Methods of assessing associated reactions of the upper limb in stroke and traumatic brain injury: A systematic review

Michelle B. Kahn1,2, Benjamin F. Mentiplay2, Ross A. Clark2, Kelly J. Bower2, & Gavin Williams1,3,4

1Department of Physiotherapy, Epworth Healthcare, Melbourne, Australia, 2Faculty of Health Sciences, Australian Catholic University, Melbourne, Australia, 3Department of Physiotherapy, University of Melbourne, Melbourne, Australia, and 4Department of Physiotherapy, La Trobe University, Melbourne, Australia

Abstract

Objective: To determine the assessment methods for upper limb (UL) associated reactions (ARs) in people with acquired brain injury (ABI).

Methods: A systematic search of 10 databases was performed for Stage 1 to identify methods that quantify ARs of the hemiplegic UL. Stage 2 searched four databases to examine the clinimetric properties and clinical utility of these methods. Two independent reviewers identified relevant articles, extracted data, assessed study methodological quality and rated the clinimetric properties and clinical utility.

Results: Eighteen articles were included. The methods used to evaluate ARs were surface electromyography (11), goniometry (5), dynamometry (5), electrogoniometry (1), subjective clinician (2) and patient rating forms (2). Electromyography, electrogoniometry and dynamometry implemented stationary, seated positions using maximal voluntary contractions of the less impaired UL as the provocative task. Standard goniometry most frequently tested ARs dynamically, using a mobility task to provoke the AR. There was limited clinimetric data available. Only half of the assessment methods were deemed clinically feasible. The most common methods were laboratory-based.

Conclusion: There were a limited number of methods used to assess ARs in people with ABI and the measurement properties of these outcomes were largely unreported. No gold standard was identified.

Keywords

Stroke, traumatic brain injury, acquired brain injury, associated reactions, upper limb, assessment, systematic review

Introduction

Associated reactions (ARs) of the upper limb (UL) are a common phenomenon in people with acquired brain injury (ABI). The person may experience involuntary UL movements that cause awkward and uncomfortable postures. Associated reactions are pathological and occur due to damaged supraspinal structures as a result of neurological injury [1]. These involuntary movements occur globally in heterogeneous muscles on the hemiplegic side, coinciding with or following effort exerted at another body site [1–3]. Incidence rates of ARs in people with stroke have been reported to range from 29–88% [4–6]. There is no current estimate in the literature of prevalence in individuals with traumatic brain injury (TBI); however, it is likely to be similar.

Associated reactions may have a substantial impact on functional abilities and quality-of-life for people with ABI. These pathological arm movements affect a diverse range of people with ABI, including individuals with relatively good UL function. Long-term abnormal UL posturing can lead to contractures [7], entrenched aberrant movement patterns and limited UL function, particularly when standing and walking [8,9]. Associated reactions can also increase the energy requirements of walking and interfere with a person’s dynamic balance, therefore increasing the risk of falling [7]. Furthermore, visibly awkward UL postures have obvious aesthetic implications, adding to the stigma of disability. The very nature of these reactions, i.e. being ‘associated’, means they happen dynamically and unintentionally in conjunction with other movements and effort [10,11]. For example, a person’s UL may exhibit an AR during effort-dependent activities such as rising from sitting and walking [10,12]. Reducing the ARs that occur during functional activities is, therefore, often a major focus of neurological rehabilitation.

Effective treatment of ARs is important to clinicians and people with ABI. Associated reactions are commonly prioritized for treatment and their presence and severity during walking is often examined as a means of monitoring patient progress or effectiveness of interventions [9,13]. Despite this, ARs remain poorly defined and consensus is lacking in relation to the contributing factors. For example, some authors suggest that ARs co-occur with spasticity [10,12,14,15] and others believe that ARs can occur independently of spasticity [16,17]. This confusion leads to a diverse range of treatment
strategies for ARs which are not evidence-based. The lack of clarity and consensus surrounding the contributing impairments may also be exacerbated by problems with taxonomy and nomenclature in describing the phenomenon of ARs. There are a number of terms for ARs, such as ‘motor overflow’, ‘associated movements’, ‘global synkineses’ and ‘mirror movements’ that are utilized interchangeably in the literature and require clarification (see Figure 1). A thorough explanation of these terms is provided in Appendix 1 of the Supplemental Material published online.

The challenge of assessing and treating ARs is further compounded by the lack of a gold standard clinical measure to accurately quantify occurrence and grade severity. Although by definition ARs are associated with movement, the current methods for assessing ARs appear to be subjective [18] or to involve relatively stationary, seated testing positions [4]. As a consequence, these tests may not adequately reflect the true extent of ARs, as they are unlikely to accurately and objectively identify these reactions. Furthermore, they may fail to identify the contributing factors when people are engaged in provocative functional tasks such as walking. Gold standard outcome measures have established and acceptable clinimetric data. Therapists need to be confident that an outcome measure is reliable and consistent with minimal error associated with repeated use or when used by different therapists [19]. The measure needs to be valid, evaluating the specific construct (i.e. ARs) and in the correct functional context to demonstrate ecological validity [20]. The measure also needs to be sufficiently responsive to detect and quantify abnormality in patient groups, demonstrate clinically meaningful change in response to treatment or recovery and discriminate between people and levels of severity [21]. Clinical utility can be defined as a multidimensional judgement about a measurement tool’s usefulness, benefits and drawbacks. It considers whether an outcome measure is appropriate, accessible, practical and acceptable to clinicians [22]. This concept is of utmost importance as it determines the practicalities of using a measurement tool in clinical practice. A tool that displays good clinical utility will be feasible for use in clinical settings, accessible, cost effective and require minimal training for users.

The overall aim of this systematic review was to evaluate existing methods for assessing ARs. Specifically, this review aimed to: (1) identify methods used to evaluate ARs in people with ABI, (2) determine their clinimetric properties and (3) assess the clinical utility of these methods. Given the problems with taxonomy and nomenclature, the working definition of ARs for this systematic review will be:

Associated reactions are an unwanted, effort dependent limb movement that occurs following cerebral damage, where there may be sensorimotor dysfunction or insufficient postural control, so that when a stimulus is applied that goes beyond the individual’s level of inhibitory or modulatory control, it results in intermittent or sustained involuntary, heterogeneous muscle activation with abnormal limb posturing, most visible in the hemiplegic UL.

Methods

This systematic review was undertaken in two stages. Stage 1 identified the methods used to evaluate ARs. Stage 2 evaluated the clinimetric properties and clinical utility of these methods. The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines were used in the presentation of results [23].

Stage 1 (Aim 1)

To identify the current methods for assessing ARs in people with stroke or traumatic brain injury.

Identification and selection of studies

Selection criteria. The selection criteria for included studies are summarized in Table I. Due to the problems with taxonomy and nomenclature for ARs, the term ‘mirror movement’ was included in the search, in spite of its apparent incongruence to ARs. This strategy ensured that any papers that used the term mirror movement to describe ARs were captured. Papers that aligned with the working definition of ARs were included. These papers had a focus on ARs in the hemiplegic
Table I. Selection criteria.

| Inclusion criteria                                                                 | Exclusion criteria                                      |
|------------------------------------------------------------------------------------|---------------------------------------------------------|
| Diagnosis of acquired brain injury (stroke or traumatic brain injury)               | Focus on children (< 18 years of age)                   |
| Adult or adolescent onset > 18 years of age                                        | Grey literature, conference abstracts, review articles, |
| An outcome measure of ARs, motor overflow or global synkinesis in the              | single case studies                                      |
| hemiplegic upper limb                                                              | Published in languages other than English               |
|                                                                                   | Focus on mirror movements                               |

ARs, Associated Reaction.

UL due to effort exerted at a remote site, i.e. the contralateral side of the body or with mobility tasks.

Search strategy. A broad search strategy was employed to ensure all related articles were retrieved. Key search terms and relevant synonyms were kept consistent across all databases and, where possible, relevant medical subject headings (MeSH terms) were used. Truncation symbols and wildcards were used to capture all suffix variations of a root word (i.e. Brain Injur* to capture Brain Injury and Brain Injuries). Search terms used to define the population were not restricted to one specific neurological diagnosis and so included stroke, traumatic brain injury, neurological and hemiplegia. The terms searched for ARs were associated reactions, associated movements, global synkineses, mirror movements and motor overflow. A detailed example of the search strategy applied to the Cumulative Index to Nursing and Allied Health Literature (CINAHL) including the medical subject heading terms is shown in Appendix 2 of the Supplemental Material published online. No limitations were placed on publication date, but language was restricted to English only. Targeted searching of the reference lists from included articles and primary authors was also performed.

Ten electronic databases were searched to find papers from inception to September 2014. The databases searched were Scopus (27 September 2014); CINAHL (30 September 2014); Sports Discus (30 September 2014); PEDro (30 September 2014); Psychinfo (28 September 2014); Cochrane Reviews (30 September 2014); PubMed (27 September 2014); MEDLINE (27 September 2014); Embase (27 September 2014) and Web of Science (27 September 2014).

Selection of articles

The initial yield included all original articles combined from each database after excluding duplicates. One reviewer (author M.K) assessed the eligibility of the title and abstract of each article in the initial yield and all irrelevant articles were removed. Full text articles were retrieved from the remaining articles and two reviewers (authors M.K and B.F.M) then independently applied the selection criteria to these articles. The final articles to be included were agreed upon by both reviewers. Where a discrepancy arose a third reviewer determined whether the article should be included (author K.B).

Data extraction and synthesis

Due to the variation in aims and methods of data collection across the studies included in this systematic review, a data extraction proforma was developed to capture the required study information. Basic study purpose and design, participant demographics, including the use of healthy controls and sample size was collated on one data extraction sheet. A separate data sheet was utilized to extract the descriptive information of each of the outcome measures utilized to assess ARs. Information was extracted and tabulated to summarize the assessment methods including the muscles or joints of the affected limb tested, whether the participant was stationary or moving dynamically during the test, participant set up position and the action or test used to induce the AR. Data extraction was performed independently by two reviewers (authors M.K and B.F.M). In cases of disagreement, a third reviewer was involved to reach consensus. Data extraction and synthesis was performed using the PRISMA Checklist and results were reported according to the PRISMA Flow Diagram [24].

Data analysis

The quality of each article was evaluated using a modified tool developed by Law and MacDermid [25], from the Law and MacDermid critical review form for quantitative studies. This method was chosen as it is appropriate for use with a range of research designs, particularly since the majority of the studies identified in this literature review were observational studies with no randomization or interventions [25]. The Law and MacDermid critical review tool was used to provide a standardized format for quality evaluation and examining the internal and external validity of each study. The Law and MacDermid tool appears to be a robust method of assessing the types of studies retrieved in relation to the aims of this review. However, the measurement properties of this tool have not yet been established and, therefore, results need to be interpreted with caution [26]. Therefore, this tool was used to highlight individual items of methodological quality and overall quality across the range of included studies for each item, without reporting of average scores for each individual study. The tool consists of 15 questions, each scored 0, 0.5 or 1, with a higher score indicating...
**Table II. Definitions of measurement properties.**

| Measurement property | Definition |
|----------------------|------------|
| Reliability          | The degree to which the instrument is free from measurement error associated with multiple tests. The extent to which the scores for patients who have not changed are the same for repeated measurement over time (test–re-test), by different persons on the same occasion (inter-rater) or by the same persons on different occasions (intra-rater) [21,27,30,50]. |
| Internal consistency Validity | The degree to which items are correlated, thus measuring the same concept [21,27,50]. |
| Face validity        | The extent to which an instrument truly measures the construct(s) it purports to measure. There are a number of different types of validity [21,50,51]. |
| Criterion validity   | Indicates that the outcomes of one instrument, the target test, can be used as a substitute measure for an established gold standard reference test [21,50,51]. |
| Content validity     | The extent to which the domain of interest is comprehensively sampled by the items and is free from the influence of other factors that are irrelevant to the purpose of the measurement [21,27,50,51]. |
| Concurrent validity  | The extent to which the scores on one instrument are compared to and relate to another in a manner that is consistent with a theoretically derived hypothesis [51]. |
| Construct validity   | The ability of an instrument to measure an abstract construct and the degree to which the instrument reflects the theoretical components of the construct [51]. |
|                      | • Known groups validity: Looks at whether a test can discriminate between individuals who are known to have the trait and those that do not or differentiates individuals of varying severity [51]. |
|                      | • Discriminant/convergent validity: Measures how a test relates to tests of different or the same constructs. Discriminant validity indicates that different results or low correlations are expected from measures that are believed to assess different characteristics. Convergent validity indicates that two measures that are believed to reflect the same underlying phenomenon will yield similar results or will correlate highly [51]. |
| Ecological validity  | The property of a scale that indicates that a test result is relevant for everyday life situations [40]. |
| Responsiveness       | The ability of an instrument to detect change over time in the construct to be measured. There are a number of different measures of responsiveness [21,50]. |
| Interpretability     | The degree to which one can assign qualitative meaning, that is clinical or commonly understood connotations to an instrument’s quantitative scores or change in scores [21,50]. |
| Floor and ceiling effects | The number of respondents who achieved the lowest or highest possible scores [51]. |

greater methodological quality. Items relate to the description or justification of (1) study purpose, (2) background literature, (3) sample size, (4) sample size justification, (5) inclusion/exclusion criteria description, (6) ethical procedures, (7) outcome measures, (8 and 9) outcome reliability and validity, (10) intervention, (11 and 12) avoidance of contamination and co-intervention, (13) analysis methods, (14) clinical importance and (15) conclusions [25]. To ensure reliable interpretation of the quality assessment items, two reviewers rated all articles independently. Where any discrepancies arose, a third reviewer was included to obtain agreement.

**Stage 2 (Aims 2 and 3)**

To determine the clinimetric properties and clinical utility of existing methods of assessing ARs.

**Search strategy and study selection**

To identify the clinimetric properties of the methods utilized in the studies to assess ARs, a second search was performed. Four databases (Medline, CINAHL, EMBASE and Scopus) were searched. These databases were chosen for Stage 2 as all of the included studies from Stage 1 were identified in these four large databases. The methods used to assess ARs in Stage 1 were searched in combination with the terms for the clinimetric properties. An example of this search is provided in Appendix 3 of the Supplemental Material published online. The reference lists of all of the included papers were hand searched for papers relating to clinimetric properties. In the absence of any published clinimetric data, the authors of each of the papers identified in Stage 1 were contacted in an attempt to gain further information.

**Data extraction and synthesis**

The clinimetric properties and their definitions have been defined in Table II. The desired clinimetric properties have been used to evaluate the methods of assessing ARs, identified in Stage 1. Two reviewers (authors M.K and B.F.M) independently rated the methods against these criteria using the rating system developed by Terwee et al. [27], that has also been used in other systematic reviews examining outcome measures [28,29]. The items received a (+) positive rating if sufficient information was available and bias unlikely, (±) indeterminate rating where available information was unclear or the method used was doubtful and a (−) negative rating if sufficient information was available but the measure did not meet the criteria. If no information was available the criterion was marked with a 0. In cases of uncertainty or disagreement, a third reviewer was involved to reach consensus. Where methods were used in multiple studies and obtained different ratings, all were presented with the corresponding reference outlined. The rating of clinical utility, based on scoring system by Tyson and Connell [30] is outlined in Table III. Data extraction and synthesis was performed...
Table III. Clinical utility rating [30].

| Criteria | Definition | Scoring* |
|----------|------------|----------|
| Time     | Time taken to administer, analyse and interpret the measurement tool | 3 = less than 10 minutes, 2 = 10–30 minutes, 1 = 30–60 minutes, 0 = > 1 hour |
| Cost     | Costs associated with purchasing and servicing the device | 3 = < £100; 2 = £100–500; 1 = £500–1000; 0 = > £1000 or unknown |
| Training and equipment | Does the measurement tool need specialist equipment and training to use? | 2 = no, 1 = yes but simple and clinically feasible, 0 = yes and not clinically feasible/unknown |
| Portability | Is the measurement tool portable? Can it be taken to the patient | 2 = yes easily (i.e. can go in pocket), 1 = yes (i.e. in a briefcase or trolley), 0 = no or very difficult |

* Addition of these scores gave a maximum score of 10. A score of 9 or greater was required for a measure to be recommended for clinical use.

using the PRISMA Checklist [24]. Where sufficient data existed, the ‘consensus-based standards for the selection of health measurement instruments’ (COSMIN) checklist was used to determine the methodological quality of the studies included and to evaluate the appropriateness of the statistical methods [21].

Results

Stage 1—Associated reaction assessment methods

Figure 2 summarizes the stages involved in identifying, screening and assessing the eligibility of the articles according to the PRISMA guidelines [24]. After the removal of duplicates, the initial yield was 767 articles. Seventeen articles were identified as meeting the selection criteria [2–6,8,10–12,14,15,17,31–35] and one additional article [36] was found during targeted searching. A total of 18 articles were included.

Description of studies

Characteristics of included studies. Study characteristics are summarized in Table IV. There were 13 observational studies, six of which were case-control studies with healthy control comparisons. There were two randomized controlled trials and three pre- and post-test case control series. Most of the studies included middle-aged adults, predominantly male and chronic in relation to the length of time since their ABI. The majority (89%) of the studies investigated stroke populations. One study investigated TBI subjects [37] and another investigated a mixed neurology cohort of stroke and traumatic brain injury and with one neuro-oncology subject [6].

Methods of assessing ARs. The methods used in each study to assess ARs are outlined in Table V. All 18 studies used different testing protocols. The methods used to assess ARs were surface electromyography (SEMG) [11], standard goniometry [5], dynamometry or load cells [5], electromyography [1], subjective clinician rating forms [2] and patient rating forms [2]. With the exception of the Associated Reaction Rating Scale [18], there were no specific names attributed to any of the methods used to specifically quantify ARs. A number of the studies utilized multiple methods to assess the ARs.

Surface electromyography was the most commonly employed method of assessing ARs. Nine of the 11 SEMG studies utilized stationary tests of ARs performed in a seated position. The majority of the seated SEMG studies employed a maximal voluntary contraction (MVC) of the intact UL as the action inducing the AR in the hemiplegic UL. The two dynamic tests with SEMG included walking as the provocative activity in one study [38] and a standing leg lift in the other [14]. The biceps brachii was the most common muscle group measured with SEMG [4,14,15,31–33,35,36,38], followed by the forearm flexors [15,17,31–33,35,39], triceps [31–33,35,36,38], brachioradialis [31–33,35,38] and forearm extensors [17,33,35,39]. Two studies [33,35] also included SEMG testing of the middle deltoid, pectoralis major and pronator teres muscles.

Standard goniometry was the next most common method overall and was most frequently used to test ARs under dynamic conditions. Four of the five studies used a dynamic mobility task including walking [12,38], sit-to-stand [3] and standing leg lift [14]. The fifth study used goniometry in a supine position where the participant was required to perform a maximal grip with their intact or less affected hand [10]. The use of goniometry to measure the AR in the hemiplegic UL was only measured at the elbow joint to evaluate the change in elbow position during the test procedure. The other joints of the hemiplegic upper limb including the shoulder, forearm, wrist or hand were not assessed, nor were any multi-planar movements.

Dynamometry or load cells were used in five studies to measure the torque produced by ARs in the hemiplegic UL. All of these studies used stationary testing methods performed in a seated position. Three studies used a maximal grip of the intact hand to induce the AR [17,31,39] and the other two studies used a MVC of the intact elbow flexors [4,37] and the intact lower limb knee extensors [4]. The abnormal finger flexion grasp of the hemiplegic UL during the AR was measured with a hand held dynamometer in two of these studies [17,39]. The elbow flexor force produced by the AR on the hemiplegic side was measured in another two studies [4,37]. One of these studies used dynamometry to measure the torque produced by the AR at the shoulder, elbow and forearm [31].

Two different clinician rating forms were used to rate ARs in two separate studies [6,18]. The Associated Reaction Rating Scale was used to subjectively rate the extent and severity of the AR in all the joints of the UL during a dynamic task of sit-to-stand [18]. The other clinician rating form [6] required clinicians to visually observe if there was any movement at any joint of the hemiplegic UL in response...
to resisted elbow and ankle movements on the intact side while in a supine position.

Patient rating forms were used in two studies [5,39]. These forms required the patient to nominate the activities of daily living that triggered their AR and then rate the severity of the AR. One study rated the arm movement on a three point scale (no, minimal and considerable arm movement) during their elected activity [5]. The other rated the activities of daily living with a 10 point Likert scale ranging from ‘no tightness’ to ‘worst tightness ever’ [39]. These activities could be stationary or involuntary (e.g. autonomic activities such as yawning, sneezing or coughing) or could be dynamic (e.g. dressing, walking, climbing stairs) depending on what the patient selected.

Quality assessment

Results of the Law and MacDermid critical review form for quantitative studies are outlined in Table VI. Areas that scored well across the studies were: purpose outline,
Table IV. Study participant characteristics.

| Reference (year) | Population | Sample size (n) | Healthy controls (n) | Male, n (%) | Age (years) (mean ± SD) | Time since injury (weeks) (Mean/Range/SD) |
|------------------|------------|-----------------|----------------------|-------------|-------------------------|-----------------------------------------|
| Ada and O’Dwyer [4] (2001) | ST | 24 | — | 13 (54%) | NR | NR/4–56/NR |
| Bhakta et al. [17] (2001) | ST | 49 | — | 16 (33%) | NR | NR/1–12.3/NR |
| Bakheit and Sawyer [12] (2002) | ST | 8 | — | 5 (63%) | 59.1 ± 12.56 | 29.5/12–51/12.09 |
| Bhakta et al. [39] (2008) | ST | 34 | — | 23 (58%) | BT-A: 60.2 ± NR | BT-A: NR/0.8–33.2/NR |
| Boissy et al. [31] (1997) | ST | 15 | 11 | 8 (53%) | 48.5 ± 13.77 | 91.2/2–80/NR |
| Boissy et al. [32] (2000) | ST | 11 | 10 | 8 (73%) | 42 ± 11.4 | 102/16.1–345.3/113.4 |
| Brunstrom [6] (1956) | ST, TBI, NO | 26 | — | 16 (62%) | 57.35 ± 12.03 | 225.2/24–1461/NR |
| Dickstein et al. [14] (1995) | ST | 9 | 6 | 4 (44%) | NR | NR/4–16/NR |
| Dickstein et al. [38] (1996) | ST | 14 | — | 8 (57%) | 62 ± 18.33 | 24.73/8–52/18.68 |
| Dvir and Panturin [3] (1993) | ST | 33 | — | 13 (39%) | 52.7 ± NR | 7.3/12–72/NR |
| Dvir et al. [10] (1996) | ST | 20 | — | 12 (60%) | 55.4 ± 16.3 | 137.6/3–132/NR |
| Honaga et al. [15] (2007) | ST | 10 | — | 7 (70%) | 65.2 ± 12.94 | 53.2/6–30/7.6 |
| Hwang et al. [33] (2005) | ST | 20 | 20 | 13 (65%) | 51.3 ± 10.42 | 12.9/1.5–24/7.37 |
| Lazarus [37] (1992) | TBI | 5 | 5 | 4 (80%) | 25.28 ± 1.23 | — |
| Macfarlane et al. [18] (2002) | ST | 19 | — | 11 (58%) | 53 ± NR | 21.6/1–27/NR |
| Mills and Quintan [36] (1985) | ST | 11 | — | 7 (64%) | 53.9 ± 19.3 | 7.4/2.3–19/1.4/7 |
| Mulley [5] (1982) | ST | 40 | — | 35 (88%) | NR | — |
| Tung et al. [35] (2011) | ST | 30 | 15 | 18 (60%) | 52.8 ± 6.4 | NR/8–104/NR |

ST: stroke; TBI, traumatic brain injury; NO, neurooncology; SD, standard deviation; NR, not reported; —, not applicable; BT-A, botulinum neurotoxin A group; Placebo, group receiving placebo.

Stage 2—Clinimetric properties and clinical utility of associated reaction measurement methods

The use of the COSMIN checklist to rate the clinimetric properties and statistical outcomes of the AR assessment methods was not possible due to the lack of data available for the methods used to quantify ARs. The searches from Stage 2 yielded no additional information on the clinimetric properties of these methods for the assessment of ARs. All 18 primary authors were contacted in an attempt to obtain further information. Only one author provided additional information in relation to their method used to assess ARs [32]. This author had an unpublished study on the test–re-test reliability of their SEMG assessment of AR, but this was not included due to the exclusion criteria of grey literature. The extent to which each AR assessment method demonstrated the clinimetric properties, extracted from the 18 included studies, is outlined in Table VII. No method of assessment of AR identified in the literature fulfils all of the required criteria for clinimetrics and clinical utility.

Reliability

Clinician rating form. Macfarlane et al. [18] investigated the inter-rater reliability of the Associated Reaction Rating Scale showing good correlations between two raters in total (r = 0.89) and modal scores (r = 0.88). Weighted Kappa scores for inter-rater agreement for each item were greater than 60% for all items except for the item titled ‘release of the AR’. This item demonstrated only moderate agreement (weighted kappa values = 0.43 for day 1 and 0.53 for day 2). Intra-rater reliability between days 1 and 2 of testing showed absolute agreement for over 70% of items for both raters for all items, except for the ‘release’ item (weighted kappa values of 0.71 and 0.61 for raters one and two, respectively). Strong correlations were also demonstrated for intra-rater reliability between the 2 days of testing for each rater in total (r = 0.91 and 0.92) and modal (r = 0.95 and 0.89) scores. However, one rater tended to rate more severely than the other for the item of ‘excursion’ (p < 0.005).

Standard goniometry. Dvir et al. [10] reported poor test–re-test reliability of standard goniometry to evaluate the AR at the elbow joint during a supine test with a large variability in scores and a coefficient of variation of 33.7%.

No data were identified in relation to the reliability of electromyography, dynamometry and load cells, electromyometry and patient rating forms for measuring ARs.

Internal consistency

No method had information regarding the internal consistency of the assessment.
Validity

None of the methods used in these studies have been formally validated to assess ARs. There appears to be no criterion standard for the assessment of ARs in the upper limb. *Criterion, content and concurrent validity* were not reported for any of the methods identified in Stage 1.

Surface electromyography. Nine studies that utilized the method of SEMG reported an aspect of construct validity,
either known groups, discriminant or convergent validity. Six studies investigated known groups validity using healthy controls as a comparator group for the people with ABI. Five of these studies [14,31–33,35] differentiated the pathological patient cohort from the healthy controls, identifying ARs only in the patient group. The remaining study [17] included healthy controls, but identified ARs in both groups, thereby detecting non-pathological associated movements in HCs. Discriminant or convergent validity are two types of construct validity. They were evaluated in six studies that utilized SEMG. Ada and O’Dwyer [4] investigated the relationship between AR and contracture ($\chi^2 = 1.19$, $p = 0.39$) and AR and spasticity ($\chi^2 = 0.36$, $p = 0.61$), showing that their method of assessing ARs was not correlated with these impairments. Similarly, another study demonstrated no relationship between peak AR using SEMG and the Modified Ashworth Score (MAS) ($r = 0.06$, $p = 0.70$) [17]. Honaga et al. [15] found there was a relationship between the biceps EMG ratio and the MAS of the elbow flexors ($r = 0.78$, $p < 0.01$) and the wrist flexors EMG ratio to the wrist flexors MAS ($r = 0.72$, $p < 0.05$), but there was no relationship to the H and T reflexes that represent hyperreflexia. The standard net excitation scores from the EMG of the AR of the hemiplegic elbow flexors were shown to correlate with three clinical assessment measures, the Barthel Index ($r = 0.49$, $p < 0.03$), Fugl-Meyer ($r = 0.64$, $p < 0.00$) and Brunnstrom scores ($r = 0.58$, $p < 0.01$), but not for the wrist and shoulder in the study by Hwang et al. [33]. Dickstein et al. [38] found that there was no correlation between the EMG activity in the biceps and brachioradialis and the angular change in elbow range of motion during walking. Dickstein et al. [14], however, failed to fulfill this criteria as they referred to multiple associations between the AR and different factors, but did not provide any statistical analyses for these.

Another study looked at the relationship between severe and moderate deficit AR groups based on their SEMG scores and the participants’ strength and spasticity; however, they also did not fulfill this criteria as they did not provide any correlations [31]. Ecological validity was only demonstrated by two of the 11 studies that used SEMG to assess ARs. These two studies evaluated ARs with SEMG during walking [38] and standing and lifting up one leg [14], whilst all of the remaining studies utilized a seated testing position.

### Standard goniometry

One study investigated known groups validity of the use of standard goniometry to measure the change in elbow range of motion when standing and lifting one leg in the air [14]. They found that the stroke patient group had earlier onset of elbow flexion before the foot was lifted compared to the healthy adults ($F = 4.76$, $p < 0.03$). Two studies that utilized standard goniometry to measure the ARs investigated convergent or discriminant validity. Both of these studies showed positive correlations between the amount of elbow flexion as measured by the standard goniometer and spasticity as measured by the MAS ($r = 0.50$, $p = 0.08$) [10,38]. There was no correlation between the change in elbow range of motion and physical function measures of the Barthel Index and Brunnstrom assessment [38]. It was also established that there was no relationship between the EMG activity in the biceps and brachioradialis and angular changes in elbow range of motion. Two of the studies did not meet the criteria for this type of construct validity, as they did not provide statistical results for their correlations [3,14]. Ecological validity was demonstrated in four of the five studies that measured the AR using standard goniometry during ‘real life’ functional tasks such as sit-stand or walking, to induce the AR [3,12,14,38].

### Table VI. Law and MacDermid quality assessment [25].

| Reference (year) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ada and O’Dwyer [4] (2001) | 1 | 1 | 1 | 1 | 0 | 1 | 0.5 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 1 |
| Bakheit and Swayer [12] (2002) | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0.5 |
| Bhakta et al. [17] (2001) | 1 | 1 | 1 | 0 | 0.5 | 1 | 1 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 1 |
| Bhakta et al. [39] (2008) | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0.5 | 1 | 1 | 0 | 1 | 1 | 1 |
| Boissy et al. [31] (1997) | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 0.5 |
| Boissy et al. [32] (2000) | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 0.5 |
| Brunnstrom [6] (1956) | 0.5 | 0.5 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | N/A (0) | N/A (0) | N/A (0) | 0 | 0.39 |
| Dickstein et al. [14] (1995) | 1 | 1 | 1 | 0 | 0.5 | 1 | N/A (0) | N/A (0) | N/A (0) | 0.4 | 0.39 |
| Dickstein et al. [38] (1996) | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 0.5 |
| Dvir and Panturin [3] (1993) | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0.5 | 0 | 1 | 0 | 1 | 0.5 |
| Dvir et al. [10] (1996) | 1 | 0.5 | 1 | 0 | 1 | 0 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 0.5 | 0.5 |
| Honaga et al. [15] (2007) | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 1 |
| Hwang et al. [33] (2005) | 1 | 1 | 1 | 0 | 0 | 0.5 | 1 | 0 | 0.5 | 1 | 0 | 1 | 1 |
| Lazarus [37] (1992) | 1 | 1 | 1 | 0 | 0 | 0.5 | 1 | 0 | 0.5 | 1 | 0 | 1 | 1 |
| Macfarlane et al. [18] (2002) | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 1 |
| Mills and Quintana [36] (1985) | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 1 |
| Mulley [5] (1982) | 0.5 | 0.5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 0.5 |
| Tung et al. [35] (2011) | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0.5 | N/A (0) | N/A (0) | N/A (0) | 1 | 0.5 |
| Average | 0.94 | 0.94 | 0.97 | 0.00 | 0.67 | 0.53 | 1 | 0.06 | 0.42 | 0.75 | 0.75 | 0.50 | 0.86 | 0.69 | 0.86 |

Items were scored 0, 0.5 or 1, where higher scores refer to high methodological quality for that item. Items related to the description or justification of (1) study purpose, (2) background literature, (3) sample, (4) sample size justification, (5) inclusion/exclusion criteria description, (6) ethical procedures, (7) outcome measures, (8 and 9) outcome reliability and validity, (10) intervention, (11 and 12) avoidance of contamination and co-intervention, (13) analysis methods, (14) clinical importance and (15) conclusions.
| Method                          | Stationary or dynamic test | Reliability | Face validity | Construct validity – known groups | Construct validity – Discriminant/ convergent validity | Ecological validity | Responsiveness | Interpretability | Floor and ceiling effects |
|--------------------------------|-----------------------------|-------------|---------------|-----------------------------------|--------------------------------------------------------|--------------------|----------------|-----------------|------------------------|
| Surface electromyography       | Stationary [4, 15, 17, 31–33, 35, 36, 39] | 0           | 0             | − [17] + [31]                     | − [31] + 33,35                                          | − [4, 15, 17, 31–33, 35, 36, 39] | − [15, 39] | − [4, 15, 17, 31–33, 35, 36, 39] | − [4, 17] |
|                               | Dynamic [14, 38]            | 0           | 0             | + [14]                           | − [14] + [38]                                          | + [14, 38]         | 0              | − [14, 38]      | 0                      |
| Standard goniometry            | Stationary [10]             | − [10]      | 0             | 0                                 | + [10]                                                 | 0                  | 0              | ± [10]          | 0                      |
|                               | Dynamic [3, 12, 14, 38]     | 0           | 0             | + [14]                           | − [3, 14] + [38]                                       | + [3, 12, 14, 38] | − [3, 12] | ± [3, 12, 14] | 0                      |
| Dynamometry                    | Stationary [4, 17, 31–37, 39] | 0           | 0             | − [17, 37] + [31]                | − [31] + [4, 17]                                       | − [4, 17, 31–37, 39] | − [37, 39] | − [17, 31–37, 39] | − [4]                 |
| Electrogoniometry              | Stationary [17]             | 0           | 0             | − [17]                           | + [17]                                                 | − [17]             | 0              | ± [17]          | 0                      |
| Clinician rating form          | Stationary [6]              | 0           | 0             | 0                                 | 0                                                     | 0                  | 0              | − [6]           | 0                      |
|                               | Dynamic [18]                | + [18]      | 0             | 0                                 | + [18]                                                 | 0                  | 0              | + [18]          | + [18]                |
| Patient rating form            | Dynamic or Stationary [5, 39]| 0           | 0             | 0                                 | + [5, 39]                                              | − [39]             | 0              | + [5, 39]       | 0                      |
Dynamometry and load cell. Boissy et al. [31] investigated known groups as a type of construct validity comparing healthy adults to the severe deficit and moderate deficit stroke groups and found that the severe deficit group had higher elbow flexion torque levels than the moderate deficit group and healthy controls ($p < 0.01$). Two other studies that used dynamometry or load cells did not positively fulfill the criteria for known groups validity as they both identified ARs in patient and healthy control groups, thereby detecting non-pathological associated movements [17,37]. From the studies included in this review, the use dynamometry or load cells to measure torque as the measure of AR did not demonstrate ecological validity, as these were all performed in a seated, stationary position.

Electrogoniometry. One study investigated a component of construct validity through measuring the impact of wrist position during the testing procedure [17]. They reported that patients with abnormal wrist flexion had a significant reduction in the measure of AR force compared to a neutral wrist position (overall median change $= −7.1$ N: Mann-Whitney test: $Z = −2.9; p < 0.01$).

Clinician rating form. Macfarlane et al. [18] reported face validity by utilizing focus groups of expert neurological physiotherapists to develop the Associated Reaction Rating Scale according to expert opinion and consensus. Ecological validity was also demonstrated in this study as the AR was assessed during the functional task of sit-to-stand.

Patient rating form. The two studies that used the patient rating forms demonstrated ecological validity, as patients were required to nominate the daily activities that provoke their AR. These activities may have been stationary or dynamic, but, given they are functionally relevant, the patient rating forms did demonstrate an aspect of ecological validity, because the nominated tasks reflect everyday life situations [5].

Responsiveness

Seven studies evaluated responsiveness by repeating the assessment at a second time point, following an intervention. These were studies that utilized methods of SEMG [15,39], standard goniometry [3,12], dynamometry [37,39] and one that used a patient rating form [39]. The criteria required to obtain a positive rating for responsiveness was not fulfilled for any of the AR assessment methods, as insufficient information was provided. Clinical measures such as minimal important difference, minimal detectable change or minimal clinical difference scores were not reported, neither were other statistical indicators such as the Liang or Guyatt index.

Interpretability

Surface electromyography. The SEMG results are presented with complex ratios and evaluation of standard net excitation scores, which are difficult to interpret without specific training and knowledge in the use of SEMG.

Standard goniometry. The standard goniometry results were reported as measures of change in elbow range of motion (degrees) during a specific task in four studies [3,10,12,14]. This score is easily interpreted by clinicians. However, there is no provision of values to determine what is considered a significant or meaningful change in elbow range of motion.

Dynamometry and load cell. Torque expressed in newton metres (Nm) as measured by dynamometry or load cells can be identified as high or low; however, the lack of specific values to rate significant or meaningful torques exerted by the hemiplegic UL AR restricts the interpretability of these scores.

Electrogoniometry. Similar to standard goniometric scores for the elbow, the wrist joint position scores from electrogoniometry, in terms of degrees, are interpretable by clinicians.

Clinician and patient rating forms. The subjective rating forms for clinicians and patients [5,18,39] also have simple interpretation of results that can easily be understood by clinicians. The results of these studies may enable clinicians to determine the presence of an AR; however, there is no provision of guidelines to determine what is considered a significant or meaningful AR or to rate the severity of the AR.

Floor and ceiling effects

There was no specific evaluation of floor and ceiling effects for any of the AR assessment methods. Two of the methods reported some information regarding this aspect of clinical metrics [4,18].

Clinician rating form. Macfarlane et al. [18] reported the scores of the Associated Reaction Rating Scale to be symmetrically distributed across the full range of the scale. However, the frequency distribution of total scores across the whole sample on 2 days of testing showed that 10 out of 76 tests yielded a score of zero, indicating that 13.2% had a floor effect. No participants scored maximally with a 12. This shows a potential floor effect for this scale.

Dynamometry and load cells. A potential floor effect was shown for the use of load cells to measure elbow flexion torque in the hemiplegic UL where a majority of the participants (83%) scored a zero (Nm) [4]. Given the high proportion of participants that failed to register a reading and no provision of other confirmatory information of the presence of an AR in this cohort, it is difficult to determine whether this method is very susceptible to a floor effect or that no AR existed in some of these participants.

There was no reference to floor and ceiling effects for any of SEMG, standard goniometry, electrogoniometry and patient rating forms.

Clinical utility

According to the criteria by Tyson and Connell [30], a clinical utility score of greater than nine implies that a tool can be
recommended for clinical use. Only three of the six (50%) methods achieved a score greater than nine, indicating that the method may be recommended for clinical use. These were standard goniometry and the clinician and patient rating forms. The other three methods, SEMG, dynamometry or load cells and electromyography were expensive, laboratory-based measures that required specific training for their use. Table VIII outlines the rating of clinical utility for each of the methods.

**Discussion**

This systematic review found that there were a number of methods used to assess ARs in people with ABI, with no gold standard, and that the measurement properties of these assessment methods were largely unreported. Associated reactions are a complex phenomenon most notable when people with ABI are engaging in tasks that are effortful or challenging for them. As such, studies have tried to replicate this in their testing protocols by either performing a functional provocative test or a MVC of the intact or less affected UL. The majority of the studies in this review used the latter, incorporating stationary, seated testing positions whilst getting the patient to perform a MVC at various joints of their intact UL to elicit a contralateral AR. An advantage of this testing procedure is the potential elimination of confounding variables. Seated testing is also a safe, secure position for the patient. However, it is unlikely to reflect what occurs on a day-to-day basis for these people. Associated reactions appear to be a multi-factorial problem related to both physical and psychological factors. To date, the contributing factors are yet to be confirmed. It appears that some factors that exist only under dynamic conditions such as postural instability, fear of falling, anxiety and motor skill ability may all contribute. Therefore, a stationary, seated testing position may fail to capture or may misrepresent such contributing factors.

Many testing protocols utilized a 100% MVC of the intact UL as the provocative task to induce the AR. This is clearly not reflective of everyday life as there are few occasions where a person is required to elicit force at these maximal levels during functional activities. The studies that used protocols with lesser levels of intensity of intact UL contraction did not manage to consistently elicit an AR [4,10,37]. This highlights that ARs are effort dependent and gradually increase with increasing effort exerted. Additionally, the multi-factorial task demands from a physical, emotional and cognitive perspective are also important and may have a direct influence on the AR. In the context of ARs, ecological validity is a measure of how well the testing protocol reflects the everyday occurrence of ARs during real-life situations [40]. Bhakta et al. [39] described 92% of stroke patients reported noticing their AR on a daily basis and 67% reporting interference with activities of daily living. This emphasizes the need to assess ARs in an ecologically valid context—i.e. during functional tasks. Only a small proportion of the studies included in this review, irrespective of measurement method, used a functional mobility activity (i.e. walking or sit-to-stand) as the provocative task to elicit the AR. Ecologically valid testing during dynamic activities that most commonly provoke ARs may provide a greater opportunity to establish the true extent of ARs experienced by people with ABI. It may enable better insights into the multifactorial nature of ARs and the contributing factors, therefore enabling improved, targeted management of these.

Clinical utility is an essential property to consider for any assessment method if the aim is for it to be ultimately used in clinical practice. Half of the methods used in the literature to evaluate ARs were laboratory-based and not feasible for use in the clinical setting. Each of electromyography, dynamometry and load cells or SEMG did not fulfill the criteria to be able to be simply used in clinical practice. Additionally, with respect to ARs, most of these methods also utilized stationary, seated testing positions.

Surface electromyography was the most common method employed in the literature to evaluate ARs in people with ABI. This technique is frequently used to evaluate muscle innervation and overactivity and is a valid method of assessing neutral spasticity [41,42]. However, it is unclear whether ARs are the same or, in fact, a completely different phenomenon to spasticity. In relation to ARs there is limited evidence to support the use of SEMG to evaluate ARs, with no standardized protocol across the 11 SEMG studies. Reporting of results are also not comparable. Studies used different methods of SEMG normalization calculations. Some used complex ratios between affected and unaffected limbs [15] and others used methods such as standard net excitation to measure the irradiated muscles relative to the background activity of the muscle group [33]. This technique was most commonly tested in a seated position and not usually performed in conjunction with a functionally provocative test. Surface EMG is known to be used in three-dimensional gait analysis and may, therefore, have the potential for use in evaluating ARs during mobility tasks. Nonetheless, only two studies [14,38] used SEMG during functional tasks and so there is very limited information on which to evaluate SEMG and form a recommendation. The reliability of SEMG has not been reported.
and results can be variable depending on the inter-electrode distance, the placement site of electrodes or the individual muscle forces [33]. The addition of dynamic body movements may further confound the accuracy and reliability of the recordings obtained due to extraneous movement. There are also problems with respect to validity of the system. Dickstein et al. [38] used SEMG to evaluate muscle activity in biceps and brachioradialis during walking. They found inconsistent activity levels in these muscles between steps and between patients with a lack of correlation with the amount of elbow flexion during walking. Surface EMG also demonstrates poor clinical utility. It requires specialist skills to set up and perform with complex, lengthy testing protocols and is typically expensive, with many systems costing in excess of $20,000. The use of SEMG, therefore, poses many challenges and is not a clinically acceptable method of testing ARs.

Goniometry was another method that was commonly used to measure the change in elbow position during an AR of the hemiplegic UL. Standard goniometry is cheap and widely available [43]. Therapists are well versed in the use of goniometers and it is already a standard part of clinical practice. Goniometry may also be used to evaluate the AR in an ecologically valid context during walking and sit-to-stand. However, ARs can occur throughout the joints of the hemiplegic UL in various patterns and across different axes. Despite this, measurement was focused at the elbow joint, specifically flexion of the elbow, in most of the studies identified. This may be due to elbow flexion being commonly identified as the principal component of the basic flexor synergy [6] and because torques from elbow flexion tend to be the most prevalent and strongest [31]. However, it may also be because the elbow is an easily accessible joint situated between two long levers, making testing simpler. Goniometry cannot measure multiple joints of the UL concurrently, nor can it measure multi-planar movements. Testing procedures are often not standardized and there is questionable reliability with potentially high measurement error rates, particularly if used during functional movements [43]. The insensitivity of standard goniometry is another major issue where it would be unable to detect the presence of small degrees of AR.

The Associated Reaction Rating Scale [18], a clinician rating form, is the method of AR assessment that has had the most investigation into its clinimetric properties and may warrant further evaluation. It was the only method developed by expert neurological clinicians with the primary aim to quantify ARs objectively in a clinical setting during a functional, dynamic task of sit-to-stand. This tool is low-cost, readily available and easily used with minimal training required for clinicians. Preliminary investigations into interrater and test–re-test reliability show promising results. Most importantly, this measure demonstrates ecological validity in the context of measuring ARs. The Associated Reaction Rating Scale requires further investigation with larger cohorts to establish reliability, concurrent validity and responsiveness. Further statistical analysis is required to determine the most accurate methods of scoring, as this is yet to be established. The investigation into the use of this measure to assess other functional tasks such as walking is also warranted. There are, however, potential issues of validity with this subjective rating scale. Observational gait analysis for lower limb and gait abnormalities has been shown to have poor accuracy with a high degree of disagreements amongst clinicians of varying levels of experience [44]. Similarly, observation and rating of the extent of an AR in a person with ABI may also pose this problem. Therefore, clinicians may report on an AR that may not actually be present, may fail to identify an AR that is there, and may not be able to detect change in the AR over time. The inaccuracy of subjective or observational rating systems may negatively impact on clinical decision-making and therapeutic service delivery.

Although no gold standard for the assessment of ARs was identified, existing methods of movement analysis in the lower limb may be utilized for the UL. For example, motion analysis systems may provide an accurate dynamic method for evaluating ARs. These systems are extensively used in the lower limb for gait analysis, but to date, are infrequently used in the UL. Three-dimensional motion analysis is the current gold standard for objective evaluation of movement kinematics [45]. It has refined the process of assessment and management of lower limb and gait abnormalities in neurological cohorts and is sometimes considered a routine part of evaluation for surgical and complex pharmacological interventions for the lower limb in people with brain injury [44–46]. The costs associated with three-dimensional motion analysis are not inconsequential and there are obvious issues with respect to clinical utility. Nonetheless, perhaps the use of such criterion reference motion analysis systems could be used to quantify ARs in a dynamic and ecologically valid context with potential use as a gold standard comparator. At the very least thorough testing using a gold standard methodology is required in order to establish the contributing factors for ARs, before comparable clinical methods can be developed and refined.

Limitations

Terminology is problematic in the field of ARs. This was addressed by attempting to clarify terminology and thoroughly defining the inclusion and exclusion criteria and only including studies that were aligned with a specific working definition of ARs. This attempted to ensure that all relevant articles were captured. A comprehensive and systematic search strategy was used which included 10 different databases. The searches were restricted to English articles only and so relevant articles in other languages may have been missed. Additionally, to try and capture all measures of ARs, a mixed neurological cohort of acquired brain injury (stroke and traumatic brain injury) was included. Given that the focus of this review was to identify outcome measures used and not evaluate treatment effects, the inclusion of mixed cohorts is unlikely to have affected the yield and scope of this review. This study did, however, exclude pediatric or juvenile onset disorders, due to the issues of normal development and the potential of presence of non-pathological association movements. Therefore, two articles that investigated ARs in cerebral palsy were excluded [47,48]. These papers did not utilize distinct methods of evaluating ARs to those identified and so their exclusion did not adversely impact this review.
The COSMIN checklist is an instrument used to determine the quality of methodological design and statistical outcomes of clinimetric studies [21]. It is the best available instrument to evaluate clinimetric studies. Whilst it was primarily developed for evaluation of patient-reported measurement tools, others have found it to be just as relevant for use in synthesizing literature for performance-based tests [49]. However, due to the lack of sufficient clinimetric data on any of the assessment methods used to evaluate ARs, this study was unable to utilize the COSMIN checklist in its full capacity.

Conclusion
No gold standard assessment for ARs was identified. This systematic review was the first undertaken to identify the assessment methods used to evaluate ARs and evaluate their clinimetric properties. Most study protocols were not designed to reflect the multi-factorial nature of ARs, with the majority including stationary, seated assessments with poor ecological validity. Few of the methods have good clinical utility and very few assess the entire UL. There is little clinimetric data to support the use of any of the methods. Further research is required to develop accurate, objective and functionally relevant methods for assessing ARs. This may enable improved insight into the contributing factors to this debilitating phenomenon and lead to better remediation of ARs for people with ABI.

Supplemental material
The following supplemental material for this article can be accessed on the publisher’s website at http://www.tandfonline.com/doi/10.3109/02699052.2015.1117657.

Appendix 1:
There are problems with taxonomy and nomenclature in describing the phenomenon of associated reactions (ARs). A number of terms exist in the literature that overlap with or are utilised interchangeably when referring to ARs. The terms most frequently used are ‘motor overflow’, ‘associated movements’, ‘global synergies’ and ‘mirror movements’. Clarification of these terms is essential for the purposes of this review, for clinical application and for effective communication amongst professionals involved in brain injury rehabilitation.

Appendix 2:
A detailed example of the first search strategy for stage 1, applied to the Cumulative Index to Nursing and Allied Health Literature (CINAHL) including the medical subject heading terms is shown in Appendix 2.

Appendix 3:
The methods used to assess associated reactions that were identified in stage 1, were searched in combination with the terms for the clinimetric properties for stage 2. An example of this search is provided in Appendix 3.

Declaration of interest
The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References
1. Walshe FMR. On certain tonic or postural reflexes in hemiplegia, with special reference to the so-called “associated movements.”. Brain 1923;46:1–37.
2. Dickstein R, Heffles Y, Abulafio N. Electromyographic and positional changes in the elbows of spastic hemiparetic patients during walking. Electroencephalography and Clinical Neurophysiology 1996;101:491–496.
3. Dvir Z, Panturin E. Measurement of spasticity and associated reactions in stroke patients before and after physiotherapeutic intervention. Clinical Rehabilitation 1993;7:15–21.
4. Ada L, O’Dwyer N. Do associated reactions in the upper limb after stroke contribute to contracture formation? Clinical Rehabilitation 2001;15:186–194.
5. Mulley G. Associated reactions in the hemiplegic arm. Scandinavian Journal of Rehabilitation Medicine 1982;14:117–120.
6. Brunnstrom S. Associated reactions of the upper extremity in adult patients with hemiplegia; an approach to training. Physical Therapy Review 1985;36:225.
7. Davies PM. Steps to follow: the comprehensive treatment of patients with hemiplegia. Heidelberg: Springer; 2000.
8. Bhakta BB, O’Connor RJ, Cozens JA. Associated reactions after stroke: a randomized controlled trial of the effect of botulinum toxin type A. Journal of Rehabilitation Medicine 2008;40:36–41.
9. Bobath B. Adult hemiplegia: evaluation and treatment. London: Butterworth-Heinemann; 1990.
10. Dvir Z, Penturin E, Prop I. The effect of graded effort on the severity of associated reactions in hemiplegic patients. Clinical Rehabilitation 1996;10:155–158.
11. Lazarus JC. Associated movement in hemiplegia: the effects of force exerted, limb usage and inhibitory training. Archives of Physical Medicine & Rehabilitation 1992;73:1044.
12. Bakheit AMO, Sawyer J. The effects of botulinum toxin treatment on associated reactions of the upper limb on hemiplegic gait—a pilot study. Disability & Rehabilitation 2002;24:519–522.
13. Stephenson RC. A survey of the assessment and grading of spasticity by physiotherapists in the Yorkshire region. Synapse 1996;1:2–6.
14. Dickstein R, Pillar T, Abulafio N. Electromyographic activity of the biceps brachii muscles and elbow flexion during associated reactions in hemiparetic patients. American Journal of Physical Medicine & Rehabilitation 1995;74:427–431.
15. Honaga K, Masakado Y, Oki T, Hirabara Y, Fujiwara T, Ota T, Kimura A, Liu M. Associated reaction and spasticity among patients with stroke. American Journal of Physical Medicine & Rehabilitation 2007;86:656–661.
16. Ada L, Q’Dwyer N. Do associated reactions in the upper limb after stroke contribute to contracture formation? Clinical Rehabilitation 2001;15:186–194.
17. Bhakta BB, Cozens JA, Chamberlain MA, Bamford JM. Quantifying associated reactions in the paretic arm in stroke and their relationship to spasticity. Clinical Rehabilitation 2001;15:195–206.
18. Macfarlane A, Turner-Stokes L, De Souza L. The Associated Reaction Rating Scale Clinical Rehabilitation: a clinical tool to measure associated reactions in the hemiplegic upper limb. Clinical Rehabilitation 2002;16:726–735.
19. Johnston MV, Findley TW, DeLuca J, Katz RT. Research in physical medicine and rehabilitation: XII. Measurement tools with application to brain injury. American Journal of Physical Medicine & Rehabilitation 1991;70:S114–S130.
20. Moseley AM, Lanzarone S, Bosman JM, van Loo MA, de Bie RA, Hassett L, Caplan B. Ecological validity of walking speed assessment after traumatic brain injury: a pilot study. Journal of Head Trauma Rehabilitation 2004;19:341–348.
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related...
patient-reported outcomes. Journal of Clinical Epidemiology 2010;63:737–745.

22. Smart A. A multi-dimensional model of clinical utility. International Journal for Quality in Health Care 2006;18:377–382.

23. Mober D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of Internal Medicine 2009;151:264–269.

24. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Götsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Annals of Internal Medicine 2009;151:W-65–W-94.

25. Law MC, MacDermid J. Evidence-based rehabilitation: a guide to practice. Thorofare, NJ: Slack Incorporated; 2008.

26. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions. The Cochrane Library 2005.

27. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 2007;60:34–42.

28. Hill BE, Williams G, Bialocerkowski AE. Clinimetric evaluation of questionnaires used to assess activity after traumatic brachial plexus injury in adults: a systematic review. Archives of Physical Medicine & Rehabilitation 2011;92:2082–2089.

29. Tyack Z, Waski J, Spinks A, Kimble R, Simons M. A guide to choosing a burn scar rating scale for clinical or research use. Burns 2013;39:1341–1350.

30. Tyson SF, Connell LA. How to measure balance in clinical practice. A systematic review of the psychometrics and clinical utility of measures of balance activity for neurological conditions. Clinical Rehabilitation 2009;23, 824–840.

31. Boissy P, Bourbonnais D, Kaegi C, Gravel D, Arsenault BA. Characterization of global synkineses during hand grip in hemiparetic patients. Archives of Physical Medicine & Rehabilitation 1997;78:1117–1124.

32. Boissy P, Bourbonnais D, Gravel D, Arsenault AB, LePage Y. Effects of upper and lower limb static exertions on global synkineses in hemiparetic subjects. Clinical Rehabilitation 2000;14:393–401.

33. Hwang I, Tung L, Yang J, Chen Y, Yeh C, Wang C. Electromyographic analyses of global synkinesis in the paretic upper limb after stroke. Physical Therapy 2005;85:755–765.

34. Macfarlane A, Turner-Stokes L, De Souza L. The associated reaction rating scale: a clinical tool to measure associated reactions in the hemiplegic upper limb. Clinical Rehabilitation 2002;16:726–735.

35. Tung LC, Yang JF, Wang CH, Hwang IS. Directional effect on post-stroke motor overflow characteristics. Chinese Journal of Physiology 2011;54:391–398.

36. Mills VM, Quintana L. Electromyography results of exercise overflow in hemiplegic patients. Physical Therapy 1985;65:1041–1045.

37. Lazarus JAC. Associated movement in hemiplegia: the effects of force exerted, limb usage and inhibitory training. Archives of Physical Medicine & Rehabilitation 1992;73:1044–1049.

38. Dickstein R, Heffes Y, Abulaffio N. Electromyographic and positional changes in the elbows of spastic hemiparetic patients during walking. Electroencephalography and Clinical Neurophysiology - Electromyography and Motor Control 1996;101:491–496.

39. Bhakta BB, O’Connor RJ, Cozens JA. Associated reactions after stroke: a randomized controlled trial of the effect of botulinum toxin type A. Journal of Rehabilitation Medicine 2008;40:36–41.

40. Platz T, Eickhof C, Nuyens G, Vuadens P. Clinical scales for the assessment of spasticity, associated phenomena, and function: a systematic review of the literature. Disability & Rehabilitation 2005;27:7–18.

41. Sorinola IO, White CM, Rushton DN, Newham DJ. Electromyographic response to manual passive stretch of the hemiplegic wrist: accuracy, reliability, and correlation with clinical spasticity assessment and function. Neurorehabilitation & Neural Repair 2008;23:287.

42. Starsky AJ, Sangani SG, McGuire JR, Logan B, Schmit BD. Reliability of biomechanical spasticity measurements at the elbow of people poststroke. Archives of Physical Medicine & Rehabilitation 2005;86:1648–1654.

43. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion review of goniometry emphasizing reliability and validity. Physical Therapy 1987;67:1867–1872.

44. Williams G, Morris ME, Schache A, McCrory P. Observational gait analysis in traumatic brain injury: accuracy of clinical judgment. Gait & Posture 2009;29:454–459.

45. Simon SR. Quantification of human motion: gait analysis–benefits and limitations to its application to clinical problems. Journal of Biomechanics 2004;37:1869–1880.

46. Fuller D, Keenan MAE, Esquenazi A, Whyte J, Mayer N, Fidler Sheppard R. The impact of instrumented gait analysis on surgical planning: treatment of spastic equinovarus deformity of the foot and ankle. Foot & Ankle International 2002;23:738–743.

47. Chiu HC, Halaki M, O’Dwyer N. Associated reactions during a visual pursuit position tracking task in hemiplegic and quadriplegic cerebral palsy. Chinese Journal of Physiology 2013;56:117–126.

48. Chiu H-C, Ada L, Butler J, Coulson S. Characteristics of associated reactions in people with hemiplegic cerebral palsy. Physiotherapy Research International 2011;16:125–132.

49. van Bloemendaal M, van de Water ATM, van de Port IGL. Walking tests for stroke survivors: a systematic review of their measurement properties. Disability & Rehabilitation 2012;34:2207–2221.