Reliability of the AMA Guides to the Evaluation of Permanent Impairment

Linda Forst, MD, MPH, Lee Friedman, PhD, UIC-SPH, EOHS, and Abraham Chukwu, MD, MPH, UIC-SPH, EOHS

Background: AMA's Guides to the Evaluation of Permanent Impairment is used to rate loss of function and determine compensation and ability to work after injury or illness; however, there are few studies that evaluate reliability or construct validity. **Objectives:** To evaluate the reliability of the fifth and sixth editions for back injury; to determine best methods for further study. **Results:** Intra-class correlation coefficients within and between raters were relatively high. There was wider variability for individual cases. Impairment ratings were lower and correlated less well for the sixth edition, though confidence intervals overlapped. **Conclusions:** The sixth edition may not be an improvement over the fifth. A research agenda should include investigations of reliability and construct validity for different body sites and organ systems along the entire rating scale and among different categories of raters.

Impairment ratings are used to describe physical or functional loss of a body part or an organ following an occupational injury or illness. The rating fixes the diagnosis and the percentage of physical and functional loss at the point of "maximal medical improvement." This percentage is frequently translated into a monetary award, and is also used as the first step in the evaluation of disability.

The Guides to the Evaluation of Permanent Impairment, developed by the American Medical Association, is the most frequently used tool in the United States to assign a permanent impairment rating. Some 40 jurisdictions use it in workers compensation systems, in personal injury litigation, and in automobile insurance systems.[1,2]

First published in 1958 and revised through the sixth edition in 2008,[3,4] *The Guides* has been criticized for following reasons: inconsistency and ambiguity in definitions; poor reliability and reproducibility; lack of content validity; failure of ratings to reflect true or perceived functional loss; lack of consistency across organ systems; poor predictive capability; an inadequate basis in scientific evidence; development by a narrow consensus of practitioners, with exclusion of other stakeholders, like attorneys and policy makers; wide variability of assigned ratings among practitioners; complexity of the system that requires many hours of study and training; a bias toward the worker; a bias toward the employer; and a marked departure in the rating protocol from prior versions.[1,2,5–9] Because of these concerns, many states have elected not to adopt the new version (Impairment Resources, 2010 http://impairment.com/PressRelease/index.htm, accessed August 5, 2010).

Despite several impassioned testimonies before workers compensation commissions and legal writings,[10–12] there is a dearth of evidence to either justify or refute claims that the fifth or sixth edition of *The Guides* is superior to the other in terms of precision, accuracy, validity, or reliability. We conducted a pilot test to determine the reliability of the fifth and of the sixth editions for low back injury; to compare ratings for individual low back cases between the fifth and sixth editions; and to sort out methodological issues for a larger, more definitive study.

From the University of Illinois at Chicago, School of Public Health, Environmental and Occupational Health Sciences, Chicago, Ill.

Address correspondence to: Linda Forst, MD, MPH, University of Illinois at Chicago, School of Public Health, Environmental and Occupational Health Sciences, 2121 W. Taylor, MC 922, Chicago, IL 60612; E-mail: ForstL@uic.edu.

DOI: 10.1097/JOM.0b013e3181fd2782

METHODS

Cases

We chose low back injuries as the study's health outcome because of the high prevalence and exorbitant costs associated with occupationally related low back disorders, and because of the lack of definitive clinical tests to aid in determining the degree of impairment of this body site. We developed 20 case histories of low back pain from summaries that we found on the Internet and in medical texts. These case summaries described the injury event, the acute findings, the laboratory tests, the medical interventions, the clinical course, and the history and physical examination at the time of "maximal medical improvement." They ranged from very mild injuries with complete resolution of signs and symptoms to severe injuries with invasive interventions and continued pain and functional impairment.

Participants

Six occupational medicine residents and two fourth-year medical students were recruited to participate. Participants were divided into two groups of three residents and one student. One group was trained to render an opinion regarding degree of impairment using the fifth edition, and one group the sixth edition. Two of the investigators (Chukwu and Forst) prepared and conducted an hour presentation and case discussion from a template they designed to assure equivalence of the training sessions. There had been 3 hours of presentations to these participants of *The Guides*, in general (the purpose, goals, development), and for body parts other than the low back prior to this study. A naïve group was chosen to avoid the influence of prior experience in conducting impairment examinations or using either version of *The Guides*. After the groups were trained, a third party, not affiliated with residency training, obtained informed consent, as approved by the IRB (Protocol #2010-0180).

Rating Impairment

Each person who agreed to participate was scheduled to sit in a room for up to 3 hours with a folder containing the 20 case descriptions and with the actual book of *The Guides*, fifth or sixth edition, depending on their assignment and training. They filled in a single impairment rating for each case on a data collection sheet.

Data Entry and Analysis

Participant code number, book edition number, and impairment ratings were entered into a MS Excel file and analyzed using Statistical Package for the Social Sciences for intra- and inter-class agreement within/between the fifth and sixth editions using nonparametric testing. We used the Shrout and Fleiss (1979)[13] formula to estimate intraclass correlations (ICC). This method assumes that all subjects are rated by the same raters who are the entire population of raters rather than a subset of raters or randomly assigned

to patients. We also tried Winer's calculations[14] for ICC and found almost no difference, so we report the former, here.

RESULTS

The average score across all raters using the fifth edition versus the average score using the sixth edition yielded a Pearson correlation coefficient of 0.88. The Pearson correlation coefficient for each rater of a given edition against the average for that edition ranged between 0.72 and 0.97 with 14 out of the 16 raters having coefficients above 0.85. In comparing the ICC for each individual rater against all other raters, the ICCs range between 0.478 and 0.85 and are all statistically significant. Table 1 shows the ICCs for comparisons within and between groups rating low back injury cases using the fifth versus the sixth edition. These range between 0.629 and 0.770.

DISCUSSION

Each state and territory in the United States has developed its own workers compensation system with the goal of protecting employers against civil law suits and exorbitant payments while providing expeditious and fair compensation to employees who have been injured or become ill at work. The impairment rating of those that have been injured in the workplace occurs within a milieu of contentiousness among employers, insurance companies, and employees, as well as attorneys who represent all three. *The Guides* has been developed in an attempt to lend objectivity to this process and to provide a logical, systematic, and fair basis for compensating those injured at work. The major changes between the fifth and the sixth editions claim to improve reliability and reproducibility. In addition, the sixth edition purportedly takes into account a broader model of disability and impairment put forth by the World Health Organization,[10] where physical impairment is only one of many determinants of participation in the workforce and in society, in general.

As each edition of *The Guides* has been published, there has been a mixture of acceptance and resistance to the recommended methodology for determining impairment citing the following: lack of a scientific basis for relating impairment to functionality; extremely limited testing of reliability, reproducibility, precision, and validity of *The Guides*; lack of inclusion of stakeholders, including attorneys, unions, legislators, along with an adequate array of clinical specialists, when writing *The Guides*; and inappropriate, or inadequately debated, acceptance of the World Health Organization model.[1,2,5,6–12,15,16] Despite the fact that there is a hot legal debate and great resistance among policy makers in state workers compensation systems to accept the sixth edition, we found no publications that compare the fifth and sixth editions of *The Guides*, and very few research publications that evaluate or compare any other versions.[2]

In this study, we found that the correlation of impairment ratings for low back injuries across raters–using the fifth edition alone, using the sixth edition alone, and comparing those using the fifth with those using the sixth–was relatively high and consistently so. This demonstrates fairly high inter-rater reliability for each edition and inter-rater reliability between the two editions.

The mean impairment ratings were somewhat lower using the sixth edition compared to the fifth. This suggests a more conservative approach to rating impairment in the newer version. Although this finding was not statistically significant, it is consistent with Colledge et al's account in 2009.[17] If accurate, injured employees stand to get less compensation when providers use the sixth edition, vs. the fifth.

There were two large outlier ratings among the raters in this study. Removing these numbers from the analysis did not alter the overall results. Furthermore, real life impairment rating is expected to yield occasional wide differences in the percentages assigned for individual cases.

There are several limitations to this study. First, we used mock cases on paper rather than real, face-to-face assessments; results might be different following an actual clinical encounter. Also, our case summaries provided information that would allow the rater to use the Diagnosis Related Evaluation, rather than the Range of Motion model in the fifth edition or the Diagnosis Based Impairment model in the sixth edition. Rater training on evaluating the back lasted only 1 hour in each edition and the participants had been introduced to the sixth edition of *The Guides* in 3 hour-long sessions, prior to being trained for this study. Results may be different if they had more extensive training, though this is probably representative of actual practice; in most states there is no formal training or certification of physicians that conduct Independent Medical Evaluations or rate impairment. We chose naïve raters and trained them to avoid study participants with a large amount of experience that might influence whether they actually followed the designated edition of *The Guides* vs. whether they were influenced by prior experience. In addition to a potential difference with experienced raters, it also is possible that different medical specialists (e.g., orthopedists, physiatrists, primary care, and occupational medicine physicians) assess impairment differently, even if they use *The Guides*.

CONCLUSIONS/RECOMMENDATIONS

The Guides fifth and sixth editions of the AMA are relatively reliable and consistent tools for rating impairment of low back injuries. The impairment ratings using the sixth edition of the AMA Guides are somewhat lower than the fifth and do not meet the claims made of improvement in reliability. Given the impact that impairment ratings, monetary settlements, and disability have on employers, employees, and society, it is critical to promote rigorous research to evaluate and refine these rating guidelines. An urgent research agenda should be implemented to answer the following questions:

1. What is the reliability and reproducibility of the rating guide for each organ system?
2. What is the "construct validity" of the ratings in terms of reflecting true functional impairment, ability to return to the same job and with the same pay, ability to return to work at all, quality of work life, and quality of life, in general? This would take a number of studies that include long-term follow-up of cases vis-à-vis quality of life and work life.
3. Does *The Guides* work well across the entire spectrum of ratings, from 0 to 100%? The construct validity for 50% to 100% is complex, given that many states judge that parties are "totally disabled" at less than 100%.

TABLE 1. Intra-Class Correlation Coefficients Comparing Impairment Raters Using the Fifth Versus Sixth Editions of AMA's Guides to the Evaluation of Permanent Impairment. Fleiss & Shrout Formulae Were Used

Comparisons	Correlation Coefficient (95% CI)
Among all raters, not controlling for version	0.629 (95% CI = 0.465, 0.795)
Among fifth edition raters, only	0.724 (95% CI = 0.56, 0.863)
Among sixth edition raters, only	0.650 (95% CI = 0.450, 0.820)
Average ratings of fifth vs. average ratings of sixth	0.770 (95% CI = 0.506, 0.902)

4. Is there a difference in ratings among physicians of the same specialty, among physicians of different specialties, or of treating physicians vs. independent medical examiners?
5. How does certification of examiners impact ratings, monetary settlements, and outcomes?
6. Do experienced physicians rigorously follow *The Guides* when evaluating patients or are they influenced by prior experience? if so, how much are they influenced and in which direction? This type of study could assess the objectivity of the rating process and evaluate, for example, whether subjective factors play a role (eg, experience, social/political view of occupational injury, etc).

REFERENCES

1. Chase K, Brigham RC, Ensalda LH, Talmage JB. *Evaluation of Impairment*. Clinics in Occupational and Environmental Medicine; 2001.
2. Brigham RC. *Study of the Impact on Changing from the American Medical Association (AMA) Guides to the Evaluation of Permanent Impairment*. Third Edition Revised to the Fourth or Fifth Editions in Determining Workers' Compensation Impairment Ratings. Performed at the request of the Labor and Employment The State of Colorado; 2002.
3. Haralson RH, III. Evaluation of the spine utilizing the AMA guides to the evaluation of permanent Impairment. 4th ed. *Instr Course Lect.* 1999;48:451–461.
4. Rondinelli RD, ed. *American Medical Association's Guides to the Evaluation of Permanent Impairment*. Chicago: American Medical Association; 2008.
5. Rondinelli RD, Changes for the new AMA guides to impairment ratings. 6th ed.: Implications and applications for physician disability evaluations. *PMR*. 2009;1:643–656.
6. McCarthy ML, McAndrew MP, MacKenzie EJ, et al. Correlation between the measures of impairment, according to the modified system of the American Medical Association, and function. *J Bone Joint Surv Am.* 1998;80:1034–1042.
7. Nitschke JE, Nattrass CL, Disler PB, Chou MJ, Ooi KT. Reliability of the AMA guide's model for measuring spinal range of motion. Its implication for whole-person impairment rating. *Spine (Phila Pa 1976).* 1999;24:262–268
8. Patel B, Buschbacher R, Crawford J. National variability in permanent partial impairment ratings. *Am J Phys Med Rehabil*. 2003;82:302–306.
9. Rondinelli RD, Katz RT. Merits and shortcomings of the AMA guides. 5th ed. a physiatric perspective. *Phys Med Rehabil Clin N Am*. 2002;13:355–370.
10. World Health Organization. International Classification of functioning, Disability, and Health (ICF). Geneva; 2001.
11. Spieler EA, Barth PS, Burton JF Jr, Himmelstein J, Rudolph L. Recommendations to guide revision of the guides to the evaluation of permanent impairment. American Medical Association. *JAMA*. 2000;283:519–523.
12. Dake M. *In Iowa Workers Compensation Commission Task Force on the AMA Guides Task Force*. Appendix D. Report of [a] Member. 2009. http://www.iowaworkforce.org/wc/amataskforce/2008amataskforcetoc.pdf, http://www.iowaworkforce.org/wc/amataskforce/dakereport.pdf. Accessed July 31, 2010.
13. Shrout PE, Fleiss JL. "Intraclass correlations: uses in assessing rater reliability." *Psychol Bull*. 1979;86:420–428.
14. Winer BJ. *Statistical Principles in Experimental Design*. New York: McGraw Hill;1971.
15. Kentucky workers compensation commissioner report [on the AMA guides. 6th ed]. 2009. Available at: http://www.labor.ky.gov/NR/rdonlyres/DC046C16-F0FF-4BB5-98B8-7C1CE7E92D02/0/CommissionerReportonAMAGuides.pdf. Accessed July 31, 2010.
16. Robinson JP, Turk DC, Loeser JD. Pain, impairment, and disability in the AMA guides. *J Law Med Ethics.* 2004;32:315–26, 191.
17. Colledge A, Hunter B, Bunkali LD, Holmes EB. Impairment rating ambiguity in The United States: the Utah impairment guides for calculating workers compensation impairments. *J Korean Med Sci.* 2009;24:S232–S241.

