

# FLAWED PROMISES: A CRITICAL EVALUATION OF THE AMERICAN MEDICAL ASSOCIATION'S GUIDES TO THE EVALUATION OF PERMANENT IMPAIRMENT

GUIDES TO THE EVALUATION OF PERMANENT IMPAIRMENT. Edited by A. Engelberg.<sup>1</sup> Chicago: American Medical Association. 1988. 3d ed. Pp. xxii, 254. \$45.00.

*Reviewed by Ellen Smith Pryor<sup>2</sup>*

Despite its medical authorship and seemingly technical subject, the American Medical Association's *Guides to the Evaluation of Permanent Impairment* deserves the careful attention of the legal community. Quietly but surely, this publication has come to play a dominant role in numerous compensation programs. For example, in recent years state after state has adopted the *Guides* as a standard for determining entitlement to workers' compensation benefits.<sup>3</sup> Perhaps more impor-

---

<sup>1</sup> Director, Department of Preventive Medicine, American Medical Association.

<sup>2</sup> Assistant Professor of Law, Southern Methodist University. I would like to thank Alan R. Bromberg, Richard J. Pierce, Jr., and Deborah A. Stone for their valuable comments. I also am grateful to C. Edwin Baker for encouraging the line of thought that led to this Review, and to Will Pryor for his indispensable help.

<sup>3</sup> In about a dozen states, the directive to use the *Guides* is statutory. See, e.g., ALASKA STAT. § 23.30.190 (Supp. 1989); GA. CODE ANN. § 114-101(5) (Supp. 1988); LA. REV. STAT. ANN. § 23:1221(4)(q) (West Supp. 1989). In many other states, administrative directives, rulings, or regulations specify use of the *Guides*. See, e.g., Gomez v. Industrial Comm'n, 148 Ariz. 565, 568, 716 P.2d 22, 25 (1986) (en banc) (discussing a rule adopted by the Arizona Industrial Commission calling for rating according to the *Guides*); Kroepflin v. North Dakota Workmen's Compensation Bureau, 415 N.W.2d 807, 808 (N.D. 1987) (discussing an administrative directive adopting the *Guides*). The states differ on whether use of the *Guides* is always, sometimes, or never mandatory. Compare MD. ANN. CODE art. 101, § 36C (Supp. 1989) (providing that until the Governor's Commission adopts impairment evaluation guidelines, physicians shall use the *Guides* "to measure all medical evaluations of permanent impairment") with N.M. STAT. ANN. § 52-2-3 (1987) (defining a "permanent physical impairment" as a condition that is "capable of being expressed in percentage terms as determined by medically or scientifically demonstrable findings" such as those presented in the *Guides* or comparable publications).

In addition, the use to which the *Guides* is put (for example, to rate whole-person impairment as opposed to organ-level impairment) varies among states. Compare ALASKA STAT. § 23.30.190 (Supp. 1989) (making the *Guides*' "whole-person" impairment ratings determinative of benefit amounts for partial disabilities) with LA. REV. STAT. ANN. § 23:1221(4)(q) (West Supp. 1989) (restricting awards of permanent partial disability to claimants whose loss of use or function is greater than 25% as evaluated in accordance with the *Guides*).

Although the *Guides* has been most heavily used by workers' compensation programs, its evaluative methods are neither specific to nor confined to workers' compensation. Instead, the *Guides*' rules can be used to calculate impairment for a variety of compensable losses. For an example of the *Guides*' use outside the workers' compensation context, see OKLA. STAT. ANN. tit. 11, §§ 50-101(11), 50-115 (West Supp. 1989), which uses the *Guides* to calculate disability retirement for certain state and municipal workers.

tantly, efforts to rate or schedule the effects of injury or disease will grow increasingly significant as the compensation debate continues to look favorably on replacing or drastically modifying the tort system.<sup>4</sup> As the dominant existing version of such a rating system, the *Guides* assumes an importance beyond its current use. The recent publication of an expanded and revised third edition<sup>5</sup> offers an opportunity to give this book the inspection it deserves and has not yet received.

I argue that the *Guides* is not the objective, medical evaluative system that it purports to be and that has been so appealing to legislators and other decisionmakers. Instead, like any impairment rating scheme, it rests in large part on important and difficult normative judgments. Yet the *Guides* obscures this from the reader; it is laden with hidden or poorly explained value judgments that frequently are gender-biased. The *Guides'* flawed promises of objectivity are especially troubling because they appeal to the craving of legislators and other decisionmakers for certainty and clarity in the difficult arena of impairment and disability assessment.<sup>6</sup> By uncritically em-

---

<sup>4</sup> Alternatives or modifications to the tort system need not include a scheduled approach to compensable losses, but many critiques look in this direction. See G. CALABRESI, *THE COSTS OF ACCIDENTS* 209-12 (1970) (discussing the possible use of schedules to establish the costs of various injuries); J. STAPLETON, *DISEASE AND THE COMPENSATION DEBATE* 163-67 (1986) (discussing options to base compensation in whole or in part on percentages of disablement); Sugarman, *Doing Away with Tort Law*, 73 CALIF. L. REV. 555, 646-47 (1985) (suggesting that future programs could follow the approach of workers' compensation by compensating partial disabilities). The appeal of scheduled approaches stems from factors such as reduced administrative costs and increased predictability of compensation, which enhance a system's ability to serve an insurance function. See AMERICAN LAW INST., *COMPENSATION AND LIABILITY FOR PRODUCT AND PROCESS INJURIES* 23 (1987) (discussing the potential advantages of following the workers' compensation approach for products or malpractice injuries).

<sup>5</sup> The *Guides* had its origin in 1956, when the AMA's Board of Trustees created an ad hoc committee to prepare "a series of practical guides for the rating of physical impairment of the various organ systems." AMERICAN MED. ASS'N, *GUIDES TO THE EVALUATION OF PERMANENT IMPAIRMENT* (2d ed. 1984) [hereinafter GUIDES II]. From 1958 to 1970, the committee published thirteen separate guides in the *Journal of the American Medical Association*. See *id.* at iii. In 1971 the AMA published the *Guides* as a single volume. See AMERICAN MED. ASS'N, *GUIDES TO THE EVALUATION OF PERMANENT IMPAIRMENT* (1971) [hereinafter GUIDES I]. The second edition, published in 1984, reflected the work of twelve expert panels that updated the clinical information supporting the ratings. See GUIDES II, *supra*, at iii. The third edition, published in 1988, is "substantially revised" (p. xvii). Reasons for the rapid revision include the growing literature on impairment, increased interest in impairment rating, and expanded use of the *Guides* (p. xvii).

The legal literature has virtually ignored the *Guides*. One thoughtful discussion of the *Guides*, however, appears in D. STONE, *THE DISABLED STATE* 108-17 (1984), in which Professor Stone notes that the *Guides* rests on "a pervading faith that a phenomenon of functional impairment, totally independent of context, can be precisely measured." *Id.* at 113.

<sup>6</sup> See FLA. STAT. ANN. § 440.15(3)(a)(3) (West Supp. 1989) (using such a schedule will "establish more certainty and uniformity in the rating of permanent impairment"); TENN. CODE ANN. § 50-6-204(d)(3) (Supp. 1988) (stating that physicians shall use the *Guides* "[t]o provide uniformity and fairness for all parties"); see also *Kroeplin v. North Dakota Workmen's Com-*

bracing the *Guides*, these decisionmakers delegate significant normative decisions to the book's medical authors.

This Review consists of three Parts: in Part I, I briefly describe the *Guides*' system. In Part II, I criticize several of its authors' key claims: that the book's system is objective and purely medical, and that evaluations made according to the *Guides* are accurate and reproducible. In addition, I explain that the book poorly identifies its underlying criteria. Finally, in Part III, I discuss those aspects of the book that are useful and defensible.

I do not contend that the effort to produce impairment ratings systems is implausible or useless. Indeed, in many contexts, such systems are justifiable.<sup>7</sup> In addition, not all the problems associated with use of the *Guides* stem from the book's own failings. The legal community frequently has misapplied the *Guides*; lawyers and judges often have misunderstood the consequences that may flow from injury or disease and the available methods of measuring those consequences.<sup>8</sup> But the *Guides*' own flawed promises, which help motivate and shape the legal community's adoption and use of the book, also deserve serious examination.

## I.

At its outset, the book defines impairment as an "alteration of an individual's health status that is *assessed by medical means*," (p. 2, emphasis in original) in contrast to disability, which is an alteration in one's capacity to meet personal, social, occupational, or legal demands, and "which is *assessed by nonmedical means*" (p. 2, emphasis in original).<sup>9</sup> The book does not purport to evaluate disability; instead, it sets out "rules" (p. 1) for measuring impairments resulting from injury or disease to virtually all bodily systems.<sup>10</sup> Each chapter from Three to Thirteen relates to a different bodily system. For example, a knee injury falls under the coverage of Chapter Three ("The Extremities, Spine, and Pelvis"). The chapter explains how to

---

pensation Bureau, 415 N.W.2d 807, 808 (N.D. 1987) (noting that the compensation bureau adopted the *Guides* "in order to establish more certainty and uniformity in the award of permanent impairment benefits"); *Corcoran v. Foster Auto GMC, Inc.*, 746 S.W.2d 452, 457 (Tenn. 1988) (observing that in adopting the *Guides*, the legislature intended to standardize impairment rating, minimize arbitrariness, and encourage predictability).

<sup>7</sup> For a more detailed discussion of the merits of schedules in various contexts, see Pryor, *Compensation and a Consequential Model of Loss*, 64 TUL. L. REV. (forthcoming 1990).

<sup>8</sup> See *id.*

<sup>9</sup> The *Guides*' glossary defines impairment somewhat differently: "the loss of, loss of use of, or derangement of any body part, system, or function" (p. 236).

<sup>10</sup> The *Guides* properly insists that it does not evaluate disability, and that disability is not a medical evaluation (p. 2). This Review argues that, contrary to the *Guides*' assertions, impairment is not a purely medical matter either. See *infra* pp. 968-70.

measure the knee impairment (evaluate flexion and extension with a goniometer) (p. 59); includes a chart that translates this data into "% Impairment of [the] Lower Extremity" (p. 61); and in turn translates this lower extremity impairment into "Impairment of [the] Whole Person" (p. 65). The rules in Chapter Six ("The Cardiovascular System") apply to a person with heart disease. A 15-25% whole-person impairment rating results if an individual has a history of myocardial infarction; has no symptoms while performing ordinary daily activities; may require medication or moderate dietary adjustment; and is able to perform an exercise test with certain findings (pp. 126-27).

These examples illustrate several points about the *Guides*' system. First, some but not all of the system-specific chapters include rules that can generate what I will term "organ-level" impairment ratings, such as the degree of impairment to the knee. Second, all of these chapters set out rules for arriving at a whole-person impairment rating. Third, some of the chapters, such as the cardiovascular system chapter, base the final whole-person rating on findings by the practitioner that include an assessment of how the patient performs daily living activities. By contrast, other chapters, such as the chapter that covers knee injuries, require findings (such as range of motion) that do not include ability to carry on daily activities. The *Guides* suggests that the whole-person rating is the most meaningful one: it "espouse[s] the philosophy that . . . all impairment ratings should be combined to be expressed as impairment of the whole person" (p. xviii). Hence, the book's primary rating objective is to generate a whole-person impairment percentage that represents the impairment consequence of injury or disease to virtually any bodily system.<sup>11</sup>

## II.

### *A. Objectivity and the Medical Nature of Impairment*

Chapter One ("Concepts of Impairment Evaluation") presents the principles and definitions that are central to the *Guides*. The chapter states that the user "must understand the concepts under which the 'rules' have been developed and the intended approach for using them to achieve objective, accurate, fair, and reproducible evaluations of individuals with medical impairments" (p. 1). Impairment, the chapter emphasizes, is a "medical matter" (p. 2), a "medical condition" (p. 6), a "medical fact" (p. 4), and is "assessed by medical means" (p. 2, emphasis in original). The book continues to stress the importance of

---

<sup>11</sup> Another key objective, discussed in Part III below, is to standardize the evaluation and reporting process.

Chapter One's concepts — virtually all the system-specific chapters begin by admonishing the reader to review those concepts.

The claims of objectivity and of the ratings' purely medical nature are deeply flawed. To understand why, one must explore the nature of impairment in more detail than the *Guides* provides. As the medical and rehabilitative literature makes clear, impairment is a meaningful concept at both the organ or system level and at the whole-person level.<sup>12</sup> For example, a knee injury may reduce the knee's functioning or scar its appearance, and also may reduce the person's ability to carry out integrated, whole-person-level activities, such as walking, bending, or dressing. The medical and rehabilitative literature, particularly that of recent years, provides many methods of measuring or classifying impairment at both the organ and whole-person levels.<sup>13</sup> Measures of organ-level impairment are based on comparisons between healthy and non-healthy appearance and function. As to whole-person impairment, most measurement methods take the form of "activities of daily living" (ADL) scales.<sup>14</sup> These scales evaluate a person's ability to carry out activities that are deemed relevant to the measurement.

This vast literature on measurement, at least implicitly, makes a critical point: it is impossible to define and measure impairment (both organ and whole-person level) except in relation to some norm.<sup>15</sup> Hence, any organ-level impairment evaluation requires (1) a decision about what organ-level qualities (*e.g.*, range of motion) are relevant to the measurement, and (2) a decision about what degree of that

<sup>12</sup> See, *e.g.*, WORLD HEALTH ORG., INTERNATIONAL CLASSIFICATION OF IMPAIRMENTS, DISABILITIES, AND HANDICAPS 27 (1980); Crewe, *Assessment of Physical Functioning*, in HANDBOOK OF MEASUREMENT AND EVALUATION IN REHABILITATION 235, 236-37 (B. Bolton 2d ed. 1987). The literature does not always use the same terminology, however. For example, under the World Health Organization's classification scheme, "impairment" refers to organ or system-level loss, see WORLD HEALTH ORG., *supra*, at 27; "disability" refers to person-level loss, see *id.* at 28; and "handicap" refers to individualized person-level losses, see *id.* at 29.

<sup>13</sup> See Frey, *Functional Assessment in the '80s*, in FUNCTIONAL ASSESSMENT IN REHABILITATION 11, 27-33 (A. Halpern & M. Fuhrer eds. 1984) (discussing the recent proliferation of functional assessment devices). The publication by the World Health Organization in 1980 of a classification scheme for the consequences of injury or disease was in part the result of greater focus by the medical community on chronic disease and injury. See WORLD HEALTH ORG., *supra* note 12, at 25.

<sup>14</sup> See, *e.g.*, M. MINOR & S. MINOR, PATIENT EVALUATION METHODS FOR THE HEALTH PROFESSIONAL 107-15 (1985); Alexander & Fuhrer, *Functional Assessment of Individuals with Physical Impairments*, in FUNCTIONAL ASSESSMENT IN REHABILITATION, *supra* note 13, at 51-52; Keith, *Functional Assessment Measures in Medical Rehabilitation: Current Status*, 65 ARCHIVES PHYSICAL MED. REHABILITATION 74, 74-77 (1984).

<sup>15</sup> See Keith, *supra* note 14, at 74 (asserting that functional measurement requires agreement on what is being measured, or "the domains of measurement"); *cf.* WORLD HEALTH ORG., *supra* note 12, at 27 (noting that impairment is "deviation from some norm in the individual's biomedical status"); *id.* at 28 (stating that disability [*i.e.*, the *Guides*' person-level impairment] represents "a departure from the norm in terms of performance of the individual").

quality will serve as the norm. Similarly, a whole-person impairment evaluation requires (1) a decision about what activities or qualities are relevant to the measurement, and (2) a decision about what levels of ability will serve as the norm.

The need for these decisions dooms the *Guides'* claims that its impairment evaluation scheme is objective and purely medical. These claims seem most appealing with respect to some organ-level impairments, such as those to the orthopedic extremities, where one might argue that a basic consensus exists on the qualities that should serve as the norm (such as range of motion). But even this argument does not justify the *Guides'* claims of objectivity and of its medical nature. Disagreement over the relevant norms is possible even for organ-level impairments. For example, the *Guides* usually does not consider chronic pain for purposes of an impairment arising from injury or disease to the extremities (pp. 37, 244). This may be a defensible choice, but it is neither a consensus view nor a purely medical decision.<sup>16</sup> In addition, the chief aim of the *Guides* is to generate whole-person ratings; even if consensus on organ-level ratings did exist, its value to the scheme would be limited.<sup>17</sup>

As to the whole-person ratings, both judgments necessary to the measurement — the relevant activities and the ability norm — are normative. Deciding which activities to count requires a value judgment about which activities are important or necessary to the person, or which activities are significant enough to warrant inclusion for compensation purposes. Hence, unsurprisingly, surveys of various ADL scales have found that only a few activities are common to all.<sup>18</sup> Measuring one's ability to perform a selected activity also requires a normative judgment: what level of ability to feed or clothe oneself, to walk, to play sports, or to have sex reflects the appropriate measurement norm?

Many of the book's examples and guidelines confirm that the ratings have a normative foundation: a strain of gender bias is fre-

---

<sup>16</sup> The *Guides'* appendix B, which addresses the topic of pain and impairment, notes that "[t]here has been little consensus and much confusion about the extent and nature of impairment that results solely from chronic pain and the chronic pain syndrome" (p. 243). One reflection of this lack of consensus is the report of the twenty-member Commission on the Evaluation of Pain, which was formed in response to the Social Security Disability Benefits Reform Act of 1984, 42 U.S.C. § 423(f)-(g) (1982 & Supp. 1989). The Commission's final report split on the issue of how disability caused by chronic pain should count for purposes of social security disability benefits. See COMM'N ON THE EVALUATION OF PAIN, REPORT OF THE COMMISSION ON THE EVALUATION OF PAIN 121-24 (1986).

<sup>17</sup> For many systems, the *Guides* gives only a whole-person impairment rating. See *supra* p. 967.

<sup>18</sup> See Frey, *supra* note 13, at 20-21 (concluding that few activities are common to all ADL scales); Keith, *supra* note 14, at 75 (explaining that although much literature agrees on the importance of measuring self-care and mobility, there is little agreement on what other elements should be measured).

quently evident. For example, in the reproductive system chapter, an impairment of the penis results in 5–10% whole-person impairment when “sexual function is possible, but there are varying degrees of difficulty of erection, ejaculation, and/or sensation” (p. 196). By contrast, the criteria for evaluating impairment of the vulva-vagina make it clear that a 0% whole-person impairment rating can result if “symptoms . . . do not require continuous treatment,” “the vagina is adequate for childbirth” during premenopausal years, and “sexual intercourse *is possible*” (p. 199, emphasis added).<sup>19</sup>

Examples in other chapters include a woman who “led a normal life caring for three children and her home” (p. 137); a woman who remained “able to do kitchen work, go shopping, and drive an automobile” (p. 139); a woman who remained “able to care for her house” (p. 128); a woman who was “able to do light housework” but did not wish to return to “her work as a seamstress” (p. 124); a woman who experienced increased “breathlessness during daily activities such as climbing stairs, mopping, or cleaning” (p. 132); and a woman who was “able to perform household duties” (p. 193). By contrast, examples of impairments of men include a man who “was an active participant in sports” (p. 138); a man who was “unable to participate in activities such as tennis and hiking” (p. 132); and a man who was able to play “18 holes of golf regularly” (p. 123).<sup>20</sup>

These passages have several implications. At a minimum, they establish that any whole-person measurement requires normative and not uncontroversial judgments about the activities or abilities that should matter for purposes of the measurement. They also illustrate that gender bias affects the *Guides*' system at two levels. The first is in the system's selection of the activities or abilities that count for purposes of arriving at the whole-person impairment ratings. Because, as I explain below,<sup>21</sup> the book does not adequately enunciate these underlying criteria, it is difficult to document gender bias at this level. But the reproductive system guideline that omits consideration

---

<sup>19</sup> The *Guides* gives an example of an “obese 38-year-old married woman” who has recurrent chronic dermatitis of the genitocrural area: “There is a remission of symptoms when her weight is controlled, when she avoids tight clothing, and when she observes careful hygienic measures. Sexual intercourse is possible if precautions are observed to avoid excessive vulvar irritation.” She has a 0% impairment of the whole person (p. 199). Later examples that give rise to impairments do mention a woman's sexual sensation (p. 199).

<sup>20</sup> These examples appear in chapters, such as that on the cardiovascular system, that require the practitioner's initial evaluation of the patient to include some form of ADL assessment. Many of the book's ADL examples are not as detailed as those cited in the text; instead, they simply refer to the patient's ability to perform usual or normal daily activities without specifying what those activities include (pp. 178–79, 181). When the book does include examples of ADL assessments, and when those examples specify some of the measured daily activities, stereotyped images appear quite frequently.

<sup>21</sup> See *infra* p. 972.

of a woman's sexual sensation is simply the most obvious example of how differing world perspectives may lead to different choices of relevant abilities or activities.

Gender bias also influences the *Guides* at another level: the stereotyped activity examples reveal the ratings rules' susceptibility to biased application. To realize why, one must understand the roles that ADL measurements play in the *Guides*. The examples related earlier — the knee injury and heart disease — can illustrate the two different ways that ADL measurement is used. As to the knee injury, the *Guides* provides initial measurement guidelines that do not require the practitioner herself to make an ADL assessment. Instead, the measurements that must be obtained relate to organ-level qualities such as range of motion. The book then provides tables that translate the organ-level data into the whole-person impairment rating. Although the book does not so state, these ratings must rest on the *authors'* assessment of how the measured losses (*e.g.*, range of motion loss) affect the person's ability to carry on activities of daily living.

By contrast, the chapter on the cardiovascular system shows that bias is also possible when the *physician* must make an ADL assessment. As to injuries covered by this chapter, the practitioner's initial evaluation of the patient — before arriving at the *Guides'* whole-person rating — requires, among other measures, at least a rough assessment of the individual's ability to perform daily living activities. To determine the whole-person impairment of a person with a history of myocardial infarction, the practitioner must assess whether the patient has symptoms while performing ordinary daily activities as well as make various physical findings. The *Guides* then translates this evaluation into a whole-person impairment rating between 15% and 25%. Hence, in some but not all chapters the practitioner's initial data-gathering calls for some form of ADL assessment.

The stereotyped activity examples show that gender bias can play a role whenever the practitioner must make some sort of ADL assessment. Suppose, for example, that a practitioner performs an impairment evaluation on a female patient with heart disease who was previously active in athletics. If the physician's image of a woman's typical daily activities corresponds to or is actually informed by those in the book — cleaning, mopping, taking care of her house, sex without sensation — she might not inquire about and evaluate the condition's effect on athletic activities. The patient thus will receive an impairment rating that does not take account of this functional loss. By contrast, when evaluating a male patient the practitioner might well include reduced athletic capacity.

The *Guides* is partly responsible for this potential for biased application of the impairment guidelines. First, the book's vagueness about its underlying criteria permits an individual practitioner's own perspective to frame the inquiry about relevant life activities. Second,

the book's examples give at least implicit and possibly direct encouragement to evaluating women and men according to stereotypical assumptions.

Given the normative nature of the judgments on which a ratings scheme depends, a book such as the *Guides* should clearly identify its underlying choices and criteria. But the book does not do so. Chapter One never plainly states that the book's whole-person ratings essentially reflect ADL measures; instead, the chapter's definition of impairment omits any reference to activities of daily living.<sup>22</sup> The glossary contains a definition of activities of daily living (p. 235),<sup>23</sup> but it leaves far too many questions unanswered. First, the specification of certain activities does not make clear what the activities include. For example, the glossary lists sexual function as an activity of daily living, but defines it as having "normal sexual function and participating in usual sexual activity," without further specifying what constitutes "normal" or "usual" sexual activity (p. 235). Second, it seems doubtful that the ratings system is solely governed by the glossary's list of activities. For example, some chapters refer to activities that do not appear on the list, such as occupational tasks (p. 170), housework (p. 128), and shopping (p. 139). In addition, some chapters refer to the individual patient's usual daily activities, in contrast to a uniform scale.<sup>24</sup> Third, the glossary gives no sense of the relevant scales of performance ability, and no explanation as to the relative weights of different activities in the overall whole-person rating. The reader does not know, for example, whether the *Guides'* final ratings give equal weight to self-care, transportation, and sexual function, or whether these activities receive different weights, and, if so, what those weights are. Finally, the *Guides* gives little or no explanation of how it "combines" two or more whole-person ratings to arrive at a single rating for a person suffering from multiple impairments.<sup>25</sup>

In sum, the *Guides* first obscures the reliance of whole-person ratings on ADL measurements, and then gives only a vague explication of its underlying criteria. What could justify this approach? The

---

<sup>22</sup> Interestingly, earlier editions of the *Guides* stated at the outset that impairment evaluation requires an ADL assessment. See GUIDES I, *supra* note 5, at iii; GUIDES II, *supra* note 5, at vii.

<sup>23</sup> The glossary lists the following: self-care and personal hygiene, communication, normal living postures, ambulation, travel, nonspecialized hand activities, sexual function, sleep, and social and recreational activities. The glossary also gives examples of these activities.

<sup>24</sup> For example, the *Guides* refers to "his or her usual activities" (p. 126) and "the patient's daily living activities" (p. 145).

<sup>25</sup> The book contains a "combined values" chart, which generates a single whole-person impairment rating when two or more impairments affect the person (pp. 246-48). The foreword refers to this as the "familiar 'Combined Values Chart,'" (p. xviii), but gives little explanation of what premises and criteria underlie the chart. For a criticism of the combined values chart in earlier editions, see D. STONE, cited above in note 5, at 115-16.

AMA cannot argue that its whole-person ratings do not rely on some assessment of ADL impact; evidence of such reliance appears throughout the book.<sup>26</sup> The AMA cannot claim that consensus on the appropriate ADL criteria renders precise identification of those criteria unnecessary: dozens of ADL scales are currently in use, and there is an active debate about their relative validity, reliability, and usefulness.<sup>27</sup> Finally, the AMA cannot argue that the book's partial aim to provide information to a nonmedical audience justifies a less detailed enunciation of its criteria. In fact, the book's use by the legal community makes careful identification of the normative foundation of the ratings scheme even more important.

### B. Accuracy and Reproducibility

The *Guides* suggests that use of its rules can achieve "accurate, fair, and reproducible" evaluations: "the user . . . must understand the concepts under which the 'rules' have been developed and the intended approach for using them to achieve objective, accurate, fair, and reproducible evaluations of individuals with medical impairment" (p. 1). It seems plausible to interpret the "accurate, fair, and reproducible" assertion to mean that with proper use the *Guides'* system will be valid and reliable. Validity essentially means that an evaluative instrument measures what it aims to measure to some acceptable degree, or, more precisely, that inferring certain conclusions from the results of the instrument is appropriate.<sup>28</sup> Reliability refers to the quality of reproducibility.<sup>29</sup>

Because an instrument may be valid for one purpose and not for another,<sup>30</sup> the validity of the *Guides* depends in part on its use. Let us suppose, then, that the *Guides* is used only for the purpose that the book itself suggests — to measure a purely "medical" condition. But the book's validity for this purpose is difficult to defend once one

---

<sup>26</sup> The digestive system chapter notes that its criteria evaluate impairment "according to a person's ability to perform activities of daily living" (p. 177). The chapters on the cardiovascular (p. 119) and urinary (p. 189) systems make a similar point.

<sup>27</sup> See Frey, *supra* note 13, at 21 (discussing studies of numerous ADL scales developed over the years); Kaufert, *Functional Ability Indices: Measurement Problems in Assessing Their Validity*, 64 ARCHIVES PHYSICAL MED. REHABILITATION 260, 260 (1983) (discussing the inadequacy of validity data in the area of disability assessment); Keith, *supra* note 14, at 74-77 (discussing the need to improve the validity and reliability of functional assessment measures).

<sup>28</sup> See Rothstein, *Measurement and Clinical Practice: Theory and Application*, in MEASUREMENT IN PHYSICAL THERAPY 1, 15-16 (J. Rothstein ed. 1985). See generally AMERICAN EDUC. RESEARCH ASS'N, AM. PSYCHOLOGICAL ASS'N & NAT'L COUNCIL ON MEASUREMENT EDUC., STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING 9-18 (1985) (discussing various means of accumulating evidence to measure the validity of a test for its intended purpose).

<sup>29</sup> See Rothstein, *supra* note 28, at 5.

<sup>30</sup> See *id.* at 16.

acknowledges that impairment is not solely a medical condition.<sup>31</sup> Problems remain even if one understands the *Guides* to make the more modest claim that its rules are accurate and fair as a measure of person-level abilities. The book's failure to explain its underlying criteria makes it difficult to ascertain its validity even for this more limited purpose.<sup>32</sup> In testing the validity of the *Guides* as a measure of the abilities that do or should matter to people, one should recall the reproductive system guideline and the stereotyped-activity examples. These passages, and the likelihood of gender bias that they imply, show that ratings performed in accordance with the book's guidelines will not always rest on a representative and fair sample of the activities or abilities that do or should matter to people.<sup>33</sup> Thus, one cannot defend the *Guides'* validity even under the more modest premise about its proper use.

The validity claim seems more defensible with respect to some of the organ-level evaluative rules. At least some data support the validity of certain organ-level evaluations, and the *Guides* defines fairly clearly the criteria underlying those evaluations. But many of the chapters do not contain organ-level ratings, and the book's primary rating objective is to express impairment at the whole-person level.<sup>34</sup>

The merits of the reproducibility claim differ among the various chapters. At least with respect to certain chapters — such as the extremities, spine, pelvis, and respiratory system — the claim of reproducibility seems plausible because the necessary evaluations require the use of tools (goniometer, inclinometer) in fairly precise and reproducible ways. Some of the chapters on other systems, however, require input that seems quite likely to vary from one evaluator to the other. For example, the cardiovascular system chapter requires an assessment of the patient's ability to perform "ordinary daily activities" (p. 121); the section on speech requires an assessment about the individual's "efficiency in using speech to make himself or herself understood in daily living" (p. 173); and the chapter on the reproductive system requires evaluation of sexual dysfunction (pp. 196–200). That the evaluation of some systems requires a less reproducible

---

<sup>31</sup> See *supra* p. 969.

<sup>32</sup> One article concluded that the *Guides* (its second edition) is acceptably valid and reliable as to measurement of hand impairment. See Gloss & Wardle, *Reliability and Validity of the American Medical Association's Guides to Ratings of Permanent Impairment*, 248 J. A.M.A. 2292 (1982). Although the article gives a persuasive analysis, using a form of criterion validation on the *Guides*, it evaluates only the hand section of the *Guides*.

<sup>33</sup> This type of validity inquiry, known as content validation, examines the measurement tool and asks whether it seems fairly to represent the content of the domain that it allegedly measures. See Rothstein, *supra* note 28, at 19–20.

<sup>34</sup> I am not suggesting that the book should provide more organ-level rules, or that organ-level impairment ratings are preferable. The question of which levels and types of impairment a ratings system should aim to measure cannot be fully addressed here.

measurement process is neither surprising nor avoidable, but it does make troubling Chapter One's undifferentiated claim of reproducibility.

Fortunately, at times the individual chapters acknowledge some of the limitations of their evaluative rules. The cardiovascular system chapter, for example, notes that information about daily living activities "is subjective, and it is open to interpretation on the part of both patient and examiner" (p. 119).

### III.

Despite these serious flaws, the book has some useful features. As explained above, it provides numerous guidelines for measuring various organ or system-level qualities.<sup>35</sup> Although the book's translations of this sort of data into impairment ratings pose the many problems discussed above, its elucidation of various evaluative techniques could be helpful in a number of contexts. For instance, in devising a loss assessment schedule, a compensation program could use these guidelines to evaluate the losses that the program deems relevant for compensation purposes. Or a system that compensates for lost wage-earning capacity in an unscheduled fashion could establish evidentiary guidelines that require, in addition to other items of evidence, the results of evaluations performed in accordance with certain of the book's guidelines.

Another useful feature of the book is its emphasis on the need to improve the report writing process. The book stresses the importance of providing a complete report of the clinical findings, and it suggests that physicians examining clinical reports specify the basis of a finding of "disability" or "impairment" (p. 7). These are useful admonitions in an area so fraught with disparate assessment definitions and techniques. But one must remember that the choice of evaluative norms is not a "medical" matter, and that many sets of relevant norms are possible in different social and legal settings. Hence, the goal of standardizing the evaluative process to the extent possible, though sensible and important, cannot displace the need to make difficult social policy judgments; at best, some standardization of evaluation and reporting is possible once those judgments have been made.

In sum, a close look at the *Guides* reveals lessons both about its use and more generally about other ratings schemes. Because the *Guides* so poorly sets out its underlying criteria, and because so many of its judgments are suspect on their surface, legal systems should

---

<sup>35</sup> For example, the book explains how to measure and record loss of motion in various orthopedic extremities with a goniometer (pp. 14, 17); how to measure spinal range of motion with an inclinometer (pp. 71-74); and how to measure loss of vision by means of particular charts and equipment (pp. 153-64).

make use of the *Guides* only in the specific and defensible ways described in this Review. On a broader level, this Review of the *Guides* shows that any effort to create a compensation schedule will involve important normative judgments about the types of functions or qualities that matter in evaluating impairments, and the significance of the loss of these functions. Medically or scientifically authored ratings schemes can offer useful insights into the nature and measurement of various sorts of losses. Hence, when developing loss assessment systems, legislatures and administrative bodies can and should learn much from consultation with the medical community. But policymakers should appreciate that the search for an "objective," "accurate," or purely "medical" system is and always will be fruitless.<sup>36</sup> And they must resist the seductive but false hope that use of a scientifically or medically authored ratings system can bypass the need to make the hard choices necessary for any loss assessment system.

---

<sup>36</sup> Unfortunately, as Professor Stone notes, "the quest for an objective method of medical evaluation of disability has a long history and continues into the present." D. STONE, *supra* note 5, at 116. She argues that, instead of attempting to find objective criteria, "one can try to examine how particular constructs and measures *systematically* exclude certain understandings and include others, how they serve the political interests of some groups at the expense of others, and how they work to produce particular types of policy results." *Id.* at 117 (emphasis in original).