Clinical Aphasiology, Vol. 22, 1994, pp. 107–118

Duke *Prognosis Profile Worksheet:* Inter- and Intraobserver Reliability

Jennifer Horner, Mary Ann Eller,
Deborah V. Dawson, and
Frances G. Buoyer

In clinical aphasiology, prognostication is a process by which clinicians assess a range of biographic, neurologic, linguistic, and behavioral data for the purpose of predicting a patient's future communication performance. Prognostication may pertain to the anticipated benefits of spontaneous recovery, the anticipated benefits of treatment, or both (Darley, 1982; Marquardt, 1982). According to Wertz (1978), "The task . . . is to make a prospective statement for each patient individually" (p. 26). Both Wertz (1978) and Darley (1982) advised clinicians to ask, *prognosis . . . for what, for whom,* and *at what point in time?* Horner and Rothi (1984) stated, "Because no single factor is sufficient for estimating prognosis, and because variables interact in complex and individual ways, we recommend rating as many variables as possible" (p. 24). One goal of a comprehensive approach is to avoid "test-driven" prognostication in favor of patient- and clinician-driven prognostication. A final introductory point is that—at least in the acute phase of recovery—prognosis for communication recovery is a dynamic phenomenon, not a static one. In Marquardt's (1982) words, "The clinician's responsibility is not always to be right; rather it is to make the best judgment possible and to be willing to change an opinion if the patient's situation warrants it" (p. 104).

The literature pertaining to prognosis is too extensive to review here, but the foundation for our worksheet can be found in Darley (1982); Hier, Mondlock, and Caplan (1983); Holland, Greenhouse, Fromm, and Swindell (1989); Horner and Rothi (1984); Kertesz (1979); Levin, Benton, and Grossman (1982); Marquardt (1982); and Rosenbek, LaPointe, and Wertz (1989). One recent study (Holland et al., 1989) reported the acute language recovery of left and right hemisphere stroke patients. Using the *Western Aphasia Battery* (Kertesz, 1979, 1982), they defined language recovery as an aphasia quotient of 93.8 or better. They submitted select variables to a multiple logistic regression analysis and found that language recovery favored younger patients, shorter hospital stays, males, hemorrhagic strokes, and right hemisphere strokes.

The purpose of this study was to assess the inter- and intraobserver reliability of a prognosis worksheet designed to help us prognosticate recovery from communication disability in adults sustaining brain damage. The prognosis worksheet was adapted from Horner and Rothi (1979, 1984). In the context of current clinical practice, the original prognosis worksheet had several shortcomings. First, the worksheet helped the clinician describe but not quantify relevant parameters. Second, there were no data on inter- or intraobserver reliability. Third, several developments in the clinic mandated a change in our perspective on the question of prognosis. The first of these developments was increased diversity in our patient caseload, from narrowly defined "aphasic" syndromes to a host of poststroke, postsurgical, posttraumatic, and post-"other" etiologies. The second development was the introduction of the concept of "cognitive-communication" disorders. Despite the 1991 position statement of the American Speech-Language-Hearing Association (ASHA), clinicians lack consensus regarding the definition or measurement of cognitive-communication disorders. The third development was the gradual accumulation of research concerning prognosis for recovery. The fourth development—a result of the increasing cost of health care—was the shorter duration of hospital stays for the purpose of rehabilitation.

All of these developments provoked questions relevant to our clinical practice and research design. First, should we use one test battery for all patients? Second, should we use one set of prognostic indicators for all patients? Third, what is the reliability of the prognostic estimate that we generate? And fourth, what is the predictive validity of the prognostic estimate?

Clinicians treating adult neurogenic communication disorders are faced with an ever-increasing diversity of theory, clinical perspectives, and caseload composition (Duffy & Myers, 1991), yet we do not yet have a consistent or satisfactory approach or nomenclature. Therefore, we decided to use a standard admission battery for all patients. Further, we decided to use a common prognosis worksheet for all patients, regardless of etiology, side of brain damage, or severity of brain damage. We deferred the question of predictive validity to a later time. In this study we addressed the clinicians' lack of a systematic, reliable approach to estimating prognosis for recovery from communication disorders during acute rehabilitation. The specific question of this study was: what is the reliability of the *Duke Prognosis Profile Worksheet*?

METHOD

Subjects

During the 12-month study period, 128 patients admitted to Duke Rehabilitation Center received evaluations by speech-language pathologists. Excluded from this study were 70 patients with isolated dysarthria or dysphagia, as well as patients who were too severely impaired to tolerate standardized language testing. Of the remaining 58 patients (Table 1), 37 were men, 21 were women, and all but one were right-handed. Stroke—left, right, or bilateral—was the most frequent etiology; the nonstroke patients were mostly neurosurgical and head trauma patients.

Table 1. Characteristics of Patients for Whom Prognostic Estimates Were Derived

	Diagnostic Category			
Patient Characteristic	*LHS*[a] (N = *18*)	*RHS*[b] (N = *14*)	*BHS*[c] (N = *10*)	*Nonstroke* (N = *16*)
Gender				
Male	14	9	6	8
Female	4	5	4	8
Handedness				
Right	17	14	10	16
Left	1	0	0	0
Education (years)				
< 6–8	5	2	3	3
9–12	7	5	2	8
13–16, college	5	6	5	4
> 16, college	1	1	0	1
Age (years)	61.8	62.4	65.6	57.5
SD	[12.9]	[13.0]	[13.2]	[16.6]
Days post onset				
Mean	39.4	52.9	18.8	40.1
Median	29.5	26.5	14.5	30.0
Range	5–190	10–335	9–40	18–138
Western Aphasia Battery				
Aphasia Quotient				
Mean	40.2	92.0	76.8	87.1
SD	[3.0]	[7.3]	[23.6]	[19.8]
Range	1.0–94.0	70.8–99.6	9.9–94.2	14.2–100.0
Cortical Quotient				
Mean	40.2	83.7	69.6	81.2
SD	[28.6]	[11.6]	[21.3]	[18.9]
Range	3.1–87.7	59.4–94.8	10.6–91.9	52.5–96.5

[a]LHS = left hemisphere stroke. [b]RHS = right hemisphere stroke. [c]BHS = bilateral hemisphere stroke.

Materials

The *Prognosis Profile Worksheet* (Appendix A) consists of two pages. The front page allows a summary of etiologic and neurodiagnostic data and notes aphasia type (if applicable), number of hospital days, former occupation, and so on. The second page, the worksheet per se, comprises 24 variables yielding four subscores. The first set of data (the top half of the form) pertains to demographic and neurologic information, each assigned a relative weight of 1, 2, or 3 points. The remaining subscores (on the bottom half) pertain to language performance, other higher cortical functions, and visuomotor functions—each assigned a relative severity rating of 1 to 5 points. By tallying these subscores, the clinician derives the total prognosis score (maximum 100 points) and then assigns an overall qualitative estimate for extent of improvement in communication.

Procedure

The standard admission battery for the *Speech and Language Pathology Program* at Duke Rehabilitation Center included the *Neurobehavioral Cognitive Status Examination* (1983) and the *Western Aphasia Battery* (Kertesz, 1979; 1982). Both tests have proven reliability and validity (Kiernan, Mueller, Langston, & Van Dyke, 1987; Shewan & Kertesz, 1980), and they allow evaluation of both language and nonlanguage abilities. A variety of clinician-selected supplementary tests and informal observations supplemented standardized testing. At the completion of the evaluation, the patient's primary clinician completed a *Prognosis Profile Worksheet*.

For the interobserver reliability question, five clinicians—*in pairs*—participated in this study of 58 patients. Each had a master's degree, a Certificate of Clinical Competence, (CCC), and a state license. The range of experience was from 1 to 9 years; the average was 4.6 years. The primary clinician (the primary observer) evaluated the patient, completed test forms in a routine manner, wrote a comprehensive diagnostic report including an addendum summarizing test scores, and then completed the two-page *Prognosis Profile Worksheet*. Subsequently, a randomly assigned independent observer who had never met or had no more than incidental knowledge of the patient reviewed background information, test data (including test record forms), and the diagnostic report. From this review, the independent observer completed a *Prognosis Profile Worksheet*.

For the intraobserver reliability question, one observer completed prognosis worksheets on 27 patients. She had appropriate qualifications and 4 years of experience. She was blinded to her original, baseline worksheet, and she was blinded to the actual treatment outcome.

RESULTS

Tables 2, 3, and 6 summarize statistical analyses for interobserver reliability; Tables 4, 5, and 7 do the same for intraobserver analyses. First, we calculated Spearman Rank Correlation Coefficients for all 58 pairs of observations for the four category subscores and the total score (see Table 2). We also looked at *intraclass* coefficients, which treated the interobserver observations as replicates and ignored the specific roles of primary and independent observers. (These were calculated using ranks for comparability with the Spearman coefficients.) All correlations were significant at $p < .0001$.

Next, we investigated possible systematic differences between observers (see Table 3). These comparisons were made by examining the differences between the scores of the primary and independent observers and testing whether the median difference was zero using the nonparametric Wilcoxon signed rank procedure. Again, this was done for the subscores and the total scores. In no instance was the result significant at the 5% level. In fact, the only result that could even be termed suggestive was for language performance ($p = .06$). However, even here the median difference was zero. The observers agreed exactly in 20 of 58 cases and within one point in 45 of 58 cases.

We then conducted parallel analyses for intraobserver reliability. In Table 4 we report Spearman and intraclass coefficients for initial and repeat scores. As in the case of interobserver reliability, we found all intraobserver correlations to be statistically significant. Table 5 presents comparisons between subscores and total scores using the Wilcoxon procedure. In only one instance—that for the Demographic/Neurologic category—was the

Table 2. Correlation Coefficients Between Primary Clinician and Independent Observer Using the *Prognosis Profile Worksheet* Contrasted With Intraclass Correlation Coefficients Based on Ranks

Category	*Spearman Correlation Coefficients**	*Intraclass Correlation Coefficients**
Demographic/lesion	.857	.857
Language performance	.909	.906
Other higher cortical	.896	.900
Visuomotor	.890	.891
Total	.968	.965

Note: $N = 58$.

*All correlations were significant at $p < .0001$. The p-values are those associated with the test of the null hypothesis of zero correlation.

Table 3. Comparisons Between Primary and Independent Observers on *Prognosis Profile Worksheet* Scores Using the Nonparametric Wilcoxon Signed Rank Procedure

Worksheet Component	p*	*Mean*	*S.E.*	*Median*
Subscore Category				
Demographic/lesion	.60	–.07	.15	0
Language performance	.06	.34	.22	0
Other higher cortical	.88	0.33	.29	0
Visuomotor	.91	–.19	.24	0
Total	.79	.09	.38	0

*p-values are those associated with the test of the null hypothesis that the median difference is zero.

Table 4. Correlation Coefficients Between Initial and Repeat Scores Using the *Prognosis Profile Worksheet* Contrasted With Intraclass Correlation Coefficients Based on Ranks

Category	*Spearman Correlation Coefficients**	*Intraclass Correlation Coefficients**
Demographic/lesion	.940	.915
Language performance	.756	.767
Other higher cortical	.812	.782
Visuomotor	.901	.901
Total	.902	.895

Note: $N = 27$.
*All correlations were significant at $p < .0001$. The p-values are those associated with the test of the null hypothesis of zero correlation.

result significant at the 5% level, though even here the median difference score was zero. Scoring agreed perfectly in this category for 10 of 27 records. Disparities suggested that the p-value actually reflected *skewness* in disagreement: that is, the initial score was less than the repeat score in 4 cases and greater than the repeat score in 13 cases.

The prognosis worksheet allows the clinician to generate a prognosis score and a prognosis estimate. The score is based on 24 categorical variables yielding 100 possible points. The qualitative prognostic estimate is based on the clinician's overall appraisal of the individual's potential for improvement, that is, as excellent, good, fair, guarded, or poor. Using these data, the kappa statistic is a measure of agreement, corrected for the agreement one would expect by chance: A kappa of 0 indicates a level of

Table 5. Comparisons Between the Initial and Repeat Scoring by One Observer on *Prognosis Profile Worksheet* Using the Nonparametric Wilcoxon Signed Rank Procedure

Worksheet Component	p*	*Mean*	*S.E.*	*Median*
Subscore Category				
Demographic/neurologic	.013	–.52	.19	0
Language performance	.480	–.07	.29	0
Other higher cortical	.139	1.04	.61	0
Visuomotor	.688	.04	.17	0
Total	.128	1.52	.87	1

*p-values are those associated with the test of the null hypothesis that the median difference is zero.

Table 6. Assessment of Interobserver Agreement for the Overall Prognostic Estimate of Improvement After Completion of the *Prognosis Profile Worksheet*

Observed Agreement	*Unweighted Kappa*	*S.E.*	p*
44.8%	.231	.075	.002
Observed Agreement	***Weighted Kappa***	***S.E.***	**p***
93.1%	.735	.128	.001

Note: Acceptable agreement is defined as either precise agreement or disparity of only one *prognosis estimate* category. The *prognosis estimate* categories are excellent, good, fair, guarded, and poor. $N = 58$.
*p-values are associated with the test of the null hypothesis that the level of agreement observed does not differ from that expected by chance alone.

agreement no different from that expected by chance alone; a kappa of 1 indicates perfect agreement.

As shown in Table 6, the first and second observer agreed on the prognostic estimate precisely 44.8% of the time (this represents 26 of 58 cases). The weighted kappa showed that the two observers agreed in 54 of 58 cases, or 93.1%. (The first observer rated "lower" in 16 instances; the second observer rated "lower" in 12.) For interobserver comparisons, both unweighted and weighted kappas were statistically significant.

As shown in Table 7, intraobserver agreement using the kappa statistic was based on 40.7% observed agreement; the weighted kappa, on 96.3% observed agreement. The weighted kappa achieved statistical significance; the unweighted did not.

Table 7. Assessment of Intraobserver Agreement for the Overall Prognostic Estimate of Improvement After Completion of the *Prognosis Profile Worksheet*

Observed Agreement	*Unweighted Kappa*	*S.E.*	p*
40.7%	.196	.105	.062
Observed Agreement	*Weighted Kappa*	*S.E.*	p*
96.3%	.860	.137	.0001

Note: Acceptable agreement is defined as either precise agreement or disparity of only one *prognosis estimate* category. The *prognosis estimate* categories are excellent, good, fair, guarded, and poor. $N = 27$.
*p-values are associated with the test of the null hypothesis that the level of agreement observed does not differ from that expected by chance alone.

DISCUSSION

After having observer pairs formed from a group of five speech-language pathologists with an average of about 5 years experience evaluate 58 acute rehabilitation patients, we compared both scores and estimates derived from the *Prognosis Profile Worksheet*. In addition, we compared both scores and estimates made by a single observer for 27 patients.

We conclude from our data that the *Prognosis Profile Worksheet* has acceptable inter- and intraobserver reliability.

We offer several caveats for this research study. First and foremost, we recognize that some of the variables comprising the Demographic/Neurologic subscore of the *Prognosis Profile Worksheet* have controversial prognostic significance—handedness, sex, age, education, and hemiplegia, to name a few. In turn, our assignment of 3-point ratings to these and other variables may be blatantly premature. Isolating these controversial variables in a logistic regression model will help us rectify this. Second, the *Prognosis Profile Worksheet,* although comprehensive, is not exhaustive. For example, we excluded from the worksheet *per se* information on lesion *site* and ratings of patient motivation because of the difficulty of fitting these complex variables into a quantitative approach. Third, we are aware of Marquardt's (1982) warning that "a determination of which patient has the best prognosis is not possible by mathematic computation" (p. 104). If this statement is true, we may be at fault for fitting our prognostic indicators into an all-too-neat 100-point package.

In closing, we suggest that research using a quantified approach to prognostication (such as the *Prognosis Profile Worksheet*) might proceed as follows. First, subjects should be stratified prospectively, by etiology, using large numbers of patients in each group. Or, as Rosenbek et al. (1989) suggested, one might "establish and study cohorts who differ only on a single variable" (p. 94). Second, rather than using a common battery (in this instance, the *Neurobehavioral Cognitive Status Examination* and the *Western Aphasia Battery*), we should explore prognosis based on "etiology-specific" batteries. Third, as Holland and colleagues did in 1989, we can subject select items catalogued on the *Prognosis Profile Worksheet* to logistic regression once sufficient subject numbers are available in each diagnostic group. (Other statistical methods are described by Rosenbek et al., 1989, p. 97.) Fourth, we envision a study that might assess cross-disciplinary prognostic estimates (e.g., asking whether neurologists' prognostic estimates agree with speech-language pathologists' prognostic estimates). Fifth, we think it would be valuable to explore the reliability of prognostication at different points in time. Rather than making estimations at roughly 1 month post onset, as we did, one might examine prognosis at 1 week, or 3 months, or 6 months post onset. Ideally, such a study would use the same group of patients. Finally, we plan to test the prognostic validity of the *Prognosis Profile Worksheet* by comparing prognostic scores and estimates with actual improvement in communication ability.

ACKNOWLEDGMENTS

Lisa A. Carr, M. H. S., Barbara K. Dodson, M. S., and Mary S. Brennan, M. S.

REFERENCES

American Speech-Language-Hearing Association (1991). Guidelines for speech-language pathologists serving persons with language, socio-communicative, and/or cognitive-communicative impairments. *Asha, 33* (Supplement 5), 21–28.

Darley, F. L. (1982). *Aphasia* (pp. 110–143). Philadelphia: W. B. Saunders.

Duffy, J. R., & Myers, P. S. (1991). Group comparisons across neurologic communication disorders: Some methodological issues. In T. E. Prescott (Ed.), *Clinical aphasiology*, Vol. 19 (pp. 1–14). Austin: PRO-ED.

Hier, D. B., Mondlock, J., & Caplan, L. R. (1983). Recovery of behavioral abnormalities after right hemisphere stroke. *Neurology, 33*, 337–343.

Holland, A. L., Greenhouse J. B., Fromm, D., & Swindell, C. S. (1989). Predictors of language restitution following stroke: A multivariate analysis. *Journal of Speech and Hearing Research, 32*, 232–238.

Horner, J., & Rothi, L. J. (1979). Aphasia recovery: Theory and treatment. *Asha, 21,* 604–605. (Abstract.)

Horner, J., & Rothi, L. J. (1984). Prognosis for recovery from aphasia: Estimating extent of improvement. *Communique* (North Carolina Speech-Language-Hearing Association Journal) (Research Edition: 20–29).

Kertesz, A. (1979). *Aphasia and related disorders.* New York: Grune & Stratton.

Kertesz, A. (1982). *Western Aphasia Battery.* New York: Grune & Stratton.

Kiernan, R. J., Mueller, J., Langston, J. W., & Van Dyke, C. (1987). The neurobehavioral cognitive status examination: A brief but differentiated approach to cognitive assessment. *Annals of Internal Medicine, 107,* 481–485.

Levin, H. S., Benton, A. L., & Grossman, R. G. (1982). *Neurobehavioral consequences of closed head injury.* New York: Oxford University Press.

Marquardt, T. P. (1982). *Acquired neurogenic disorders* (pp. 67–104). Englewood Cliffs, NJ: Prentice-Hall.

Neurobehavioral Cognitive Status Examination Manual (1983). Fairfax, CA: Northern California Neurobehavioral Group.

Rosenbek, J. C., LaPointe, L. L., & Wertz, R. T. (1989). *Aphasia: A clinical approach* (pp. 55–103). Austin, TX: PRO-ED.

Shewan, C. M., & Kertesz, A. (1980). Reliability and validity characteristics of the *Western Aphasia Battery* (WAB). *Journal of Speech and Hearing Disorders, 45,* 308–324.

Wertz, R. T., (1978). Neuropathologies of Speech and Language: An introduction to patient management. In D. F. Johns (Ed.), *Clinical management of neurogenic communicative disorders* (pp. 1–101). Boston: Little, Brown.

APPENDIX A: DUKE PROGNOSIS PROFILE WORKSHEET

ESTIMATING POTENTIAL FOR EXTENT OF IMPROVEMENT
(Horner et al., Experimental Version, 1989)
(Adapted from Horner & Rothi, 1984)

Patient:
Date of Onset:

Date of Evaluation:
Clinician:

ETIOLOGY

STROKE

..... Ischemic
  ..... Embolic
  ..... Thrombotic
..... Hemorrhagic
  ..... Intracerebral (ICH)
  ..... Subarachnoid (SAH)
    ..... Aneurysm
    ..... Arteriovenous malformation (AVM)

TRAUMA

..... Closed head injury (CHI)
  ..... Concussion
  ..... Contusion
  ..... Subdural hematoma (SDH)
..... Intracerebral hemorrhage (ICH)
  ..... Coup ..... Contre coup
  ..... Open head injury (OHI)/with skull fracture
..... Concussion
  ..... Contusion
  ..... SDH
  ..... ICH
  ..... Coup ..... Contre-coup

TUMOR

..... Glioblastoma Multiforme
  ..... Astrocytoma
  ..... Meningioma
..... Metastatic
..... Other:

..... Loss of consciousness—or coma RLA I, II, III—and duration:

Neuroradiology:

..... Head CT Scan ..... Head MRI ..... Angiography ..... PET or SPECT ..... Other

Aphasia Profile (W.A.B.-A.Q.) [if applicable]

..... Anomic
..... Conduction
..... Wernicke
..... Transcortical Sensory

..... Broca
..... Transcortical motor
..... Global
..... Mixed transcortical

..... Unclassifiable; NOT aphasic; normal
..... Unclassifiable; NOT aphasic; ABnormal

ADDITIONAL VARIABLES:

Race: Caucasian Black Hispanic Other:

Occupation (now or in past):

Number of TOTAL hospital days (i.e., from acute hospital admission through discharge from rehabilitation):

Appendix A. (continued)

RATING:		3	2	1
HANDEDNESS:		..... Left	..... Right	..... Ambidextrous
SEX:		..... Male	..... Female	
AGE:		..... ≤ 40	 41–69	..... ≥ 70
EDUCATION:		..... > 12	 9–12	..... < 9
LESION TYPE:	Stroke	..... Hemmorhagic	..... Ischemic	..... Hemorrhagic infarct
	Trauma	..... Focal	..... Diffuse	..... Focal and diffuse
	Tumor	..... Slow growing	..... Rapid growing	
LESION NUMBER:		.... Single	..... Multiple	..... Multiple with atrophy
LESION SIDE:		..... Right	..... Left	..... Bilateral
NEUROLOGIC SIGNS:		..... No weakness	..... L or R arm/leg	..... Bilateral arm/leg
VISUAL FIELD DEFECT:		..... No	..... Yes	..... (Neglect precludes unequivocal test)
MONTHS POST ONSET:		 0–1	 1–6	..... > 6
				{ /30}

RATING:	5	4	3	2	1	0 = Not testable	
LANGUAGE PERFORMANCE:	Normal	Mild	Moderate	Severe	Profound		
Auditory	.....	.....	.....	.....	.....	{ /5}	
Oral-verbal	.....	.....	.....	.....	.....	{ /5}	
Reading	.....	.....	.....	.....	.....	{ /5}	
Writing	.....	.....	.....	.....	.....	{ /5}	
							{ /20}
OTHER HIGHER CORTICAL FUNCTIONS:							
Construction (Copying, Drawing, Blocks)	.....	.....	.....	.....	.....	{ /5}	
Calculations	.....	.....	.....	.....	.....	{ /5}	
Abstract reasoning	.....	.....	.....	.....	.....	{ /5}	
Verbal memory	.....	.....	.....	.....	.....	{ /5}	
Visual memory	.....	.....	.....	.....	.....	{ /5}	
							{ /25}
VISUAL-MOTOR FUNCTIONS:							
Praxis: Buccofacial	.....	.....	.....	.....	.....	{ /5}	
Praxis: Limb	.....	.....	.....	.....	.....	{ /5}	
Praxis: Verbal	.....	.....	.....	.....	.....	{ /5}	
Dysarthria	.....	.....	.....	.....	.....	{ /5}	
Visual neglect	.....	.....	.....	.....	.....	{ /5}	
							{ /25}
TOTAL:							{ /100}

OVERALL PROGNOSTIC ESTIMATE FOR EXTENT OF IMPROVEMENT
IN APHASIA/COGNITION/COMMUNICATION

..... Excellent ..... Good ..... Fair ..... Guarded ..... Poor