Journal of Clinical Neuroscience 19 (2012) 423–427

Contents lists available at SciVerse ScienceDirect

Journal of Clinical Neuroscience

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

Clinical Study

The diagnostic accuracy of selected neurological tests

S.J. Sullivan [a,*], G.D. Hammond-Tooke [b], A.G. Schneiders [a], A.R. Gray [c], P. McCrory [d]

[a] Centre for Physiotherapy Research, School of Physiotherapy, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand
[b] Department of Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand
[c] Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand
[d] Centre for Health, Exercise and Sports Medicine, University of Melbourne, Melbourne, Victoria, Australia

ARTICLE INFO

Article history:
Received 27 May 2010
Accepted 28 September 2011

Keywords:
Brain lesion
Neurological examination
LR(+)
LR(-)
Reliability
Sensitivity
Specificity

ABSTRACT

The diagnostic value and reliability of selected neurological clinical tests was studied in control subjects with normal neuroimaging ($n = 42$), and subjects with a focal brain lesion ($n = 38$). The items were studied by two examiners blinded to group membership and using standardized protocols, and subsequently by a neurologist who was not blinded to diagnosis. The positive likelihood ratios ranged from 1.06 (pronator drift) to 22.11 (single leg stance with eyes open, while the negative likelihood ratios ranged from 0.47 (tandem gait) to 0.97 (pupil symmetry). Three items (single leg stance – eyes closed – firm surface; single leg stance – eyes open – foam surface; and tandem gait) successfully distinguished between the two groups (odds ratio $p < 0.05$). The inter-rater reliability was generally poor, with only tandem gait showing excellent agreement (kappa [K] = 0.92). Tandem gait was the only item to show noteworthy agreement (K = 0.93) between the examiners and the neurologist. The tests varied considerably in their ability to detect radiologically demonstrated structural brain lesions, and several items were poorly reproducible, questioning their value as part of a routine neurological examination.

1. Introduction

The clinical neurological examination is an essential element of neurological practice. Steeped in tradition, it depends on the knowledge, training, skilled observations, experience and reasoning of the neurologist.[1,2] The examination yields a raft of diagnostic information, and contributes to the overall diagnosis. The need for evidence-based practice has led recently to the study of diagnostic properties such as reliability, sensitivity, specificity and validity in bedside clinical tests.[3–5] Although the key neurological examination tests are broadly defined, the exact technique and scoring protocols are variable, and in some there is only minimal evidence of their reliability and/or diagnostic ability.

In the few published studies, the reliability of individual items varied;[4] and in one report trainee neurologists showed agreement similar to, and for some cases better than, that obtained by more senior neurologists.[6] Studies of the sensitivity and specificity of examination items have yielded similar mixed results.[7,8]

The purpose of this study was to examine the diagnostic accuracy of selected elements of the neurological examination in identifying persons with a structural neurological lesion and to document the reliability (inter-rater and intra-rater) of the selected elements when assessed by a physiotherapist, and to compare the reliability of a neurologist (traditional approach) and a physiotherapist (standardised approach) in the administration and scoring of these components of the neurological examination.

2. Methods

2.1. Patients

Patients with one or more lesions on structural neuroimaging, and control subjects with symptoms such as headache and normal neuroimaging, were recruited from the neurology outpatient clinic of Dunedin Hospital (New Zealand) between March 2006 and January 2007. Patients of both sexes between the ages of 18 years and 80 years were considered for inclusion in the study. The Lesion group included patients with a diagnosis of a structural cerebral lesion(s) (for example, stroke; multiple sclerosis; tumour) confirmed by either CT scans or MRI. The Control group consisted of patients attending the neurology clinic who had a normal neurological examination (in the opinion of the treating neurologist), and normal structural neuroimaging.

Patients were excluded if they exhibited obvious hemiplegia, movement disorder, aphasia or gait disorder, drowsiness or cognitive impairment, had known disease of the peripheral nervous system or spinal cord or were unable to follow instructions.

* Corresponding author. Tel.: +64 03 479 5429; fax: +64 03 479 8414.
E-mail address: sjohn.sullivan@otago.ac.nz (S.J. Sullivan).

0967-5868/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jocn.2011.09.011

2.2. Sample size

The sample size calculation indicated that 43 participants were needed for each group to obtain precision for sensitivity and specificity estimates of ±15%. Also, each group needed 21 participants to achieve 80% power to show adequate reliability, defined by an intra-class correlation coefficient (ICC) of at least 0.7, using a two-sided alpha level of 0.05, and assuming a true ICC of at least 0.9.

2.3. Neurological tests

The items included were selected to provide a central nervous system (CNS) screening examination broadly covering the neuroaxis and easy to perform outside the clinical environment, and potentially by non-clinicians. The items were selected, following expert consultation and review of core neurology texts, from items commonly included in the neurological examination, and from items used by other professional groups to measure aspects of neurological function. Where possible, items with standardised administration and scoring protocols were used. For other items we developed protocols, taking into account the need to standardise the measurement of the items and their suitability to assessment by a range of health professionals. The items were scored using either a categorical (yes/no [Y/N]) scale or as a continuous measurement such as time.

These items generally followed the protocols described in neurology textbooks,[1,2] except where indicated:

(i) Speech production: This item was adapted from a standardised protocol[9] and required the patient to repeat the short phrase "The sky is blue in Cincinnati" without slurring or using incorrect words. The response was scored as normal (negative) or abnormal (positive).
(ii) Pupil symmetry: The relative difference between the size of the two pupils was scored as normal (negative) or abnormal (positive).
(iii) Pronator drift (PD): Patients were required to stand with their eyes closed and their arms extended at shoulder level (parallel to the ground) with their palms facing upwards for 30 s. Any noticeable drift of the arms or wrist pronation was coded as a positive response.
(iv) Finger-to-nose (FTN): The patients were seated and the dominant arm extended forward at shoulder level with the index finger extended and eyes open. The patient was required to touch the tip of their nose with their index finger and return their arm to the outstretched position as fast as possible while keeping their head still. Five successive repetitions with each arm were performed and the total time for the completion of the five repetitions constituted a trial. This technique was based on published protocols.[10,11]
(v) Single leg stance (SLS): The patient was instructed to maintain balance while standing on their preferred leg with their eyes closed for up to a maximum of 30 s. The patient's performance was measured while standing, with their eyes closed, on a firm (SLS eyes closed – firm) and a foam surface (SLS eyes closed – foam). Patients were instructed to keep their hands on their hips, look straight ahead and keep their eyes closed throughout the task. A third condition required the patient to stand in a similar manner on the foam surface with their eyes open (SLS eyes open – foam). Three trials were performed for each condition. The time(s) the patient was able to maintain their balance was recorded and the trial with the longest time with censoring at 30 s was used for the analysis.
(vi) Tandem gait (TG): The patient walked along a 3-m line with a heel-to-toe gait. The patient was instructed to turn 180° at the end of the line and return to the start. Patients were allowed to use their arms to maintain balance during the walk. Sustained deviations from the line, repeated stumbles or a fall led to the termination of the trial. The time (s) to successfully complete the trial was recorded.

2.4. Procedures

Test data were collected in a clinical laboratory environment by two physiotherapists trained in the standardised administration and scoring of the assessment tasks. The inter-rater and intra-rater reliability component of the study was performed over two weeks. Where the patient was assessed independently by both physiotherapists (inter-therapist reliability) the assessment order was randomised. The tests were assessed in the same order at each session (same morning/afternoon) and the pace of the session was adjusted to each patient's needs. Both physiotherapists were blinded to the diagnostic grouping of the patients.

The same test items, with the exception of standing balance on foam, were assessed by a single neurologist (GHT) within two weeks of the initial assessment. This examination was carried out in a standard manner and each test item was scored subjectively on a scoring system ranging from "0" (normal) to "3" (very abnormal). Due to the study design, the neurologist could not be blinded to the patients' status.

The study procedures were approved by the Lower South Regional Ethics Committee (New Zealand) and all patients gave written informed consent.

2.5. Data analyses

Data from the initial assessment was used in the diagnostic analysis and in the comparisons between physiotherapist and neurologist. When there was more than one trial for a particular item, the score indicating the "best" performance was selected; and timed variables (for example, FTN, and TG) were classified into "poor" or "normal" performance using the following decision rule. Poor performance was classified by taking the 2.5th or 97.5th percentile with reference to the Control group performance, and the choice of the cut-off point (either the 2.5th or 97.5th percentile) was based on the value that indicated worst performance.

These were then used to classify both patients and controls and to calculate sensitivity, specificity, positive likelihood ratio (LR[+]) and negative likelihood ratio (LR[−)]. All four statistics are presented to provide a comprehensive profile of the data and to facilitate comparisons to other studies with the statistics of interest being LR(+) and LR(−). These statistics are interpreted according to the criteria that LR(+) values > 5 and LR(−) values < 0.2 are of clinical importance.[12] Where sufficient data were present, logistic regression was used to calculate the diagnostic odds ratio (OR) controlling for age and sex. A Bonferroni adjustment was made to all p values to maintain a 0.05 Type I error rate per diagnostic measurement. Reliability of the neurological test items was determined using the Kappa (K) statistic. The K was also used to assess agreement between the breakpoint-based classifications and a neurologist's assessment. A K value between 0 and 20 represents "slight"; 21 and 40, "fair"; 41 and 60 "moderate"; or 0.61 and 0.80, "substantial" agreement, while a value > 0.81 indicates "almost perfect" agreement.[13] All analyses were performed using Statistical Analysis Software 9.1.2 (Cary, NC, USA).