Disability of Arm, Shoulder and Hand and Michigan Hand Outcomes Questionnaires: Exploring Responsiveness and Diagnostic Performance in a Sample of Outpatients with and without Hand and Wrist Complaints*

Questionários Disability of the Arm Shoulder and Hand e Michigan Hand: Explorando a responsividade e performance diagnóstica em amostra de pacientes ambulatoriais com e sem queixas nas mãos e punho

Vinícius Ynoe de Moraes[1,2]   Jamile Caroline Velasques Faria[1]   Marcela Fernandes[1,2]
Jorge Raduan-Neto[1,2]   Aldo Okamura[1]   João Carlos Belloti[1,2]

[1] Hand Surgery Service, Hospital Alvorada, Moema, São Paulo, SP, Brazil
[2] Orthopedics and Traumatology Department, Universidade Federal de São Paulo, São Paulo, SP, Brazil

Address for correspondence Vinícius Ynoe de Moraes, PhD, Rua Borges Lagoa, 786, São Paulo, SP, 04041-050, Brazil (e-mail: vymoraes@gmail.com).

Abstract

Objective   The present study aimed to verify whether, in an adult population with nontraumatic complaints in the upper limbs, (1) the Disability of the Arm, Shoulder and Hand (DASH) questionnaire and the Michigan Hand Outcomes Questionnaire (MHQ) are susceptible to a "ceiling effect" when compared with a sample of healthy subjects; and (2) to determine cutoff points for diagnostic performance and the intercorrelation for DASH and MHQ in both samples.

Methods   This was a prospective, comparative, nonrandomized study. In total, 150 subjects were included, with 75 in the case group (with disease) and 75 in the control group (without disease). This was a sample of patients recently admitted to a hand surgery outpatient clinic. Controls were matched to clinical cases according to inclusion. The ceiling effect was determined by a maximum response rate ($> 15\%$); receiver operating characteristic (ROC) curves determined cutoff points for sickness definition, and DASH and MHQ sensitivity and specificity. Statistical significance was set at $p < 0.05$.

Keywords
► hand trauma/surgery
► shoulder
► arm
► disability assessment
► surveys and questionnaires

Results The DASH and MHQ questionnaires had no ceiling effect for the case group. In this group, 18 (24%) patients had the maximum DASH score, but none (0%) had the maximum MHQ score. For the control group, 1 (1.33%) subject had the maximum DASH score, but none scored for MHQ. For case determination, DASH scores of 7.1 had 80% sensitivity and 60.3% specificity, whereas MHQ scores of 76.9 had 56.2% sensitivity and 97.3% specificity.

Conclusion The DASH and MHQ questionnaires are reliable tools to measure the impact of hand and wrist morbidities on daily activities, and they are not susceptible to ceiling effects. The DASH questionnaire is more sensitive for patient identification, whereas the MHQ is more specific. As such, the MHQ seems more appropriate when a more specific functional increase is expected.

Resumo

Objetivo Verificar se, em uma população adulta com queixa não traumática dos membros superiores, (1) os questionários Disability of the Arm, Shoulder and Hand (DASH, na sigla em inglês) e Michigan Hand Questionnaire (MHQ, na sigla em inglês) estão suscetíveis ao "efeito de teto", comparando com amostra de não-doentes; (2) determinar pontos de corte de performance diagnóstica e correlação interquestionários para DASH e MHQ em ambas as amostras.

Método Estudo prospectivo, comparativo e não randomizado. Incluímos 150 pacientes, 75 no grupo caso (com doença) e 75 no grupo controle (sem doença). Trata-se de amostra de pacientes recém-admitidos em ambulatório de cirurgia da mão. Os controles foram pareados de forma balanceada de acordo com a inclusão dos casos. Determinamos a presença de efeito de teto por meio da taxa de respostas máximas ($> 15\%$) e associamos curvas receiver operating characteristic (ROC, na sigla em inglês) para a determinação de pontos de corte para a determinação de doentes, associados a medidas de sensibilidade e especificidade. Consideramos $p < 0.05$ para significância estatística.

Resultados Os questionários DASH e MHQ não demonstraram o efeito de teto para o grupo com doença. A porcentagem de pacientes do grupo caso com nota máxima foi de $n = 18$ (24%) no DASH e de 0% no MHQ. Para o grupo sem doença, 1 (1,33%) dos participantes pontuou com nota máxima para DASH, enquanto nenhum pontuou para o MHQ. Na determinação de casos, escores de DASH de 7,1 apresentam sensibilidade de 80% e especificidade de 60,3%. Para o MHQ, um escore de 76,9 apresenta sensibilidade de 56,2% e especificidade de 97,3%.

Palavras-chave
► traumatismos da mão/cirurgia
► ombro
► braço
► avaliação da deficiência
► inquéritos e questionários

Conclusão Os questionários DASH e MHQ são ferramentas confiáveis na mensuração do impacto das morbidades das mãos e dos punhos nas atividades diárias dos pacientes e não são suscetíveis a efeito de teto. O questionário DASH é mais sensível para a identificação de doentes, enquanto o MHQ é mais específico. Em situações nas quais se espera um incremento funcional mais discreto (ou mais específico), o MHQ parece mais adequado.

Introduction

The hand is one of the most important segments of the human body for daily activities. It is known that upper limb conditions impact function and quality of life. However, there is a huge number of diseases to be explored, with different presentations and morbidity. Hand and upper limb function can be assessed using several tools. In hand surgery, the Disability of the Arm, Shoulder and Hand (DASH) questionnaire[1,2] and the Michigan Hand Outcomes questionnaire (MHQ) are widely used.[3]

The DASH questionnaire, created in 1994 by the Institute for Work and Health of Ontario, was developed to measure the impact resulting from several upper limb conditions. It is a self-administered questionnaire in which the patient evaluates his/her upper limb function with 30 items scored from zero to 5. The maximum DASH score is 100 and corresponds to maximum disability, whereas a zero score indicates the lack of upper limb disability.[2,4]

The MHQ is a specific tool for hands and wrists to assess important factors pointed out by patients, hand therapists and hand surgeons. It consists of 100 hypothetical questions

divided into 6 parts: 1) overall hand function; 2) activities of daily living; 3) pain; 4) work performance; 5) aesthetics; 6) patient satisfaction with hand function.[3,5]

One way to assess the responsiveness to a questionnaire is to check for a ceiling effect.[6] The ceiling effect occurs when a large number of people reach the maximum test score even when some pathology is present.[4] In this situation, there is a concern as to whether a certain questionnaire is able to distinguish between sick and nonsick patients.

McHorney et al.[7] define a ceiling effect when > 15% of the patients reach the highest possible test score. For Wang et al.,[6] a ceiling effect requires a percentage > 20%. The present study is based on the hypothesis that these questionnaires may not be suitable for measuring differences between groups with elective hand conditions, as their impact is very specific or minimal.

Due to the importance of an objective hand and wrist function assessment, the present study aimed to measure any ceiling effect for the DASH and MHQ in subjects with common hand surgical conditions from a hand surgery outpatient clinic in the city of São Paulo, SP, Brazil. In addition, the present study aimed to determine cutoff points for sickness definition and the respective diagnostic performance measures (sensitivity and specificity) for the DASH and MHQ.

Methods

Prospective, comparative, nonrandomized study developed at the hand surgery medical residency service from our institution. All participants signed an informed consent form after the study proposal was read and explained to them. The present study was approved by the ethics committee under the number 3.512.219. Adult patients from the outpatient clinic of our hospital, located in the city of São Paulo, SP, Brazil, were included in the study. Control subjects were companions of the patients or volunteers from the hospital assistance group.

Inclusion Criteria

Consecutive adult subjects, aged 18 to 70 years old, from both genders, belonging to the following groups:

Group 1: Patients with clinical complaints in the hands/wrists (pain, paresthesia, nodules, and nontraumatic deformities), admitted to the study before any therapeutic (both surgical and nonsurgical) procedure.

Group 2: Subjects with no hand and wrist complaints.

Noninclusion Criteria

Previous surgical treatment of any hand and/or wrist condition.

Subjects who refused to sign the informed consent form.

Subjects unable to answer the questionnaires (DASH and MHQ) due to cognitive or educational impairments.

Statistical Analysis

The descriptive analysis consisted of mean and standard deviation (SD) values for quantitative variables. Proportions were shown along their 95% confidence intervals (CIs). For inferential analysis, the means were compared using the Mann-Whitney U test, and proportions were compared using the Fisher exact test. Receiver operating characteristic (ROC) curves determined the sensitivity and specificity of the DASH and MHQ cutoff points. The optimal point was at the right uppermost region of the curve (the presumed point with the greatest tool accuracy). Additionally, the consistency of ROC curves for DASH and MHQ was determined by calculating the area below the curve; the closer the area is to one, the more robust the ROC curve. The correlation between tools was defined using the Spearman method. Altman criteria were employed to categorize this correlation, which was poor if < 20%, reasonable if ranging from 21 to 40%, moderate if ranging from 41 to 60%, good if ranging from 61 to 80%, and very good if ranging from 81 to 100%.[8] All inferential analyzes considered an alpha value of 5%.

Sample Size Calculation

Using the minimum difference of 20% for ceiling effect detection, 73 patients are required in each group considering a one-tailed chi-squared test, a beta value of 80% and an alpha value of 5%.

Interventions

After inclusion, both questionnaires (DASH and MHQ) were applied at the same time, following demographic data collection (including age, gender, educational level, profession, comorbidities, and illness motivating the visit).

Outcomes

DASH
The DASH is a region-specific, self-administered questionnaire to measure upper limb dysfunction. Its score considers the total activity of both limbs. Since the DASH measures dysfunction, the higher the score, the greater the self-reported disability. It has additional modules regarding playing sports and musical instruments. For the purposes of the present study, three modules were considered: DASH-1, general module; DASH-2, physical activities and playing musical instruments; DASH-3, work activities.

MHQ
The MHQ is a region-specific, self-administered questionnaire indicated for the general evaluation of hand conditions. It evaluates pain, function, aesthetics, and satisfaction. This assessment differentiates between left and right hands. As the present study presents a comparison group without disease, the average between hands was considered as the measure of interest. Since the MHQ measures function/satisfaction, the higher the score, the greater the self-reported ability/satisfaction.

Primary Outcomes

Ceiling Effect Determination

The DASH and MHQ scores in both populations with their respective SD values. To determine the presence or not of a ceiling effect, the proportion of "maximum scores" for the DASH and MHQ were compared in both populations and in economically active, very young (up to 40 years old) subjects from both populations. The general module (DASH-1) alone was considered for DASH ceiling effect determination because the remaining modules are complementary activities (playing sports or musical instruments, work) (►Table 1).

Diagnostic Performance

Case and control groups were compared, and ROC curves were created to determine performance values (sensitivity and specificity) to define cutoff points. The cutoff point was determined by analyzing the right uppermost point on the graph (visual analysis of the ROC curve). Positive (LR +) and negative (LR-) likelihood ratios are also shown. Positive likelihood ratio values range from 1 to infinite. Higher LR+ values increase the probability of disease when the test is positive. Negative likelihood ratio values range from 0 to 1. Smaller values (close to 0) increase the probability of non-disease if the test is negative.

Secondary Outcomes

The DASH (DASH-1, DASH-2, and DASH-3) and MHQ (average value from both hands) scores were compared in both populations.

Results

There were 75 patients in the case group (20 men and 55 women) and 75 subjects in the control group (32 men and 43 women). The case group presented tenosynovitis ($n = 34$; 45.3%), carpal tunnel syndrome ($n = 24$; 32%), synovial cysts of the wrist/hand ($n = 16$; 21.3%), and wrist/hand arthritis ($n = 8$; 10.6%).

Exploring the ceiling effect

Only one patient from the control group had a maximum DASH-1 score, whereas 18 patients from the case group had the maximum score (1.3% versus 24%; Fisher exact tests; $p = 0.00$). For the MHQ, no subjects from either group presented the maximum score. This demonstrates that the MHQ, even in the absence of disease, identifies potential limitations that may not be perceived by the patients.

Instruments performance: defining affected and nonaffected subjects

The cutoff point for the general DASH module (DASH-1) was 7.1, with 80% sensitivity and 60.3% specificity (►Figure 1). For DASH-3, the cutoff point was 9.4, with 78.1% sensitivity and 65.8% specificity. For the MHQ, the cutoff point was 76.9, with 56.2% sensitivity and 97.3% specificity (►Figure 2). For the physical activity and work module (DASH-2), preliminary analysis of data consistency (calculation of the area under the curve) indicated that there was not enough consistency to determine cutoff points. Likelihood ratios demonstrated the substantial ability of the MHQ in case definition (positive test, high disease probability), while the DASH had a good ability to identify the absence of disease (negative test, low disease probability) (►Table 2).

## Correlation between the DASH and MHQ in case and control groups

The Spearman method determined the correlation between the DASH and MHQ scores. ►Table 3 shows that there was a correlation between the tools both in the case and control

Table 1 Demographic data and functional scores

| Variable	With disease	Without disease	p-value
Age (mean value, SD)	41 (14.6)	38.7 (13.3)	0.33
College-educated (n, %)	39 (53.4)	41 (56.2)	0.86
Income (median value, IQ)	2.9 (2.8–5.7)	3.8 (2.8–6.8)	0.61
Female gender (n, %)	53 (72.6)	41 (56.2)	0.04
College-educated (n, %)	39 (53.4)	41 (56.2)	0.86
Income (median value, IQ)	2.9 (2.8–5.7)	3.8 (2.8–6.8)	0.61
DASH-1 (mean value, SD)	22.2 (18)	7.9 (8.4)	< 0.001
DASH-2 (mean value, SD)	4.3 (14.9)	3.1 (10.8)	0.76
DASH-3 (mean value, SD)	29.8 (23.7)	8.8 (12.9)	< 0.001
MHQ (mean value, SD)	60.1 (14.8)	71.3 (17.1)	< 0.001

Abbreviations: DASH, Disability of the Arm, Shoulder and Hand questionnaire; IQ, interquartile range; MHQ, Michigan Hand Outcomes questionnaire.
*In minimum wages.
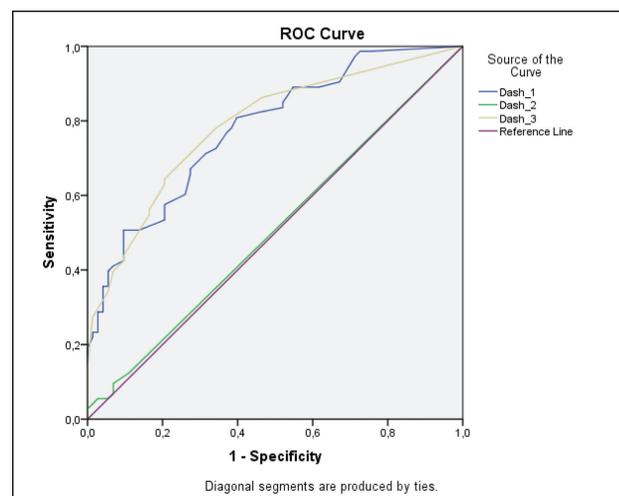
Fig. 1 Receiver operating characteristic (ROC) curve for the Disability of the Arm, Shoulder and Hand (DASH) questionnaire.
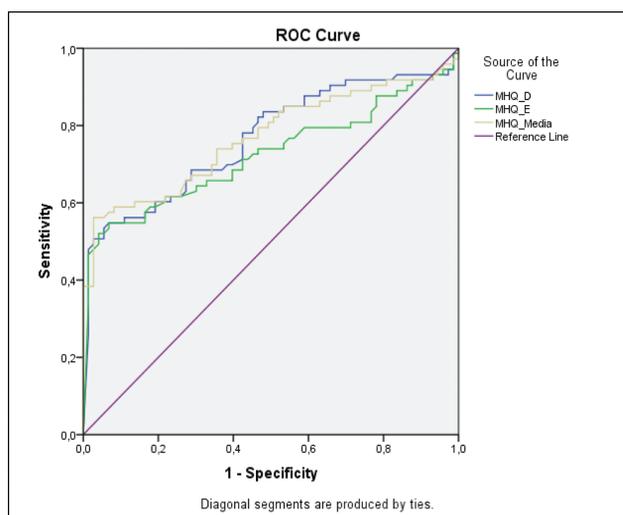
Fig. 2 Receiver operating characteristic (ROC) curve for the Michigan Hand Outcomes questionnaire (MHQ).

groups, and that correlation was better for the DASH-1 score and the average MHQ score in both groups.

In addition, the results demonstrate a greater correlation for the case group compared with the control group.

Discussion

The present study evaluated established tools for self-reported upper limb function and tested responsiveness properties and correlation. The hypothesis that a ceiling effect could underestimate the DASH and MHQ scores, raised at the beginning of the present study, was not confirmed, demonstrating the good specificity of both tools for diseases, even those related with low morbidity and debilitation rates, such as synovial cysts and tenosynovitis. Our hypothesis was based on the fact that nonsevere hand conditions may not have a significant impact on daily living activities, which is the focus of the DASH and of the MHQ.[9] Authors report that a ceiling effect requires more than 15%[4] or 20%[6] of subjects with maximum scores in tests performed on sick patients.[10]

In this context, we identified 1 case (1.3%) with maximum score in the nonsick population. In the case group, 18 (24%) patients had a maximum DASH score, showing the psychometric robustness of this tool to differentiate patients and nonpatients. As for the MHQ, no one from either group reached the maximum score. A recent review suggests that the MHQ, due to its high internal consistency, has a better ability to identify lower functional gains and, consequently,

Table 3 Spearman correlation in groups with or without disease

	With disease	Correlation	Without disease	Correlation
DASH-1 x DASH-2	-0.14	poor	0.13	poor
DASH-1 x DASH-3	0.74*	very good	0.76*	very good
DASH-2 x DASH-3	0.07	poor	0.09	poor
MHQ x DASH-1	-0.77*	very good	-0.61*	good
MHQ x DASH-3	-0.77	very good	-0.49	moderate

Abbreviations: DASH, Disability of the Arm, Shoulder and Hand questionnaire; MHQ, Michigan Hand Outcomes questionnaire.

lower treatment effects.[11] In contrast, the MHQ is a more redundant tool when the treatment effect seems more evident. In the clinical practice, when modest functional gain is expected, such as a small cord release in Dupuytren disease, the MHQ presents better specificity. In conditions where the treatment effect seems more exuberant, such as median nerve decompression in carpal tunnel syndrome, this MHQ specificity seems diluted, making the DASH a better option, since it is more responsive and easier to apply in this scenario. It is also relevant to consider the minimally relevant clinical difference in this choice, that is, in the clinical practice, how much a score needs to change for the patient to perceive an event/intervention. This phenomenon has already been explored by the same authors.[12]

Our results demonstrate the good or very good correlation between the DASH and the MHQ. The DASH modules (both general and work modules) also correlate in a very satisfactory way. As for the sports and playing musical instruments module, correlations were always poor since these practices are very heterogeneous among populations. In practice, the guidelines of the DASH developers report that this module must only be used for specific subpopulations consisting of athletes and musicians. A positive highlight is the more robust correlation coefficients in the case group, which strengthens the psychometric capacity of these tools.

There were some difficulties regarding the application of the questionnaires, such as the large number of questions and the different scale for the MHQ, causing demotivation and confusion, respectively, when filling out each form. McMillan et al.[13] compared the responsiveness for the DASH, the MHQ and the Patient-Specific Functional Scale (PSFS) questionnaires and found a greater responsiveness for the DASH in patients with tumors, for the MHQ in patients

Table 2 Cutoff values and likelihood ratio for each instrument

Variable	Cutoff value	Positive LR	Negative LR	Area under the curve	*p-value*
DASH-1	7.1	2.0	0.33	0.77	< 0.001
DASH-2	Not calculated	Not calculated	Not calculated	0.51	0.86
DASH-3	9.4	2.28	0.33	0.78	< 0.001
MHQ	76.9	20.8	0.45	0.77	< 0.001

Abbreviations: DASH, Disability of the Arm, Shoulder and Hand questionnaire; LR, Likelihood ratio; MHQ, Michigan Hand Outcomes questionnaire.

with carpal tunnel syndrome and wrist pain, and for the PSFS in patients with finger contracture.

Although different from each other, both questionnaires proved to be important tools for measuring the impact of hand and wrist conditions on the daily activities of the patients. We believe that the questionnaires are partially complementary: the MHQ is better to discriminate small changes in health status, whereas the DASH[14] is better in identifying "cases" and in excluding healthy subjects (low negative likelihood ratio). In the clinical practice, it seems sensible to choose the MHQ for a very specific comparative assessment (for example, to compare two alternative therapies for trapeziometacarpal arthritis), and the DASH for broader studies with a population bias. As such, the use of these tools in their quick form may be an alternative. The Brief-MHQ seems to be a good option compared with its original version. Studies have already demonstrated the equivalent psychometric ability of the Brief-MHQ,[15,16] the Quick-DASH, and other validated tools.[17] These alternatives have special appeal when a researcher wants to measure multiple outcomes, as it shortens the process with no potential loss of quality.

Conclusion

The DASH and the MHQ are robust, responsive tools, and they are not susceptible to a ceiling effect, not confirming the hypothesis initially raised in the present study. Both are well correlated, mainly in sick patients. The DASH questionnaire is more sensitive for identifying patients, while the MHQ is more specific. In situations in which a more specific functional increase is expected, the MHQ is more appropriate. In situations in which population assessments are planned, the DASH is more appropriate.

Conflicts of Interest
The authors declare no conflicts of interest.

References

1 Orfale AG, Araújo PM, Ferraz MB, Natour J. Translation into Brazilian Portuguese, cultural adaptation and evaluation of the reliability of the Disabilities of the Arm, Shoulder and Hand Questionnaire. Braz J Med Biol Res 2005;38(02):293–302

2 Hudak PL, Amadio PC, Bombardier CThe Upper Extremity Collaborative Group (UECG) Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. Am J Ind Med 1996;29(06):602–608

3 Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am 1998l23(04):575–587

4 Kim SJ, Lee BG, Lee CH, Choi WS, Kim JH, Lee KH. Comparison of ceiling effects between two patient-rating scores and a physician-rating score in the assessment of outcome after the surgical treatment of distal radial fractures. Bone Joint J 2015;97-B(12):1651–1656

5 Ribeiro NP, Schier AR, Silva ACO, Nardi AE. Traducao e adaptacao transcultural do Michigan Hand Outcomes Questionnaire. J Bras Psiquiatr 2011;60(02):99–110

6 Wang L, Zhang Z, McArdle JJ, Salthouse TA. Investigating Ceiling Effects in Longitudinal Data Analysis. Multivariate Behav Res 2009;43(03):476–496

7 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4(04):293–307

8 Altman DG. Practical statistics for medical research. New York: Chapman and Hall; 1991

9 de Moraes VY, Ferrari PM, Gracitelli GC, Faloppa F, Belloti JC. Outcomes in orthopedics and traumatology: translating research into practice. Acta Ortop Bras 2014;22(06):330–333

10 Nolte MT, Shauver MJ, Chung KC. Normative Values of the Michigan Hand Outcomes Questionnaire for Patients with and without Hand Conditions. Plast Reconstr Surg 2017;140(03):425e–433e

11 Shauver MJ, Chung KC. The Michigan hand outcomes questionnaire after 15 years of field trial. Plast Reconstr Surg 2013;131(05):779e–787e

12 Maia MV, de Moraes VY, Dos Santos JB, Faloppa F, Belloti JC. Minimal important difference after hand surgery: a prospective assessment for DASH, MHQ, and SF-12. SICOT J 2016;2:32

13 McMillan CR, Binhammer PA. Which outcome measure is the best? Evaluating responsiveness of the Disabilities of the Arm, Shoulder, and Hand Questionnaire, the Michigan Hand Questionnaire and the Patient-Specific Functional Scale following hand and wrist surgery. Hand (N Y) 2009;4(03):311–318

14 Paranaíba VF, Santos JB, Raduan Neto J, Moraes VY, Belotti JC, Faloppa F. Aplicação do PRWE na fratura da extremidade distal do rádio: comparação e correlação com desfechos consagrados. Rev Bras Ortop 2017;52(03):278–283

15 Figueiredo S, Pereira AD, Silva MS, Leite L, Costa G, Silva C. Correlações clínicas, funcionais e imagiológicas após artroplastia da articulação metacarpofalangiana com implante de silicone na mão reumatoide. Rev Bras Ortop 2018;53(02):208–212

16 Waljee JF, Kim HM, Burns PB, Chung KC. Development of a brief, 12-item version of the Michigan Hand Questionnaire. Plast Reconstr Surg 2011;128(01):208–220

17 Ibrahim S, Meleppuram JJ. Estudo retrospectivo da placa anterior superior como tratamento para fraturas instáveis da clavícula distal (tipo 2 de Neer). Rev Bras Ortop 2018;53(03):306–313