# Orthopaedic special tests and diagnostic accuracy studies: house wine served in very cheap containers

Eric J Hegedus,[1] Alexis A Wright,[1] Chad Cook[2]

House wine is the abstruse alcoholic drink sold by restaurants to thirsty, unquestioning individuals who are looking for the path of least resistance. House wine is easily accessible, simplifies an oft-complex process and 'feels' like its highbrow cousins. House wine is alluring to sellers of wine because it appeals to the masses and is profitable. House wine meets a need; but its packaging (usually a box) is generally looked down upon by discerning consumers since it signals lack of quality.

> Unfortunately, orthopedic special tests are the house wine of the research community and diagnostic accuracy studies are the cheap containers in which they are served.

Before you object too strenuously, we submit the following case (or box, sticking with the theme). Like cheap wine, orthopaedic special tests are ubiquitous. Our quick search of PubMed using 'diagnostic accuracy and orthopaedic tests' revealed 539 articles (most of them relevant). Beyond primary literature, a Google search for orthopaedic special tests elicited 967 000 results including textbooks and websites. Textbooks frequently introduce a litany of tests; often including similar tests with variable names and no discriminating properties. YouTube is laden with demonstrations of selected tests. This popularity begs the question, 'Why?'

For the sellers of the 'house wine', for example, researchers and journals, orthopaedic tests and diagnostic accuracy studies have mass appeal. This mass appeal results in higher visibility and revenue. For researchers and journals, capital comes in the form of publications that add to eminence (ie, eponymous (self-named) tests) and citations that improve impact factor. We know first-hand of the popularity of this topic, as our own systematic reviews of orthopaedic special tests are frequently cited.[1 2] Lastly, just as cheap house wine lurks in the most distinguished establishments, there is a track record of biased diagnostic accuracy publications in many high-impact journals.[3]

## WHAT IS THE ALLURE OF SPECIAL TESTS AND DIAGNOSTIC ACCURACY STUDIES?

What is the allure of special tests and diagnostic accuracy studies for the clinician? We suggest a one word answer: simplicity. First, these tests appear to simplify clinical diagnosis. Much as the average wine drinker can eschew the five-page wine list and ask for the house red, many clinicians opt for the easy decision that a dichotomous test result provides. Diagnosis, a complex, dynamic, iterative process with multiple interacting variables, is converted into a simplified 'yes/no' decision by special tests.

Special tests appear to make evidence-based practice easy. The statistics governing diagnostic accuracy are presented in a 2×2 table and are simple to calculate. Further, because orthopaedic special tests are so endemic in the literature, the practising clinician assumes that when a new test with great metrics is published, the test is evidence based and can be transferred easily into practice. To continue with our metaphor, despite their provenance, orthopaedic special tests taste just fine to the consumer.

## WHY BUST THE MYTH?

So why complain when everyone seems happy? We feel people are only happy because they do not understand the nuances that make a diagnostic accuracy test or a quality wine, truly good. Critical analysis reveals that although many eponymous tests appear magical in the hands of the inventors and their followers, they prove as diagnostic as a coin flip when examined by independent groups. These tests routinely become embedded into clinical practice and are preserved by continued study in a vain hope that somehow, someday, the test will show its 'true' worth, will somehow become palatable.

In addition to the wine (special test) being bad, the container is often worse. Diagnostic accuracy studies are mostly of poor

[1]Department of Physical Therapy, High Point University, High Point, North Carolina, USA
[2]Doctor of Physical Therapy Division, Duke University, Durham, North Carolina, USA

**Correspondence to** Dr Eric J Hegedus, Department of Physical Therapy, High Point University, One University Parkway, High Point, NC 27268, USA; ehegedus@highpoint.edu

quality/heavily influenced by bias.[4 5] These studies are commonly retrospective and, we would argue, possess a questionable gold standard.[6 7] If the gold standard is faulty, then any comparison to it as the 'true' diagnosis will also be faulty. Surgery is often the gold standard but very often surgical and imaging results fail to correlate with the clinical presentation, especially when the term 'syndrome' is part of the diagnostic label.[1]

We emphasise that diagnostic studies often examine tests in an unrealistic way. As an example, case–control studies[8 9] will compare diagnostic ability of a rotator cuff tear test in a group of patients with shoulder pain versus a group of patients with low back pain, and the results are all recorded as 'yes/no' completely ignoring the patient response of 'sort of'. Of course, the test performs beautifully in this scenario but is useless in clinical practice where we are seeing patients with shoulder pain and attempting to differentiate among the many competing causes of that shoulder pain while responding to a patient's non-committal response to a pain-based test.

## TAKE HOME MESSAGE AND CALL TO ACTION

So what should researchers and clinicians to do? Researchers can follow published quality scales[7 10] during the study design process that act as guidelines to produce better manuscripts. This parallels knowing what soil, amount of rain and elevation contribute to the best grape and wine.

Clinicians should quit looking for overly simplistic answers. Clinical diagnosis, like producing a great wine, is complex and requires an appreciation of the data that can be gathered within the nuances of patient interaction. Like a good wine connoisseur who understands what varietal matches each selected food, the clinician can refine his or her examination by using meaningful tests and measures that may serve a variety of purposes. These tests may not be the traditional 'special tests/magic bullets' that allow a short cut towards diagnosis; instead, they may be components of the physical examination that provide context to the underlying problem at hand (eg, lack of strength, lack of mobility or lack of integrity) or may be combined with valuable information from epidemiology and patient history.

Researchers should also study special tests in this context, a context that examines more complex clinical reasoning in favour of reducing this fine wine to the vinegar that most special tests represent. The best clinicians use fewer tests and make decisions on refined data. Knowledgeable wine drinkers know that large volumes at cheap prices are generally not the way to a great experience. As clinicians and researchers, we feel that clinicians have long been served an inferior product in a cardboard container. It is time to order top shelf! That product exists and our community has matured.

**Contributors** Drs CC and EJH collaborated on the original manuscript and Dr AAW edited the manuscript.

**Competing interests** None declared.

**Provenance and peer review** Not commissioned; internally peer reviewed.



**To cite** Hegedus EJ, Wright AA, Cook C. *Br J Sports Med* 2017;**51**:1578–1579.

Revised 12 April 2017
Accepted 17 April 2017
Published Online First 4 May 2017

*Br J Sports Med* 2017;**51**:1578–1579.
doi:10.1136/bjsports-2017-097633

## REFERENCES

1 Cook C, Mabry L, Reiman MP, *et al*. Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review. *Physiotherapy* 2012;98:93–100.
2 Hegedus EJ, Goode AP, Cook CE, *et al*. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med* 2012;46:964–78.
3 Korevaar DA, Cohen JF, Hooft L, *et al*. Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies. *J Clin Epidemiol* 2015;68:708–15.
4 Maas ET, Juch JN, Ostelo RW, *et al*. Systematic review of patient history and physical examination to diagnose chronic low back pain originating from the facet joints. *Eur J Pain* 2017;21:403–14.
5 Reiman MP, Mather RC, Hash TW, *et al*. Examination of acetabular labral tear: a continued diagnostic challenge. *Br J Sports Med* 2014;48:311–9.
6 Cook C. Challenges with diagnoses: sketchy reference standards. *J Man Manip Ther* 2012;20:111–2.
7 Cook C, Cleland J, Hegedus E, *et al*. The creation of the diagnostic accuracy quality scale (DAQS). *J Man Manip Ther* 2014;22:90–6.
8 Lijmer JG, Mol BW, Heisterkamp S, *et al*. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
9 Rutjes AW, Reitsma JB, Di Nisio M, *et al*. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469–76.
10 Whiting P, Rutjes AW, Reitsma JB, *et al*. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.