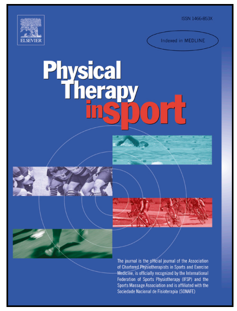
# Accepted Manuscript




The reliability of shoulder range of motion measures in competitive swimmers

Dr Helen Walker, Tania Pizzari, Henry Wajswelner, Peter Blanch, Clinical Services Manager, Laura Schwab, Kim Bennell, Professor and NHMRC Principal Research Fellow, Belinda Gabbe

PII: S1466-853X(16)00031-6

DOI: 10.1016/j.ptsp.2016.03.002

Reference: YPTSP 712

To appear in: *Physical Therapy in Sport*

Received Date: 9 September 2015

Revised Date: 10 February 2016

Accepted Date: 3 March 2016

# The reliability of shoulder range of motion measures in competitive swimmers

Helen Walker*, Tania Pizzari, Henry Wajswelner, Peter Blanch, Laura Schwab, Kim Bennell, Belinda Gabbe.

Corresponding author:

Dr Helen Walker

Physio4athletes

Address- 301/370 St Kilda Rd Melbourne 3004

Phone- +61 417 369 278, Fax- +613 96904043

Email- helen@physio4athletes.com

# The reliability of shoulder range of motion measures in competitive swimmers


## Abstract

***Objectives:*** Investigate reliability of shoulder internal and external rotation (IR, ER), abduction in internal rotation (ABIR) and combined elevation (CE) range of motion tests in competitive swimmers.

***Design:*** Within participants, inter- and intra-examiner reliability

***Setting:*** Physiotherapy Department, University of Melbourne, Australia.

***Participants**:* 17 competitive swimmers (aged 12-24 years) who participate in at least 5 weekly swimming sessions and two physiotherapy examiners.

***Main outcome measures:*** Inter- and intra-examiner reliability of IR, ER, ABIR and CE.

***Results:*** Good to excellent intra-examiner reliability across tests (ICCs: 0.85-0.96) with standard error of measurement (SEM)and minimal detectable change at 90% confidence interval ($MDC_{90}$) ranging from 2-5, and 5-12 degrees, respectively. Good to excellent inter-examiner reliability for all tests (ICCs: 0.77-0.94) except left IR (ICC: 0.65). Inter-examiner SEM and $MDC_{90}$ ranged from 2-5 degrees and 5-12 degrees, respectively.

***Conclusion:***

Shoulder range of motion tests were reliable when applied by the same examiner. Inter-examiner reliability was acceptable for all tests except IR, which was affected by inconsistent manual scapula stabilisation between examiners.

**Keywords**

Shoulder, range of motion, reliability, swimming

## Introduction

Reliable measurement tools are essential for interpreting research investigating risk factors for sports injuries, or evaluating the effectiveness of treatment interventions (Bahr & Holme, 2003; Valentine & Lewis, 2006; Cools, Wilde, Tongel, Ceyssens, Ryckewaert & Cambier, 2014). Studies investigating shoulder range of motion (ROM) measurement reliability suggest that no single method affords superior measurement reproducibility (Hoving et al., 2002; Valentine & Lewis, 2006; Cools et al., 2014, Mullaney, McHugh, Johnson, & Tyler, 2010)**.** Methodologies and results vary with the movement type (active or passive ROM), measurement device, test position, and study population (symptomatic or asymptomatic) investigated (Riddle, Rothstein & Lamb, 1987; Andrews & Bohannon, 1989; Croft, Pope, Boswell, Rigby & Silman, 1994; Green, Buchbinder, Glazier & Forbes, 1998; Sabari, Maltzev, Lubarsky & Homel 1998; Boon & Smith, 2000; Hayes, Walton, Szomor & Murrell, 2001; Hoving et al. 2002; de Winter et al., 2004; Valentine & Lewis 2006; Kobler, Vega, Widmayer & Cheng, 2009; Muir, Corea & Beaupre, 2010; Cools et al. 2014;). Nonetheless, the measurement of shoulder ROM is an important component of the clinical examination of the shoulder joint, as alterations in ROM have been implicated in the aetiology of pathology (Ellenbecker et al., 2002; de Winter et al. 2004; Valentine & Lewis, 2006; Cools et al., 2014). In competitive swimmers, the incidence of reported shoulder pain is high, and some authors have suggested a correlation between shoulder ROM and injury (McMaster, Roberts & Stoddard, 1998; Ozcaldiran 2002, Walker, Gabbe, Wajswelner, Blanch & Bennell, 2012). Competitive swimmers regularly undergo pre-participation screening to identify possible intrinsic risk factors for shoulder injury (Walker et al. 2012; Blanch, 2004). These tests include shoulder rotational measurements, as well as the sport-specific tests abduction in internal rotation (ABIR) and combined elevation (CE) (Blanch, 2004). The reliability of measurement of these tests in competitive swimmers has not been reported to date. This study sought to investigate the intra and inter-examiner reliability of tests of range of shoulder internal (IR) and external rotation (ER) at 90 degrees abduction, ABIR and CE in competitive swimmers.

## 2. Method

*2.1 Setting and study design*

A reliability study was conducted involving two physiotherapist examiners.

*2.2 Participants*

A convenient sample of 17 swimmers was recruited from five competitive swimming clubs in Melbourne, Australia. Swimmers participating in at least five swim sessions per week were eligible for this study. Exclusion criteria included a history of shoulder surgery or dislocation, shoulder pain on the day of testing or the presence of any other injury that would interfere with the test procedures. Examiners were two physiotherapists with twelve (A) and five years (B) clinical experience who were skilled in manual examination of the shoulder. Examiner A had post graduate qualifications in sports physiotherapy and extensive prior experience with screening competitive swimmers.

*2.3 Procedures*

Swimmers completed a baseline questionnaire regarding demographics, anthropometric features, swimming training and injury history. Shoulder range of motion testing was conducted by two physiotherapists with a Dualer Inclinometer (J-Tech Medical, Salt Lake City), which was zeroed before each test. Each swimmer was tested twice by both examiners for all ROM tests, with an intervening 30 minute rest break. Swimmers were randomised to an examiner and this order was maintained for the second test. Shoulder ROM test order was standardised and the right shoulder was tested first. Three trials of each test were conducted and results averaged for analysis. Examiners were blinded to all test results.

*2.4 Shoulder Internal Rotation and External Rotation*

Active IR and ER ROM measurements were conducted in supine on a plinth, from a start position of 90 degrees shoulder abduction and elbow flexion, as determined by standard goniometry techniques (Riddle et al., 1987). The forearm and wrist were maintained in a neutral position and the inclinometer was attached 5cm distal to the olecranon process of the elbow. During the shoulder rotation movement, the examiner stabilised the scapula with one hand using a caudal and posteriorly directed force, to prevent unwanted shoulder girdle motion.

*2.5 Combined Elevation*

CE was measured in prone on a plinth with the arms outstretched, elbows extended, the thumbs linked together and palms down. The chin, torso and legs were maintained in contact with the plinth during testing such that only the arms were lifted (Blanch, 2004). . The inclinometer was attached to the right arm just below the deltoid insertion and swimmers actively elevated both arms as far as possible, keeping the elbows straight. The start and finish positions of bilateral arm elevation were recorded.

*2.6 Shoulder Abduction in Internal Rotation*

Active ABIR was measured from an upright standing position with the back, head and heels against a wall. The arms were positioned by the side, with the elbows flexed to 90 degrees and forearms pronated. The inclinometer was attached just below the deltoid insertion with the face of the inclinometer in the coronal plane of movement. Bilateral shoulder abduction was then performed maintaining the elbows at 90 degrees flexion and the forearms perpendicular to the plane of abduction, in order to standardise shoulder internal rotation. Contact between the upper arm and wall was maintained at all times during testing. The finish position was recorded.

*2.7 Data Analysis*

Data were analysed using SPSS 11.0 for Windows. All variables were examined for normality using Kolomogorov-Smirnov normality tests. Measurement reliability was examined using intraclass coefficient (ICC) models 2,3 and 3,3 for intra and inter-examiner reliability, respectively. Standard error of measurement (SEM= Average Standard deviation x $\sqrt{(1 - ICC)}$ and minimal detectable change at the 90% confidence level were calculated ($MDC_{90}$= 1.65*SEM*$\sqrt{2}$) (Haley & Fragala-Pinkham, 2006; Portney & Watkins, 2009; Cools et al, 2014). Good to excellent reliability was defined a priori as an ICC >0.75 (Portney & Watkins, 2009). Paired t-tests and Wilcoxon signed ranks test were conducted to detect any systematic differences between data sets (Portney & Watkins 2009; Cools et al., 2014).

## 3. Results

Swimmers ranged in age from 12 to 24 years and participated in six weekly swim sessions, on average, as summarized in Table 1. Means and standard deviations for the ROM tests are shown in Table 2. A priori examination of the raw ROM data revealed extreme outlying scores for right-sided IR recorded by Examiner B, for one participant. The entire ROM data set for this participant was deleted to prevent any undue influence of this data on results (Tabachnick & Fidell 1989), leaving 16 participants for analysis.

**Table 1: Swimmer characteristics**

| Characteristics | Male N=8 | Female N=8 | Total N=16 |
|---|---|---|---|
| Age (years) | 18 ± 3 | 16 ± 3 | 17 ± 3 |
| Age started swim competition (years) | 10 ± 3 | 9 ± 3 | 9 ± 3 |
| Height (cm) | 182 ± 5 | 166 ± 6 | 174 ± 10 |
| Average swimming distance/week (km) | 59 ± 11 | 53 ± 13 | 56 ± 12 |
| Number of Swim sessions/week | 6 ± 1 | 8 ± 1 | 6 ± 1 |
| Top two preferred strokes | Freestyle, Backstroke | Backstroke, Freestyle | Freestyle, Backstroke |

**Table 2 : Mean ROM for each shoulder test**

| ROM Test | | Trial 1 Examiner A | | Trial 2 Examiner B | |
|---|---|---|---|---|---|
| **IR (°)** | Right | 49 ±10 | 48 ±9 | 58 ±13 | 55 ±11 |
| | Left | 53 ±7 | 53 ±7 | 63 ±8 | 60 ±9 |
| **ER (°)** | Right | 97 ±10 | 99 ±11 | 101 ±12 | 102 ±11 |
| | Left | 94 ±9 | 94 ±8 | 98 ±10 | 100 ±9 |
| **ABIR(°)** | Right | 119 ±11 | 120 ±10 | 123 ±13 | 124 ±12 |
| | Left | 123 ±10 | 126 ±9 | 124 ±12 | 126 ±11 |
| **CEF (°)** | | -5 ±9 | -3 ±8 | -3 ±9 | -0.4 ±8 |
| **CEX (°)** | | 16 ±7 | 17 ±7 | 15 ±7 | 17 ±7 |

CEX= Combined elevation test total excursion

*3.1 Inter-examiner Reliability*

As detailed in Table 3, there was good to excellent inter-examiner reliability for all shoulder ROM tests (ICCs: 0.77-0.94) with the exception of left IR (ICC: 0.65). Inter-examiner SEMs ranged from two to five degrees and $MDC_{90}$ ranged from five to 12 degrees. The mean differences in results between examiners for each ROM test are shown in Table 4. Significant differences between examiners were found for IR and ER both times they were measured and for right ABIR and CEF on the second testing occasion (p<0.05). The greatest mean differences (7 to 9 degrees) were recorded for IR testing. This indicated a systematic testing difference between examiners for IR measurements.

**Table 3: Inter-examiner comparison of mean difference in ROM results for each trial**

| ROM Test | | Test | Trial 1 z or t | P value | Mean difference (○) | Trial 2 z or t | p value | Mean difference (○) |
|---|---|---|---|---|---|---|---|---|
| **IR** | Right | T- test | -4.5 | 0.00* | 8 | -5.8 | 0.00* | 7 |
| | Left | T-test | -7.1 | 0.00* | 9 | -3.9 | 0.00* | 7 |
| **ER** | Right | T- test | -2.1 | 0.05 | 4 | -2.7 | 0.02* | 3 |
| | Left | T- test | -3.8 | 0.00* | 4 | -4.3 | 0.00* | 6 |
| **ABIR** | Right | T-test | -1.7 | 0.1 | 4 | -2.4 | 0.03* | 4 |
| | Left | T-test/Wilcoxon# | -0.9 | 0.4 | 1 | -0.1# | 0.50 | 1 |
| **CES** | | T-test | -1.6 | 0.1 | 2 | -2.2 | 0.05 | 2 |
| **CEF** | | T-test/Wilcoxon# | -0.9 | 0.4 | 2 | -2.7# | 0.01#* | 3 |
| **CEX** | | T-test | -0.4 | 0.7 | 1 | -0.7 | 0.5 | 1 |

*significance level p<0.05, # =Wilcoxon signed ranks test

*3.2 Intra-examiner Reliability*

Overall, intra-examiner reliability was higher than inter-examiner reliability. There was good to excellent intra-examiner reliability for both examiners across all ROM tests

(ICCs: 0.85-0.96). The SEM and $MDC_{90}$ for the intra-examiner condition ranged from two to five degrees, and five to 12 degrees, respectively. The inter-examiner reliability results for the first trial are summarised in Table 4.

**Table 4: Intra-examiner reliability**

| **ROM TEST** | | **EXAMINER A** | | | | **EXAMINER B** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **ICC** | **95% CI** | **SEM (°)** | **$MDC_{90}$ (°)** | **ICC** | **95% CI** | **SEM (°)** | **$MDC_{90}$ (°)** |
| **IR** | Right | 0.96 | 0.90,0.990 | 2 | 5 | 0.85 | 0.85,0.97 | 5 | 12 |
| | Left | 0.90 | 0.72,0.97 | 2 | 5 | 0.93 | 0.77,0.97 | 2 | 5 |
| **ER** | Right | 0.95 | 0.85,0.98 | 2 | 5 | 0.94 | 0.82,0.98 | 3 | 7 |
| | Left | 0.94 | 0.83,0.98 | 2 | 5 | 0.90 | 0.71,0.96 | 3 | 7 |
| **ABIR** | Right | 0.94 | 0.82,0.98 | 3 | 7 | 0.91 | 0.76,0.97 | 4 | 9 |
| | Left | 0.90 | 0.69,0.96 | 3 | 7 | 0.89 | 0.69,0.96 | 4 | 9 |
| **CEF** | | 0.91 | 0.76,0.97 | 2 | 5 | 0.95 | 0.56,0.99 | 2 | 5 |
| **CEX** | | 0.92 | 0.77,0.97 | 2 | 5 | 0.93 | 0.73,0.98 | 2 | 5 |

**4. Discussion**

This study quantified the reliability of several shoulder ROM tests using relative and absolute indices. Overall, the ROM tests demonstrated good to excellent intra- and inter-examiner reliability in competitive swimmers. To date, there is no research available for comparison of this study's results for ABIR measurement. Measurement reliability of CE has been previously investigated in cricketers with similarly excellent intra and inter-examiner reliability reported (ICCs: 0.87-0.97) although CE ROM was measured with a tape measure and they tested on the floor (Dennis et al., 2007).

Shoulder ROM reliability is reported to vary with the movement direction and type, measurement device and population group investigated (Valentine & Lewis, 2006; Cools et al., 2014). There is consensus that intra-examiner shoulder ROM testing affords improved reliability compared with inter-examiner testing (Riddle et al., 1987; Green et al., 1998; Boon & Smith, 2000; Hayes et al., 2001; Awan, Smith & Boon, 2002; de Winter et al., 2004; Mullaney et al, 2010; Kolber & Hanney, 2012; Cools et al., 2014). Certainly, this study demonstrated superior reliability, across all ROM tests for the intra-examiner comparisons.

The intra and inter-examiner reliability of ER and IR ROM in the current study was higher than, or equivocal to, previously reported results for these tests (Ellenbecker, Roetert, Piorkowski & Schulz, 1996; Green et al., 1998; Boon & Smith, 2000; Awan et al., 2002; Muir et al., 2010; Cools et al., 2014). Nevertheless, in this study, the measurement of ER afforded overall superior reliability when compared with IR ROM. Previous authors have reported that the measurement of shoulder IR ROM is problematic, due to difficulties in isolating movement to the glenohumeral joint (Boon & Smith, 2000; Awan et al., 2002; Muir et al., 2012; Cools et al., 2014). In order to address this issue, manual scapula stabilisation was employed in the current study. The reliability of IR ROM has been investigated both with and without the scapula stabilised, with varied results (Boon & Smith 2000; Awan et al., 2002; Muir et al., 2010; Cools et al., 2014). Boon and Smith (2000) reported that scapula stabilisation afforded improved IR reliability for passive goniometric measurements. Conversely, Awan et al. (2002) found IR passive ROM to be more reliable without scapula stabilisation, when assessed with an inclinometer. Nevertheless, the IR reliability in the current study was higher than that previously reported by both studies investigating passive IR testing (Boon & Smith, 2000; Awan et al., 2002). More recent investigations of active IR ROM goniometric measurement reliability without the scapula stabilised, reported lower ICCs and larger SEMs than the current study (Muir et al., 2010). The current results concur with the notion that manual scapula stabilisation affords better IR ROM measurement reliability (Cools et al., 2014, Muir et al., 2010). Currently, innovative shoulder ROM measurement tools such as smart phone inclinometers and camera visual systems (Kinect, Microsoft, Seattle, WA, USA) are showing promising reliability for shoulder ROM measurement (Werner et al., 2014; Huber, Seitz, Leeser & Sternad., 2015). These evolving tools may negate shoulder-positioning issues and become suitable, cost effective tools in the swimming screening environment.

Clinically, measurement of shoulder rotation ROM is deemed an important component of shoulder joint examination, as alterations in ROM have been implicated in the aetiology of pathology (Ellenbecker et al., 2002; de Winter et al. 2004; Valentine & Lewis, 2006; Cools et al., 2014). In this study, the use of manual scapula stabilisation during rotational measurement is a potential limitation, given that it limits, as much as possible, shoulder motion to the glenohumeral joint and does not reflect the combined glenohumeral and scapulothoracic motion occurring at the shoulder complex during daily activities or sport.. Nonetheless, shoulder IR ROM measurement is one of several methods for measuring posterior shoulder tightness, a proposed risk factor for shoulder pain in overhead athletes (Myers, Laudner, Pasquale, Bradley and Lephart, 2005; Myers, Oyama, Wassinger, Ricci, Abt, Conley, and Lephart, 2007). Previous authors have reported that deficits in shoulder IR ROM, measured with the scapula stabilised, are significantly correlated with posterior shoulder tightness and contend that accurate posterior shoulder tightness measurement can only be obtained with the scapula fixed (Myers et al, 2005, Myers et al, 2007). Therefore, in this study, the use of manual scapula stabilisation during rotational measurements, with the intention to limit motion to the glenohumeral joint is reasonable.

This study investigated the measurement error associated with each ROM test by calculating both SEM and $MDC_{90}$. Intra-examiner SEMs ranged from two to 5 degrees, indicating that a true change in ROM on re-testing may be detected with a difference of four to 10 degrees. Intra-examiner $MDC_{90}$ results indicated that to be 90% confident that change is not due to measurement error or intra-examiner variability, a difference of greater than five to 7 degrees for examiner A, and 5 to 12 degrees for examiner B, is required depending on the test conducted (Carter, Lubinski, & Domboldt, 2011). Inter-examiner SEMs ranged from four to five degrees (95 % CI: 8 to 10 degrees) and $MDC_{90}$ from five to 12 degrees. IR measurement results showed the greatest measurement error and highest minimum detectable change for both intra-examiner and inter-examiner conditions. Previous studies calculated $MDC_{90}$ for shoulder ROM testing in varying directions and reported values ranging from 2.8 to 9 degrees, although different measurement methods were employed (Kolber et al., 2011; Kolber et al., 2012; Cools et al., 2014). Nonetheless, the current results for MDC90 are in line with these previous reports, with the exception of right IR measurement ($MDC_{90}$ =12 degrees). A difference of greater than 10 degrees between examiners on shoulder ROM testing has been defined as unacceptable for clinical purposes (de Winter et al., 2004). Most ROM tests in this study fall within this recommendation with the exception of right shoulder IR testing.

Additionally, this study demonstrated a systematic testing difference for IR between examiners. Examiner B consistently recorded IR measurements that were seven to nine degrees higher than Examiner A. It is unlikely that this was due to the sequence of examiner testing as this was randomly allocated and the effect was not present for the other ROM tests (de Winter et al., 2004). In particular, SEM and $MDC_{90}$ were higher for Examiner B when testing right shoulder IR. Whilst examiners were consistent within themselves, these findings likely demonstrate the difficulty of standardising manual scapular stabilisation between examiners and perhaps between sides. This concurs with prior reports that IR ROM measurement is less reliable (Muir et al., 2010; Cools et al., 2014). The training process for this study utilised a physiotherapist as a subject to provide feedback regarding scapula stabilisation. Despite ROM testing being an entry level skill for physiotherapists, Examiner A was considerably more experienced with the testing methods and the results may suggest a learning effect. The reliability of IR measurement may be improved for future studies by practising scapular stabilisation with objective feedback, such as a pressure biofeedback unit.

One possible explanation for the good to excellent reliability findings across tests in this study is that measurement of active ROM is more reliable than passive ROM. The measurement of passive ROM introduces potential errors associated with examiners' determination of end of range and differences in the amount of applied passive force

(Green et al., 1998; Boon & Smith, 2000; de Winter et al., 2004; Cools et al., 2014). Similarly, inclinometers may improve measurement reliability, compared with goniometry, due to the elimination of errors associated with locating joint centres of rotation and/or bony landmarks (Green et al., 1998; Awan et al., 2002; Mullaney et al., 2010). A further explanation for this study's results is that the mean of three measurements was used for analysis, which may have enhanced reliability, although these procedures are commonly utilised (de Winter et al., 2004; Cools et al., 2014). Additionally, this study employed ROM testing methods that were standardised for participant position, joint stabilisation and the determination of end of ROM, all of which are purported to minimise shoulder ROM variation (Hoving et al., 2002).

As previously discussed, the ROM data set for one participant was excluded from analysis due to outlying scores for right sided shoulder IR recorded by Examiner B. These extreme scores were 22 and 17 degrees greater than Examiner B's mean for right sided IR, and 31 and 24 degrees greater than Examiner A's mean, in the first and second trial, respectively. Outlying scores can be true scores that appear extreme because the sample size is too small, or may be caused by measurement or recording errors (Portney & Watkins, 2009). Given that the IR scores recorded for this participant were extreme for only one examiner, it is unlikely that these scores represented true values. The most plausible explanation for these extremely high ROM scores was insufficient scapula stabilisation applied by Examiner B, which allowed unwanted scapulo-thoracic motion to occur, although the exact cause is unknown. Nevertheless, it is unlikely that exclusion of this data set considerably affected the reliability results across tests.

A power analysis was not conducted prior to commencement of this study. Previous authors reported that a sample size of 19 participants was required, at 80% power ($\beta$=20) and 5% significance, for two examiners or testing occasions, when investigating the reliability of shoulder abduction, flexion, IR and ER ROM measurement in subjects with and without symptoms using inclinometry, tape measure (ER) and visual estimation (IR) (Valentine & Lewis, 2006). Despite different testing methods, the sample size in the current study falls below this estimate. As such, it is possible that the statistical power in this study was diminished by the small sample investigated and the lack of variability in results.

The extent to which the results of this study can be generalised to other populations is limited, given the specialised sample investigated. Competitive swimmers undergo training that optimises upper limb positioning during the stroke and, as such, may perform consistently when asked to self-determine active ROM limits. Therefore, these results may not be applicable to other sporting disciplines or symptomatic patient groups.

**5. Conclusions**

The accuracy of measurement of the variables of interest is crucial to a study's ability to detect potential associations between risk factors and injury (Krosshaug, Anderson, Olsen, Myklebust & Bahr, 2005). This study provides meaningful information regarding the reliability and standard errors of measurement for several shoulder ROM measurements, such that the results of future prospective cohort studies employing these tests can be assessed with confidence with respect to effect size and clinical significance. The good to excellent reliability of shoulder ROM measurement methods employed in this study indicates that these measures are suitable assessment tools for competitive swimmers. In most instances, the shoulder ROM tests can be applied confidently by more than one examiner. When measuring shoulder IR, it is recommended that one examiner only conducts testing, until such a time that scapula stabilisation methods, and subsequently intra and inter-examiner variability is reduced. These assessment tools can be used for baseline testing in future prospective investigations of risk factors for shoulder pain in competitive swimmers.

**References**


Andrews, A. W. & Bohannon R. W. (1989). Decreased Shoulder Range of Motion on Paretic Side after stroke. *Physical Therapy*, 69(9), 768-772.

Awan, R., Smith, J., & Boon, A.J. (2002). Measuring shoulder internal rotation range of motion: a comparison of three techniques. *Archives of Physical Medicine and Rehabilitation.* 83(9): 1229-1234.

Bahr, R. & I. Holme (2003). Risk factors for sports injuries- a methodological approach. *British Journal of Sports Medicine*, 37(5), 384-392.

Blanch, P. (2004). Conservative management of shoulder pain in swimming. *Physical Therapy in Sport, 5*, 109-124.

Boon, A.J. & Smith J. (2000). Manual Scapular stabilisation: Its effect on Shoulder Rotational Range of Motion. *Archives of Physical Medicine and Rehabilitation,* 81(7), 978-983.

Carter, R.E., Lubinski, J. & Domboldt, E. (2011) *Rehabilitation principles: Research and applications*. St Louis, MO: Saunders.

Cools, A.M., De Wilde, L., Van Tongel, A.V., Ceyssens, C., Ryckewaert, R. and Cambier, D.C. (2014). Measuring shoulder external and internal rotation strength and range of motion: comprehensive intra-rater and inter-rater reliability study of several testing protocols. *Journal of Shoulder and Elbow Surgery*, 23, 1454-1461.

Croft, P., Pope, D., Boswell, R., Rigby, A & Silman, A. (1994). Observer variability in measuring elevation and external rotation of the shoulder. *British Journal of Rheumatology,* 33(10), 942-946.

Dennis, R.J., Finch, C.F., Elliott, B.C. & Farhart, P.J. (2008). The reliability of musculoskeletal screening tests used in cricket. *Physical Therapy in Sport*. 9(1), 25-33.

de Winter, A.F., Heemskerk, M.A., Terwee, C.B., Jans, M.P., Deville, W., van Schaardenburg, D.J., … Bouter, L.M. (2004). Inter-observer reproducibility of measurements of range of motion in patients with shoulder pain using a digital inclinometer. *BMC Musculoskeletal disorders,* 5, 18-26.

Ellenbecker, T.S., Roetert, E.P., Piorkowski, P.A. & Schulz, D.A. (1996). Glenohumeral joint internal and external rotation range of motion in elite junior tennis players. *The Journal of Orthopaedic and Sports Physical Therapy*, 24(6), 336-341.

Ellenbecker, T.S., Bailie, D.S., Mattalino, A.J., Carfagno, D.G., Wolff, M.W., Brown, S.W. & Kulikowich, J.M. (2002). Intra-rater reliability of a manual technique to assess anterior humeral head translation on the glenohumeral joint. *Journal of Shoulder and Elbow Surgery*, 11(5), 470-475.

Green, S., Buchbinder R., Glazier, R. & Forbes, A.(1998). Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment and efficiacy. *British Journal of Medicine*. 316(7128), 354-360.

Haley, S.M. & Fragala-Pinkham, M.A. (2006). Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy*, 86(5), 735–743.

Hayes, K., Walton, J.R., Szomor, Z.R. & Murrell, G.A. (2001). Reliability of five methods of assessing shoulder range of motion. *Australian Journal of Physiotherapy*, 47(4), 289-294.

Hoving, J.L., Buchbinder, R, Green, S., Forbes, A., Bellamy, N., Brand, C., … Stockman, A.(2002). How reliably do rheumatologists measure shoulder movement? *Annals of the Rheumatic Diseases*, 61(7), 612-616.

Huber, M.E., Seitz, A.L., Leeser, M. & Sternad, D. (2015). Validity and reliability of Kinect skeleton for measuring shoulder joint angles: a feasibility study. *Physiotherapy*, In press, http://dx.doi.org/10.1016/j.physio.2015.02.002

Kolber, M.J., Vega, F. Jr, Widmayer, K. & Cheng, M.S. (2011). The reliability and minimal detectable change of shoulder mobility measurements using a digital inclinometer. *Physiotherapy Theory and Practice*, 27(2), 176-184.

Kolber, M.J. & Hanney, W.J. (2012). The reliability and concurrent validity of shoulder mobility measurements using a digital inclinometer and goniometer: a technical report. *International Journal of Sports Physical Therapy*, 7(3), 306-313.

Krosshaug, T., Andersen, T.E., Olsen, O.E., Myklebust, G. & Bahr, R. (2005). Research approaches to describe the mechanisms of injuries in sport: limitations and possibilities. *British Journal of Sports Medicine*, 39(6), 330-339.

McMaster, W.C., Roberts, A., & Stoddard, T. (1998). A correlation between shoulder laxity and interfering pain in competitive swimmers. *American Journal of Sports Medicine*, 26(1), 83-86.

Muir, S.W., Corea, C.L. & Beaupre, L. (2010). Evaluating change in clinical status: reliability and measures of agreement for the assessment of glenohumeral joint range of motion. *North American Journal of Sports Physical Therapy*. 5(3), 98-110.

Mullaney, M.J., McHugh, M.P., Johnson, C.P. & Tyler, T.F. (2010). Reliability of shoulder range of motion comparing goniometer to a digital level. *Physiotherapy Theory and Practice.* 26(5), 327-333.

Myers, J.B., Oyama, S., Wassinger, C.A., Ricci, R.D., Abt, J.P., Conley, K.M., Lephart, S.M. (2007). Reliability, precision, accuracy, and validity of posterior shoulder tightness assessment in overhead athletes. *American Journal of Sports Medicine*. 35(11), 1922-1930.

Myers, J.B., Laudner, K.G, Pasquale, M.R., Bradley, J.P., Lephart, S.M. (2005). Glenohumeral range of motion deficits and posterior shoulder tightness in throwers with pathologic internal impingement. *American Journal of Sports Medicine*. 34(3), 385-391.

Ozcaldiran, B. (2002). A relation between static flexibility and shoulder pain in competitive age group swimmers. *The Pain Clinic*, 14(3), 159-163.

Portney, L.G. & Watkins, M.P., (2009) *Foundations of Clinical Research. Applications to Practice* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Riddle, D.L., Rothstein, J.M. & Lamb, R.L. (1987). Goniometric reliability in a clinical setting. Shoulder measurements. *Physical Therapy*, 67(5), 668-673.

Sabari, J. S., Maltzev, I., Lubarsky, D. & Homel, P. (1998). Goniometric assessment of shoulder range of motion: comparison of testing in supine and sitting positions." *Archives of Physical Medicine and Rehabilitation***,** 79(6), 647-651.

Tabachnick, B.G., & Fidel, L.S. (1989). *Using multivariate statistics.* (2nd ed.). New York: Harper-Collins.

Valentine, R. E. & J. S. Lewis (2006). "Intraobserver relaibility of 4 physiologic movements of the shoulder in subjects with and without symptoms." *Archives of Physical Medicine and Rehabilitation*, 87(9), 1242-1249.

Walker, H., Gabbe, B., Wajswelner, H., Blanch P. & Bennell, K. (2012). Shoulder pain in swimmers: A 12-month prospective cohort study of the incidence and risk factors. *Physical Therapy in Sport*. 13(4), 243-249.

Werner, B.C., Holzgrefe, R.E., Griffin, J.W., Lyons, M.L., Cosgrove, C.T., Hart, J.M. & Brockmeier, S.F. (2014). Validation of an innovative method of shoulder range of motion measurement using a smartphone clinometer application. *Journal of Shoulder and Elbow Surgery*, 23(11), e275-282.

**Acknowledgements**

Professor Kim Bennell is supported by a National Health and Medical Research Council Principal Research Fellowship

**Highlights**

- Shoulder active range of motion measurement reliability was examined in swimmers.
- Supine internal and external rotation range was tested with the scapula stabilised.
- Abduction in internal rotation and combined elevation range were tested.
- Intra and inter-examiner reliability of all but one test was good to excellent.
- Internal rotation should be measured by the same examiner when examining swimmers.

**Ethical Statement**

The study was approved by the University of Melbourne Human Research Ethics Committee and all participants or their parents/guardians provided written informed consent.