JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1990, 23, 99–110 NUMBER 1 (SPRING 1990)

THE SELF-INJURY TRAUMA (SIT) SCALE: A METHOD FOR QUANTIFYING SURFACE TISSUE DAMAGE CAUSED BY SELF-INJURIOUS BEHAVIOR

BRIAN A. IWATA
UNIVERSITY OF FLORIDA

GARY M. PACE
NEW MEDICO REHABILITATION CENTER OF FLORIDA

ROBERT C. KISSEL
BEHAVIOR CONSULTANTS, INC., BLOOMINGTON, INDIANA

PAUL A. NAU
NEW MEDICO REHABILITATION CENTER OF FLORIDA

AND

JON M. FARBER
ALEXANDRIA-LAKERIDGE PEDIATRIC CENTERS, ALEXANDRIA, VIRGINIA

A method is described for classifying and quantifying surface tissue damage caused by self-injurious behavior. The Self-Injury Trauma Scale permits differentiation of self-injurious behavior according to topography, location of the injury on the body, type of injury, number of injuries, and estimate of severity. Fifty pairs of independently scored records were subjected to interrater reliability analyses, and the following mean (median) percentage agreement scores were obtained: overall agreement, 97% (98%); location of injury, 99% (100%); type of injury, 96% (100%); number of injuries, 89% (100%); and severity of injury, 94% (100%). Percentage agreement also was calculated for three summary scores: Number Index, 90%; Severity Index, 92%; and Estimate of Current Risk, 100%. Potential applications and limitations of the scale are discussed.

DESCRIPTORS: assessment, injury measurement, self-injurious behavior, Self-Injury Trauma Scale

The dependent variable of choice in assessment and treatment research on self-injurious behavior (SIB) is based on some dimension of responding, such as frequency, duration, or percentage of time intervals during which the SIB occurs. These measures are preferred over subjective ratings of improvement because they are accurate indicators of behavior and are highly sensitive to changes in responding over time. Nevertheless, measures based on the observable outcome of SIB, rather than the ongoing behavior itself, can be useful in the course of conducting research or in evaluating the effects of clinical programs.

By definition, SIB "produces injury to the individual's own body" (Tate & Baroff, 1966, p. 281), and the social relevance of the behavior therefore lies in its traumatic outcome. The measurement of physical injuries prior to treatment can establish the fact that a client or subject actually displays behavior warranting serious attention, because a given response topography occurring at a certain frequency and intensity should produce physical damage in order to be considered self-injurious. Conversely, injury measurement following treatment can corroborate observed changes in behavior, because reduction of an injury-producing response below a certain level should be reflected in the

This research was supported by Grant HD16052 from the National Institute of Child Health and Human Development. Reprints and instructions on scale administration may be obtained from Brian A. Iwata, Psychology Department, University of Florida, Gainesville, Florida 32611.

eventual disappearance of observable trauma. In both of these instances, data on injuries provide a means of assessing social validity (Wolf, 1978).

Injury measures also can be used to estimate the risk associated with SIB, which may provide an additional basis for the selection of intervention procedures. For example, high-risk behavior must be prevented through restraint or other protective means and may require treatment with quick-acting but potentially intrusive interventions either immediately or when less intrusive methods have failed. The continued occurrence of low-risk behavior, on the other hand, might be tolerated until an effective yet nonintrusive solution to the problem is identified. These views are common in the literature on SIB, yet methods by which a discrimination can be made between high- and low-risk behaviors have never been identified. Risk is typically inferred from information about the frequency of a behavior, but frequency per se may correlate poorly with actual physical damage, which also is a function of intensity. Measures that take into account both frequency and intensity would be helpful in determining risk, and the extent of trauma reflects the interaction of these dimensions of behavior.

Finally, in situations where researchers and clinicians rely on data collected by others (e.g., out-of-clinic or follow-up reports by teachers or parents), periodic measurement of injuries provides a secondary means (the preferred being periodic direct observation) of assessing maintenance and/or compliance with treatment recommendations. Parental reports of compliance (i.e., low occurrences of behavior) given the continued *absence* of injury in their child may or may not be accurate but are consistent; such reports given the continued *presence* of injury (or further physical deterioration) clearly require direct verification.

In spite of these potential advantages, procedures for measuring injury are virtually nonexistent in research on SIB. For example, we were able to find only two studies in which a subject's injuries were actually counted (Carr & McDowell, 1980; St. Lawrence & Drabman, 1983). The dependent variable—number of skin abrasions directly observed or noted in a medical chart—appeared to be an adequate outcome measure, but it did not provide information on the severity of the behavior, nor is it generally applicable across response topographies. In the present study, we describe the development and evaluation of the Self-Injury Trauma (SIT) Scale, which provides an objective estimate of location, type, number, and severity of visible (surface) tissue damage caused by SIB.

DEVELOPMENT AND USE OF THE SIT SCALE

Scale Development

Construction of the SIT Scale was based on four sources of input: published material on injury classification, our previous experience in the evaluation and treatment of over 200 self-injurious individuals, application of preliminary versions of the scale with 9 pilot subjects, and consultation from physicians. These processes are described briefly.

Injury measurement is common in several medical specialties, particularly emergency medicine. We initially examined the most frequently used trauma scales, including the Abbreviated Injury Scale (Committee on the Medical Aspects of Automotive Safety, 1971), the Anatomic Index (Champion et al., 1980), the Anatomical Injury Code (Milholland & Crowley, 1979), the Comprehensive Research Injury Scale (Committee on the Medical Aspects of Automotive Safety, 1972), the Injury Severity Score (Baker, O'Neill, Haddon, & Long, 1974; Barancik & Chatterjee, 1981), the Modified Injury Severity Scale (Mayer, Matlak, Johnson, & Walker, 1980), the Trauma Chart (Greenspan, McLellan, & Greig, 1985), and the section on injury codes contained in the International Classification of Diseases—Adapted (U.S. Department of Health, Education, and Welfare, 1968). All of these instruments are well suited to the collection of epidemiological data on injuries seen in hospital emergency rooms and shock-trauma centers, where the extent of injury varies widely and the severity ranges from minor to fatal. Consequently, the scales are comprehensive but they lack the sensitivity needed to differentiate SIB, because most SIB clusters at the "minor" end of the continuum. For

example, many of the items found in trauma scales involve determination of whether an anatomical part is ruptured, crushed, or completely severed, and trauma of this type is rarely produced by SIB. The primary information gathered from reviewing these scales was classification of injuries by location on the body. Because the SIT Scale was intended for noninvasive use, all of the selected locations were externally visible.

Surface (skin) tissue damage generally is classified in one of four categories: abrasions, burns (chemical or thermal), contusions, or lacerations (Baker, O'Neill, & Karpf, 1984; Grossman, 1984; Trott, 1983). Although abrasions and lacerations differ in origin, their appearance often is similar in that an extensive but localized abrasion is functionally the same as a laceration (i.e., both produce a distinct break in the skin). These two categories were therefore collapsed into one. Another category, burns, was deleted because of its extremely low prevalence as SIB. The resulting classification by type of injury yielded two categories: abrasions/lacerations (AL) and contusions (CT).

Although number of wounds per se does not greatly alter the risk of an injury, to some degree it reflects the extent of a disorder such as SIB. Our initial attempts at quantifying this aspect of SIB involved counting each wound found on pilot subjects. This practice was not very useful for two reasons. First, number of wounds is somewhat confounded with type of wound. Forceful and repeated banging to the same location on the head may produce only one large contusion; by contrast, a single bite to the skin usually leaves several small lacerations, and an abrasion may produce a dozen or more small lesions. Second, although an individual who exhibits multiple topographies of SIB produces a large number of wounds overall, we have found that the injury at a given location usually is limited. Therefore, we assigned rankings based on the number of wounds we have observed typically: 1 = one wound (common in a mild case of SIB but rare in a severe case), 2 = two to four wounds (common), and 3 = five or more wounds (rare).

Injury severity usually is scored on a relatively subjective basis with labels such as "mild," "moderate," and "severe," accompanied by descriptions of the observed state of the anatomy. More quantitative methods were explored in the development of the SIT Scale but eventually were abandoned. For example, we initially used flexible plastic rulers to measure the length of abrasions/lacerations, but it was impossible to measure depth, which is a major factor determining the severity of this class of injuries, noninvasively. Similarly, we estimated the area of contusions by placing cutouts of known dimensions over the affected body part, but this measure yielded no information on height (degree of swelling) or tissue rupture, which contribute to the severity of injuries produced by pressure. Although descriptions of the state of a wound contain some arbitrary elements, they are no more subjective than behavioral definitions in that they specify observable criteria for the occurrence of an "event." Thus, the determination of severity for abrasions/lacerations and contusions was based on descriptions developed by a physician (one of the authors), which were consistent with definitions found in standard texts (Grossman, 1984; Trott, 1983) and those used in common practice. These descriptions were revised based on pilot observations and input from other physicians.

Scoring Procedures

The current version of the SIT Scale is reproduced in Figure 1. Part I involves identifying the topographic aspects of SIB and is similar to information obtained for most self-injurious individuals. In addition, evidence of healed injuries is documented for two reasons. First, we have found that many self-injurious individuals are either restrained or maintained in situations where SIB is prevented. This artificially reduces observable trauma, in which case healed injuries may be a more accurate reflection of the extent of damage to be expected if the self-injurious individual were unrestrained, unattended, or unmedicated. Second, a healed injury may suggest a previous form of SIB that no longer occurs but might reappear in the future. Only the first five categories of injuries listed in Part I—forceful contact (two types), scratching, biting, and

THE SELF-INJURY TRAUMA (SIT) SCALE

Patient:_____ Examiner:_____ Date:_____

PART I. GENERAL DESCRIPTION AND SUMMARY OF HEALED INJURIES

Check each type of self-injurious behavior exhibited by the patient. Next, note any physical evidence of healed injuries (scars, permanent disfigurement, missing body parts), along with the specific site.

Self-Injurious Behaviors:

___ Forceful contact with head or face
___ Forceful contact with other body part
___ Scratching, picking, rubbing skin
___ Biting
___ Eye gouging
___ Ingestion of inedible materials (pica)
___ Vomiting or rumination
___ Air swallowing (aerophagia)
___ Hair pulling (trichotillomania)
___ Other:_____

Healed Injuries:

1 _____
2 _____
3 _____
4 _____
5 _____

PART II. MEASUREMENT OF SURFACE TRAUMA

For each area of the body containing a current (unhealed) injury, identify the location and number of wounds, and note the type and the severity of the worst wound at that particular location.

Number: Score: 1)--One wound
2)--Two-four wounds
3)--Five or more wounds

Type: Abrasion or Laceration (AL): A break in the skin, either superficial or deep, caused by tearing, biting, excessive rubbing, or contact with a sharp object.

Contusion (CT): A distinct area marked by abnormal discoloration or swelling, with or without tissue rupture, caused by forceful contact.

Severity: Score AL as: 1)--Area is red or irritated, with only spotted breaks in the skin.
2)--Break in the skin is distinct but superficial; no avulsion.
3)--Break in the skin is deep or extensive, or avulsion is present.

Score CT as: 1)--Local swelling only or discoloration without swelling.
2)--Extensive swelling.
3)--Disfigurement or tissue rupture.

(scoring chart on next page)

Figure 1. The Self-Injury Trauma (SIT) Scale.

PART II (CONTINUED)

LOCATION		NUMBER	TYPE	SEVERITY	COMMENT
Head:	Scalp	1 2 3	AL CT	1 2 3	
	Ear L/R	1 2 3	AL CT	1 2 3	
	Eye L/R	1 2 3	AL CT	1 2 3	
	Eye Area L/R	1 2 3	AL CT	1 2 3	
	Face	1 2 3	AL CT	1 2 3	
	Nose	1 2 3	AL CT	1 2 3	
	Lips/Tongue	1 2 3	AL CT	1 2 3	
	Neck/Throat	1 2 3	AL CT	1 2 3	
Upper Torso:	Shoulder L/R	1 2 3	AL CT	1 2 3	
	Chest/Stomach	1 2 3	AL CT	1 2 3	
	Back	1 2 3	AL CT	1 2 3	
Lower Torso:	Abdomen/Pelvis	1 2 3	AL CT	1 2 3	
	Hips/Buttocks	1 2 3	AL CT	1 2 3	
	Genitalia	1 2 3	CL CT	1 2 3	
	Rectum	1 2 3	AL CT	1 2 3	
Extremities:	Upper Arm/Elbow L/R	1 2 3	AL CT	1 2 3	
	Lower Arm/Wrist L/R	1 2 3	AL CT	1 2 3	
	Hand/Finger L/R	1 2 3	AL CT	1 2 3	
	Upper Leg/Knee L/R	1 2 3	AL CT	1 2 3	
	Lower Leg/Ankle L/R	1 2 3	AL CT	1 2 3	
	Foot/Toe L/R	1 2 3	AL CT	1 2 3	

PART III. SCORING SUMMARY

A. Number Index (NI)

From Part II, add all of the scores under the Number column and enter the total here:____

	NI Score		Part II Total
(circle)	0	→	No injuries
	1	→	1 - 4
	2	→	5 - 8
	3	→	9 - 12
	4	→	13 - 16
	5	→	17 or more

B. Severity Index (SI)

From Part II, enter the frequency of scores from the Severity Column: 1:____; 2:____; 3:____

	SI Score		Severity Scores from Part II
(circle)	0	→	No injuries
	1	→	All severity scores are 1's
	2	→	One 2; No 3's
	3	→	Two or more 2's; No 3's
	4	→	No more than one 3
	5	→	Two or more 3's

C. Estimate of Current Risk Based on Location and Severity

☐	LOW →	No injuries or: Any AL-1, CT-1, or AL-2 except near eyes
☐	MODERATE →	Any AL-2 near eyes, Any CT-2 except on head
☐	HIGH →	Any CT-2 on head, Any AL-3 or CT-3

eye gouging—are relevant to later scoring. Pica, vomiting, and air swallowing produce damage that is not externally visible; and hair pulling, although classified as a self-injurious response in the research literature, produces negligible trauma.

Part II involves the actual observation and recording of current injuries. For each location indicated, the examiner first notes whether an injury is present, counts the injuries at a given location, and assigns a ranking (1, 2, or 3) that corresponds to the number of wounds present. Next, the most severe injury at each location is identified as an abrasion/laceration (AL) or contusion (CT), and then is given a rank score (1, 2, or 3) for severity, based on separate criteria for AL and CT wounds. Classification by type and severity is limited to the worst wound because less serious wounds are not considered in determining overall severity and risk.

Part III summarizes the data obtained in Part II. The first score, the Number Index (NI), is based on the total number of injuries observed. The index is calculated by adding all of the scores in the "Number" column in Part II and assigning a ranked score from 0 to 5. The second score is the Severity Index (SI), which is based on the relative occurrence of severity scores in Part II. Severity scores limited to 1 result in an SI value of 1, severity scores of 2 result in SI values of 2 or 3, and severity scores of 3 result in SI values of 4 or 5.

The final summary index in Part III is the Estimate of Current Risk. In lieu of using weighted or transformed summary scores as the basis for determining risk, Part III allows a conservative estimate based on the single occurrence of certain types of injuries. For example, an extensive abrasion or deep laceration (AL-3), contusion resulting in skin rupture (CT-3), or extensive swelling on the head (CT-2) places the individual in a category of high risk, which requires immediate medical attention and subsequent prevention or rapid elimination of the behavior. Injuries falling into the categories of moderate or low risk, on the other hand, require medical attention as needed and periodic monitoring, but probably can be allowed to occur when conducting behavioral assessments or exposing self-injurious individuals to therapeutic contingencies during treatment. We do not suggest, however, that low-risk injuries can be ignored; these also require intervention to prevent the occurrence of further and perhaps more serious trauma.

PRELIMINARY APPLICATION OF THE SCALE

Method

Subjects

Thirty-five subjects participated. Their ages ranged from 3 to 19 years and their functional abilities varied considerably. All but one exhibited either multiple topographies of SIB or the same topography on several locations on the body. Across all subjects, eye gouging was the only form of SIB amenable to scoring that was not represented. Subjects were examined in several contexts: preliminary screening for possible inclusion in treatment, screening prior to the start of a treatment program, or examination during the course of treatment or at follow-up.

Procedure

Each subject was examined one to three times by two observers (authors), yielding a total of 50 pairs of records. Repeated examinations were scheduled at least 1 week apart and averaged 24 days apart (range, 7 to 69 days). This schedule decreased the likelihood that an observer would score the same injuries during subsequent examinations.

The examination procedure was conducted as described previously, with three exceptions. First, Part I was eliminated because it did not enter into any of the reliability analyses. Second, none of the subjects had injuries to the buttocks, genitalia, or rectum upon initial examination, and these areas were skipped when a subject underwent an additional examination. Third, all injuries (not just the worst) at a given location were scored for type and severity in order to assess scoring reliability. The two observers worked independently, with the reliability exam conducted immediately following the primary exam. At least one observer recorded the start and stop times to obtain an estimate of the

amount of time required to conduct an examination. These times averaged 11.7 min per exam.

Evaluation of Interrater Reliability

The 50 pairs of records were compared on an item-by-item basis, and percentage agreement scores were calculated on observers' overall scoring, as well as on individual categories (location, type, number, and severity) and summary scores (number index, severity index, current risk). In each case, we selected the most stringent formula available.

Overall reliability. To obtain a measure of observer reliability across the entire scale, each scorable item from Part II (location, type of injury, number of injuries, and severity of injury) was examined for agreement or disagreement. For each location examined, five opportunities for agreement were possible: presence or absence of injury (one item), the score for number (one item), type of injury (two items), and the score for severity (one item). Therefore, reliability was calculated by dividing the number of exact agreements by the number of locations examined ($\times 5$) and multiplying by 100. This calculation, although a conservative estimate of overall agreement, did not reflect the fact that observers might be highly reliable in one scoring category (e.g., number of injuries) but not in another (e.g., type of injury). Therefore, separate reliability percentages were calculated for each scoring category.

Reliability on location of injury. This category was examined to determine whether or not observers agreed on the presence or absence of an injury at a given location. Reliability was calculated by dividing the number of agreements on the presence or absence of injury by the number of locations examined and multiplying by 100.

Reliability on number of injuries. Reliability was calculated by dividing the number of agreements on identical ranking by the number of locations at which an injury was scored and multiplying by 100. Instances in which both observers agreed that there was no injury at a given location (blank box agreements) were deleted; in this respect the calculation is equivalent to occurrence or scored interval reliability (Repp, Deitz, Boles, Deitz, & Repp, 1976). In addition, because an agreement was scored only if both observers entered the same ranking (1, 2, or 3), this calculation was equivalent to exact agreement for scored intervals (Repp et al., 1976).

Reliability on type of injury. If an observer scored the presence of an injury, two entries were possible (AL and/or CT). Therefore, reliability for this category was scored by dividing the number of agreements on type of injury by the number of locations at which an injury was scored ($\times 2$) and multiplying by 100. Again, instances where no injury was scored by both observers were deleted.

Reliability on severity of injury. Reliability was calculated in a manner similar to that for number of injuries by dividing the number of agreements on identical ranking by the number of locations at which an injury was scored and multiplying by 100.

Reliability on summary scores. Because the three summary scores obtained in Part III—Number Index, Severity Index, and Estimate of Current Risk—are further reductions of data obtained from the actual scoring of injuries done in Part II, it was expected that high reliability for injury scores would be reflected in the summary scores. To verify this assumption, reliability was calculated for each of the three summary scores by dividing the number of agreements on identical summary ranking by the number of paired records scored and multiplying by 100.

Results

Summary of Scored Records

Although the actual scores obtained from applying the SIT Scale in this study do not necessarily reflect injury prevalence among the general population of developmentally disabled individuals who exhibit SIB, they do provide information about a sample of individuals whose SIB was judged to be serious enough to warrant referral for treatment. Figure 2 shows the distribution of summary scores for NI, SI, and current risk based on the first application of the SIT Scale to each of the 35 subjects (scores from the 15 repeated examinations were excluded). As indicated by the NI and SI

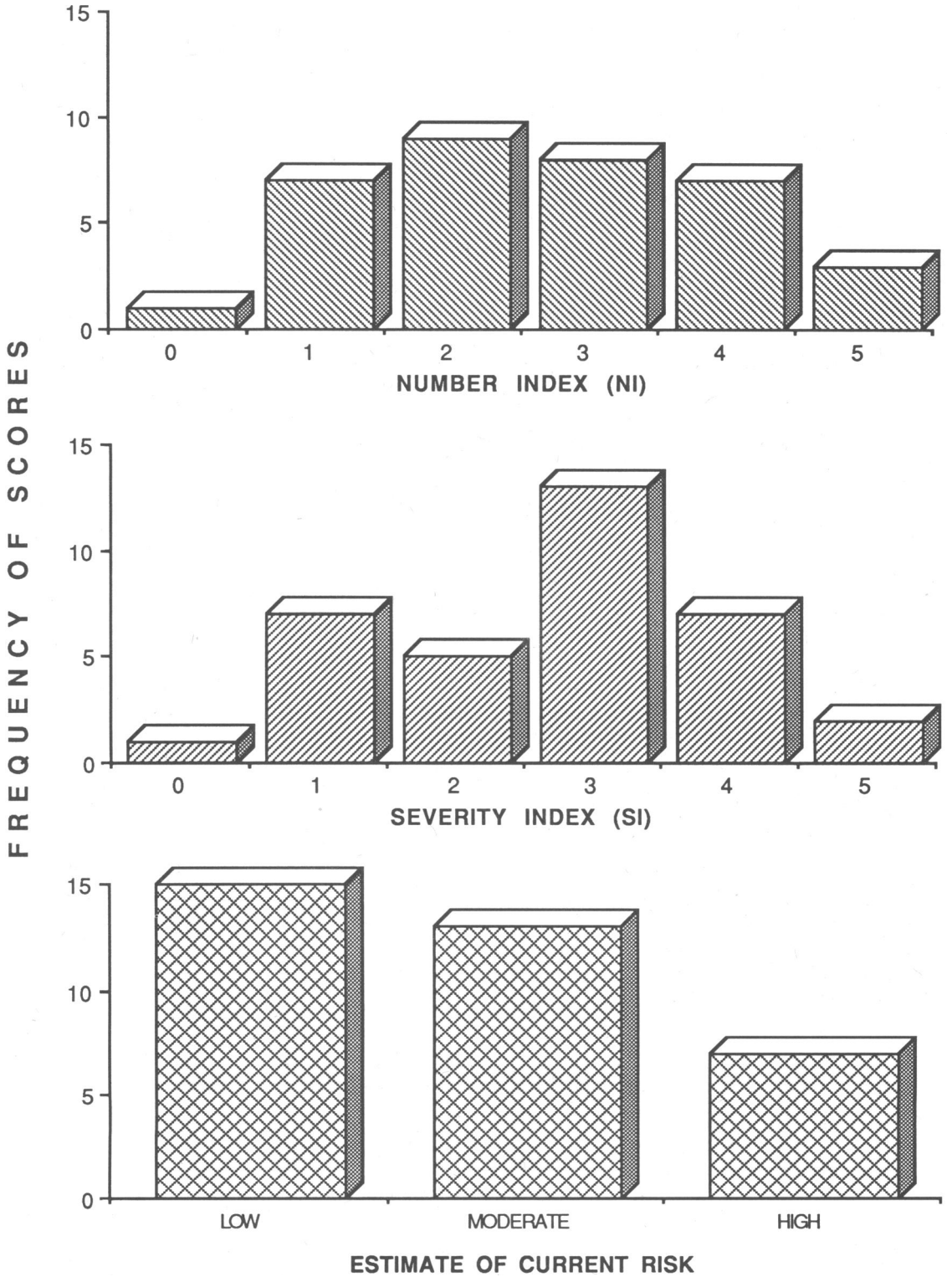

Figure 2. Frequency distributions across 35 scored SIT records for Number Index (NI), Severity Index (SI), and Estimate of Current Risk.

distributions, this group of self-injurious individuals showed wide variation in terms of both number and severity of SIB. One individual, for example, exhibited mild face slapping that produced no observable trauma; by contrast, other individuals (often those exhibiting multiple forms of SIB) had significant and numerous injuries on as many as 5 to 10 body locations. The results for current risk indicate that most subjects' (28 of 35) SIB was judged to pose low to moderate risk.

Figure 3 shows a scatter plot of obtained SI values as a function of obtained NI values for the 35 subjects. There is a noticeable relationship between injury number and severity, in that subjects who had more injuries also tended to have more serious ones. For example, 16 individuals received NI scores of 1 or 2; none of these received an SI score higher than 3. Similarly, 10 individuals received NI scores of 4 or 5, all of whom received an SI score of 3 or higher. Although these data suggest that injury number and severity are related in a positive way, they should be interpreted cautiously because there is no evidence to indicate any causal relationship. Thus, it is quite possible that, for example, an individual who exhibits severe headbanging will show relatively few injuries.

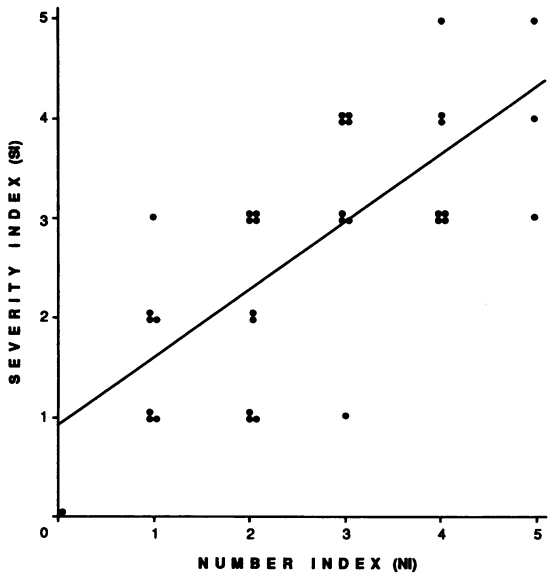

Figure 3. Scatter diagram (with a line-of-best-fit curve) of obtained SI (severity) values as a function of obtained NI (number) values across 35 scored SIT records.

Reliability Analyses

The results obtained for all reliability calculations are presented in Table 1. All of the category (Part II) mean scores were above 85%, which is considered an acceptable level of agreement given the stringent methods of comparison that were used (e.g., exact agreements on occurrence). Examination of the ranges, however, reveals the existence of some low scores for the categories of type, number, and severity. Had we included blank boxes (agreement on nonoccurrence) in the calculations, the low numbers would have disappeared. As is the case sometimes with occurrence agreement, the low scores were obtained when observers recorded very few injuries, and one of them either missed a wound because it was small or gave it a different ranking. The fact that these discrepancies occurred rarely is reflected in the median reliability scores, which indicate that over half of all records were in perfect agreement. Of a total of 250 reliability calculations (overall plus the four individual categories × 50 records), only 15 yielded an agreement score below 80%.

Agreement results obtained for the three summary scores (NI, SI, and current risk) were 90% or higher. For NI and SI, there were five and four disagreements, respectively, and in every case these were one number apart (e.g., a subject received an SI of 2 based on one observer's scores and an SI of 3 based on the other observer's scores). There were no disagreements for current risk.

DISCUSSION

Results of this study indicate that the Self-Injury Trauma (SIT) Scale is a reliable method for collecting data on surface tissue damage caused by SIB. The advantages of the SIT Scale include its objectivity, its applicability to the most commonly seen forms of SIB, and its provision for documenting multiple aspects of surface injury (location, type, number, and severity). Given these characteristics, the SIT Scale may be helpful in documenting the extent of damage caused by SIB, in determining initial risk, as a secondary source of

Table 1
Percentage Agreement Scores Obtained Across 50 Pairs of Scored SIT Scale Records

Scoring Category (Part II)	Mean	Median	Range	Summary Category (Part III)	Agreement score
Overall Reliability (all categories)	97	98	86–100	Number Index (NI)	90
Reliability on Location of Injury	99	100	86–100	Severity Index (SI)	92
Reliability on Type of Injury	96	100	70–100	Estimate of Current Risk	100
Reliability on Number of Injuries	89	100	40–100		
Reliability on Severity of Injury	94	100	50–100		

data in substantiating successful treatment, or perhaps as a primary source of data in epidemiological studies or follow-up situations in which it might be impossible to conduct direct observation.

The data obtained from administration of the SIT Scale in this study are not surprising, because it is common knowledge that the extent of SIB varies considerably. The advantage of using a measurement instrument such as the SIT Scale lies in its ability to define variation quantitatively. Although the difference between a scaled score of 1 versus 2 is arbitrary, it is based on numerical and objective criteria; this represents a clear improvement over current practice.

The results for current risk were interesting in that most of the self-injurious individuals in a referred population were judged to pose low or moderate risk. These results suggest that a significant proportion of SIB does not produce serious injury. Thus, institutional policies requiring that all SIB must be prevented, blocked, or immediately followed by medical treatment may not always be in clients' best interest. Although such practices are consistent with cultural notions of humane care, they often preclude implementation of therapeutic programs in which a contingency is arranged for the presence or absence of SIB; that is, if the behavior cannot occur, it (or its absence) will not contact the contingency. Furthermore, well-intentioned actions by parents, therapists, and others may have the inadvertent effect of reinforcing the unwanted behavior. Response prevention, comfort, and other forms of contingent attention may serve as positive reinforcement; similarly, interrupting ongoing activities in order to provide minor wound care may serve as negative reinforcement (i.e., escape). An alternative method of providing routine medical care for self-injurious individuals whose injuries are minor might consist of arranging checks and wound care on a fixed- or variable-time schedule that is independent of behavior.

Although the results presented here support a number of potential uses for the SIT Scale, the instrument has several limitations that must be acknowledged. First, measurement of SIB based on outcome alone cannot be substituted for data obtained through direct observations of behavior. Physical trauma caused by SIB and reflected in SIT scores is a function of several factors: response topography, intensity, and frequency, as well as the affected location on the body. Consequently, trauma indices provide little information about any single dimension of behavior and may be insensitive to either immediate or relatively small but important changes in the frequency or intensity of SIB. We did not examine formally the relationship between short-term changes in behavior and observable injury, but our experience has been that trauma is visible for hours and even days after all SIB has ceased completely. It is also possible for SIB to increase, within certain limits, without necessarily producing more trauma. This lack of close correspondence between SIB as a response versus an outcome makes SIT scores more suitable as a basis for evaluating follow-up rather than short-term treatment effects.

A second limitation of the SIT Scale is its insensitivity to internal injuries. Some forms of SIB

(e.g., pica) are known to cause internal damage, and other forms produce internal as well as external damage. For example, head banging, in addition to producing observable trauma, can result in skull fractures, hemorrhage, or retinal damage. Individuals who exhibit SIB that may affect internal anatomy and physiology (e.g., aerophagia, chronic vomiting, rumination, pica, and injuries resulting from forceful contact) should be referred to specialists for further evaluation.

Finally, the assignment of risk based on SIT scores is determined solely by the extent of damage that has resulted from past episodes of SIB. In this respect, predictions of risk based on the SIT Scale are no different than those arrived at through more informal means or even objective data on response frequency and intensity. None of these methods takes into account the possibility that a future episode of behavior may either increase in intensity, as in mild head banging that becomes more severe, or alter slightly in topography, as in face punching that shifts to the eyes. These and other changes may be very difficult or impossible to predict and present a compelling case for continued involvement of medical professionals in the management of severe SIB.

Future research on the use of the SIT Scale might focus on several issues related to development and application. For example, although the Number Index and Severity Index are straightforward numerical summaries, the Estimate of Current Risk is based on judgments of damage likely to result from certain types of injuries. Another method would be to assign risk based on the actual consequences of the behavior (e.g., no treatment required, minor wound care [topical dressing or bandage], outpatient physician care [minor surgery or casting], and inpatient care). In practice, however, most policies require immediate action—through restraint or medication—to prevent serious injuries resulting from SIB. Thus, risk markers based on medical consequences can be derived only from instances in which preventive efforts failed; this would require an extremely large data base.

Another area of research involves determining the relationship between behavior change and observable trauma. Our informal experience has been that the onset of baseline or treatment conditions produces an increase in injuries due to the fact that many individuals, prior to the implementation of an intensive treatment program, are typically restrained. In addition, as noted previously, there appears to be a delay between observed reductions in behavior and the disappearance of trauma. More systematic documentation of these anecdotal observations is required.

SIB is a complex disorder and is somewhat unique from a behavioral standpoint in that it poses both behavioral and medical risks. As such, its assessment and treatment must be based on the most direct and complete methods of observation, and the SIT Scale is not designed to serve as a replacement for either behavioral data or medical screening. On the other hand, the present data indicate that the SIT Scale is an appropriate alternative, general, and more systematic method for arriving at informal decisions that will be made inevitably.

REFERENCES

Baker, S. P., O'Neill, B., Haddon, W., & Long, W. B. (1974). The Injury Severity Score: A method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma,* **14,** 187–196.

Baker, S. P., O'Neill, B., & Karpf, R. S. (1984). *The injury fact book.* Lexington, MA: Lexington Books.

Barancik, J. I., & Chatterjee, B. F. (1981). Methodological considerations in the use of the Abbreviated Injury Scale in trauma epidemiology. *Journal of Trauma,* **21,** 627–631.

Carr, E. G., & McDowell, J. J. (1980). Social control of self-injurious behavior of organic etiology. *Behavior Therapy,* **11,** 402–409.

Champion, H. R., Sacco, W. J., Lepper, R. L., Atzinger, E. M., Copes, W. S., & Prall, R. H. (1980). An Anatomic Index of injury severity. *Journal of Trauma,* **20,** 197–202.

Committee on the Medical Aspects of Automotive Safety. (1971). Rating the severity of tissue damage: The Abbreviated Scale. *Journal of the American Medical Association,* **215,** 277–280.

Committee on the Medical Aspects of Automotive Safety. (1972). Rating the severity of tissue damage: The Comprehensive Scale. *Journal of the American Medical Association,* **220,** 717–720.

Greenspan, L., McLellan, B. A., & Greig, H. (1985). Abbreviated Injury Scale and Injury Severity Scale: A scoring chart. *Journal of Trauma,* **25,** 60–64.

Grossman, J. A. (1984). *Minor injuries and disorders: Surgical and medical care.* Philadelphia: J.B. Lippincott.

Mayer, T., Matlak, M. E., Johnson, D. G., & Walker, M. L. (1980). The Modified Injury Severity Scale in pediatric multiple trauma patients. *Journal of Pediatric Surgery,* **15,** 719–726.

Milholland, A. V., & Crowley, R. A. (1979). Anatomical Injury Code. *American Surgeon,* **45,** 93–100.

Repp, A. C., Deitz, D. E. D., Boles, S. M., Deitz, S. M., & Repp, C. F. (1976). Differences among common methods for calculating interobserver agreement. *Journal of Applied Behavior Analysis,* **9,** 109–113.

St. Lawrence, J. S., & Drabman, R. S. (1983). Interruption of self-excoriation in a pediatric burn victim. *Journal of Pediatric Psychology,* **8,** 155–159.

Tate, B. G., & Baroff, G. S. (1966). Aversive control of self-injurious behavior in a psychotic boy. *Behaviour Research and Therapy,* **4,** 281–287.

Trott, A. (1983). *Principles and techniques of minor wound care.* New Hyde Park, NY: Medical Examining Publishing Co.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis,* **11,** 203–214.

U.S. Department of Health, Education, and Welfare. (1968). *International classification of diseases (adapted), 8th edition.* (DHEW Publication No. PHS 1693). Washington, DC: U.S. Government Printing Office.

Received June 26, 1989
Initial editorial decision August 23, 1989
Revision received September 19, 1989
Final acceptance October 31, 1989
Action Editor, David P. Wacker