

Concordância Inter e Intraobservador de Fraturas de Úmero Antes e Após a Exposição à Classificação AO com um Esquema Cognitivo

Anderson Reus Trevisol Bárbara Mendes Boppré Gustavo ou Renan Gabriel El-Kouba Ana Paula Bonilauri Ferreira Ricardo Pietrobon

Resumo

Introdução

As fraturas de úmero são o terceiro tipo mais comum de fraturas em pessoas acima de 65 anos, depois de fraturas do rádio distal e fêmur proximal (court2006epidemiology). Os diversos sistemas de classificação de fraturas ortopédicas são propostos para melhorar a sua descrição, baseados na sua morfologia, no seu comportamento biológico e mecânico, e para orientar diretrizes terapêuticas. Apesar de termos evoluído nas últimas décadas na correlação da cognição e diagnóstico médico, ainda existem lacunas do conhecimento a serem respondidas em relação à essa intersecção no que diz respeito às classificações ortopédicas.

A classificação de Neer é amplamente utilizada na prática clínica e em pesquisa (S Brorson et al. (2002)), porém, a literatura sugere existir uma baixa concordância entre observadores utilizando este sistema de classificação (Siebenrock and Gerber (1993); Sidor et al. (1993); Bernstein et al. (1996); Sjöden et al. (1997); S Brorson et al. (2002); Shrader et al. (2005); Stig Brorson et al. (2009)). Este baixo índice de concordância pode ter associação ao uso inadequado do esquema cognitivo. O conhecimento prévio construído em um “esquema” facilita muito o conhecimento contextual se a nova informação é assimilada no esquema (Ruiter, Kesteren, and Fernandez (2012)). Assim sendo, a interpretação das escalas deveria estar associada a um esquema cognitivo coerente (Ruiter, Kesteren, and Fernandez (2012); A. P. R. B. Ferreira et al. (2010)), o que poderia contribuir para o aumento da concordância entre observadores.

O objetivo deste estudo foi investigar a confiabilidade intra e interobservador da classificação de Neer para fraturas de úmero com um esquema cognitivo.

Métodos

Comitê de Ética

Este estudo foi aprovado pelo comitê de ética em pesquisa do Hospital Municipal São José (Joinville-SC) previamente ao início deste projeto (número do

protocolo 239.511). Todos os participantes assinaram o termo de consentimento previamente à participação no estudo.

Participantes

Um total de 15 residentes de ortopedia participaram deste estudo, 5 eram do primeiro ano, 6 do segundo ano e 4 do terceiro ano. A média de idade era 28 anos e todos eram do gênero masculino.

Image selection

Two second-year residents that were not enrolled directly with the study and an orthopedic surgeon, especialista in membro superior selected 20 images containing fractures with a wide pattern variation chosen to cover the full spectrum of the Neer classification for humerus fractures. As imagens eram na incidência ântero-posterior e perfil. As imagens foram obtidas de arquivos dos Hospital São Municipal São José (Joinville – SC). Qualquer sinal de identificação de pacientes foi removido. Imagens radiográficas mal posicionadas que poderiam gerar problemas na interpretação foram excluídas. Imagens de baixa qualidade ou com artefatos ou outros defeitos técnicos também foram excluídas.

Neer Classification

O úmero proximal é dividido em 4 partes: cabeça, diáfise, tubérculo maior e tubérculo menor. É considerado desvio quando o segmento desloca-se mais de 1 cm (0,5 cm para o tubérculo maior) ou angulação maior do que 45°. As fraturas em uma parte são aquelas com pouco ou nenhum desvio, independentemente do número de traços de fratura; as fraturas em duas partes apresentam desvio de apenas uma parte (colo cirúrgico, colo anatômico, tubérculo maior ou tubérculo menor); as em três partes são fraturas da diáfise com um dos tubérculos, separados da cabeça; e as fraturas em quatro partes apresentam dissociação entre as quatro partes anatômicas do úmero proximal e alta incidência de necrose avascular (NEERII (1970)). [<http://radiopaedia.org/articles/proximal-humeral-fracture-classification-neer>]

Situated schemata extraction using Cognitive Task Analysis

For the purposes of our paper, we define a situated schema as the collection of concepts and situations (e.g., narratives) that an expert hand and upper extremity surgeon relates to each classification category. In order to extract the situated schema from our expert hand and upper extremity surgeon (GE), we used the following sequence.

First, the Neer classification was presented to the hand and upper extremity surgery expert in an electronic format combining text and graphics for each classification category. Second, we asked the surgeon to “think aloud” about what they thought when finding a case in their daily practice. After an initial description, we specifically asked the expert to discuss any diagnostic, biomechanical or related therapeutic decision if it had not yet been previously mentioned. We also encouraged the expert to provide any narratives that might occur to him while thinking aloud about each classification category. The entire process was recorded in a video.

Second, the video was analyzed and a graph constituted by nodes and edges was built using [Graphviz](#). Each node represented either a concept or a situation, while edges connected relationships among diagnostic, biomechanical and therapeutic nodes.

Procedimentos e Logística do Estudo

Avaliação Inicial Na avaliação inicial, todos os residentes independentemente classificaram as 20 imagens de acordo com a Classificação de Neer. Todos os participantes simultaneamente, alocados na mesma sala, foram instruídos a não olhar as respostas dos outros ou discutir os casos clínicos. As imagens a serem classificadas estavam armazenadas na Plataforma online Edx [<http://code.edx.org/>]. As respostas eram armazenadas na própria plataforma online. Os residentes podiam consultar na internet o sistema de classificação de Neer. Não foi estipulado um tempo limite para classificar as imagens. Os autores do estudo não participaram como sujeitos do estudo.

Avaliação em 30 dias Após 30 dias, cada residente independentemente classificou novamente as 20 imagens em uma ordem alterada na Plataforma Edx. Todos os outros procedimentos foram executados exatamente como descrito na sessão da Avaliação Inicial.

Intervenção The educational intervention was constituted by weekly sessions where participants completed 20 exercises related to the diagnosis, biomechanics and therapeutic planning of shoulder fractures. Os residentes, após terem respondido cada exercício, eram apresentados à resposta correta, juntamente com uma explicação justificando tal resposta. Os exercícios estavam organizados, na Plataforma online Edx, em “blocos” de 20 perguntas cada, ou seja, foram criados 4 blocos correspondentes as quatro semanas de intervenção. Os blocos de perguntas foram programados para serem liberados semanalmente, isto é, os residentes tinham acesso somente a um bloco de perguntas por semana. The full spectrum described by the Neer classification was covered based on cognitive schema from our expert hand and upper extremity surgeon (GE) as described in the previous session - Situated schemata extraction using Cognitive Task Analysis.

Pós-Intervenção, Avaliação em 60 dias After the 4-week intervention period, all participants classified the same 20 images following the same protocol.

Medidas avaliadas

Intra-observer agreement was measured by comparing ratings by the same participant between baseline and the thirty 30-day assessment. Baseline inter-observer agreement was measured at the 30-day assessment. The pre-post intervention evaluation was conducted by comparing the 30-day pre-intervention assessment with the 60-day post-intervention assessment.

Análises dos dados

All data were extracted directly from [MySQL](#) and [MongoDB](#) databases connected to the [Open edX](#) platforms. Data sets were then merged, also undergoing an exploratory graphical analysis to verify distributions, percentages, means and frequencies/percentages as well as rates of missing data.

Apenas observadores que haviam completado um determinado grupo de observações (dia 1, dia 30 ou dia 60) foram considerados na análise. Porcentagens de concordância assim como valores de Kappa Fleiss foram reportados. Kappa Fleiss é uma medida de concordância para variáveis categóricas que leva em consideração a possível concordância ao acaso. Fleiss and Cohen (1973) Por fim, a comparação entre os valores de Kappa pré e pós intervenção (dias 30 e 60, respectivamente) foram estimadas através da computação de erros padrão e intervalos de confiança (95%) utilizando bootstrap.

Resultados

Resultados descritivos

Quando todos os 9 observadores completando as 3 avaliações foram considerados, não houve nenhuma instância em que houvesse concordância completa entre todos. O mesmo ocorreu quando a avaliação se restringiu ao dias isolados 1 (15 observadores), 30 (11 observadores) e 60 (13 observadores). No entanto, a porcentagem de concordância entre os observadores foi aumentada quando a classificação foi simplificada para incorporar apenas o número de partes e com 1 grau de tolerância: 21.4% para o dia 1 (15 observadores), 40% para o dia 30 (11 observadores) e 45% para o dia 60 (13 observadores).

Em relação a porcentagem de concordância com o mesmo indivíduo entre os dias 1 e 30, a taxa foi de 28.5%, chegando a 65.7% quando um grau de tolerância foi adicionado.

Valores de kappa

Os valores de Kappa Fleiss foram de 0.302 ($p < 0.001$) para o estudo como um todo (9 observadores), 0.256 ($p < 0.001$) para o dia 1, 0.277 ($p < 0.001$) para o dia 30, 0.315 ($p < 0.001$) para o dia 60. A diferença entre os dias 30 (pré-intervenção) e 60 (pós-intervenção) não foi estatisticamente significativa.

Quando a avaliação se restringiu ao número de partes, os valores de Kappa Fleiss chegaram a 0.279 ($p < 0.001$) para o grupo como um todo, 0.26 ($p < 0.001$) para o dia 30 e 0.332 para o dia 60. A diferença entre os dias 30 (pré-intervenção) e 60 (pós-intervenção) não foi estatisticamente significativa. A concordância entre os mesmos indivíduos, comparando os dias 1 e 30, foi de 0.01 ($p = 0.809$)

Discussão

Até onde sabemos, esse é o primeiro estudo avaliando uma intervenção na tentativa de melhorar o grau de concordância entre observadores para a classificação de Neer de fraturas do ombro. Nossos resultados indicaram uma concordância que se manteve relativamente estável durante o estudo, não tendo sido modificada em função da intervenção. Também encontramos que o grau de concordância é melhorado quando (1) a classificação se faz simplesmente por partes ao invés de feita com detalhes de cada fratura e (2) quando o número de partes é diminuído.

O baixo grau de concordância entre observadores nesse estudo corrobora com artigos anteriores sobre classificações de fraturas ortopédicas (Siebenrock and Gerber (1993); Sidor et al. (1993); Bernstein et al. (1996); Sjödén et al. (1997); S Brorson et al. (2002); Shrader et al. (2005); Stig Brorson et al. (2009)). A complexidade das classificações pode conduzir à essa baixa concordância, levando à diversas interpretações, principalmente por médicos em treinamento e com menos experiência (Kristiansen et al. (1988); Foroohar et al. (2011)). Apesar de que, a intenção de se criar uma classificação que seja “completa” é louvável, a sua praticidade pode ser desperdiçada em função da alta carga cognitiva exigida dos profissionais que a irão utilizar. A criação de escalas mais condizentes com diferentes níveis de experiência na interpretação radiográfica poderiam ser melhor aproveitadas. Ou seja, dependendo da área de atuação do profissional, escalas mais simplificadas ou mais complexas poderiam ser utilizadas. Por exemplo, médicos sub-especialistas utilizariam escalas mais detalhadas, enquanto médicos que trabalhem em pronto socorros utilizariam escalas mais simplificadas. O grau de detalhamento em cada uma destas escalas distintas seria determinado através de estudos que identifiquem o grau de concordância obtido na prática clínica diária.

Apesar do nosso treinamento ter sido baseado em esquemas cognitivos estabelecidos na literatura (Regehr and Norman (1996); Ruiter, Kesteren, and Fernandez (2012)), não houve uma melhora da concordância como nós esperávamos. Esquemas cognitivos situados partem do princípio de que o cérebro raciocina não

apenas através de informações armazenadas nele, mas utilizando também fatores ambientais como tecnologias, contatos sociais, experiências prévias, entre outros fatores (Van Merriënboer and Sweller (2010)). O fato do tempo do treinamento não ter sido muito longo pode ter relação com o não aumento da concordância entre observadores, assim como a exposição não ter acontecido em um contexto clínico, e sim, em um ambiente educacional artificial. Se a intervenção educacional tivesse ocorrido durante a prática clínica diária, ou seja, no momento de atendimento a pacientes com fraturas de ombro, talvez os resultados poderiam ter sido superiores.

Nosso estudo apresenta limitações, mesmo até onde sabemos, ser o primeiro a inserir uma intervenção, baseada em esquemas cognitivos, na tentativa de melhorar o grau de concordância entre observadores. Primeiro, a nossa intervenção educacional utilizou um desenho pré-pós ao invés de um estudo randomizado. Na ausência de randomização, não se pode fazer afirmações sobre relações causais entre a intervenção e a ausência de impacto sobre o grau de concordância. No entanto, amostras significativamente maiores são requeridas em estudos randomizados, o que pode dificultar sua execução devido à barreiras logísticas. Segundo, o fato deste estudo ter sido realizado em uma única instituição, não podemos generalizar os resultados. A participação de várias instituições é sempre desejável, porém há um envolvimento logístico para a realização de estudos dessa natureza, o que dificulta o processo. Por último, um tempo maior de treinamento, bem como a contextualização na prática clínica diária são aspectos ideais a serem ineridos no estudo. Porém, como nas limitações anteriores, fatores logísticos dificultam a sua realização.

Concluímos que intervenções educacionais curtas e descontextualizadas da prática clínica não são recomendadas no aprendizado de classificações complexas. No entanto, consideramos que uma personalização de escalas relacionada a graus de experiências diferentes poderia ser de maior aproveitamento. Em relação a estudos futuros, recomendaríamos estudos randomizados que permitam investigações causais, assim como a contextualização e personalização das intervenções educacionais.

Bernstein, Joseph, Louis M Adler, John E Blank, Robert M Dalsey, Gerald R Williams, and Joseph P Iannotti. 1996. "Evaluation of the Neer System of Classification of Proximal Humeral Fractures with Computerized Tomographic Scans and Plain Radiographs*." *The Journal of Bone & Joint Surgery* 78 (9). The Journal of Bone; Joint Surgery: 1371–5.

Brorson, S, Jens Bagger, Annette Sylvest, and A Højbjartsson. 2002. "Improved Interobserver Variation After Training of Doctors in the Neer System a RANDOMISED TRIAL." *Journal of Bone & Joint Surgery, British Volume* 84 (7). British Editorial Society of Bone; Joint Surgery: 950–54.

Brorson, Stig, Jens Bagger, Annette Sylvest, and Asbjørn Hrobjartsson. 2009. "Diagnosing Displaced Four-Part Fractures of the Proximal Humerus: a Review of Observer Studies." *International Orthopaedics* 33 (2). Springer: 323–27.

Ferreira, Ana Paula Ribeiro Bonilauri, Rodrigo Fernando Ferreira, Dimple Rajgor, Jatin Shah, Andrea Menezes, and Ricardo Pietrobon. 2010. "Clinical Reasoning in the Real World Is Mediated by Bounded Rationality: Implications for Diagnostic Clinical Practice Guidelines." *PloS One* 5 (4). Public Library of Science: e10265.

Fleiss, Joseph L, and Jacob Cohen. 1973. "The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability." *Educational and Psychological Measurement*. Sage Publications.

Foroohar, Abtin, Rick Tosti, John M Richmond, John P Gaughan, and Asif M Ilyas. 2011. "Classification and Treatment of Proximal Humerus Fractures: Inter-Observer Reliability and Agreement Across Imaging Modalities and Experience." *J Orthop Surg Res* 6: 38.

Kristiansen, Bjarne, Ulrich LS Andersen, Claus A Olsen, and Jens-Erik Varmarken. 1988. "The Neer Classification of Fractures of the Proximal Humerus." *Skeletal Radiology* 17 (6). Springer: 420–22.

NEERII, CHARLES S. 1970. "Displaced Proximal Humeral Fractures Part I. Classification and Evaluation." *The Journal of Bone & Joint Surgery* 52 (6). The Journal of Bone; Joint Surgery: 1077–89.

Regehr, Glenn, and Geoffrey R Norman. 1996. "Issues in Cognitive Psychology: Implications for Professional Education." *Academic Medicine* 71 (9). LWW: 988–1001.

Ruiter, Dirk J, Marlieke TR van Kesteren, and Guillen Fernandez. 2012. "How to Achieve Synergy Between Medical Education and Cognitive Neuroscience? An Exercise on Prior Knowledge in Understanding." *Advances in Health Sciences Education* 17 (2). Springer: 225–40.

Shrader, M Wade, Joaquin Sanchez-Sotelo, John W Sperling, Charles M Rowland, and Robert H Cofield. 2005. "Understanding Proximal Humerus Fractures: Image Analysis, Classification, and Treatment." *Journal of Shoulder and Elbow Surgery* 14 (5). Elsevier: 497–505.

Sidor, Michael L, JD Zuckerman, T Lyon, K Koval, F Cuomo, and N Schoenberg. 1993. "The Neer Classification System for Proximal Humeral Fractures." *J Bone Joint Surg Am* 75: 1745–50.

Siebenrock, Klaus A, and Christian Gerber. 1993. "The Reproducibility of Classification of Fractures of the Proximal End of the Humerus." *Journal of Bone and Joint Surgery-A-American Volumes* 75 (12): 1751–55.

Sjödén, Göran OJ, Tomas Movin, Peter Güntner, Peter Aspelin, Leif Ahrengart, Hans Ersmark, and Anders Sperber. 1997. "Poor Reproducibility of Classification of Proximal Humeral Fractures: Additional CT of Minor Value." *Acta Orthopaedica* 68 (3). Informa UK Ltd UK: 239–42.

Van Merriënboer, Jeroen JG, and John Sweller. 2010. "Cognitive Load Theory

in Health Professional Education: Design Principles and Strategies.” *Medical Education* 44 (1). Wiley Online Library: 85–93.