Correspondence

Inter-rater variability in the interpretation of the head impulse test results

Alexander Cuculiza Henriksen, Peter Hallas

Department of Emergency Medicine, Zealand University Hospital, Køge, Denmark

Correspondence to: Peter Hallas
Department of Emergency Medicine,
Zealand University Hospital,
Lykkebækvej 1, 4600 Køge, Denmark
E-mail: hallas@rocketmail.com

Dear Editor,

The head impulse test (HIT) is recommended for bedside evaluation of patients with dizziness, as part of the Head-Impulse—Nystagmus—Test-of-Skew (HINTS) test.[1] Previous studies on the inter-rater reliability of the HIT have only included 2 raters or used an advanced, eye-tracking technique that is not commonly available for use at the bedside.[2,3] We estimated the inter-rater variability for HIT, for multiple raters, without the use of advanced eye-tracking equipment.

Videos of the HIT were sent to 46 doctors (37 were subscribers to an intra-departmental newsletter and 9 were doctors from the neurology department). Three publicly available educational videos of HITs were included: 2 abnormal examples and 1 normal. Text and sounds in the videos were removed. Responders were asked if the HITs in the videos were normal.

The response rate was 57% (n = 26). One responder reported technical difficulties and was excluded. Of the remaining 25 responders, 15 were at intern level (< 1-year postgraduate), 4 were in specialist training, and 6 were at consultant level. Further, 20% (n = 5) of the participants had formal education in HIT or HINTS, 44% (n = 11) had read about HINTS or watched instructive videos about HINTS; 36% (n = 9) had used HIT/HINTS in a clinical setting; 6 responders were unaware of HIT before this survey.

The overall kappa value was 0.46, using free-marginal multi-rater kappa (Online Kappa Calculator, http://justus.randolph.name/kappa). The overall agreement was 72.9% for all responders. Excluding the group of responders without previous experience resulted in a kappa value of 0.73, and an overall agreement of 86%, i.e., a moderate level of agreement.

In their original online postings, the three videos were labeled as normal or abnormal by the individuals who had posted them. If one defines these as the correct interpretations of the tests, an increasing trend of more correct answers was observed with increasing clinical acumen, although this was not statistically significant (intern level, 78% correct; specialist training, 80%; consultant level, 94%; chi-square test for trend, P = 0.28).

Our approach has some limitations. Firstly, videos were selected because they show a "classic" HIT response. Real-life cases may show a less obvious response. Other limitations include the small number of responders, and the lack of experience of many responders. From our study, we could not identify reasons for the disagreements in interpretation. We suggest that for most responders the disagreement may be due to difficulties in accurately tracking subtle eye-movements. This would be an argument in support of using advanced eye-tracking equipment for routine HITs.[4]

A previous study reported a kappa value of 0.73 (two doctors evaluating multiple patients in an emergency department).[2] In the present study, we evaluated the opposite scenario (multiple raters and few patients), and found a similar value of 0.72 among clinicians experienced with HIT. Thus, this study supports the notion that the HIT has a moderate level of inter-rater agreement.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Kohn MA. HINTS to identify stroke in ED patients with dizziness. Acad Emerg Med 2014;21:347.
2. Vanni S, Nazerian P, Casati C, et al. Can emergency physicians accurately and reliably assess acute vertigo in the emergency department? Emerg Med Australas 2015;27:126-31.
3. Ross LM, Helminski JO. Test-retest and interrater reliability of the video head impulse test in the pediatric population. Otol Neurotol 2016;37:558-63.
4. Alhabib SF, Saliba I. Video head impulse test: a review of the literature. Eur Arch Otorhinolaryngol 2017;274:1215-22.