European Research in Telemedicine/La Recherche Européenne en Télémédecine (2012) 1, 111—114

Available online at
SciVerse ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

REVIEW

Review: Reliability of NIHSS by telemedicine

Revue : la fiabilité du NIHSS par télémédecine

E. Berthier*, P. Decavel, F. Vuiller, C. Verlut, T. Moulin, E. Medeiros de Bustos

Service de neurologie, CHU de Besançon Jean-Minjoz, 3, boulevard Fleming, 25030 Besançon cedex, France

Received 15 August 2012; accepted 29 August 2012

KEYWORDS
NIHSS;
Reliability;
Reproducibility;
Stroke;
Telestroke;
Telemedicine;
Non-neurologists

Summary Telestroke, specialist area within telemedicine, has progressed significantly over the recent decades. Numerous technological advancements have resulted in the phase of transition that stroke medical practice is currently undergoing. Certain tests, assessments and practices within this field are particularly suited to telemedicine developments. Included in this is NIHSS assessment, the focus of this article. There have been numerous contributions to the literature concerning the reliability of NIHSS: by telemedicine, by non-neurologists and more recently by non-neurologists via telemedicine. This article aims to establish a review of the reliability of NIHSS assessment by telemedicine via an analysis and comparison of the results, methods and conclusions of research previously conducted in the literature.

MOTS CLÉS
NIHSS ;
Fiabilité ;
Reproductibilité ;
AVC ;
Téléstroke ;
Télé-AVC ;
Télémédecine ;
Non neurologues

Résumé Le télé-AVC, domaine spécialiste au sein de la télémédecine, a progressé de façon significative pendant les décennies récentes. De nombreuses avances technologiques ont entraîné une phase de transition que subit actuellement la practice médicale en AVC. Certains tests, évaluations et pratiques dans le domaine des AVC sont particulièrement adaptés aux développements dans la télémédecine. Cela comprend l'évaluation NIHSS, le sujet de cet article. Il existe une multitude de contributions à la littérature concernant la fiabilité du NIHSS : par la télémédecine, par les non neurologues, par télémédecine, et, plus récemment, par les non neurologues via la télémédecine. Cet article a comme objectif l'établissement d'une revue de la fiabilité de l'évaluation du NIHSS par la télémédecine à travers une analyse et une comparaison des résultats, méthodes et les conclusions des recherches précédemment réalisées dans la littérature.

* Corresponding author.
E-mail address: med_eli@hotmail.com (E. Berthier).

2212-764X/$ — see front matter © 2012 Published by Elsevier Masson SAS.
http://dx.doi.org/10.1016/j.eurtel.2012.09.001

Introduction

Telemedicine is an area in constant development and has progressed significantly over the recent decades. Within the area of telestroke and NIHSS assessment, the focus of this article, there have been numerous contributions to the literature on the reliability of NIHSS by non-neurologists [1—3], by telemedicine [1,4—9], and more recently by non-neurologists via telemedicine [1]. This review aims to establish a review of the reliability of NIHSS assessment by telemedicine via an analysis and comparison of the results, methods and conclusions of research previously conducted in the literature.

Study design in the literature

In terms of study design, five studies were very similar in particular [4—7]. Shafqat et al. [4] reported a series of 20 patients with real-time NIHSS evaluation in the acute phase and, in certain cases, the chronic phase (until day 73 after stroke onset). Wang et al. [5] also presented the results of 20 patients in the acute phase but with limited methodological analysis, and no use of weighted Kappa tests or CI calculations. Handschu et al. [6] published a series of 41 cases of stroke in the acute phase only. Meyer et al. [7] presented the results of NIH scoring at bedside and by telemedicine in 25 patients by two NIHSS certified neurologists, and compared junior and senior stroke neurologist examinations to assess reliability especially for telemedicine-naive stroke practitioners. The study by Berthier et al. [1] reported the results of 28 patients with confirmed stroke in the acute phase, with weighted Kappa tests and CI calculations. Overall, except for the exclusion of five TIA patients in Berthier et al., there were no significant differences in terms of sample size and recruitment of stroke patients between these studies (Table 1).

NIHSS assessment

NIHSS assessment is very similar in all the studies, with generally one remote and one local neurologist, both NIHSS certified. The process in one study [1] was slightly different in that the NIHSS scores were obtained by two local bedside examiners (neurologist and emergency physician), and two remote examiners (junior and senior neurologists), whereas in the other studies there were mainly two neurologists as raters. Non-neurologists who were not NIHSS certified were still informed of how to perform the assessment. In all studies, the local and remote neurologists and non-neurologists, were blinded to each others' scores and an average score was obtained both remotely and locally. Furthermore, at the end of each session, the consensual NIHSS score served as a reference, which was validated by the participating examiners together. The NIHSS scores carried out within each study were standardised.

Technical aspects

The technical characteristics across all studies are comparable [6,8,10]. Nevertheless, simultaneous evaluation when four different examiners involved [1] caused the average time taken to perform the NIHSS to be 3—5 minutes longer [4] compared to the average of the other four studies.

The technical equipment used was also generally similar across the studies, with the audiovisual equipment in Handschu et al.'s study being particularly good [6]. Different camera angles were also possible across the studies in order to facilitate NIHSS scoring and provide the best visual information for evaluating items such as lower limb motricity, neglect and, in particular, ataxia items, LOC, language function and dysarthria. This was also essential in obtaining close up shots to accurately evaluate facial paresis and ocular motor function.

Reliability of NIHSS in different contexts

In terms of mean NIHSS values, these were higher in the more recent studies by [1,7], compared to the mean values of the earlier studies [5,6]. However, quantification of interrater variance in NIHSS scores for each patient would appear to be greater in the earlier studies.

Shafqat et al.'s results demonstrate a slight or average correlation for several items that is explained by the notion of inherent variability in interrater reproducibility studies [4]. Handschu et al.'s results show high weighted Kappa coefficients for all items and good or excellent results for the majority of items, which can be attributed to the superior technical quality of their audiovisual equipment [6].

The lowest correlation rates across the studies were ataxia (0.61), dysarthria (0.67) and facial paresis (0.75) and could be explained by poor image quality and sound quality, as well as difficulty in distinguishing between dysarthria and speech disturbances by phonetic paraphasia. For the non-neurologists, items 2—4, 7 and 11 also presented the lowest levels of concordance [9]. In the literature, with the exception of Handschu et al.'s study, ataxia has proven to be a difficult item to score for neurologists too, both at bedside and by telemedicine [6,9]. While it is possible that these items could be removed from the scale, this would be unwise as they are still essential for characterising stroke study populations [11,12].

Non-neurologist scores for all items in the studies were fairly comparable, but for all items but one (neglect), scores for the non-neurologists were higher than those in Shafqat et al.'s study [4]. Most results from the studies demonstrated displayed excellent agreement, good agreement, and in some studies, no items displayed poor or no agreement at all [7,8]. Less than 15% of the patients had a significant variance of two points and no variance of superior to 2 [1] as against 45% of patients in two studies [4,5].

Scores were more strongly correlated for non-neurologists when the calculation of results was based on a comparison the scores of two junior neurologists, not one junior and one emergency physician [7]. Furthermore, each patient was not tested separately, nor the entire spectrum of NIH items. Performing the NIHSS at bedside, remotely and simultaneously by examiners is crucial in guaranteeing the reliability and coherence of the test and thus avoid a potential marking bias, as has arisen in previous studies [4,6].