hand assessment tool (Q-HAT) – a discriminative and evaluative approach. J Spinal Cord Med 2004;27(2):164.
6 Sollerman C, Ejeskar A. Sollerman hand function test: a standardized method and its use in tetraplegic participants. Scand J Plast Reconstr Hand Surg 1995;29:167–76.
7 Velstra IM, Curt A, Frotzier A, Abel R, Kalsi-Ryan S, Rietman JS, Bolliger M. Changes in strength, sensation and prehension in acute cervical spinal cord injury: European multicenter responsiveness study of the GRASSP. Neurorehabil Neural Repair 2015 Sept; 29(8):755–66. PMID: 25567122.
8 Kalsi-Ryan S, Beaton D, Ahn H, Askes H, Drew B, Curt A, Popovic M, Wang J, Verrier M, Fehlings M. Responsiveness, sensitivity and minimally detectable difference of the graded and redefined assessment of strength, sensibility, and prehension, version 1.0 (GRASSP V1). J Neurotrauma 2016;33(3):307–14.
9 Rice D, Faltynek P, McIntyre A, Mehta S, Foulon B, Teasell R. Upper limb rehabilitation following spinal cord injury. www.scireproject.com. 2016.
10 Anderson KD. Targeting recovery: priorities of the spinal cord-injured population. J Neurotrauma 2004;21(10):1371–83.
11 Snoek GJ MJIJ, Hermens HJ, Maxwell D, Biering-Sorensen F. Survey of the needs of patients with spinal cord injury: impact and priority for improvement in hand function in tetraplegics. Spinal Cord 2004;42(9):526–32.
12 Steeves JD, Lammertse DP, Kramer JL, Kleitman N, Kalsi-Ryan S, Jones L, Curt A, Blight A, Anderson K. Outcome measures for acute/subacute cervical sensorimotor complete (AIS-A) spinal cord injury during a phase 2 clinical trial. Top Spinal Cord Inj Rehabil 2012;18(1):1–14.
13 Velstra I-M, Fellinghauer C, Abel R, Kalsi-Ryan S, Rupp R, Curt A. The graded and redefined assessment of strength, sensibility,

and prehension version 2 provides interval measure properties. J Neurotrauma 2018;35(6):854–63.
14 Kalsi-Ryan S, Albisser U, Fehlings M, Curt A, Verrier M, Fellinghauer C, Velstra IM. GRASSP version 2: a comprehensive SCI upper limb outcome measure. Annual Meeting of the American Spinal Injury Association; May 2018; Rochester, USA.
15 Portney LG, Watkin MP. Foundation of Clinical Research: Applications to Practice. Chapter 26: Statistical Measures of Reliability; Second ed. Norwich, CT: Appleton and Lange; 2000. p. 557–86.
16 Streiner DL, Norman GL. Health Measurement Scales: A Practical Guide to Their Development and Use. Reliability Chapter 8 & 10. Second ed. New York: Oxford University Press Inc.; 1995. p. 104–27.
17 Beckerman H, Roebroeck M, Lankhorst G, Becher J, Bezemer PD, Verbeek A. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res 2001;10:571–78.
18 Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–9.