MDT World Press

LITERATURE REVIEW: GUEST COMMENTARY

When the Foundation is Shaken
Georg Supp, PT, MT, Dip. MDT

Quo Vadis Reliability?
In the preface of Robin McKenzie's book *The Lumbar Spine: Mechanical Diagnosis and Therapy* (2003), Nik Bogduk wrote about MDT: "…studies have sought the evidence for its reliability, validity and efficacy. Its reliability is now beyond doubt…".

The evidence concerning the topic of reliability in regards to MDT has been overwhelmingly positive. In an area where other systems bob up and down in the Stone Age, McKenzie has an impressive track record with convincing evidence. The superiority in terms of reliability is easy to explain. Within MDT, palpation doesn't play any role and inspection just a marginal role. A standardized history taking and repeated movements performed by the patients themselves lead onto classification and management.

Evidence supporting the reliability of MDT has been established for many years (Razmjou et al. 2000, Kilpikoski et al. 2002, Clare et al. 2003, Clare et al. 2005, May and Ross 2009, Abady et al. 2014, May et al. 2006). Since spring 2014, everything seems to be different. The 'MDT tribe' is slowly waking up from a study published by Mark Werneke and colleagues that seemingly unsettles the foundation pillars of the system, and I hope this article will encourage discussion.

The Study

McKenzie Lumbar Classification: Inter-rater Agreement by Physical Therapists with Different Levels of Formal McKenzie Postgraduate Training
Werneke MW, Deutscher D, Hart DL, Stratford P, Ladin Joel, Weinberg J, Herbowy S, Resnik L. Spine (Phila Pa 1976). 2014 Feb 1; 39(3):E182-90. doi: 10.1097/BRS.0000000000000117

Study Design: Inter-rater chance-corrected agreement study.

Objective: Mark Werneke (USA), Daniel Deutscher (Israel) and their team examined whether there is a relationship between the level of formal McKenzie postgraduate training – without the Credentialling Exam (CE) – and reliability of educated clinicians in evaluating back pain patients. The researchers examined reliability concerning the following classification variables:
- main McKenzie syndromes
- presence of lateral shift
- derangement reducibility
- directional preference
- centralization

Summary of Background Data: Minimal level of McKenzie postgraduate training needed to achieve acceptable agreement of McKenzie classification system is unknown.

Methods: 47 Israeli physical therapists with an average professional experience of 14 years completed multiple sets of two independent successive examinations. Therapists conducted history taking and clinical examination in a one-to-one situation. A ten minute break had been scheduled between the two examinations.

The authors gathered data according to training level in three groups: Part A+B, Part C, and Part D. Agreement was assessed with Kappa coefficients and associated 95% confidence intervals. A minimum Kappa threshold of 0.60 was used as a predetermined criterion for level of agreement acceptable for clinical use.

Results: Raters examined 1662 patients in 25 clinics (mean age = 51 ± 15, range 18–91, 57% females, 57 % chronic). Eventually, the researchers were able to analyze data of 1587 patients. Data distributions were not even and were highly skewed for all classification variables. No training level studied had acceptable agreement for any McKenzie classification variable. Agreements for all levels of McKenzie postgraduate training were higher than expected by chance for most of the classification variables except

Parts A + B training level for judging lateral shift and centralization and Part D training level for judging reducibility. Agreement between training levels Parts A+B, Part C, and Part D were similar with overlapping 95% confidence intervals.

An overview of the Kappa values:

<u>Three syndromes and OTHERS:</u>	0.40 (Part A/B), 0.44 (Part C) and 0.37 (Part D)
<u>Lateral Shift:</u>	0.11 (A/B), 0.37 (C) and 0.43 (D)
<u>Reducibility of Derangement:</u>	0.26 (A/B), 0.26 (C) and 0.11 (D)
<u>Directional Preference:</u>	0.39 (A/B), 0.27 (C) and 0.33 (D)
<u>Centralization:</u>	0.11 (A/B), 0.35 (C) and 0.39 (D)

Conclusion: Results indicate that the level of inter-rater chance-corrected agreement of the McKenzie classification system was not acceptable for therapists at any level of formal McKenzie postgraduate training. This finding raises concerns about the clinical utility of the McKenzie classification system at these training levels. Additional studies are needed to assess agreement levels for therapists who receive additional training or experience at the McKenzie credentialed or diploma levels.

Comment
Running a reliability study with more than 1600 patients is an impressive task. Mark Werneke, Daniel Deutscher and their team deserve the greatest respect for this accomplishment. I'm not aware of any reliability study that has used such high numbers. The design of former studies did not follow recommended blinded assessment and number of raters and patient guidelines for reporting reliability. But what are the consequences of these results on the McKenzie system and its users? There may be a short answer favoring a killer argument, and there are multiple, slightly longer, but helpful answers.

The Short Answer:

Qualification of therapists
At the end of a driving school, after having taken all the required lessons, a driving test has to be attended. Only those who pass the test fulfill a minimal standard and receive a driver license. During the driving lessons, it may be obvious that some student drivers are fast learners, are able to operate confidently, and rapidly become trustworthy. Other student drivers need more time and more instruction. A successful driving test certifies who is allowed to drive a car on his own. No matter how the individual driving lessons went, the exam ensures the quality. This can be compared quite logically with MDT training. Not until the CE have participants proven their competency. For the training from Part A to CE, it's not the principle of 'addition' but the principle of 'multiplication' that applies. If one factor is zero, the result is consequently also zero – at least in terms of the completeness of the training. Of course, there are some participants who are willingly adjusting their clinical reasoning according to MDT as they work through the courses. However, crucially there are more clinicians who are considering MDT just as a "tool in their PT toolbox"; they are mixing their new knowledge with their old, combining approaches and systems. When someone passes the CE, the probability will increase that he will value MDT not as a tool, but as the toolbox itself.

For that reason, former studies on prevalence (Hefford 2008, May and Aina 2012), reliability (Razmjou et al. 2000, Kilpikoski et al. 2002, Clare et al. 2003, Clare et al. 2005, May and Ross 2009, Abady et al. 2014) and effectiveness (Long et al. 2004, Rosedale et al. 2014, Machado et al. 2010) did well by using the CE or the Diploma in MDT as a minimal standard for the inclusion of clinicians. Researchers who utilized therapists without this minimal level got comparably poor results as Werneke and Deutscher (Riddle and Rothstein 1993, Kilby et al. 1990). The authors of the current study wisely considered this aspect when coming to their conclusion. Their doubts about the utility of the McKenzie classification are referring to the examined training levels and not to the system in general.

Whether a driving school generates good drivers should be judged by examining the successful examinees and not by the student drivers during their first driving lessons.

The Longer Answers:

The study represents daily clinical practice
The results are hardly surprising. They mirror very well daily clinical practice:

Classification

One strength of the McKenzie system is the immediate and consequent involvement of the patient – not only in terms of management but also concerning diagnosis. The provisional classification on day one is only a snapshot. Very often, it is only after a trial of initial treatments over the first few sessions that the clinician is able to be sure about the classification and management. Classification by multiple sessions has shown superiority regarding final diagnosis and prognosis compared to a single session based classification (Werneke and Hart 2003). This accounts to a special degree for chronic patients, which account for the majority in Werneke's study. The fact that clinicians categorize differently in the first session would only be clinically relevant and disappointing if they would not be able to adjust their classification due to the clinical presentation at certain points in time. Audrey Long has shown that a change of the treatment strategy according to the responses on trial treatments will lead to a significant improvement of long term results (Long et al. 2008).

Lateral Shift

Reliable visual detection of a lateral shift had been revealed as wishful thinking by many studies before (Donahue et al. 1996, Clare et al. 2003, Seymour et al. 2002, Kilpikoski et al. 2002). To accept this as a fact should be even easier after this paper.

Directional Preference (DP)

The same thoughts as for classification apply here. Of course, it is desirable to find the particular DP on day one but in everyday clinical practice it commonly needs one or multiple trial treatments to achieve this.

Reducibility

Fortunately, future generations of MDT clinicians will not have to struggle with the oxymoron "Irreducible Derangement". Since summer 2014, there exists only "Reducible Derangements" or just "Derangements". The McKenzie Institute International had adjusted and updated the classification due to the current "state of the clinical art". The category 'OTHER' now contains subcategories for the discontinued model "Irreducible Derangement".

Raters in the Israeli study did classify 83.5 % of all back pain patients as Derangements with 93 % of them as "reducible". This aspect should raise concerns as the available evidence doesn't support such a high prevalence of Derangements (May and Aina 2012). A certain overemphasis of Derangement and Directional Preference in MDT training may play an important role here.

Reliability: statistical versus clinical?

Reliability is usually appreciated as the foundation pillar of a system for diagnosis and management. But reliability is a dull sword if the classification system is not valid. The subgroup of stabilization within the Treatment Based Classification (Delitto et al. 1995) is a nice example for this disproportionateness. Tests for identification of this subgroup may be reliable (Rabin et al. 2013) but the validity of the Clinical Prediction Rule for stabilization is questionable (Rabin et al. 2014), the subgroup shows substantial overlap with other subgroups (Werneke et al. 2010, Apeldoorn 2014) and its ultimate impact on outcomes is in dispute (Henry et al. 2014).

In clinical research, the term reliability is commonly used in the context of classification and refers to particular variables which are determined in the initial evaluation – let's call this "statistical reliability". BUT in daily clinical practice, what's happening over time, during the course of treatment, is more important. How reliably do clinicians handle their patients? Do they draw the same conclusions from symptomatic, mechanical and functional responses? It seems that there also exists "clinical reliability".

Beyond reliability: Why does McKenzie work?

Reliable classification within the initial evaluation is ONE thing; a successful course of treatment is another one. The recent publication of Daniel Deutscher is showing this quite impressively (Deutscher et al. 2014). After reliability, Deutscher investigated the efficiency of McKenzie trained physical therapists in terms of functional status and number of required treatment sessions. He analyzed data of more than 20,000 back pain patients. Results: McKenzie trained therapists work more cost effectively and achieved better functional outcomes than therapists without MDT training. But only marginal differences between training levels became apparent. Therapists who had attended a Part C needed significantly less treatment sessions than clinicians that attended only Parts A and B.

So, why does McKenzie work?
Directional Preference and Centralization are effective guidelines for the treatment of LBP patients (Surkitt et al. 2012, Long et al. 2004, 2008) but finally mechanical classification and directional oriented treatment represents only a cutout of MDT management. MDT emphasizes self-responsibility of patients. MDT trained clinicians provide their patients with understandable and empowering movement and posture strategies instead of intimidating them with patho-anatomic explanations. A clearly structured, active management approach functions in the framework of a biopsychosocial thought model (Takasaki et al. 2014). The mechanical examination using repeated movements should be also appreciated as an effective entry point to a fear releasing, movement oriented management. It's not only a standard procedure for the definition of a diagnostic category. Therapeutic alliance instead of sole instructions for exercise programs has a better chance of success (Ferreira et al. 2013).

Should all this have an impact on the MDT education?
Besides all thoughts on the strength of MDT management and the reflection on "clinical reliability" of therapists, achieving better "statistical reliability" should be on the topic list for future MDT training programs. The current MDT educational program is outstanding, involving the participants and promoting an interactive learning experience.

There may be some potential in using more role plays instead of patient vignettes and forcing the participants to go actively through the evaluation process instead of analyzing a provided case study. Patient demonstrations on MDT courses are an integral part of the training but watching an experienced clinician examining a patient can't replace the learning process of conducting a history taking and clinical examination on your own.

My heartfelt thanks to Mark and Daniel and the great Dennis Hart who passed before the study had been published. They rouse the McKenzie people with their brilliant studies. It is movement in the right direction – very much in line with Robin McKenzie.

References
Abady, Afshin Heidar; Rosedale, Richard; Overend, Tom J.; Chesworth, Bert M.; Rotondi, Michael A. (2014): Inter-examiner reliability of diplomats in the mechanical diagnosis and therapy system in assessing patients with shoulder pain. In: *The Journal of manual & manipulative therapy* 22 (4), S. 199–205. DOI: 10.1179/2042618614Y.0000000068.

Apeldoorn, Adri (2014): Centralization and clinical signs of lumbar instability, an observational study. McKenzie Institut D / CH / A und FOMT. 4. Deutsches Symposium für Mechanische und Manuelle Therapie, Stuttgart, 25.01.2014.

Clare, Helen A.; Adams, Roger; Maher, Christopher G. (2003): Reliability of detection of lumbar lateral shift. In: *Journal of manipulative and physiological therapeutics* 26 (8), S. 476–480. DOI: 10.1016/S0161-4754(03)00104-0.

Clare, Helen A.; Adams, Roger; Maher, Christopher G. (2005): Reliability of McKenzie classification of patients with cervical or lumbar pain. In: *Journal of manipulative and physiological therapeutics* 28 (2), S. 122–127. DOI: 10.1016/j.jmpt.2005.01.003.

Delitto, A.; Erhard, R. E.; Bowling, R. W. (1995): A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. In: *Physical therapy* 75 (6), S. 470-85; discussion 485-9.

Deutscher, Daniel; Werneke, Mark W.; Gottlieb, Ditza; Fritz, Julie M.; Resnik, Linda (2014): Physical Therapists' Level of McKenzie Education, Functional Outcomes, and Utilization in Patients With Low Back Pain. In: *The Journal of orthopaedic and sports physical therapy* 44 (12), S. 925–936. DOI: 10.2519/jospt.2014.5272.

Donahue, M. S.; Riddle, D. L.; Sullivan, M. S. (1996): Intertester reliability of a modified version of McKenzie's lateral shift assessments obtained on patients with low back pain. In: *Physical therapy* 76 (7), S. 706-16; discussion 717-26.

Ferreira, Paulo H.; Ferreira, Manuela L.; Maher, Christopher G.; Refshauge, Kathryn M.; Latimer, Jane; Adams, Roger D. (2013): The therapeutic alliance between clinicians and patients predicts outcome in chronic low back pain. In: *Physical therapy* 93 (4), S. 470–478. DOI: 10.2522/ptj.20120137.

Hefford, Cheryl (2008): McKenzie classification of mechanical spinal pain: profile of syndromes and directions of preference. In: *Manual therapy* 13 (1), S. 75–81. DOI: 10.1016/j.math.2006.08.005.

Henry, Sharon M.; Van Dillen, Linda R; Ouellette-Morton, Rebecca H.; Hitt, Juvena R.; Lomond, Karen V.; DeSarno, Michael J.; Bunn, Janice Y. (2014): Outcomes are not different for patient-matched versus nonmatched treatment in subjects with chronic recurrent low back pain: a randomized clinical trial. In: *The spine journal : official journal of the North American Spine Society* 14 (12), S. 2799–2810. DOI: 10.1016/j.spinee.2014.03.024.

Kilby, Julia; Stigant, Mark; Roberts, Andrew (1990): The Reliability of Back Pain Assessment by Physiotherapists, Using a 'McKenzie Algorithm'. In: *Physiotherapy* 76 (9), S. 579–583. DOI: 10.1016/S0031-9406(10)63053-2.

Kilpikoski, Sinikka; Airaksinen, Olavi; Kankaanpää, Markku; Leminen, Päivi; Videman, Tapio; Alen, Markku (2002): Interexaminer reliability of low back pain assessment using the McKenzie method. In: *Spine* 27 (8), S. E207-14.

Long, Audrey; Donelson, Ron; Fung, Tak (2004): Does it matter which exercise? A randomized control trial of exercise for low back pain. In: *Spine* 29 (23), S. 2593–2602.

Long, Audrey; May, Stephen; Fung, Tak (2008): Specific directional exercises for patients with low back pain: a case series. In: *Physiotherapy Canada. Physiothérapie Canada* 60 (4), S. 307–317. DOI: 10.3138/physio.60.4.307.

Machado, Luciana A C; Maher, Chris G.; Herbert, Rob D.; Clare, Helen; McAuley, James H. (2010): The effectiveness of the McKenzie method in addition to first-line care for acute low back pain: a randomized controlled trial. In: *BMC medicine* 8, S. 10. DOI: 10.1186/1741-7015-8-10.

May, Stephen; Aina, Alessandro (2012): Centralization and directional preference: a systematic review. In: *Manual therapy* 17 (6), S. 497–506. DOI: 10.1016/j.math.2012.05.003.

May, Stephen; Littlewood, Chris; Bishop, Annette (2006): Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. In: *The Australian journal of physiotherapy* 52 (2), S. 91–102.

May, Stephen; Ross, Jenny (2009): The McKenzie classification system in the extremities: a reliability study using Mckenzie assessment forms and experienced clinicians. In: *Journal of manipulative and physiological therapeutics* 32 (7), S. 556–563. DOI: 10.1016/j.jmpt.2009.08.007.

Rabin, Alon; Shashua, Anat; Pizem, Koby; Dar, Gali (2013): The interrater reliability of physical examination tests that may predict the outcome or suggest the need for lumbar stabilization exercises. In: *The Journal of orthopaedic and sports physical therapy* 43 (2), S. 83–90. DOI: 10.2519/jospt.2013.4310.

Rabin, Alon; Shashua, Anat; Pizem, Koby; Dickstein, Ruthy; Dar, Gali (2014): A clinical prediction rule to identify patients with low back pain who are likely to experience short-term success following lumbar stabilization exercises: a randomized controlled validation study. In: *The Journal of orthopaedic and sports physical therapy* 44 (1), S. 6-B13. DOI: 10.2519/jospt.2014.4888.

Razmjou, H.; Kramer, J. F.; Yamada, R. (2000): Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. In: *The Journal of orthopaedic and sports physical therapy* 30 (7), S. 368-83; discussion 384-9. DOI: 10.2519/jospt.2000.30.7.368.

Riddle, D. L.; Rothstein, J. M. (1993): Intertester reliability of McKenzie's classifications of the syndrome types present in patients with low back pain. In: *Spine* 18 (10), S. 1333–1344.

Rosedale, Richard; Rastogi, Ravi; May, Stephen; Chesworth, Bert M.; Filice, Frank; Willis, Sean et al. (2014): Efficacy of exercise intervention as determined by the McKenzie System of Mechanical Diagnosis and Therapy for knee osteoarthritis: a randomized controlled trial. In: *The Journal of orthopaedic and sports physical therapy* 44 (3), S. 173-81, A1-6. DOI: 10.2519/jospt.2014.4791.

Seymour, Ron; Walsh, Thom; Blankenberg, Chris; Pickens, Anthony; Rush, Heath (2002): Reliability of Detecting a Relevant Lateral Shift in Patients with Lumbar Derangement: A Pilot Study. In: *Journal of Manual & Manipulative Therapy* 10 (3), S. 129–135. DOI: 10.1179/106698102790819193.

Surkitt, Luke D.; Ford, Jon J.; Hahne, Andrew J.; Pizzari, Tania; McMeeken, Joan M. (2012): Efficacy of directional preference management for low back pain: a systematic review. In: *Physical therapy* 92 (5), S. 652–665. DOI: 10.2522/ptj.20100251.

Takasaki, Hiroshi; Saiki, Takeshi; Iwasada, Yoshihiro (2014): McKenzie Therapists Adhere More to Evidence-Based Guidelines and Have a More Biopsychosocial Perspective on the Management of Patients with Low Back Pain than General Physical Therapists in Japan. In: *OJTR* 02 (04), S. 173–181. DOI: 10.4236/ojtr.2014.24023.

Werneke, Mark; Hart, Dennis L. (2003): Discriminant validity and relative precision for classifying patients with nonspecific neck and back pain by anatomic pain patterns. In: *Spine* 28 (2), S. 161–166. DOI: 10.1097/01.BRS.0000041580.02041.37.

Werneke, Mark W.; Hart, Dennis; Oliver, Dave; McGill, Troy; Grigsby, David; Ward, Jason et al. (2010): Prevalence of classification methods for patients with lumbar impairments using the McKenzie syndromes, pain pattern, manipulation, and stabilization clinical prediction rules. In: *The Journal of Manual & Manipulative Therapy* 18 (4), S. 197–204. DOI: 10.1179/10669811