



Clinical Outcome Scales in Neurology

HAROLD P. ADAMS, JR., MD
DIVISION OF CEREBROVASCULAR DISEASES
DEPARTMENT OF NEUROLOGY
UNIVERSITY OF IOWA

Conflicts of Interest

- ▶ No conflicts of interest

Outcome Measures

- ▶ Important component of a clinical trial
- ▶ Selected measures reflect the aims
 - ▶ Clinical outcome measures
 - ▶ Surrogate outcome measures
- ▶ Disease-specific outcome measures used in neurologic research
- ▶ For the purposes of this presentation, I will emphasize clinical research in stroke as a model

Choices of Outcomes

- ▶ General terms
 - ▶ Impairment
 - ▶ Handicap /activity limitations
 - ▶ Disability /participation restriction
- ▶ Mortality
- ▶ Quality of life
- ▶ Healthcare costs
- ▶ Recurrent events/ neurologic worsening
- ▶ Adverse events

Surrogate Markers

- ▶ Imaging
 - ▶ Brain imaging: size, location, and evolution of lesion
 - ▶ Brain functional imaging
 - ▶ Vascular imaging: recanalization
- ▶ Biomarkers
 - ▶ Variety of options: inflammatory, biochemical, genetic
- ▶ Physiological: EEG, NCV, vital capacity
- ▶ High potential for surrogate markers that complement clinical outcomes in many areas of neurology

Clinical vs Surrogate Outcomes

- ▶ Clinical outcomes are paramount
- ▶ Surrogate outcomes used to buttress clinical outcomes
- ▶ A trial that demonstrates improvement in surrogate measures, but no clinical benefit must be considered negative

Issues in Design

- ▶ Prevention
 - ▶ Avoiding initial or recurrent events
 - ▶ Halting neurological worsening
- ▶ Acute treatment
 - ▶ Limiting injury
 - ▶ Prevention or treatment of medical/neurological complications
 - ▶ Reducing mortality
 - ▶ Improving outcomes
- ▶ Recovery and rehabilitation
 - ▶ Maximizing recovery and limiting neurological sequelae

Issues in the Design of Clinical Trials in Ischemic Stroke

- ▶ Broad spectrum of vascular diseases
- ▶ Wide variations in the extent and locations of brain injuries
- ▶ Epidemiological variables and the presence of comorbid diseases
- ▶ Use of multiple concomitant therapies – “best medical care”
- ▶ Treatment goals and the nature of the intervention that is being tested in the trial

Clinical Rating Instruments

- ▶ Fundamental component of clinical research now used in practice because they provide important information for both researchers and clinicians
 - ▶ Types and severity of neurological impairments
 - ▶ Changes in neurological status
 - ▶ Decisions about acute and long-term management
 - ▶ Responses to treatment
 - ▶ Outcomes

Requirements for a Useful Clinical Rating Instrument

- ▶ Must have inherent credibility- face validity
 - ▶ Germane to the clinical situation
 - ▶ Widely used and clinically useful
- ▶ Makes sense to both health care providers and the public
- ▶ Understandable
- ▶ A knowledgeable person should have a mental image of the patient's status when given the "score"

Steps in Development of a Clinical Rating Instrument

- ▶ Complex process
 - ▶ Purpose of scale and information to be gained
 - ▶ Relevant to the assessment of patients
 - ▶ Generally based on the patient's performance
 - ▶ Items assessed by history or examination
 - ▶ Define how the scoring of a new scale will interdigitate with other rating instruments
- ▶ Need for a clear plan for testing and validating the instrument

Attributes of a Useful Clinical Rating Instrument

- ▶ Easy to administer for patients and assessors
 - ▶ Should not be time-consuming or burdensome
- ▶ Performance and scoring are straightforward
 - ▶ Clear instructions on the use
 - ▶ Administering and scoring of the scale
- ▶ Tested for reliability and reproducibility
 - ▶ Inter-rater agreement
 - ▶ Intra-rater reproducibility
- ▶ Educational and certification programs

Quality Control Measures in Clinical Trials I

- ▶ Extra requirement in research studies, especially true in multi-center clinical trials
- ▶ Requirements
 - ▶ Scale is administered correctly
 - ▶ Scoring is accurate and consistent

Quality Control Measures in Clinical Trials II

- ▶ Well-validated scales should be used
 - ▶ Comparison with other research programs
 - ▶ Requirement of funding agents and regulators
- ▶ Programs to increase reliability and reproducibility
 - ▶ Education and certification
 - ▶ Central adjudication

Enthusiasm for New Clinical Rating Instruments

- ▶ Researchers often have the desire to develop a new rating instrument
- ▶ Process is time-consuming and may not be successful
- ▶ Delays the primary goal of the project
- ▶ Best to adopt/adapt current scales



General Organization of Clinical Rating Instruments

- ▶ Usually based on history and direct examination
- ▶ Generally, two types of scales
 - ▶ Numerical scale – total of scoring of components of assessment
 - ▶ Single score scale – based on an aggregate of all information rather than scoring individual items of the assessment

Numerical Scales

- ▶ Several items assessed and scored
- ▶ Scores of each item added to give a total score
- ▶ Total score may represent a different combination of items
- ▶ Depending on the scale, a high score can be good or bad
- ▶ Example: NIH Stroke Scale

Measuring Neurological Impairments in Stroke

- ▶ Goals
 - ▶ Assess baseline severity of stroke
 - ▶ Affects prognosis and decisions for treatment
 - ▶ Assess for improvement or worsening of the patient's neurological status
 - ▶ May be used as an outcome measure
- ▶ NIH Stroke Scale most commonly used instrument in ischemic stroke

NIH Stroke Scale

Measurements of Acute Cerebral Infarction: A Clinical Examination Scale

Thomas Brott, MD, Harold P. Adams Jr., MD, Charles P. Olinger, MD,
John R. Marler, MD, William G. Barsan, MD, José Biller, MD, Judith Spilker, RN,
Renée Holleran, RN, Robert Eberle, Vicki Hertzberg, PhD, Marvin Rorick, MD,
Charles J. Moomaw, PhD, and Michael Walker, MD

We designed a 15-item neurologic examination stroke scale for use in acute stroke therapy trials. In a study of 24 stroke patients, interrater reliability for the scale was found to be high (mean $\kappa=0.69$), and test-retest reliability was also high (mean $\kappa=0.66-0.77$). Test-retest reliability did not differ significantly among a neurologist, a neurology house officer, a neurology nurse, or an emergency department nurse. The stroke scale validity was assessed by comparing the scale scores obtained prospectively on 65 acute stroke patients to the patients' infarction size as measured by computed tomography scan at 1 week and to the patients' clinical outcome as determined at 3 months. These correlations (scale-lesion size $r=0.68$, scale-outcome $r=0.79$) suggested acceptable examination and scale validity. Of the 15 test items, the most interrater reliable item (pupillary response) had low validity. Less reliable items such as upper or lower extremity motor function were more valid. We discuss methods for improving the reliability and validity of brief examination scales to be used in stroke therapy trials. (*Stroke* 1989;20:864-870)

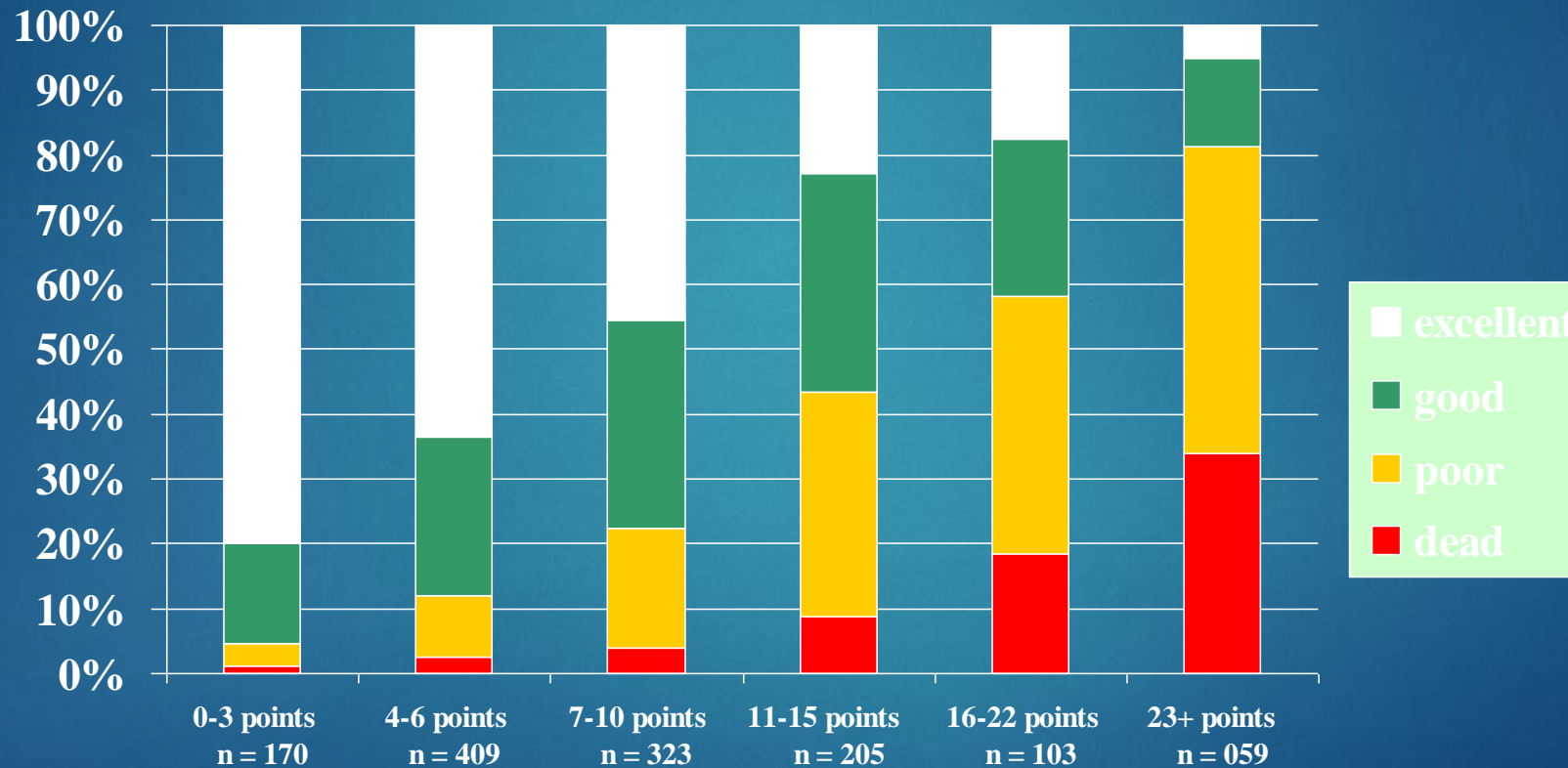
- ▶ 15 items of the neurological examination
- ▶ Each item independently scored
- ▶ Give a baseline severity of neurological impairments
- ▶ Could be used sequentially to monitor for worsening or improvement
- ▶ Range of scores 0 – 42
- ▶ Higher scores more severe stroke

Initial Validation NIH Stroke Scale

- ▶ Initial testing – high inter-rater agreement ($k = 0.69$) and test – retest reliability ($k = 0.66 - 0.77$)
- ▶ Prospectively assessed and total scores were compared to size of infarctions on CT and outcomes at 3 months
 - ▶ Acceptable scale validity
 - ▶ Scores correlated well with size of lesions and outcomes
- ▶ Tested in several other venues
- ▶ Now used internationally in wide range of stroke research

Brott et al, Stroke, 1989: 20: 864

Prognostic Importance NIH Stroke Scale score



Improving Reliability of Scores NIH Stroke Scale

- ▶ Certification process using videotapes
 - ▶ Education, testing, remediation, and reliability assessment
 - ▶ Moderate to excellent agreement on most items
 - ▶ Facial paresis and ataxia perform weakly

Albanese et al, Stroke, 1994; 25: 1746

Lyden et al, Stroke, 1994: 25: 2220

Advantages of NIH Stroke Scale

- ▶ Well-validated measure of stroke severity that can be performed rapidly by a wide range of health care professionals
- ▶ Good correlation with outcomes and used for planning acute and long-term care
- ▶ High inter-rater agreement and intra-rater reproducibility
- ▶ Adapted for multiple languages and cultures
- ▶ Can be administered via telemedicine
- ▶ Educational and certification programs exist

Disadvantages of NIH Stroke Scale

- ▶ “Bias” towards the dominant hemisphere
 - ▶ With similarly sized lesions in similar locations, scores are higher with left hemisphere lesions
 - ▶ Result of orientation and commands linked to language
- ▶ Range of scores among raters
- ▶ Moderate-to-excellent agreement in most items with the following exceptions:
 - ▶ Ataxia, facial paresis, and aphasia

Current Status

NIH Stroke Scale

- ▶ Modifications of NIH Stroke Scale have been attempted but original version remains the standard
- ▶ Most widely used clinical assessment scale of stroke severity in research and clinical care
 - ▶ Entry criterion for trials and in the selection of interventions
 - ▶ Used in inter-physician communications in a way that is similar to the Glasgow Coma Score in patients with trauma
 - ▶ Likely will not be replaced in the near future

Glasgow Coma Scale

Glasgow Coma Scale		
Best eye response (E)	Spontaneous – open with blinking at baseline	4
	Opens to verbal command, speech, or shout	3
	Opens to pain, not applied to face	2
	None	1
Best verbal response (V)	Oriented	5
	Confused conversation, but able to answer questions	4
	Inappropriate responses, words discernible	3
	Incomprehensible speech	2
	None	1
Best motor response (M)	Obeys commands for movement	6
	Purposeful movement to painful stimulus	5
	Withdraws from pain	4
	Abnormal (spastic) flexion, decorticate posture	3
	Extensor (rigid) response, decerebrate posture	2
	None	1

- ▶ Used in patients with head injuries
- ▶ 3 items to assess
- ▶ Scores 3-15
- ▶ Strongly correlated with outcomes
- ▶ Used in patients with other severe neurologic events

Montreal Cognitive Assessment

- ▶ Brief screening tool to detect mild cognitive impairments
 - ▶ Approximately 10 minutes to assess by direct observation
 - ▶ Similar to what is observed in a clinical setting
 - ▶ Incorporates some widely used neuropsychology tests such as the Trail-Making Test
- ▶ Has been extensively tested in patients with cognitive impairments from a variety of causes
 - ▶ Educational and certification programs are not available
- ▶ May be more sensitive than the Mini-Mental Status Exam but may be less specific

Barthel Index

- ▶ Scale used to assess disability
- ▶ Simple system based on historical reporting from subject or caregiver that does not require much training
 - ▶ Exams 10 items of activities of daily living: scores 0, 5, 10 or 15
 - ▶ Dependent, partially independent, totally independent
 - ▶ Scores range from 0 – 100 by adding individual items
 - ▶ Virtually no disability: 95 – 100, institutionalized care: < 60
 - ▶ Heavily weighted towards motor function
- ▶ Has ceiling and floor effects and is relatively insensitive
- ▶ Scores are well recognized by health care providers

Migraine Specific Quality of Life Questionnaire

- ▶ 14 Questions about severity and frequency of migraine headaches
- ▶ Patient completes the questionnaire and is asked to answer each question
- ▶ Each item scored independently
- ▶ 1□ None of the time 2□ A little bit of the time 3□ Some of the time 4□ A good bit of the time 5□ Most of the time 6□ All of the time
- ▶ Used in clinical trials to test therapies to prevent migraine

Unified Parkinson's Disease Rating Scale

- ▶ Four major categories of evaluation with multiple questions in each category
 - ▶ Intellectual function, mood, and behavior
 - ▶ Activities of daily living
 - ▶ Motor examination
 - ▶ Motor complications
- ▶ Involves history or findings on examination
- ▶ Each question scored 0 no problems, 1 minimal problems, 2 mild problems, 3 moderate problems, 4 severe problems
- ▶ Range in scores 0-199

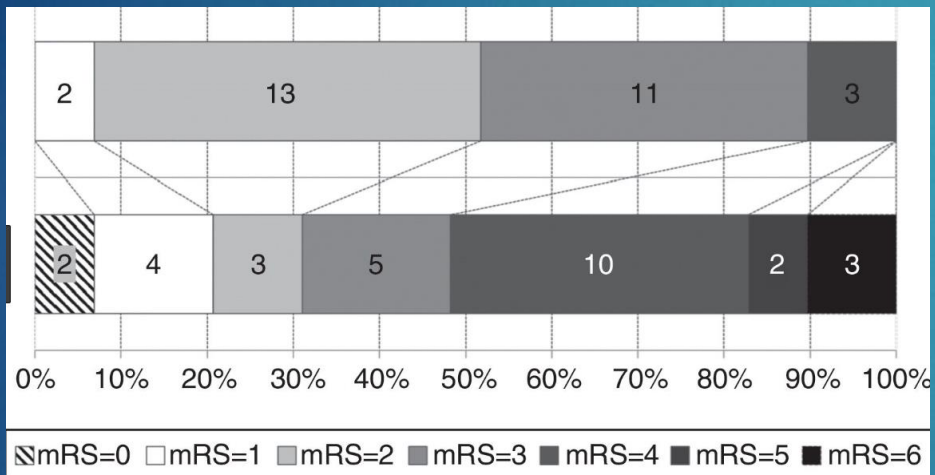
Global Measures of Outcomes

- ▶ Scales widely accepted by medical community, funding authorities, and governmental regulators
 - ▶ Broadly differentiate favorable from unfavorable outcomes
 - ▶ Used in both acute and recovery trials
 - ▶ Measure impact on multiple neurological impairments or disabilities
- ▶ May miss important neurological issues
 - ▶ Discrete areas of neurological disability
 - ▶ Over-emphasize some components of recovery
 - ▶ Often have ceiling- and floor- effects
- ▶ Require larger clinical trials

Overall Assessment with a Single Score

- ▶ Most common scales in neurology
- ▶ Example in stroke: Modified Rankin Scale
- ▶ All components of the assessment are summarized in a single score
- ▶ Ranges to separate the good from the bad
- ▶ Each score has specific and defined criteria
- ▶ Generally, the higher the score, the poorer the situation

Modified Rankin Scale



- ▶ Global outcome scale that is internationally accepted and used widely in stroke studies
- ▶ Information about the status of the patient with an emphasis on motor limitations and walking
- ▶ Can be performed by a broad spectrum of health care providers
- ▶ Different scores (levels of recovery) are understood by physicians and governmental bodies

Reliability of Scores Modified Rankin Scale

- ▶ Paired assessments among researchers
 - ▶ 100 paired assessments, inter-rater agreement ($K = 0.57$)
- ▶ Review of 10 international trials
 - ▶ Reliability varied (weighed $K = 0.25$ to $K = 0.95$)
- ▶ Educational program and structured interview
 - ▶ Mass video-based training
 - ▶ 90% achieved certification on first time, 85% of remainder were subsequently certified
 - ▶ Heterogeneity across countries but native English language did not affect outcomes
 - ▶ Need new strategies to improve reliability

Quinn et al, Stroke, 2007; 38: 2257

Quinn et al, Stroke, 2008; 39: 2975

Quinn et al, Stroke, 2009; 40: 762 and 3393

Kurtzke Expanded Disability Status Scale

- ▶ Used in assessing patients with multiple sclerosis
- ▶ Involves 20 levels of neurologic impairment
- ▶ 0.0 no symptoms, 10.0 dead
- ▶ 0.5 increments describing progressive disability
- ▶ Example:
- ▶ Level 3.0
Moderate disability in one Functional System (one FS grade 3, others 0 or 1) or mild disability in three or four FS (three or four FS grade 2, others 0 or 1) though fully ambulatory.
- ▶ Has been validated for clinical studies

Modality-Specific Scales

Recovery After Stroke

- ▶ Evaluate responses to an intervention aimed at a specific impairment/disability
 - ▶ Used extensively in rehabilitation research
 - ▶ Particularly useful for testing a device or local intervention
 - ▶ Some neurological impairments may improve at different rates and degrees
- ▶ Collecting data from a small number of subjects
- ▶ Lack of clear data on overall outcome
- ▶ Scales may not be well understood by clinicians or the public and results may be widely accepted

Fugl-Myer Assessment of Motor Recovery after Stroke

- ▶ Internationally accepted scale to assess motor recovery after stroke
 - ▶ Several domains are assessed for a total of 226 points
 - ▶ Each item assessed 0: cannot do, 1: partial, 2: fully performs
 - ▶ Motor: 100 points (66 arm,) sensory: 24 points, balance: 14 points, joint movement: 44 points, joint pain: 44 points
- ▶ 45 minutes to administer by a trained physical therapist
- ▶ Not widely used in clinical stroke trials
- ▶ Physicians do not have a good understanding of the meanings of the scores

Quality of Life Measures

- ▶ Used in a broad range of research studies testing promising therapies
- ▶ Covers a broad range of functioning
 - ▶ Physical
 - ▶ Psychological
 - ▶ Social
 - ▶ General health
- ▶ Influenced by person's experiences, beliefs, expectations and perceptions
- ▶ Generally have not been the primary way to measure success of treatment
- ▶ Euro-QOL and Stroke Impact Scale

Euro-QOL

- ▶ Euro-QOL (EQ-5D) is a simple and brief self-administered instrument in two parts
- ▶ Five dimensions – each in three grades
 - ▶ 1 – no problem, 2- moderate problems, 3 severe problems
 - ▶ Mobility
 - ▶ Self-care
 - ▶ Usual activities
 - ▶ Pain/discomfort
 - ▶ Anxiety/depression
- ▶ Visual analogue scale
 - ▶ 0 – worst imaginable
 - ▶ 100 – best possible

Importance of Selecting Existing Scales

- ▶ Information available about usefulness
- ▶ Educational and certification programs
- ▶ Results are understandable by medical community
- ▶ Allows comparison among clinical trials
- ▶ Important to meet regulatory body acceptance

Primary Stroke Rating Instruments

NINDS Recommendations

- ▶ Neurological impairment
 - ▶ NIH Stroke Scale
- ▶ Functional status
 - ▶ Modified Rankin Scale
 - ▶ Barthel Index
- ▶ Emotional and cognitive status
 - ▶ Center for Epidemiological Studies – Depression Scale
 - ▶ Montreal Cognitive Assessment
 - ▶ Trail-making A & B Tests
- ▶ Participation and quality of life
 - ▶ European Quality of Life Scale
- ▶ Performance
 - ▶ Walking speed

Clinical Stroke Investigation

European Agency Evaluation of Medicinal Products

- ▶ Functional outcomes
 - ▶ Barthel Index
- ▶ Global outcome scales
 - ▶ Modified Rankin Scale
 - ▶ Glasgow Outcome Scale
- ▶ Neurological deficit scales
 - ▶ Scandinavian Stroke Scale
 - ▶ Canadian Neurological Scale
 - ▶ NIH Stroke Scale
 - ▶ Unified Stroke Scale

Conclusions I

- ▶ Wide variety of clinical scales to use in neurologic research
- ▶ Have used example of stroke
- ▶ Similar measures available for other neurologic diseases
- ▶ Choice of scales influenced by
 - ▶ Face value
 - ▶ Reproducibility
 - ▶ Internal construct

Conclusions II

- ▶ Used to select patients for enrollment in the trial
- ▶ Attributes
 - ▶ Reliable
 - ▶ Precise
 - ▶ Valid
 - ▶ Feasible
 - ▶ Acceptable

Conclusions III

- ▶ Choice of scales also influenced by the primary aims of the research
 - ▶ Acute vs long-term intervention
 - ▶ Duration of follow-up
 - ▶ Nature of the intervention
 - ▶ Primary hypothesis
 - ▶ Progression of disease
 - ▶ Favorable outcomes, unfavorable, mortality
 - ▶ Adverse events, related to intervention, not related
 - ▶ New events

Conclusions IV

- ▶ Trials must assure accuracy of the clinical assessments
 - ▶ Selection, follow-up, endpoints, outcomes
 - ▶ Education and certification of investigators
 - ▶ Central assessments of outcomes
 - ▶ In person, telephone, videos, teleconference
 - ▶ Adjudication of endpoints and outcomes

Conclusion V

- ▶ Provide a quantitative element to a complex clinical situation
- ▶ Foster communication
- ▶ Results of clinical research are described using these instruments
- ▶ Both researchers and clinicians should have an understanding of the information conveyed by the use of the instruments