Manual Therapy 33 (2018) 11–17

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/locate/math

Original article

Consistency of commonly used orthopedic special tests of the shoulder when used with the McKenzie system of mechanical diagnosis and therapy

Afshin Heidar Abady[a,*], Richard Rosedale[b], Bert M. Chesworth[a,c], Michael A. Rotondi[d], Tom J. Overend[a]

[a] School of Physical Therapy, Western University, London, Canada
[b] London Health Sciences Centre, London, Canada
[c] Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, Canada
[d] School of Kinesiology and Health Sciences, York University, Toronto, Canada

ARTICLE INFO

Keywords:
McKenzie
Mechanical Diagnosis And Therapy (MDT)
Orthopedic Special Tests (OSTs)
Shoulder

ABSTRACT

Background: Shoulder Orthopedic Special Tests (OSTs) are used to assist with diagnosis in shoulder disorders. Issues with reliability and validity exist, making their interpretation challenging. Exploring OST results on repeated testing within Mechanical Diagnosis and Therapy (MDT) shoulder classifications may offer insight into the poor performance of these tests.

Objectives: To investigate in patients with shoulder complaints, whether MDT classifications affect the agreement of OST results over the course of treatment.

Methods: An international group of MDT clinicians recruited 105 patients with shoulder problems. Three commonly used OSTs (Empty Can, Hawkins-Kennedy, and Speed's tests) were utilized. Results of the OSTs were collected at sessions 1, 3, 5 and 8, or at discharge from an MDT classification-based treatment. The *Kappa* statistic was utilized to determine the agreement of the OST results over time for each of the MDT classifications.

Results: The overall *Kappa* values for Empty Can, Hawkins-Kennedy and Speed's tests were 0.28 (SE = 0.07), 0.28 (SE = 0.07) and 0.29 (SE = 0.07), respectively. The highest level of agreement was for Articular Dysfunction for the Empty Can test (0.84, SE = 0.19). For shoulder Derangements, there was no agreement for any of the OSTs (*P* values > 0.05).

Conclusion: The lack of agreement when the OSTs were consecutively tested in the presence of the MDT Derangement classification contrasted with the other MDT classifications. The presence of Derangement was responsible for reducing the overall agreement of commonly used OSTs and may explain the poor consistency for OSTs.

1. Introduction

Shoulder pain is one of the primary reasons for referral to physiotherapy with an annual prevalence of 100–160 per 1000 patients in the general population (Winters et al., 1999). It has been shown to be relentless and recurring, with half of all cases remaining unresolved after 18 months (Croft et al., 1996). Complexity of the shoulder, and absence of uniformity in diagnostic labeling (Schellingerhout et al., 2008) hinder accurate diagnosis. This can have significant implications for conservative management where ideally the diagnosis should directly guide clinical reasoning and decision making (Lange et al., 2017; McClure and Michener, 2015). These diagnostic challenges may inadvertently lead to inappropriate and perhaps more costly interventions (Cook, 2010).

For physical examination of the shoulder, Orthopedic Special Tests (OSTs) are commonly used (Sciascia et al., 2012) and despite a heavy reliance on their use, demonstrate only limited utility for informing diagnosis (Schellingerhout et al., 2008; Lange et al., 2017; Cook, 2010; Hegedus et al., 2008). Studies have revealed conflicting diagnostic performance for the majority of OSTs used in the assessment of common shoulder disorders such as rotator cuff pathology, sub-acromial impingement and superior labrum anterior-to-posterior (SLAP) lesions (Hegedus et al., 2008; Kuhn et al., 2007; Dessaur and Magarey, 2008; Hughes et al., 2008; Powell et al., 2008; Walsworth et al., 2008; Walton and Sadi, 2008; Beaudreuil et al., 2009; Lewis, 2009; Meserve et al., 2009; Munro and Healy, 2009; Nomden et al., 2009; May et al., 2010; McFarland et al., 2010; Gadogan et al., 2011; Tucker et al., 2011). Considering the shortcomings of commonly used OSTs, a

* Corresponding author. School of Physical Therapy, 1201 Western Road, Rm 1588 Elborn College, Western University, London, Ontario, N6G 1H1, Canada.
E-mail address: aheidara@uwo.ca (A. Heidar Abady).

http://dx.doi.org/10.1016/j.msksp.2017.10.001
Received 13 July 2017; Received in revised form 25 September 2017; Accepted 4 October 2017

growing body of opinion favours the implementation of an approach that is different than a patho-anatomical based assessment and diagnosis of musculoskeletal disorders (Schellingerhout et al., 2008; Beaudreuil et al., 2009; May and Rosedale, 2012). In principle, the use of a reliable form of classification should decrease practice variation, and enhance the effectiveness of treatment by matching that intervention to a specific subgroup (Guide to physical therapist practice, 2001; Koes et al., 2006). The McKenzie system of Mechanical Diagnosis and Therapy (MDT) is one alternative method that has been proposed to assist the clinician in formulating a classification that enables an appropriate management strategy (Heidar Abady et al., 2017). The MDT system was initially described in 1981 as a new method for classification and treatment of patients with back pain (McKenzie, 1981). The system uses a non-pathology specific classification approach that consists of a thorough history and physical examination monitoring the effects of repeated movements, sustained positions and loading strategies on patients' clinical presentations (May, 2009).

Several systematic reviews show varying degrees of support for the utilization of the MDT system when treating patients with acute and chronic low back pain (Clare et al., 2004; Cook et al., 2005; Machado et al., 2006; Hettinga et al., 2007; Slade and Keating, 2007; May and Donelson, 2008; Aina et al., 2004; May et al., 2006). The MDT system has also demonstrated acceptable reliability (Razmjou et al., 2000; Kilpikoski et al., 2002; Clare, 2005; Dionne et al., 2006; Werneke et al., 2011a) and varying degrees of validity (Long, 1995; Sufka et al., 1998; Werneke et al., 1999; Werneke and Hart, 2001; Wetzel and Donelson, 2003; Berthelot et al., 2007; Long et al., 2008; Werneke et al., 2011b; Takasaki and May, 2014; Petersen et al., 2011; Machado et al., 2010) when used in patients with spinal disorders. A growing body of evidence supports the application of the MDT system when treating patients with musculoskeletal disorders of the extremity (Heidar Abady et al., 2017; Aina and May, 2005; May, 2006; Kelly et al., 2008; May and Ross, 2009; Shouta et al., 2009; Scott et al., 2011; Littlewood et al., 2012; Heidar Abady et al., 2014; Rosedale et al., 2014; Willis et al., 2017). Although reliability varies considerably between different study designs (Takasaki et al., 2017; Takasaki, 2016), very good inter-examiner reliability has been reported specifically for the shoulder (Heidar Abady et al., 2014).

In the McKenzie system, extremity disorders include the following syndromes and subgroups (McKenzie and May, 2000):

- Derangement, identified by the presence of a directional preference which will give a rapid and lasting improvement in symptoms, in range of movement and in function;
- Articular Dysfunction, identified by intermittent pain consistently produced only at a restricted end range of motion with no rapid change of symptoms or range;
- Contractile Dysfunction, identified by intermittent pain, consistently produced by loading the musculo-tendinous unit, for instance, with an isometric contraction against resistance;
- Postural syndrome, identified by intermittent pain only produced by sustained loading, with movements and activities being unaffected;
- OTHER subgroups are considered when none of the above syndrome patterns are present. Each has a definition and specific criteria that together complete the classification for all remaining presentations. Examples include Trauma, Peripheral Nerve Entrapment and Inflammatory (Appendix A).

Although there are clear issues with the validity and clinical interpretation of OSTs, their use is still widespread, with many clinicians continuing to utilize these tests as a basis for diagnosis in shoulder disorders (Sciascia et al., 2012). One common observation by MDT clinicians and reported in various case studies (Kaneko et al., 2009; Kidd, 2013; Lynch and May, 2013) is that the results of OSTs can change depending upon the MDT classification. For example, in one case study (Kidd, 2013), the initial treatment of a patient with a shoulder Derangement was reported to have an immediate effect on the 'Empty can' test, the 'Lift off' test and the 'Hawkins-Kennedy' test, with test results shifting from positive to negative within the first session and remaining negative until discharge. It is possible that the insights from this case may give one possible explanation as to why these OSTs appear inherently unreliable and of questionable validity. Derangement has a variable nature in terms of movement loss, direction of preference and pain behavior. Hence, at times a patient may be experiencing severe symptoms, considerable loss of motion and limited function; at other times the symptoms may be milder, with greater range and better function. This may happen either naturally in response to the patient's daily movements and loading of the joint or in response to the therapeutic intervention e.g. repeated end range movements in the directional preference. The implication for OSTs when tested in the presence of Derangement is that at times, when the Derangement is more severe they may test positive and at other times when the Derangement is milder they may test negative. The OSTs are intended to gauge the presence or absence of a particular pathology or diagnosis, however, in the presence of Derangement, the OST results may be dependent upon the current behavior of the Derangement rather than reflecting the specific pathology they are proposed to identify. This can be particularly apparent when the Derangement is treated with directional preference exercises, where it can be taken from a more painful and limited state to a much less severe state in a short period of time. The classification of Derangement is reported to be a prevalent cause of shoulder pain (May and Rosedale, 2012; Heidar Abady et al., 2017) as it is with other musculoskeletal problems (May and Rosedale, 2012). Hence its presence could be a factor underlying the historic lack of accuracy of the OSTs.

The aim of our study was to investigate, in patients with shoulder complaints, whether MDT classifications and their subsequent treatment regime affects the agreement of commonly used OSTs over time. To determine if shoulder Derangement interferes with the results of OSTs, we hypothesized that over the course of treatment, there would be lower agreement between consecutive OST results in patients with shoulder Derangement compared to patients with shoulder Articular or Contractile Dysfunction. This would be the first study to explore the consistency of OST results within the MDT classification system of the shoulder.

2. Methodology

2.1. Study design and setting

This was a multi-centre prospective longitudinal study that ran concurrently with a study that explored the clinical application of the MDT system in patients with shoulder disorders (Heidar Abady et al., 2017). An international group of 15 McKenzie Institute International diploma and credential holders recruited and collected data from consecutive patients visiting their clinics for treatment of a shoulder problem. These study collaborators were licensed physiotherapists with over one year of experience in applying the MDT system to patients who presented with an upper extremity problem.

Instructions, consent forms and data collection sheets were distributed to the study collaborators. To minimize bias, participating physiotherapists had no awareness of the study objectives and hypotheses. In addition, different orthopedic clinicians who were unaware of the patients' MDT classifications performed and recorded the OST results. The patients were followed up until their discharge from their treatment program, and the completed data collection forms were sent to the primary investigator for analysis. Ethics approval for the study was obtained from the Health Sciences Research Ethics Board of Western University.

A confidence interval (CI) approach for sample size estimation of *Kappa* was used (Rotondi and Donner, 2012). Assuming a preliminary estimate of *Kappa* = 0.7, with a 95% CI of 0.2, we decided that 89

participants were needed for five MDT classifications to ensure a reasonable number of cases across subcategories. Considering a 10% dropout rate, a total of 100 participants was calculated to be a sufficient number for our primary outcome; however, by the time the primary investigators received sufficient data from the study collaborators and declared the end of the study, five additional patients were already recruited and their data were collected. Therefore, clinical data for a total of 105 patients were collected from March 2013 to November 2014.

2.2. Participants

To be included in the study, participants were required to be over the age of 18, English speaking and with a shoulder disorder for which they were pursuing physiotherapy intervention. No specific shoulder diagnosis was required for inclusion. Patients were excluded if they had a surgical intervention on their shoulder within six months before the beginning of their physiotherapy program. No specific shoulder diagnoses were excluded, as one of the intentions of our concurrent study (Heidar Abady et al., 2017) was to classify all patients presenting with shoulder pain using the MDT system.

2.3. Examination and classification

A "treatment-*as*-usual" approach was utilized, and patients were assessed and treated following MDT methods and principles. Patients were allocated to one of the following five subgroups: Derangement, Articular Dysfunction, Contractile Dysfunction, OTHER and Spinal; the latter was recognized as patients referred with "shoulder pain" but the cervical spine was confirmed as the source of symptoms. Spinal classification was accepted to be a cervical spine Derangement and was expected to demonstrate a similar treatment response as shoulder Derangement when the cervical spine was treated. OTHER subgroups included all patients who failed to meet the criteria for any one of the previously described classifications.

2.4. Intervention and outcomes

Treatment followed recognized procedures for each MDT classification; patients were treated with distinctively matched exercises and the relevant progression of forces were pursued as per the MDT method (McKenzie and May, 2000). As there would have been numerous individualized MDT exercise programs depending on each patient's diagnosis and response to treatment, the specific intervention and progression of forces were left to the discretion of the treating practitioners.

Three commonly used OSTs documented in systematic reviews of shoulder tests (Dessaur and Magarey, 2008; Hughes et al., 2008; Powell et al., 2008; Walton and Sadi, 2008; Beaudreuil et al., 2009; Meserve et al., 2009; Munro and Healy, 2009; May et al., 2010) were utilized: Empty Can, Hawkins-Kennedy, and Speed's. In the Empty Can test, resistance is given to abduction in two different positions – 90 degrees of arm abduction with neutral (no) rotation, and 90 degrees of abduction with the shoulder medially rotated and angled forward 30° (empty can position), so that the patient's thumb points toward the floor in the plane of the scapula (Cook and Hegedus, 2008). Examiners look for weakness or pain, which reflects a positive test (Cook and Hegedus, 2008). In the Hawkins-Kennedy test, with the elbow in 90 degrees of flexion, the examiner forward flexes the arm to 90° then quickly medially rotates the shoulder (Magee, 2008). As the indicator of a positive test, examiners look for a sharp pain in the superior aspect of the shoulder (Magee, 2008). The Speed's test consists of resisted forward flexion of the arm while the elbow is fully extended and the patient's forearm is first supinated, and then pronated (Magee, 2008). A positive test induces increased tenderness in the bicipital groove, particularly with the arm supinated (Magee, 2008).

The treating practitioner classified the patients into one of the five MDT classifications. To avoid any potential bias from the treating clinician, a second practitioner with education and training in applying the above named OSTs, was blinded to the patients' MDT classifications and administered the OSTs. The patients were followed up until their discharge from physiotherapy, or after 4 weeks or 8 treatment sessions, whichever came first. The patients' clinical information was collected at the initial assessment, and data on the OST results were collected at sessions 1, 3, 5 and 8, or at their discharge from physiotherapy treatment, whichever came first.

2.5. Data analysis

Descriptive statistics were calculated for the MDT classifications, and patient characteristics. Based on whether the compared variable was continuous or nominal, one-way analysis of variance (ANOVA) or Chi square analysis was performed to compare the following baseline characteristics and potential confounding factors among the MDT classifications: Upper Extremity Functional Index (UEFI) (Stratford et al., 2001), and Numeric Pain Rating Scale (NPRS) (Jensen et al., 1986) scores at baseline, age, sex, hand dominance of the affected shoulder, duration of symptoms, the history of previous episodes with the same condition, medication use and the physical demands of work/daily activities. There were fewer participants in the Articular and Contractile Dysfunction categories and since both types of Dysfunctions have significant similarities, such as their consistent response to examination procedures and slower recovery time, the two groups were merged into a single broad classification of Dysfunction. This allowed for a more equivalent sample size in comparison to the Derangement and Spinal classifications. However, an additional analysis was also conducted whereby the two Dysfunction classifications were analysed as separate groups.

The *Kappa* coefficient and standard error (SE) were calculated to determine the level of agreement of OST results on repeated testing during treatment within each MDT classification. Repeated OST test results were included in the analysis when they were available for at least 3 out of 4 data collection points. The participants, with less than three sets of data, were excluded from the main analysis. The MAGREE macro in Statistical Analysis System (SAS) version 9.3 for Windows was used for data analysis. Traditional thresholds of *Kappa* values were utilized for interpretation as follows: Less than 0.40 = Poor; 0.41–0.60 = Moderate; 0.61–0.80 = Good; and 0.81–1.00 = Very Good (Streiner and Norman, 2003).

3. Results

The flow of patient enrolment and MDT diagnoses is presented in Fig. 1. Of the 105 patients enrolled in the study, 12 patients dropped out for the following reasons: shoulder manipulation done by specialist (n = 1); treatment continued in another centre closer to patient (n = 1); change in insurance coverage urged switching to another physiotherapy clinic (n = 1); failure to complete data collection due to emergency leave of absence by treating physiotherapist (n = 2); sudden travel out-of-town for lengthy period of time (n = 3); decline to return for follow up visit following initial session (n = 4).

Of the 93 participants who completed the study, 11 patients were excluded as they had either two concurrent MDT classifications, or were diagnosed as one of the OTHER MDT subgroups. Of the remaining 82 patients, we decided to run the analysis by including patients who had OST results for at least three of the four data collection points. This allowed us to include 75 eligible participants (only 36 participants had their OST test results for all four data collection points).

Distribution of the MDT classifications and patient characteristics are presented in Table 1. There was no statistically significant difference among the three main MDT subgroups of Derangement, Dysfunction, and Spinal for the patient characteristics and outcome scores

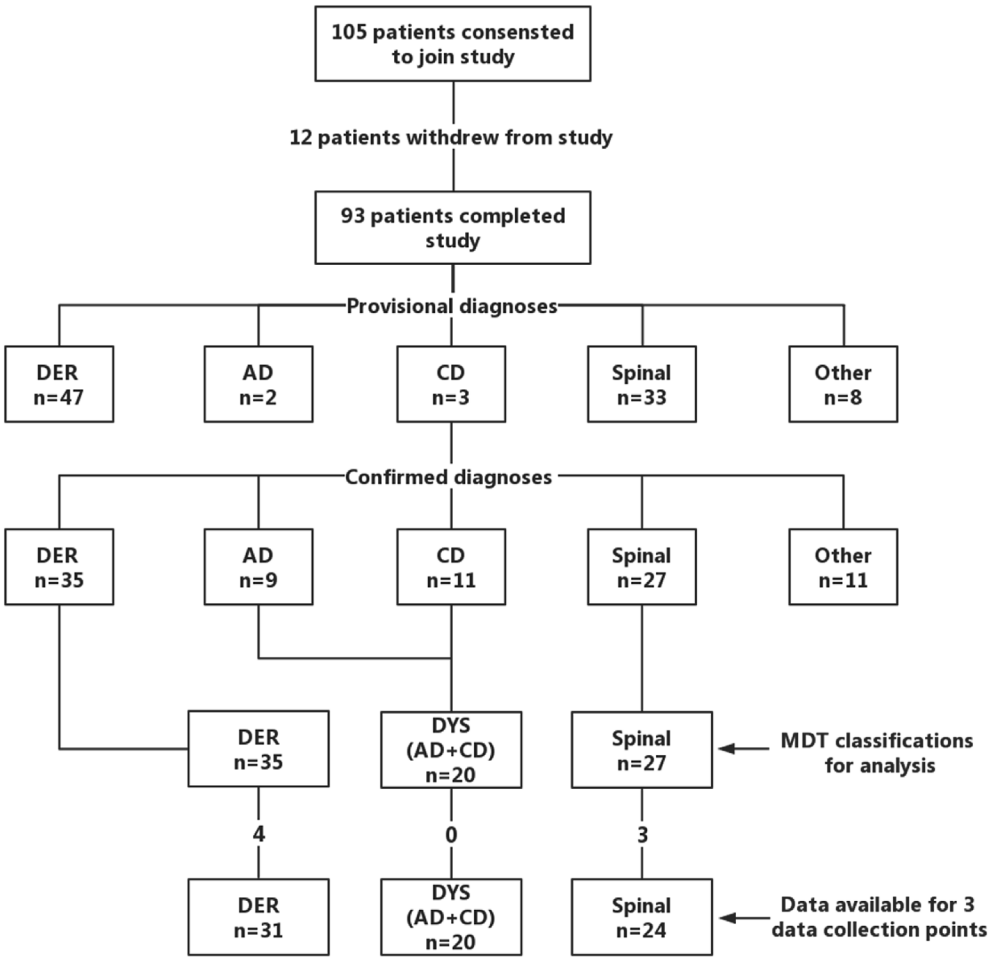

Fig. 1. Flow of patients and MDT classifications. Abbreviations: AD, Articular Dysfunction; CD, Contractile Dysfunction; DER, Derangement; DYD, Dysfunction; MDT, Mechanical Diagnosis and Therapy.

at baseline (Table 1).

Values of agreement within each one of the MDT classifications for the Empty Can test are shown in Table 2. The overall *Kappa* value (i.e. regardless of MDT classification) was 0.28 (SE = 0.07). The highest level of agreement was in the Dysfunction category with *Kappa* = 0.67 (SE = 0.13); with 0.84 (SE = 0.19) for Articular, and 0.49 (SE = 0.17) for Contractile Dysfunction. There was no agreement within Spinal and Derangement categories (equivalent to zero) as *P* values were greater than 0.05 ($P = 0.13$, and $P = 0.44$ respectively).

Values of agreement within each one of the MDT classifications for the Hawkins-Kennedy test are shown in Table 3. The overall *Kappa* value (i.e. regardless of MDT classification) was 0.28 (SE = 0.07). The highest level of agreement was again in the Dysfunction category with *Kappa* = 0.60 (SE = 0.13); with 0.42 (SE = 0.19) for Articular, and 0.59 (SE = 0.17) for Contractile Dysfunction. The agreement level within the Spinal classification was *Kappa* = 0.26 (SE = 0.12), and there was no agreement within the Derangement category (equivalent to zero) as the *P* value was greater than 0.05 ($P = 0.50$).

Values of agreement within each one of the MDT classifications for the Speed's test are shown in Table 4. The overall *Kappa* value (i.e.

Table 1
Patient characteristics and outcome scores at baseline.

Variable		MDT Classification (n, %)			*P*-Value
		Derangement (31, 41.3%)	Dysfunction (20, 26.7%)	Spinal (24, 32%)	
Age, mean (SD)		47.7 (15.6)	54.1 (15.8)	50.8 (18.7)	0.42[a]
Sex, n (% female)		11 (35.5)	8 (40.0)	14 (58.3)	0.22[a]
NPRS, mean (SD)		5.6 (1.9)	4.7 (2.1)	5.6 (1.6)	0.15[a]
UEFI, mean (SD)		54.7 (15.5)	54.2 (16.0)	51.9 (16.8)	0.80[a]
Hand Dominancy, n (% dominant)		21 (67.7)	13 (65.0)	15 (62.5)	0.92[a]
Previous episodes, n (% yes)		11 (35.5)	8 (40.0)	13 (54.2)	0.37[a]
Medication use, n (% yes)		12 (38.7)	6 (30.0)	8 (33.3)	0.80[a]
Duration of symptoms	≤12 weeks	18 (58.1)	7 (35.0)	14 (58.3)	0.21[a]
	> 12 weeks	13 (41.9)	13 (65.0)	10 (41.7)	
Physical activities	Sedentary-light	18 (58.1)	11 (55.0)	17 (70.8)	0.50[a]
	Medium-heavy	13 (41.9)	9 (45.0)	7 (29.2)	

Abbreviations: MDT, Mechanical Diagnosis and Therapy; NPRS, Numeric Pain Rating Scale; UEFI, Upper Extremity Functional Index; SD, standard deviation.
[a] Not significant.

Table 2
Agreement findings for Empty Can test by MDT classification.

MDT Classification	*Kappa*	Standard Error	*P*-Value
Articular Dysfunction	0.84	0.19	< 0.0001
Contractile Dysfunction	0.49	0.17	0.0023
Overall agreement	**0.28**	**0.07**	< **0.0001**
Spinal	0.13	0.12	0.13[a]
Derangement	0.02	0.10	0.44[a]
Dysfunction (AD + CD)	0.67	0.13	< 0.0001

Abbreviations: MDT, Mechanical Diagnosis and Therapy; AD, Articular Dysfunction; CD, Contractile Dysfunction.
[a] Not significant.

Table 3
Agreement findings for Hawkins-Kennedy test by MDT classification.

MDT Classification	*Kappa*	Standard Error	*P*-Value
Articular Dysfunction	0.42	0.19	0.01
Contractile Dysfunction	0.59	0.17	0.0003
Overall agreement	**0.28**	**0.07**	< **0.0001**
Spinal	0.26	0.12	0.01[a]
Derangement	−0.0005	0.10	0.50[a]
Dysfunction (AD + CD)	0.60	0.13	< 0.0001

Abbreviations: MDT, Mechanical Diagnosis and Therapy; AD, Articular Dysfunction; CD, Contractile Dysfunction.
[a] Not significant.

Table 4
Agreement findings for Speed's test by MDT classification.

MDT Classification	*Kappa*	Standard Error	*P*-Value
Articular Dysfunction	0.47	0.19	0.008
Contractile Dysfunction	0.45	0.17	0.005
Spinal	0.37	0.12	0.0007
Overall agreement	**0.29**	**0.07**	< **0.0001**
Derangement	0.09	0.10	0.19[a]
Dysfunction (AD + CD)	0.46	0.13	0.0002

Abbreviations: MDT, Mechanical Diagnosis and Therapy; AD, Articular Dysfunction; CD, Contractile Dysfunction.
[a] Not significant.

regardless of MDT classification) was 0.29 (SE = 0.07). The highest level of agreement was again in the Dysfunction category with *Kappa* = 0.46 (SE = 0.13); with 0.47 (SE = 0.19) for Articular, and 0.45 (SE = 0.17) for Contractile Dysfunction. The agreement level within the Spinal classification was *Kappa* = 0.37 (SE = 0.12), and there was no agreement within the Derangement category (equivalent to zero) as the *P* value was greater than 0.05 (*P* = 0.19).

4. Discussion

To our knowledge, this was the first study to explore the agreement across repeat testing of three OSTs within MDT classifications of the shoulder. This is perhaps not surprising, as in principle the OSTs are oriented towards gaining a patho-anatomical diagnosis whereas the MDT classification is symptom-based. Hence, OSTs would not normally be an integral part of the MDT assessment. However, many MDT trained clinicians still choose to use OSTs as baseline measures.

The main finding of our study was poorer agreement with repeated testing of the OSTs in patients with Derangement compared to patients with either Contractile or Articular Dysfunction. This is consistent with a case study of a patient with shoulder Derangement (Kidd, 2013) that reported test results for the Empty Can, Lift off, and Hawkins-Kennedy tests during a standard MDT assessment and treatment protocol. These tests changed from positive to negative during the initial treatment session and remained negative until discharge. This inconsistency of the OSTs has been a frequent observation by MDT practitioners among patients with Derangement. Specifically, what is noted is that positive OSTs will often become negative as soon as the treatment process is initiated, hence the assumption in these cases is that the tests were initially false positives and not truly indicative of the patho-anatomical condition they were being used to diagnose. In our study, inconsistent test results for OSTs performed in patients assigned to the Derangement classification were revealed by poor agreement statistics across repeat testing from the initial assessment through three to four treatment sessions. This may be due to the variable and quickly changing nature of the Derangement classification especially as it rapidly responds to intervention. Reproduction of these findings in another cohort would provide confirmatory evidence that some OST results are impacted by the nature of the MDT classification.

The overall agreement for Empty Can, Hawkins-Kennedy, and Speed's tests were almost identical with a *Kappa* = 0.28 (SE = 0.07) for Empty Can and Hawkins-Kennedy tests, and a *Kappa* = 0.29 (SE = 0.07) for Speed's test. However, as shown in Tables 2–4 when values for Derangement and Spinal (a cervical spine Derangement) were removed from the analyses, the agreement level increased dramatically with *Kappa* values of 0.67 (SE = 0.13), 0.60 (SE = 0.13), and 0.46 (SE = 0.13) for Empty Can, Hawkins-Kennedy, and Speed's tests respectively. Furthermore, *P*-values for the Derangement classification were greater than 0.05 for all the three OSTs studied. The *P*-value was similarly greater than 0.05 for the Spinal classification for the Empty Can test. This indicates that the agreement was no greater than zero for the above listed analyses, while agreement varied between moderate-to-good for either Dysfunction classification when the Derangement and Spinal categories were eliminated from the analyses. In the case of Articular Dysfunction for the Empty Can test, the agreement was the highest with *Kappa* = 0.84 (SE = 0.19) which indicates a very good agreement.

The low agreement or no agreement with repeated testing of the OSTs in patients with Derangement classification, including spinal Derangements, may be due to the variable and quickly changing nature of the classification especially as it rapidly responds to intervention. Therefore, the presence of Derangement may explain the poor consistency recorded for the majority of the OSTs and was certainly responsible for reducing the overall agreement in the OSTs used in this study. These results would give additional support for the position taken that clinicians should not rely on these OSTs as diagnostic and prognostic tools (Schellingerhout et al., 2008; Lange et al., 2017; Cook, 2010; Beaudreuil et al., 2009). However, there is a clear difference in their consistency in the presence of a Derangement as compared to when Derangements were absent. A rationale could be made for an initial MDT screening of shoulder patients to ensure that shoulder and cervical Derangements have been ruled out before any other testing is performed. This may then enhance the value of the OSTs and perhaps lead to their improved diagnostic capability, if indeed a patho-anatomical diagnosis is still sought.

Alternatively, these OSTs could be used as baseline tests in the differentiation between the MDT classifications of Derangement and Articular and Contractile Dysfunctions. If the OSTs change from positive before, to negative after a repeated movement exam or the initiation of treatment then this would be consistent with a Derangement being present.

The major limitations of this study were as follows: As a "treatment-*as*-usual" approach was followed, a pre-determined number of treatment sessions was not feasible for each one of our patients. Thus, it is possible that the study participants received a variable number of treatment sessions, potentially influencing treatment results. However, the treating physiotherapists were unaware of the study objectives, minimizing any inclination to influence the outcome of each classification category. In addition, a second practitioner, blinded to the patients' MDT classifications administered the OSTs to avoid any potential bias from the treating clinician. A second limitation due to following a

"treatment as usual" approach was that some patients did not have their data available for all four data collection points; therefore, analysis was done on data from three data collection points to avoid weakening power of our analysis. A third limitation of the study was that only three OSTs were evaluated in the study as it was not feasible to include all the numerous OSTs used for shoulder assessment. Therefore, no extrapolations can be made to other OSTs not investigated in the current study. Finally, the MDT method was followed; therefore, the study results may not be generalizable to other methods of practice.

As a next step, future studies could investigate other OSTs utilized for shoulder assessment, and use a pre-set and equal number of treatment sessions for all patients so that data would be available for all data collection points. Due to the presence of a clear pattern in our findings indicating that the Derangement classification could be the reason for inconsistent OST results, further investigations are warranted on the OSTs utilized in the assessment of other musculoskeletal disorders in both spinal and peripheral conditions.

In conclusion, due to the ability of the Derangement classification to rapidly change, it clearly has the capacity to compromise the reliability of OSTs potentially reducing their clinical utility. Thus, being aware of this characteristic of Derangement prior to the use of these shoulder OSTs could assist clinicians in their interpretation of the test results.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authorship statement

Afshin Heidar Abady carried out the literature search and review, study design and planning, data collection, statistical analysis, and interpretation / synthesis of results. He took the lead role in the preparation of the manuscript including the initial draft, coordination of revisions, and submission. Richard Rosedale contributed to the literature search and review, study design and planning, coordination of data collection, interpretation of results, and manuscript review and revision. Tom J Overend and Bert M Chesworth contributed to study design and planning, interpretation of results, and review and revision of the manuscript. Michael A Rotondi supervised and guided the statistical analysis, sample size calculation, interpretation of results and review and revision of the manuscript. All authors read and approved the final article.

Acknowledgement

We thank the following MDT trained physical therapists who volunteered as study collaborators for this study: Susan Bamberger, Dip. MDT; Jane Borgehammar, Dip. MDT; Chris Chase, Dip. MDT; Jonathan Doulton, Dip. MDT; Anja Franz, Dip. MDT; Keith Fernandes, Cert. MDT; Kim Greene, Dip. MDT; Josh Kidd, Dip. MDT; Melissa Kolski, Dip. MDT; Audrey Long, Dip. MDT; Joe Maccio, Dip. MDT; Brian McClenahan, Dip. MDT; John Salituri, Cert. MDT; Allan Sawyer, Cert. MDT; Allicia Wilson, Cert. MDT.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.msksp.2017.10.001.

References

Aina, A., May, S., 2005. A shoulder derangement. Man. Ther. 10, 159–163.
Aina, A., May, S., Clare, H., 2004. The centralization phenomenon of spinal symptoms—a systematic review. Man. Ther. 9, 134–143.
Beaudreuil, J., Nizard, R., Thomas, T., Peyre, M., Liotard, J.P., Boileau, P., Marc, T., Dromard, C., Steyer, E., Bardin, T., Orcel, P., Walch, G., 2009. Contribution of clinical tests to the diagnosis of rotator cuff disease: a systematic review. Jt. Bone Spine 76, 15–19.
Berthelot, J.M., Delecrin, J., Maugars, Y., Passuti, N., 2007. Contribution of centralization phenomenon to the diagnosis, prognosis, and treatment of diskogenic low back pain. Jt. Bone Spine 74, 319–323.
Clare, H.A., 2005. PT, Adams R, Maher CG. Reliability of McKenzie classification of patients with cervical or lumbar pain. J. Manip. Physiol. Ther. 28, 122–127.
Clare, H.A., Adams, R., Maher, C.G., 2004. A systematic review of efficacy of McKenzie therapy for spinal pain. Aust. J. Physiother. 50, 209–216.
Cook, C., 2010. The lost art of the clinical examination: an overemphasis on clinical special tests. J. Man. Manip. Ther. 18 (1), 3–4.
Cook, C.E., Hegedus, E.J., 2008. Orthopedic Physical Examination tests: an Evidence-based Approach. Pearson Prentice Hal, Upper Saddle River, NJ.
Cook, C., Hegedus, E.J., Ramey, K., 2005. Physical therapy exercise intervention based on classification using the patient response method: a systematic review of the literature. J. Man. Manip. Ther. 13, 152–162.
Croft, P., Pope, D., Silman, A., 1996 Sep 7. The clinical course of shoulder pain: prospective cohort study in primary care. Primary Care Rheumatology Society Shoulder Study Group. BMJ 313 (7057), 601–602.
Dessaur, W.A., Magarey, M.E., 2008. Diagnostic accuracy of clinical tests for superior labral anterior posterior lesions: a Systematic Review. J. Orthop. Sports Phys. Ther. 38 (6), 341–352.
Dionne, C.P., Bybee, R.F., Tomaka, J., 2006. Inter-rater reliability of McKenzie assessment in patients with neck pain. Physiotherapy 92, 75–82.
Gadogan, A., Laslett, M., Hing, W., McNair, P., Williams, M., 2011. Interexaminer reliability of orthopaedic special tests used in the assessment of shoulder pain. Man. Ther. 16, 131–135.
Guide to physical therapist practice, 2001. American physical therapy association. second ed.. Phys. Ther. 81, 9–746.
Hegedus, E.J., Goode, A., Campbell, S., Morin, A., Tamaddoni, M., Moorman, C.T., 2008. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. Br. J. Sports Med. 42, 80–92.
Heidar Abady, A., Rosedale, R., Overend, T.J., Chesworth, B.M., Rotondi, M.A., 2014. Inter-examiner reliability of diplomats in the Mechanical Diagnosis and Therapy system in assessing patients with shoulder pain. J. Man. Manip. Ther. 22 (4), 199–205.
Heidar Abady, A., Rosedale, R., Chesworth, B.M., Rotondi, M.A., Overend, T.J., 2017. Application of the McKenzie system of Mechanical Diagnosis and Therapy (MDT) in patients with shoulder pain; a prospective longitudinal study. J. Man. Manip. Ther. http://dx.doi.org/10.1080/10669817.2017.1313929.
Hettinga, D.M., Jackson, A., Klaber Moffett, J., May, S., Mercer, C., Woby, S.R., 2007. A systematic review and synthesis of higher quality evidence of the effectiveness of exercise interventions for nonspecific low back pain of at least 6 weeks duration. Phys. Ther. Rev. 12, 221–232.
Hughes, P., Taylor, N.F., Green, R.A., 2008. Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review. Aust. J. Physiother. 54, 159–170.
Jensen, M.P., Karoly, P., Braver, S., 1986. The measurement of clinical pain intensity: a comparison of six methods. Pain 27 (1), 117–126.
Kaneko, S., Takasaki, H., May, S., 2009. Application of mechanical diagnosis and therapy to a patient diagnosed with de Quervain's disease: a case study. J. Hand Ther. 22, 278–284.
Kelly, E., May, S., Ross, J., 2008. The reliability of the McKenzie classification system using extremity McKenzie assessment forms. Int. J. Mech. Diagn. Ther. 3, 3–6.
Kidd, J., 2013. Treatment of shoulder pain utilizing mechanical diagnosis and therapy principles. J. Man. Manip. Ther. 21 (3), 168–173.
Kilpikoski, S., Airaksinen, O., Kankaanpaa, M., Leminen, P., Videman, T., Alen, M., 2002. Interexaminer reliability of low back pain assessment using the McKenzie method. Spine 27, 207–214.
Koes, B.W., van Tulder, M.W., Thomas, S., 2006. Diagnosis and treatment of low back pain. BMI 332, 1430–1434.
Kuhn, J.E., Dunn, W.R., Ma, B., Wright, R.W., Jones, G., Spencer, E.E., Wolf, B., Safran, M., Spinder, K.P., McCarty, E., Kelly, B., Holloway, B., 2007. Interobserver agreement in the classification of rotator cuff tears. Am. J. Sport Med. 35 (3), 437–441.
Lange, T., Matthijs, O., Jain, N.B., Schmitt, J., Lutzner, J., Kopkow, C., 2017 Mar. Reliability of specific physical examination tests for the diagnosis of shoulder pathologies: a systematic review and meta-analysis. Br. J. Sports Med. 51 (6), 511–518.
Lewis, J.S., 2009. Rotator cuff tendinopathy/subacromial impingement syndrome: is it time for a new method of assessment? Br. J. Sports Med. 43, 259–264.
Littlewood, C., Ashton, J., Chance-Larsen, K., May, S., Sturrock, B., 2012. Exercise for rotator cuff tendinopathy: a systematic review. Physiotherapy 98, 101–109.
Long, A., 1995. The centralisation phenomenon. Its usefulness as a predictor of outcome in conservative treatment of chronic low back pain. Spine 20, 2513–2521.
Long, A., May, S., Fung, T., 2008. The comparative prognostic value of directional preference and centralization: a useful tool for front-line clinicians? J. Man. Manip. Ther. 16, 248–254.
Lynch, G., May, S., 2013. Directional preference at the knee: a case report using mechanical diagnosis and therapy. J. Man. Manip. Ther. 21 (1), 60–66.
Machado, L.A.C., de Souza, MvS., Ferreira, P.H., Ferreira, M.L., 2006. The McKenzie method for low back pain. A systematic review of the literature with a meta-analysis approach. Spine 31, E254–E262.
Machado, L.A., Maher, C.G., Herbert, R.D., Clare, H., McAuley, J.H., 2010. The effectiveness of the McKenzie method in addition to first-line care for acute low back pain: a randomized controlled trial. BMC Med. 8, 10.
Magee, D.J., 2008. Orthopedic Physical Assessment, fifth ed. Saunders Elsevier, St.

Louis, MO.
May, S., 2006. Classification by McKenzie mechanical syndromes; Report on directional preference and extremity patients. Int. J. Mech. Diagn. Ther. 1 (3), 7–11.
May, S., 2009 Sep. The McKenzie classification system in extremities: a reliability study using McKenzie assessment forms and experienced clinicians. J. Manip. Physiol. Ther. 32 (7), 556–563.
May, S., Donelson, R., 2008. Evidence-informed management of chronic low back pain with the McKenzie method. Spine J. 8, 134–141.
May, S., Rosedale, R., 2012. A survey of the McKenzie classification system in the extremities: prevalence of mechanical syndromes and preferred loading strategies. Phys. Ther. 92 (9), 1175–1186.
May, S., Ross, J., 2009. The McKenzie classification system in the extremities: a reliability study using McKenzie assessment forms and experienced clinicians. J. Manip. Physiol. Ther. 32, 556–563.
May, S., Littlewood, C., Bishop, A., 2006. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. Aust. J. Physiother. 52, 91–102.
May, S., Chance-Larsen, K., Littlewood, C., Lomans, D., Saad, M., 2010. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. Physiotherapy 96, 179–190.
McClure, P.W., Michener, L.A., 2015. Staged approach for rehabilitation classification: shoulder disorders (STAR-Shoulder). Phys. Ther. 95, 791–800.
McFarland, E.G., Garzon-Muvdi, J., Jia, X., Desai, P., Petersen, S.A., 2010. Clinical and diagnostic tests for shoulder disorders: a critical review. Br. J. Sports Med. 44 (5), 328–332.
McKenzie, R.A., 1981. The Lumbar Spine. Mechanical Diagnosis and Therapy. Spinal Publications, Waikanae, New Zealand.
McKenzie, R.A., May, S., 2000. The Human Extremities: Mechanical Diagnosis and Therapy. Spinal Publications, Waikanae, New Zealand.
Meserve, B.B., Cleland, J.A., Boucher, T.R., 2009. A meta-analysis examining clinical test utility for assessing superior labral anterior posterior lesions. Am. J. Sport Med. 37 (11), 2252–2258.
Munro, W., Healy, R., 2009. The validity and accuracy of clinical tests used to detect labral pathology of the shoulder - a systematic review. Man. Ther. 14, 119–130.
Nomden, J.G., Slagers, A.J., Bergman, G.J.D., Winters, J.C., Kropmans, T.J.B., Dijkstra, P.U., 2009. Interobserver reliability of physical examination of shoulder girdle. Man. Ther. 14, 152–159.
Petersen, T., Larsen, K., Nordsteen, J., Olsen, S., Fournier, G., Jacobsen, S., 2011. The McKenzie method compared with manipulation when used adjunctive to information and advice in low back pain patients presenting with centralization or peripheralization: a randomized controlled trial. Spine 36 (24), 1999–2010.
Powell, J.W., Huijbregts, P.A., Jensen, R., 2008. Diagnostic utility of clinical tests for SLAP lesions: a systematic literature review. J. Man. Manip. Ther. 16 (5), E58–E79.
Razmjou, H., Kramer, J.F., Yamada, R., 2000. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low back pain. J. Orthop. Sports Phys. Ther. 30, 368–389.
Rosedale, R., Rastogi, R., May, S., Chesworth, B.M., Filice, F., Willis, S., Howard, J., Naudie, D., Robbins, S.M., 2014. Efficacy of exercise intervention as determined by the McKenzie system of Mechanical Diagnosis and Therapy for knee osteoarthritis: a randomized controlled trial. J. Orth. Sports Phys. Ther. 44 (3), 173–181.
Rotondi, M.A., Donner, A., 2012 Jul. A confidence interval approach to sample size estimation for studies of interobserver agreement with multiple raters and outcome. J. Clin. Epidemiol. 65 (7), 778–784.
Schellingerhout, J.M., Verhagen, A.P., Thomas, S., Koes, B.W., 2008. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. Man. Ther. 13, 478–483.
Sciascia, A.D., Spigelman, T., Kibler, W.B., et al., 2012. Frequency of use of clinical shoulder examination tests by experienced shoulder surgeons. J. Athl. Train. 47, 457–466.
Scott, A., Huisman, E., Khan, K., 2011. Conservative treatment of chronic Achilles tendinopathy. Can. Med. Assoc. J. 183, 1159–1165.
Shouta, K., Hiroshi, T., May, S., 2009. Application of Mechanical Diagnosis and Therapy to a patient diagnosed with a de Quervain's disease: a case study. J. Hand Ther. 22, 278–284.
Slade, S.C., Keating, J., 2007. Unloaded movement facilitation exercise compared to no exercise or alternative therapy on outcomes for people with non-specific chronic low back pain: a systematic review. J. Manip. Physiol. Ther. 30, 301–311.
Stratford, P.W., Binkley, J.M., Stratford, D.M., 2001. Development and initial validation of the upper extremity functional index. Physiother. Can. 53 (4), 259–267.
Streiner, D.L., Norman, G.R., 2003. Health Measurement Scales. A Practical Guide to Their Development and Use, thi[rd] ed. Oxford University Press, Oxford.
Sufka, A., Hauger, B., Trenary, M., et al., 1998. Centralisation of low back pain and perceived functional outcome. J. Orthop. Sports Phys. Ther. 27, 205–212.
Takasaki, H., 2016. Agreement of Mechanical Diagnosis and Therapy classification in people with extremity conditions. Phys. Ther. 96 (10), 1525–1553.
Takasaki, H., May, S., 2014. Mechanical diagnosis and therapy has similar effects on pain and disability as 'wait and see' and other approaches in people with neck pain: a systematic review. J. Physiother. 60 (2), 78–84.
Takasaki, H., Okuyama, K., Rosedale, R., 2017. Inter-examiner classification reliability of Mechanical Diagnosis and Therapy for extremity problems - systematic review. Musculoskelet. Sci. Pract. 27, 78–84.
Tucker, S., Taylor, N.F., Green, R.A., 2011. Anatomical validity of the Hawkins-Kennedy test - a pilot study. Man. Ther. 16, 399–402.
Walsworth, M.K., Doukas, W.C., Murphy, K.P., Mielcarek, B.J., Michener, L.A., 2008. Reliability and diagnostic accuracy of history and physical examination for diagnosing glenoid. Am. J. Sport Med. 36 (1), 162–168.
Walton, D.M., Sadi, J., 2008. Identifying SLAP lesions: a meta-analysis of clinical tests and exercise in clinical reasoning. Phys. Ther. Sport 9, 167–176.
Werneke, M., Hart, D.L., 2001. Centralisation phenomenon as a prognostic factor for chronic pain or disability. Spine 26, 758–765.
Werneke, M., Hart, D.L., Cook, D., 1999. A descriptive study of the centralization phenomenon. A prospective analysis. Spine 24, 676–683.
Werneke, M., Hart, D.L., Deutscher, D., Stratford, P.W., 2011a. Clinician's ability to identify neck and low back interventions: an inter-rater chance corrected agreement pilot study. J. Man. Manip. Ther. 19, 172–181.
Werneke, M., Hart, D.L., Cutrone, G., Oliver, D., McGill, M., Weinberg, J., Grisby, D., Oswald, W., Ward, J., 2011b. Association between directional preference and centralization in patients with low back pain. J. Orthop. Sports Phys. Ther. 41 (1), 22–31.
Wetzel, F.T., Donelson, R., 2003. The role of repeated end-range/pain response assessment in the management of symptomatic lumbar discs. Spine J. 3, 146–154.
Willis, S., Rosedale, R., Rastogi, R., Robbins, S., 2017. Inter-rater reliability of the McKenzie system of mechanical diagnosis and therapy in the examination of the knee. J. Man. Manip. Ther. 25 (2), 83–90.
Winters, J.C., Sobel, J.S., van der Windt, D.A.W.M., Jonquiere, M., de Winter, A.F., van der Heijden, G.J., et al., 1999. NHG standaard schouderklachten (versie 1999) (guidelines for shoulder complaints of the Dutch college of general practitioners (version 1999). Huisarts Wet. 42, 222e31.

Afshin Heidar Abady completed his Bachelor of Science and Master of Science programs in physiotherapy with Iran University of Medical Sciences, Tehran, Iran. Afshin is currently pursuing his PhD program in Health and Rehabilitation Sciences at Western University in London, Canada. He is simultaneously pursuing his transitional Doctor of Physical Therapy (t-DPT) program through Andrews University in Michigan, USA. Afshin is certified in McKenzie system of Mechanical Diagnosis and Therapy (MDT) through The McKenzie Institute Canada. He is licensed to practice physiotherapy through the college of physiotherapists of Ontario, Canada.

Dr. Tom Overend is an Associate Professor at Western University. He does research and supervises graduate students in the School of Physical Therapy. Dr. Overend teaches a course in the School of Health Studies and looks after the PT Field Seminar in the Health and Rehabilitation Sciences graduate program. He is also the Acting Chair of the Health and Rehabilitation Sciences graduate program for 2016-17. He was a Director of the School of Physical Therapy from 2009-2014.

Richard Rosedale graduated from Guy's Hospital School of Physiotherapy in 1992 and soon after immigrated to Canada to start work at London Health Sciences Centre. He has over 20 years of experience in orthopaedics and occupational health. After obtaining his Diploma in Mechanical Diagnosis and Therapy in 1997 he became Faculty of the McKenzie Institute in 2003 and is currently an International Instructor. Richard is an MDT Diploma examiner and member of the McKenzie Institute's International Education Committee. He has authored and co-authored numerous publications exploring MDT, including an RCT of MDT for patients with knee OA.

Dr. Bert Chesworth is an Associate Professor at Western University. He does musculoskeletal rehabilitation research and supervises graduate students in the School of Physical Therapy. Dr. Chesworth teaches a course in the Graduate Program in Health and Rehabilitation Sciences and is the Director of the School of Physical Therapy.

Dr. Michael Rotondi is an Associate Professor of Biostatistics at York University. In addition to his methodological work in inter-observer agreement studies and cluster randomized trials, he is an active collaborative health researcher. Current major projects include: Leading a study to develop new statistical methods that will enable examination of risk factors for chronic health conditions in the urban Aboriginal community in Canada; and Co-leading a randomized trial to investigate the effectiveness of an iPhone-based diabetes monitoring system for improving glycemic control in teenagers with type I diabetes.