Disentangling classification systems from their individual categories and the category-specific criteria: an essential consideration to evaluate clinical utility

Julie Fritz

College of Health, The University of Utah, USA

Invited Commentary on: 'Prevalence of classification methods for patients with lumbar impairments using the McKenzie syndromes, pain pattern, manipulation, and stabilization clinical prediction rules', Werneke et al.

The classification of patients with low back pain (LBP) into subgroups for the purpose of directing treatment decision-making has been a particular focus among physical therapists for several decades. The study by Werneke and colleagues examines the mechanical diagnosis and therapy (MDT) classification paradigm popularized by McKenzie,¹ and aspects of the treatment-based classification (TBC) paradigm, originally described by Delitto and others.² These two paradigms have likely been researched more extensively than other physical therapy-based classification strategies for patients with LBP, and are more often applied in clinical practice as well.³ Evaluating the translation of research findings related to these classification paradigms into routine clinical care is an important consideration for understanding the potential of classification-based decision-making strategies towards improving the outcomes of physical therapy practice.

Numerous criteria have been described for examining classification systems and their ability to favorably impact clinical practice, including feasibility, generalizability, and repeatability.⁴⁻⁶ Essential considerations for evaluating the clinical utility of classification systems are discrimination, comprehensiveness, and mutual exclusivity. These principles essentially require that a classification system be able to discriminate between important categories of patients, and be able to place every individual patient into one, and only one, of these categories. A classification system that cannot identify meaningful difference between

patients, permits a sizable portion of patients to be unclassified, or places an individual patient into multiple categories concurrently will be of limited use in assisting clinical decision-making. Although clinicians and researchers have recognized how critical the principles of discrimination, comprehensiveness, and mutual exclusivity are, there is clearly a need for more empirical data evaluating these characteristics for the MDT or TBC paradigms. This study begins to examine these important issues in a large sample of patients drawn from clinical practice. As such the study provides an opportunity to begin a discussion on issues related to the clinical utility of popular classification paradigms.

A comprehensive classification paradigm, such as the MDT or TBC system, is composed of several categories or subgroups, each with a corresponding set of clinical criteria that defines membership within the category, and each criterion is operationally-defined with respect to performance and interpretation. The MDT classification paradigm encompasses three primary categories (derangement, dysfunction, and posture); while the TBC paradigm describes four primary categories (manipulation, specific exercise, stabilization, and traction). Both the MDT and TBC systems include the clinical finding of centralization as an important criterion for assigning an individual patient to a category. The study by Werneke and colleagues examines the prevalence of the overall MDT classification paradigm, two categories from the TBC system, and one clinical criterion related to both — centralization. Evaluating the prevalence results reported in this study, and their potential clinical implications, requires consideration of the level of classification being examined (overall paradigm, individual category or specific criterion).

The principles of discrimination, comprehensiveness, and mutual exclusivity apply at the level of the overall paradigm, and are therefore important considerations

Correspondence to: Julie Fritz, College of Health, The University of Utah, Salt Lake City, UT, USA. Email: julie.fritz@hsc.utah.edu

for examining the prevalence results for the McKenzie syndromes reported by Werneke. Over a quarter (27%) of the patients in the current study could not be categorized within the MDT system, and somewhat higher percentage than other studies.⁷⁻⁹ Of those patients classifiable within the MDT system in the current study, the vast majority (92%) were in the derangement category, similar to other reports in the literature.⁷⁻¹⁰ Considering the results from Werneke along with the preponderance of evidence in the literature, the primary concern with respect to the MDT classification paradigm may be one of discrimination. If a classification system places nearly all patients into a single category, the efficiency, i.e. the value added relative to the amount of time required, of the classification paradigm in clinical practice, may be questionable.

The centralization phenomenon was originally described as one criterion used to discriminate patients into categories within the MDT classification system.¹¹ Increasing evidence for the prognostic relevance of centralization¹² has focused attention on this single examination finding as a possible classification paradigm unto itself.¹³ As Werneke and colleagues discuss, there is a need for additional research to determine if a finding of centralization (or non-centralization) should be conceptualized as a paradigm for classification, or if this finding should represent one criterion within a larger paradigm. If it is the latter, and centralization is most effectively used as a criterion within a classification system, it is also clear that more research is needed to evaluate how centralization should be prioritized relative to other criteria. What is apparent from the evidence to date is that centralization is an important clinical finding that is relevant to treatment decision-making. It does not seem to fit as a system of classification on its own based on the research currently available. While there is evidence to support a specific treatment approach for patients who do centralize,¹⁴ it is unclear that there is a single approach appropriate for all patients who do not centralize, suggesting a need for further discrimination based on additional criteria for these patients.

With respect to the TBC paradigm, Werneke and colleagues choose to examine two categories from within the larger system — manipulation and stabilization, and elect to use the criteria defined by studies evaluating clinical prediction rules to define membership in these categories. Based on these criteria, 13 and 7% of the patients in this study fit the manipulation and stabilization categories respectively. These values are somewhat lower than those reported in previous studies that have reported the prevalence of these classifications using the criteria defined in the prediction rule studies. In samples

drawn from physical therapy practice, we found that 28% of patients fit the prediction rule criteria for the manipulation category,¹⁵ and in a separate sample also found that 28% fit the prediction rule criteria for the stabilization classification.¹⁶ Another sample drawn from a primary care physician setting reported that 59% of patients fit the prediction rule criteria for the manipulation category,¹⁷ highlighting the importance of practice setting relative to the prevalence of the manipulation category.

The relatively low prevalence of patients fitting the prediction rule criteria for these categories also points to the importance of understanding the methodology used to develop a prediction rule, and how the findings from these prediction rule studies are incorporated into the comprehensive TBC paradigm. The prediction rule studies referenced by Werneke and colleagues used similar methodology, which sought to maximize the positive likelihood ratio for predicting a successful treatment outcome.^{16,18} Statistically, this approach will minimize the possibility of false-positive classification decision, resulting in a set of criteria that more narrowly define those patients who are very highly likely to succeed with the treatment being studied. In other words, the prediction rule criteria developed in these studies were intentionally designed to be narrow — identifying the patients who appear to definitely need the treatment being considered. It is possible, and indeed likely, that additional patients also need the treatment being considered, but the prediction rule criteria identify those patients for whom the opportunity to achieve success with the treatment under consideration should not be missed. We could have taken a different methodological approach to defining prediction rule criteria by seeking to minimize the negative likelihood ratio for predicting a successful outcome. This would have resulted in capturing a larger percentage of all patients, but would lead to more false-negative classifications (i.e. patients who appear to need the treatment but do not have a successful outcome). When the methodology used to develop clinical prediction rule criteria is considered, the low prevalence rates are not surprising.

The large percentage of patients left uncategorized by the prediction rule criteria would clearly violate the need for comprehensiveness in a classification system, but this criticism confuses the distinction between criteria for an individual category and an overall classification paradigm. We use the clinical criteria identified in these prediction rule studies to inform a comprehensive classification decision-making algorithm. A comprehensive algorithm is essential for dealing with the issues of comprehensiveness and mutual exclusivity when translating the TBC paradigm into clinical practice. The algorithm we have

Figure 1 TBC algorithm²³ (reproduced with kind permission from the *Journal of Orthopedic Sports Physical Therapy*).

used to in previous translational research studies as well as ongoing quality improvement efforts as pictured in Fig. 1 (note that the traction subgroup is removed because these patients were excluded from the particular studies). The algorithm is designed to incorporate the findings from clinical prediction rule studies, while managing the inevitable problem of comprehensiveness by requiring that all patients appropriate to the algorithm ultimately be placed into the category that best fits their clinical presentation. In clinical research studies that have used this algorithm, the prevalence of a manipulation categorization has ranged from 23 to 45%, and stabilization 24 to 41%.^{19–21}

Werneke and colleagues also discuss the important issue of mutual exclusivity, recognizing that a sizable

proportion of patients will inevitably demonstrate clinical criteria suggestive of more than one category within a classification system. Given the complexity of LBP and the diversity of clinical presentations, the issue of exclusivity is unlikely to be overcome by defining sufficiently narrow categories. This approach would probably result in categories so numerous as to be completely unfeasible for clinical practice. Instead, the problem of mutual exclusivity likely requires a better understanding of which clinical criteria take precedence over others. Werneke and colleagues highlight an example of this issue with respect to the overlap between centralization and prediction rule criteria for either manipulation or stabilization in many patients. I disagree with Werneke's contention that the TBC algorithm in some way implies that this

overlap would not be possible. This criticism appears to ignore the fact that the hierarchical nature of the classification algorithm was developed specifically in recognition of the reality that overlap does exist. If there were no overlaps among the various criteria, a hierarchical algorithm would be unnecessary. The TBC algorithm prioritizes a clinical finding of clear centralization as the most important criterion, placing a patient into a category labeled specific exercise within the TBC system. This patient may also have clinical criteria indicative of other categories such as manipulation or stabilization, but would be placed into the specific exercise category. The algorithm then prioritizes the criteria related to manipulation classification followed by stabilization. It is important to note that these prioritization decisions are based largely on clinical judgment, and have not been evaluated with appropriate research studies. Finally, if a patient fails to completely demonstrate any of the criteria outlined in the first part of the algorithm, the therapist is led to select the most appropriate category to avoid unclassified patients. I would add again that although we have some evidence that physical therapy management based on this TBC algorithm results in better clinical outcomes than physical therapy that is not based on this decision-making paradigm,^{19,22} there is a continual need for more research towards improving decision-making and maximizing the benefits to patients.

I agree with Werneke and his colleagues in advocating the need for more research into critical topics related to the classification of patients with LBP. This study points out several specific areas for research. In particular I agree that the issue of patients who present with overlap between clinical criteria and questions of treatment prioritization, is an area that needs to be addressed, as it is critical for translating research findings into clinical practice. Issues of comprehensiveness, mutual exclusivity, and discrimination can often be removed through the methodology of controlled research studies, but are part of the reality of clinical practice. Finally, it is important to remember that the ultimate goal of any classification paradigm is to improve the efficiency and effectiveness of clinical care provided to patients, and it is on these criteria that usefulness will be determined.

References

- McKenzie RA, May S. The lumbar spine: mechanical diagnosis and therapy. 2nd ed. Minneapolis, MN: Orthopedic Physical Therapy Products; 2003.
- Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative management. *Phys Ther* 1995;75:470–89.
- Spoto MM, Collins J. Physiotherapy diagnosis in clinical practice: a survey of orthopaedic certified specialists in the USA. *Physiother Res Int* 2008;13:31–41.
- Billis EV, McCarthy CJ, Oldham JA. Subclassification of low back pain: a cross-country comparison. *Eur Spine J* 2007;16:865–79.
- Buchbinder R, Goel V, Bombardier C, Hogg-Johnson S. Classification systems of soft tissue disorders of the neck and upper limb: do they satisfy methodological guidelines? *J Clin Epidemiol* 1996;49:141–9.
- Riddle DL. Classification and low back pain: a review of the literature and critical analysis of selected systems. *Phys Ther* 1998;78:708–37.
- Hefford C. McKenzie classification of mechanical spinal pain: profile of syndromes and directions of preference. *Man Ther* 2008;13:75–81.
- Kilpikoski S, Airaksinen O, Kankaanpää M, Leminen P, Videman T, Alen M. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine (Phila Pa 1976)* 2002;27:E207–14.
- Razmjou H, Kramer JF, Yamada R. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. *J Orthop Sports Phys Ther* 2000;30:368–83.
- May S. Classification by McKenzie mechanical syndromes: a survey of McKenzie-trained faculty. *J Manip Physiol Ther* 2006;29:637–42.
- McKenzie RA. The lumbar spine: mechanical diagnosis and therapy. Waikanae: Spinal Publications New Zealand Ltd; 1981.
- Aina A, May S, Clare H. The centralization phenomenon of spinal symptoms — a systematic review. *Man Ther* 2004;9:134–43.
- Werneke MW, Hart DL, Resnik L, Stratford PW, Reyes A. Centralization: prevalence and effect on treatment outcomes using a standardized operational definition and measurement method. *J Orthop Sports Phys Ther* 2008;38:116–25.
- Browder DA, Childs JD, Cleland JA, Fritz JM. Effectiveness of an extension oriented treatment approach in a subgroup of patients with low back pain: a randomized clinical trial. *Phys Ther* 2007;87:1608–18.
- Fritz JM, Childs JD, Flynn TW. Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain likely to respond quickly to spinal manipulation. *BMC Fam Pract* 2005;6:29.
- Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil* 2005;86:1753–62.
- Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. *Eur Spine J* 2008;17:936–43.
- Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short term improvement with spinal manipulation. *Spine (Phila Pa 1976)* 2002;27:2835–43.
- Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying sub-groups of patients with 'non-specific' low back pain: results of a randomized clinical trial. *Spine (Phila Pa 1976)* 2006;31:623–31.
- George SZ, Bialosky JE, Donald DA. The centralization phenomenon and fear-avoidance beliefs as prognostic factors for acute low back pain: a preliminary investigation involving patients classified for specific exercise. *J Orthop Sports Phys Ther* 2005;35:580–8.
- George SZ, Zeppieri G, Jr, Cere AL, Cere MR, Borut MS, Hodges MJ, et al. A randomized trial of behavioral physical therapy interventions for acute and sub-acute low back pain (NCT00373867). *Pain* 2008;140:145–57.
- Fritz JM, Delitto A, Erhard RE. Comparison of a classification-based approach to physical therapy and therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine (Phila Pa 1976)* 2003;28:1363–72.
- Fritz JM, Cleland JA, Childs JD. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *J Orthop Sports Phys Ther* 2007;37:290–302.