Table 1. AMA Guides- Accepted articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Validity	Reliability	Predictive Value
Patel B et al, 2003[75]	Cross-sectional study of physiatrists' ratings of permanent partial impairment across the USA	52 US physiatrists (out of 200) responded to a survey with 3 clinical vignettes for which they were asked to provide a diagnosis, an impairment rating and how the impairment rating was obtained	3 clinical vignettes (2 of which were obtained from the AAPM&R Disability Evaluation Program handbook) with the first being a case of radiculopathy, the second a case of radial neuropathy and the third a case for total hip replacement	Such a wide range of ratings suggests a lack of validity	The range of impairment values were 0-14% for radiculopathy, 3-60% for radial neuropathy and 8-100% for total hip replacement; responses did not correlate with region, years of experience or type of practice except in the first case where there was a trend of younger respondents giving lower ratings; respondents used different sources for obtaining their results with 5th than 4th edition of the AMA guides being the most common	Predictive value would be affected by lack of validity and reliability of developing ratings
van Oosterom FJT et al, 2007[95]	Validation study comparing the functional module of the DASH to the AMA Guides ratings in patients with phalangeal fractures	78 patients with 228 fractures were enrolled from the Academic Hospital Rotterdam from July 1987 to July 1997 with an average of 3 fractures per patient	Patients with multiple phalangeal fractures within the same hand treated surgically in the Academic Hospital Rotterdam (now Erasmus Medical Centre) between July 1987 and July 1997 with a minimum follow-up period of 2 years	Pearson correlation between AMA Impairment Ratings and DASH function module was low even after controlling for comorbidities and the follow-up time; No correlation for the hand DASH score (r=0.09;p=0.42); Weak partial correlation coefficients for the hand (r=0.35;p=0.002) and total body (r=0.34;p=0.003) DASH function module scores and the work DASH score (r=0.24;p=0.045)	n/a	n/a

Table 2. AMA GUIDELINES - 4TH Ed. – Excluded Articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Validity	Reliability	Predictive Value
Gloss DC et al, 1982[37]	Cross-sectional validity and inter-rater reliability study	118 patients (96 male; 22 female) treated surgically by a hand surgeon and/or plastic surgeon for an injury or disease condition affecting one hand	Patients 9-80 years of age treated surgically by a hand surgeon and/or plastic surgeon for an injury or disease condition affecting one hand	Product moment correlation between hand impairment and finger dexterity test (r=-0.528; p<0.01), hand dynamometer test (r=-0.198; p>0.01), tapping test (r=-0.331; p<0.01), placing test (r=0.487; p<0.01), displacing test (r=0.534; p<0.01), one-hand test (r=0.582; p<0.01), two-hand test (r=0.598; p<0.01)	Inter-rater correlation (Cronbach's alpha?)=0.75	n/a
Nattrass CL et al, 1999[71]	Cross-sectional validation study	34 subjects (21 women; 13 men) from an outpatient department in a Rehabilitation Medicine Unit in Melbourne, Australia	Participants between 20 and 65 years of age with chronic low back with or without leg pain of at least 6 months; participants excluded if had cervical or thoracic involvement or any acute muscle spasm	n/a	Range of Motion measured using dual inclinometer: Pearson's r=0.12-0.38 with the Wadell Disability Index; Pearson's r=0.22-0.38 with Oswestry Disability Index; Pearson's r=0.35-0.54 with Waddell Physical Impairment Scale	n/a

Table 2. AMA GUIDELINES - 4TH Ed. – Excluded Articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Validity	Reliability	Predictive Value
Nitschke JE et al, 1999[72]	Repeated measures design for intra and inter-rater reliability study over 1 week	34 subjects (21 women; 13 men) from an outpatient department in a Rehabilitation Medicine Unit in Melbourne, Australia	Participants between 20 and 65 years of age with chronic low back or leg pain of at least 6 months; participants excluded if had cervical or thoracic involvement or any acute muscle spasm	n/a	Inter-rater ICC=0.13-0.52 for dual inclinometer ROM measure (4th ed AMA Guides) with best values for flexion followed by extension and lateral flexion; Intra-rater ICC=0.70-0.90 with best values for flexion and right lateral flexion, worst for extension	n/a
Rondinelli RD, 2009[82]	Narrative review; suggests to validate individual scales used in AMA	n/a	n/a	no quantitative results	no quantitative results	no quantitative results

Table 2. AMA GUIDELINES - 4TH Ed. – Excluded Articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Validity	Reliability	Predictive Value
Ryu SG et al, 2009[83]	Qualitative comparison of the Korean Academy of Medical Sciences Guideline for rating the impairment in Mental and Behavioural Disorders with the 6th edition AMA Guides	none; qualitative analysis of differences between the two guidelines	none; qualitative analysis of differences between the two guidelines	n/a	n/a	n/a

12.2 Glasgow Outcome Scale

Purpose of the Literature Review

The purpose of the review is to determine the reliability and validity (face, construct and predictive) of the Glasgow Outcome Scale (GOS). The results of the review will be submitted to the Catastrophic Impairment Expert Panel to inform their deliberation of the Glasgow Outcome Scale.

Methodology

We conducted a review of the recent literature. The title and abstract of the articles were screened for relevance by Drs. Craig Jacobs and Pierre Côté. The relevant articles were critically appraised by Craig Jacobs using the SIGN instrument for cohort studies and a modified version of the QUADAS instrument for validity and reliability studies .[2;98]

Search Strategy

We searched Pubmed from January 2000 to December 2010. The search strategy included the following key terms: MeSH term 'glasgow outcome scale' combined using an 'AND' connector with '(valid* OR reliab* OR predict* OR prognos*)'. Limits activated included 'Humans', 'English', and 'Published in the last 10 years'. The reference lists of the selected articles were reviewed for other potentially relevant articles.

Pubmed Search Strategy

Glasgow outcome scale [MESH] AND (valid* OR reliab* OR predict* OR prognos*).

Results

Search Results

The Pubmed search yielded 337 results. Three hundred twenty five (325) articles were not relevant and excluded from the review. Dr. J. David Cassidy, one of the Catastrophic Impairment Expert Panel members provided an additional six articles. One additional article was selected for review from the reference lists of the previously identified articles and one newly published article that was emailed from Pubmed was included. Therefore, a total of 20 articles were critically appraised for methodological quality.

Following the critical appraisal, nine articles were deemed scientifically admissible[46;54;59;68;78;79;90;99;100](Table 1) and eleven were excluded. [6;13;21;26;57;60;77;87;88;92;101] The articles were excluded for the following reasons: narrative reviews[88;92]; no blinding of assessor[57;77]; significant loss to follow-up[57;87]; inadequate reporting of results[13] (we contacted the corresponding author of the article for clarification-no response was obtained); not relevant to the review[60]; selection bias and no validation of translated questionnaires[21]; small cell size in multivariable regression leading to poor precision of estimates[6]; and incorrect scoring of GOS.[26]

Reliability of the Glasgow Outcome Scale

Four articles addressed the reliability of the Glasgow Outcome Scale. One study addressed inter-rater reliability of an alternative method to determine GOSE scores (adding a quality control system).[59] One study addressed the reliability in-person and telephone structured interviews for the GOS and GOSE.[78] One study addressed the reliability of postal questionnaires for the GOS.[99] Finally, one study addressed the inter-rater reliability of the GOS and GOSE using structured interviews.[100]

Lu et al. (2010) studied the inter-rater reliability study of the 6-month GOSE outcome using a new GOSE rating system in a sample of six cases that covered the range of GOSE outcomes.[59] The new method used three strategies to improve the ratings: (1) a quality-control system with a central reviewer, (2) an algorithm to compute the GOS score, and (3) use questions to distinguish between the upper and lower categories of a specific GOS score to arrive at the GOSE score. The reliability of the new method was compared to a variant of the new method (no quality control) and to the conventional method of computing the GOSE. The inter-rater reliability was k_w=0.97 (95% CI 0.91, 1.00) for the new method; k_w=0.79 (95% CI 0.69, 0.89) for the new method without quality control; and k_w=0.70; 95% CI 0.60, 0.81 for the conventional method (Table 1) The authors conclude that the new method reduces inter-rater variations.[59] The improvement in the reliability can be attributed the use of a quality control system.

Pettigrew et al. (2003) conducted a test-retest and inter-rater reliability study of structured in-person and telephone interviews for the GOS and GOSE.[78] For the test-retest reliability study, 30 head-injured participants were first interviewed in person and then re-interviewed a few days later by telephone by the same rater. For the inter-rater reliability, 56 head-injured participants were interviewed by telephone and then in person up to 1 month later by a different person who was blinded to the previous interview. The test-retest reliability statistics were k_w=0.92 (95% CI .57–1.00) for both GOS and GOSE. Weighted kappa for the interrater reliability study were k_w=0.85 (95% CI .59–1.00) for the GOS and k_w=0.84 (95% CI .58–1.00) for the GOSE (Table 1). The authors conclude that structured interviews for the GOS and GOSE are reliable even when using two different methods of contact (in-person and telephone). They also conclude that that inter-rater reliability is high using structured interviews through two different methods of data collection.[78]

Wilson et al., (2002) studied the test-retest reliability of a postal questionnaire version of the GOS and the GOSE. [99] The study included participants with a diagnosis of head-injury for whom a 6-month GOS score was available. Participants were randomly allocated to either be part of a test-retest study of the postal questionnaire (two week interval), or to answer questions through a telephone interview. The questionnaire was completed by the person with a head injury or by a proxy. Patients received either the GOS or GOSE. Of the 174 participants who completed the initial questionnaire 141 (81%) responded to the second one. The mean interval between completion of the first and second questionnaires for the mailed test-retest study was 14.7 days for the GOS and 14.4 days for the GOSE. In the telephone study, the mean interval between the telephone interview and completion of the postal questionnaire was 5.7 days for the GOS and 6.4 days for the GOSE. The authors do not state whether the telephone interview and scoring of the postal questionnaire were performed by different individuals. The test-retest reliability was k_w=0.94 (95% CI 0.60-1.00) for the GOS questionnaire and k_w=0.98 (95% CI

0.66-1.00) for the GOSE. For the comparison with the telephone interview, k_w= 0.67 (95% CI = 0.35–1.00) for the GOS and k_w= 0.92 (95% CI 0.59–1.00) for the GOSE (Table 1). The authors conclude that there is good test-retest reliability as well as good agreement between postal questionnaires and telephone interviews.[99] However, the reliability of the GOSE is more stable across methods of administration.

Wilson et al., (1998) conducted an inter-rater reliability study of the GOS and GOSE using structured interviews.[100] Fifty head-injured participants were recruited from a regional neurosurgical unit and were interviewed between 5-17 months post-injury by a research psychologist and then by one of two research nurses. Interviews occurred in person on the same day using a structured GOSE questionnaire. The weighted kappa was k_w= 0.89 for the GOS and k_w = 0.85 for the GOSE. The authors concluded that the assessment of the GOS and GOSE using standard interviews is practical and reliable. (The structured interview with explanations as well as accompanying notes and definition of terms are provided as appendices.)[100]

In summary, the evidence suggests that both the GOS and the GOSE are reliable measures. One study strongly supports the use of quality control to improve reliability and reduce the inter-rater variability in the GOSE scoring.

Face/Construct validity of the Glasgow Outcome Scale

Three cohort studies provide information on the construct validity of the Glasgow Outcome Scale. One study investigated the association between clinical predictors and poor outcome (measured with the GOS).[54;90] The second study correlated the GOS/GOSE with the Functional Status Evaluation (FSE).[46] Finally, we included a study of the predictors of recovery (measured with a modified GOS) in a cohort of young children.[79]

Thornhill et al., (2000) conducted a large prospective hospital-based cohort study of head-injured patients (aged 14 years or older) in Glasgow.[90] Participants (n=2962) were enrolled at five hospitals over a one year period and were followed for one year. Baseline data was collected through chart reviews. Follow-up status was measured through telephone interviews or postal questionnaires one year after the injury. Severity of injury was determined by GCS score at baseline (13-15 = mild, 9-12 = moderate, 3-8 = severe). The outcome was determined with the GOS. All severe and moderately injured participants were followed for one year and a random sample of the mild-injured group was followed for one year. Overall follow-up rate was 71%, ranging between 71-73% for the three groups. The incidence of disability from head-injury is reported at one year. Independent predictors of poor outcome (death or disability) were reported using multivariate logistic regression for the mild-injury group. The incidence of head-injured patients with severe or moderate disability at one year was 154 per 100,000 population (95% CI: 138-169). Disability (severe or moderate GOS scores) at one year was present in 51% of mild head injury participants, 54% of moderate head injury participants, and 78% of severe head injury participants. Predictors of death or disability (moderate disability on the GOS or worse) at one-year in the mild injured group are: >40 years of age (OR=1.80, CI 1.11 to 2.91), pre-existing physical limitations (OR= 2.24, CI 1.33 – 3.86), and a history of brain illness (OR=2.07, CI 1.33 – 3.21).[90] This study supports the validity of the GOS because it is associated with those with a worse risk profile at baseline. It also suggests that the status of patients changes over a one year period.

Hudak et al., (2005) conducted a cohort study of head injury patients to measure the correlation between the GOSE and the Functional Status Evaluation (FSE). Participants (n=177) with head injuries had both the FSE and the GOSE administered at six and twelve months post-injury. The FSE measures outcome in 10 domains including personal care, ambulation, major activity involving work or school, home management, leisure and recreation, travel, social integration, standard of living, financial independence and executive functioning. The FSE was administered through a structured interview whereas the GOSE was administered by self-administered questionnaire. Linear regression was used to determine the correlation between FSE and GOSE scores. The authors rep ort a strong correlation between the two outcome measures (r=-0.83, R^2=0.70). GOSE was moderately correlated with ICU stay (r=-0.44) and hospital stay (r=-0.44).[46] This study supports the contract validity of the GOSE as a measure of functional status.

Prasad et al. (2002) studied the predictors of recovery from traumatic brain injury in a cohort of 60 children less than six years of age who sustained either inflicted or noninflicted traumatic brain injury. The outcome was measured three and 12 months post-injury with a modified version of the GOS. *"Good outcome referred to the return to age-appropriate or preinjury levels of functioning and the return to full-time classes with no special education services. Moderate disability was assigned based on: (1) a significant reduction in cognitive functioning from estimated premorbid levels; (2) motor deficits including hemiparesis interfering with activities of daily living; (3) referral to outpatient rehabilitation therapies, and (4) attending special education or resource classes. Severe disability was assigned if (1) cognitive functioning was in the deficient range, (2) severe motor deficits were present, such as lack of age appropriate postural control or ambulation, and/or (3) there was referral for inpatient rehabilitation. The criteria for persistent vegetative state were unchanged and reflected total dependence for daily care. For the purposes of this study, moderate disability, severe disability and persistent vegetative state were classified as 'poor outcome'."*[79]
Regression analysis suggests that the modified Glasgow Coma Scale score, the duration of impaired consciousness and the number of intracranial lesions visualized on CT/MRI accounted for a significant amount of the variance in the Glasgow Outcome Scale (GOS) 3- and 12-month evaluations. Inflicted brain injury adversely affected the GOS. Age at injury and the Injury Severity Score were not associated with GOS scores at follow-up. It is important to note that this was a small study and the final selection of variables in the model was likely influenced by the power of the study.[79]

In summary, the evidence supports that the GOS and GOSE have good face validly. Although we only retrieved three studies, the reviewed literature suggests that the GOS and GOSE have adequate construct validity.

<u>Predictive validity of the Glasgow Outcome Scale</u>
Two studies addressed the predictive validity of the Glasgow Outcome Scale.

King et al., (2005) conducted a study to measure the association between demographic and clinical variables (including the 3-month GOS) and the GOS measured at 12-months post-injury.[54] A total of 159 participants with severe closed traumatic head injury (GCS ≤8) were enrolled at the University of Pittsburgh Medical Center. No participants were lost to follow-up. Three participants died. The GOS was dichotomized into good outcome (GOS 4-5) or poor

outcome (GOS 1-3). Multivariate logistic regression was used to assess the relationship of the variables with a poor GOS at 12-months. The GOS score at 3-months was the best predictor of a poor outcome (OR = 15.22, $p<0.001$). Presence of prolonged hypotension, diffuse axonal injury and fixed/dilated pupils on admission also predicted the outcome (Table 1). Probabilities of poor outcome were then generated for each GOS score (adjusting for the other significant predictors of poor outcome in the model). A GOS score of 2 at 3-months had an adjusted risk of 89.4% for a poor outcome at 12-months. A GOS score of 5 at 3-months had an adjusted risk of 0.11% for a poor outcome at 12-months. The authors conclude that the 3-month GOS score is a powerful predictor of long-term outcome in patients with severe traumatic brain injury.[54]

Miller et al., (2005) studied the association between the GOS score measured within three months and patient's outcome 15 months post-injury.[10] One hundred and twenty-one participants with traumatic brain injury were enrolled at 7 military or Department of Veterans Affairs hospitals in the US. Injury severity was classified by length of loss of consciousness; those who lost consciousness for more than 24 hours were deemed severely injured. The majority of the study participants were young, male, employed, active military personnel. 88% of patients with a GOS of 5 at baseline retained a score of 5 at 15 months post-injury, with the remainder having a GOS of 4 at 15 months. Two thirds (66.7%) of those participants with a GOS score of 4 at baseline improved to a score 5 at 15 months with the remainder still at a GOS 4 at 15 months. Forty percent of those with a GOS score of 3 at baseline improved to a score of 5 at 15 months, 50 percent improved to a score of 4 at 15 months, and 10 percent remained at a score of 3.[68]

The authors report that the injury severity at baseline did not seem to affect the likelihood of improved GOS score at 15 months since 59% of severely injured participants with a baseline GOS score of 3 or 4 increased to a score of 5 at 15 months. Also, 54.5% of non-severely injured participants with a baseline GOS score of 3 or 4 increased their score to a 5 at 15 months.[68]

In summary, the reviewed evidence suggest that the GOS score measured three months after an injury is a strong predictor of the GOS score measured one year after the injury. The evidence also suggest that the status of head injury patients changes significantly over time and that an important proportion of patients improve within the first 12 to 15 months post-injury.

Table 1. Glasgow Outcome Scale- Accepted articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Reliability	Face/Construct Validity	Predictive Validity
Hudak et al, 2005[46]	Prospective study. Comparison between FSE and GOS-E (both used as outcomes).	177 participants with TBI with FSE and GOS-E done at 6-12 months post-injury.	Brain injury requiring hospitalization confirmed either by presence of abnormal neuroimaging or altered mental status greater than 30 min.	n/a	Linear regression for correlation between FSE and GOS-E scores: r= -0.83, r2=0.70. Spearman correlation for correlation between variables and outcome (no multivariate analysis done). Report moderate correlation between length of ICU stay (FSE:r=0.36,p,0.001;GOSE:r=-0.44,p=0.003) and hospital stay (FSE:r=0.50,p,0.001; GOSE: r=-0.44,p,0.003). Report smaller relationships between GCS and outcomes (FSE:r=-0.18, p=0.025; GOSE:0.18,p=0.014). No difference for patients who went through craniotomy or not. No correlation between CT scan findings and GOSE. No correlation between Abbreviated Injury Scale and Injury Severity Scale and either GCS, FSE, or GOS-E.	n/a
King et al, 2005[54]	Prospective cohort study examining association between GOS scores at 3 months with GOS at 12 months.	159 patients with severe, closed traumatic brain injury in level 1 trauma center.	Severe traumatic brain injury = GCS ≤8. Open or penetrating wounds excluded. Patients dying prior to 3 month analysis excluded from analysis.	n/a	n/a	Association of predictor variables and poor outcome at 12 months: GOS at 3 months, OR = 15.2 (95% CI 5.3 - 45.6); Prolonged hypotension, OR = 3.7 (CI 1.0 - 13.7); Diffuse axonal injury, OR = 5.5 CI 2.0 - 15.6); Fixed and dilated pupils on admission, OR = 12.1 (1.2 - 118.2). Probability of poor outcome at 12

Table 1. Glasgow Outcome Scale- Accepted articles						
Author, Publication Year	**Study Design**	**Setting & Subjects, Number (n) Enrolled**	**Case Definition**	**Reliability**	**Face/Construct Validity**	**Predictive Validity**
						months after TBI is: 89.4% (CI = 85.7 – 93.2%) for GOS 2 at 3 months; 32.1% (CI = 27.0 – 37.3) for GOS 3 at 3 months; 5.0% (CI = 0.0 – 10.7%) for GOS 4 at 3 months; and 0.11% (95% CI: 0.0 – 0.14%) for GOS 5 at 3 months.
Lu et al, 2010[59]	Inter-rater reliability study comparing a new alternate method of assessing GOSE and GOS scores with conventional structured interviews. Inter-rater reliability determined between an expert and untrained raters for each group.	32 raters from different trauma centers.	6 transcripts of structured outcome interview with patients used. Transcripts contained real patient data originating from previous studies. Covered the range of GOSE outcomes from lower severe disability to lower good recovery (as assigned by an expert).	GOSE: New alternative system: Overall agreement 97%, κw (weighted kappa) 0.97, 95% CI 0.91 - 1.00; Alternative system w/o central monitoring: Overall agreement 76%, κw=0.79, CI 0.69 - 0.89; Conventional structured interview: Overall agreement 63%, κw = 0.70, CI 0.60 - 0.81. GOS: Alternative method: Overall agreement 97%, κw= 0.95, CI 0.89 - 1.00; Alternative method w/o central monitoring: Overall agreement 83%, κw=0.81, CI 0.69 - 0.92, CI 0.69 - 0.92; Conventional structured interview: Overall agreement 83%, κw=0.76, CI 0.63 - 0.89.	n/a	n/a

Table 1. Glasgow Outcome Scale- Accepted articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Reliability	Face/Construct Validity	Predictive Validity
Miller et al, 2005[68]	Prospective cohort. GOS measured w/in 3 months post-injury and at 15 months.	121 participants hospitalized for TBI in 7 military and Department of VA medical centers in US.	TBI with neuro evaluation w/in 3 months of injury, adult survivors of closed head injury with GOS score recorded during evaluation, LOC info available to classify injury by severity, and return for subsequent post-injury evaluations when GOS assessed.	n/a	n/a	88% of patients with GOS at 5 at baseline retained score of 5 at 15 months postinjury, with the remainder having a GOS of 4 at 15 months. 66.7% of those participants with a GOS score of 4 at baseline improved to a score 5 at 15 months with the remainder still at a GOS 4 at 15 months. 40% of those with a GOS score of 3 at baseline improved to a score of 5 at 15 months, 50 percent improved to a score of 4 at 15 months, and 10 percent remained at a score of 3. Injury severity did not seem to affect likelihood of improving GOS score at 15 months. 59% of severely injured participants with a baseline score of 3 or 4 increased to a score of 5 at 15 months; 54.5% of non-severely injured participants with a baseline score of 3 or 4 increased their score to a 5 at 15 months.

Table 1. Glasgow Outcome Scale- Accepted articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Reliability	Face/Construct Validity	Predictive Validity
Pettigrew et al, 2003[78]	Study 1: Test-retest (same rater, first in person, then phone). Study 2: Inter-rater reliability (different rater, first by telephone, then in person).	Study 1: 30 head injury participants. Study 2: 56 head injury participants.	Head injury not defined. Cases obtained from a study that examined apolipoprotein E genotype and recovery after head injury. GCS scores reported.	Study 1: GOSE - overall agreement 77%. κ (kappa) = 0.72 (CI 0.55 - 0.88). κw (weighted kappa) = 0.92 (CI 0.56-1.00). GOS - overall agreement = 90%. κ = 0.85 (CI 0.59 - 1.00); κw = 0.92 (CI 0.57 - 1.00). Study 2: GOSE - overall agreement = 71%. κ = 0.64 (CI 0.51 - 0.77). κw = 0.84 (CI 0.58 - 1.00). GOS - overall agreement = 86%. κ = 0.78 (CI 0.59 - 0.97); κw = 0.85 (CI 0.59 - 1.00).	n/a	n/a
Prasad et al, 2001[79]	Prospective cohort. GOS measured at baseline and 12 months.	Children aged 1 to 6 years hospitalized for inflicted (n=31) or non-inflicted (n=29) TBI in two children's hospitals in Texas, USA.	Moderate TBI (GCS 9-12 or 13-15 with +ve neuroimaging on admission) and severe TBI (GCS 3-8), gestational age ≥32 weeks and age at injury < 6 years. GOS was modified for children. Severity Factor Score (SFS): Duration of impaired consciousness, lowest GCS score, and number of intracranial lesions on CT/MR imaging.	n/a	Severity factor score (SFS) predicted level of recovery (e.g. a severity factor score 1 standard deviation (SD) above average (indicating high severity) predicts a 0.045 chance of good recovery. An SFS 1 SD below average (indicating low severity) predicted a 0.726 likelihood of good recovery.)	n/a

Table 1. Glasgow Outcome Scale- Accepted articles						
Author, Publication Year	**Study Design**	**Setting & Subjects, Number (n) Enrolled**	**Case Definition**	**Reliability**	**Face/Construct Validity**	**Predictive Validity**
Thornhill et al, 2000[90]	Prospective cohort study. F/u with GOS at 1 year.	Five acute hospitals in Glasgow. 2962 participants with head injuries enrolled. Excluded those who lived outside Glasgow. Aimed to follow up all patients with severe or moderate head injuries and a random sample of patients with mild or unclassified head injury. F/u rate was 71% overall.	Patients admitted with head injury >14 yoa.	n/a	Factors independently associated with death or disability (GOS 1-4) in mild head-injured participants: >40 yoa (OR, 1.80, CI 1.11 - 2.91), pre-existing physical limitations (OR, 2.24, CI 1.33 – 3.86), history of brain illness (OR 2.07, CI 1.33 – 3.21).	n/a
Wilson et al, 2002[99]	Reliability study of GOS and GOSE postal questionnaire. Test/retest for postal questionnaire and comparison between postal questionnaire and structured telephone interviews.	174 participants recruited from a head injury group from a previous study. 16 yoa or older, not currently in hospital or nursing home.	Previous admission to neurosurgical unit with dx of head injury.	Test/retest for postal questionnaire: 1. GOS κw (weighted kappa) = 0.94 (95%CI:0.60 - 1.00); 2. GOSE κw =0.98 (CI:0.66 - 1.00). Telephone interview vs. postal questionnaire: 1.GOS κw = 0.67 (CI: 0.35 - 1.00); 2. GOSE κw =0.92 (CI: 0.59 - 1.00).	n/a	n/a

Table 1. Glasgow Outcome Scale- Accepted articles

Author, Publication Year	Study Design	Setting & Subjects, Number (n) Enrolled	Case Definition	Reliability	Face/Construct Validity	Predictive Validity
Wilson et al, 1998[100]	Inter-rater reliability study of GOS and GOSE with structured interview guidelines.	50 patients (42M; 8F) recruited from regional neurosurgical unit.	No explicit case definition. Seems that inclusion was head injury with GCS<16. Participants recruited from "head injury admissions" at regional neurosurgical unit. 30% had GCS score of 3-8 (severe), 14% had GCS of 9-12(moderate), and 56% had GCS of 13-15 (mild). Restricted to conscious survivors.	Inter-rater reliability: Overall agreement 92% for GOS and 78% for GOSE. κw (weighted kappa) for GOS = 0.89 and GOSE = 0.85.	n/a	n/a

12.3 Glasgow Coma Scale – Literature Review

1. Purpose of the Literature Review

The purpose of the review is to determine the reliability and validity (face, construct and predictive) of the Glasgow Coma Scale (GCS). The results of the review will be submitted to the Catastrophic Impairment Expert Panel to inform their deliberation of the Glasgow Coma Scale.

2. Methodology

We conducted a review of the recent literature. The title and abstract of the articles was screened for relevance by Dr. Pierre Côté. The relevant articles were critically appraised by Drs. Heather Shearer, Craig Jacobs and Maja Stupar using a modified version of the QUADAS instrument for reliability and validity studies and the SIGN tool for predictive validity/cohort studies.[2;98]

2.1 Search Strategy

A Medline (Pubmed) search was performed from January 2000 to December 2010. The search strategy included key terms (Glasgow coma scale [MESH]) AND (valid* OR reliab* OR predict* OR prognos*). References of selected articles were reviewed for other potentially relevant articles. Finally, the 'Find Similar' option in Pubmed was used to select further relevant articles using the most recent relevant result. The option to have any new articles with the term 'Glasgow Coma Scale' emailed to the reviewer was selected.

3. Results

3.1 Search Results

The Pubmed search yielded 1031 results. Because of the large number of articles, it was decided that only articles dated from January, 2007 to December 2010 would be reviewed. Again, references of selected articles were reviewed for other potentially relevant articles and five additional articles were selected. This limitation reduced the number of articles to 462. Following a title and abstract review for relevance to the topic of validity, reliability or predictive value of the Glasgow Coma Scale, 286 articles were excluded. An additional 14 articles dated prior to 2007 were selected for review based on references lists of selected articles. Of these remaining 190 articles, another 143 articles were excluded based on an abstract review for their relevance to the topic. Forty-seven complete articles were reviewed for quality using the QUADAS tool and to obtain quantitative results relevant to the topic. Twenty-one articles were included and 26 excluded from the final selection of forty-seven articles.

The reasons for the exclusion of 26 articles are provided below:

- Five studies did not address reliability, validity or predictive value of the GCS. [8;10;19;53;73]
- Two review articles were narrative rather than systematic reviews.[55;103]
- One article had a very low participation rate.[97]
- One article provided no inclusion/exclusion criteria, used inappropriate statistical methodology, and did not blind the assessors.[89]
- One article was excluded due to inappropriate analysis.[86]
- One article had small cell sizes in multivariable regression which lead to imprecision of estimates.[34]

applicable to those individuals who are functionally able to complete the examination, meaning that this study is not applicable to those with lower GCS scores.[38]