UNIVERSITY OF IOWA HEALTH CARE

# Where Do Clinical Outcome Scales Come From?

Harold P. Adams, Jr., MD
Division of Cerebrovascular Diseases
Department of Neurology
UIHC Comprehensive Stroke Center
University of Iowa

# Conflicts of Interest

UNIVERSITY OF IOWA HEALTH CARE

- No commercial conflicts of interest
- Do receive grant support from NINDS
- Will not discuss any therapies not approved or cleared by the FDA

# Evidence-Based Medicine

UNIVERSITY OF IOWA HEALTH CARE

- Evidence-based medicine has changed the management of patients with neurologic disease
- Guidelines, which provide the foundation of evidence-based medicine, have created the standards for care
- Information for the guidelines largely comes from the results of modern clinical trials
- Responses by regulatory bodies and third party payers also are influenced by the results of clinical trials

# Outcome Measures

UNIVERSITY OF IOWA HEALTH CARE

- Important component of a clinical trial
- Selected measures reflect the goals and design of the trial
  - Clinical outcome measures
  - Surrogate outcome measures
- Reflect the primary and secondary aims of the trial
- For the purposes of this presentation, I will use clinical research in stroke as a model

# Issues in Design of Trials in Stroke

UNIVERSITY OF IOWA HEALTH CARE

- Primary goals of stroke trials vary and affect the design of the research program
- Prevention
  - Avoiding stroke or recurrent stroke
  - Reducing mortality
- Acute treatment
  - Limiting the brain injury
  - Prevention or treatment of medical/neurological complications
  - Improving outcomes
- Recovery and rehabilitation
  - Maximizing recovery and limiting neurological sequelae

# Issues in the Design of Clinical Trials in Ischemic Stroke

UNIVERSITY OF IOWA HEALTH CARE

- Stroke encompasses a 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

# Surrogate Markers Trials in Ischemic Stroke

UNIVERSITY OF IOWA HEALTH CARE

- Imaging
  - Brain imaging: size, location, and evolution of stroke
  - Brain functional imaging
  - Vascular imaging: recanalization
- Biomarkers
  - Variety of options: inflammatory, biochemical, genetic
- Currently, no surrogate marker has gained acceptance as a primary outcome measure
- Clinical outcomes remain the measure of success of any treatment in stroke

# Surrogate Markers
# Other Areas of Neurology

UNIVERSITY OF IOWA HEALTH CARE

- Imaging
  - Brain imaging: size, evolution, new lesions
  - Brain functional imaging
- Biomarkers
  - Physiological: EEG, NCV, vital capacity
  - Other: inflammatory, biochemical, genetic
- High potential for surrogate markers that complement clinical outcomes in many areas of neurology

# Clinical Outcome Measures Other Areas of Neurology

UNIVERSITY OF IOWA HEALTH CARE

- Outcomes in non-stroke research
  - Mortality
  - Improving neurological status
  - Maintaining independence
  - Slowing progression
  - Prevention of recurrent events
- Primary outcome measure determined by the aim of the trial

# Prevention Trials

UNIVERSITY OF IOWA HEALTH CARE

- Goal is to prevent recurrent events
  - New disease-specific recurrences
  - New comorbid events
  - Mortality
- Goal is to prevent progression
  - New recurrences, new impairments, or worsening
  - Neurological disability
  - Global outcome

# Clinical Rating Instruments (Stroke Scales)

UNIVERSITY OF IOWA HEALTH CARE

- Fundamental component of clinical research that now is used in practice because they provide important information for both researchers and clinicians
  - Assessing the types and severity of neurological impairments
  - Monitoring changes in neurological status
  - Influencing decisions about acute and long-term management
  - Examining responses to treatment
  - Determining outcomes

# Requirements for a Useful Clinical Rating Instrument

UNIVERSITY OF IOWA HEALTH CARE

- Must have inherent credibility- face validity
  - Germane to the clinical situation
  - Widely used and clinically useful
- Results are believable and make 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" on a stroke scale

# Steps in Development of a Clinical Rating Instrument

UNIVERSITY OF IOWA HEALTH CARE

- Complex process that requires thought
  - Purpose of scale and information to be gained
  - Relevant to the assessment of patients
  - Generally based on the patient's performance
  - Items to be 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

UNIVERSITY OF IOWA HEALTH CARE

- 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

UNIVERSITY OF IOWA HEALTH CARE

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

UNIVERSITY OF IOWA HEALTH CARE

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

# General Organization of Clinical Rating Instruments

UNIVERSITY OF IOWA HEALTH CARE

- Grading the severity of the illness, responses to therapy, and outcomes
- Usually based on history and direct examination
  - Impairments, disability, handicap
  - Some scales include results of diagnostic tests
- 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

UNIVERSITY OF IOWA HEALTH CARE

- 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

# Overall Assessment with a Single Score

UNIVERSITY OF IOWA HEALTH CARE

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

# Modality-Specific Scales

UNIVERSITY OF IOWA HEALTH CARE

- Large number of rating instruments that are most commonly used in rehabilitation and recovery research
- Emphasize recovery or compensation in a specific activity
  - Language and speech
  - Walking
  - Hand function
- Do not provide an assessment of the patient's autonomy

# Global Measures of Outcomes

UNIVERSITY OF IOWA HEALTH CARE

- 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

# Measuring Neurological Impairments

UNIVERSITY OF IOWA HEALTH CARE

- Goals
  - Assess baseline severity of stroke
  - Affects prognosis and decisions for treatment
  - Assess for improvement or worsening of the patient's neurological status
- NIH Stroke Scale most commonly used instrument
- Developed by researchers at University of Cincinnati, University of Iowa, and NINDS

# National Institutes of Health 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

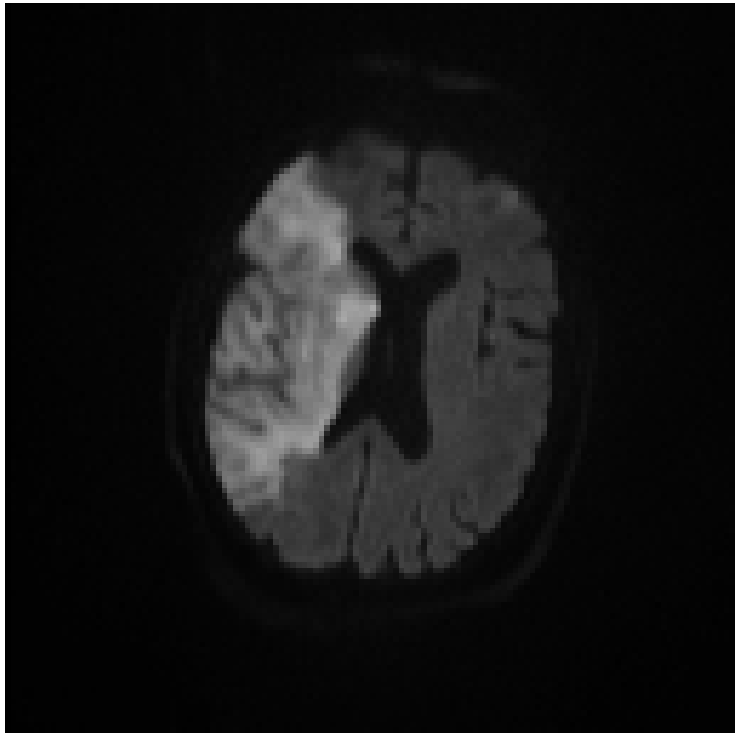
UNIVERSITY OF IOWA HEALTH CARE

- 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

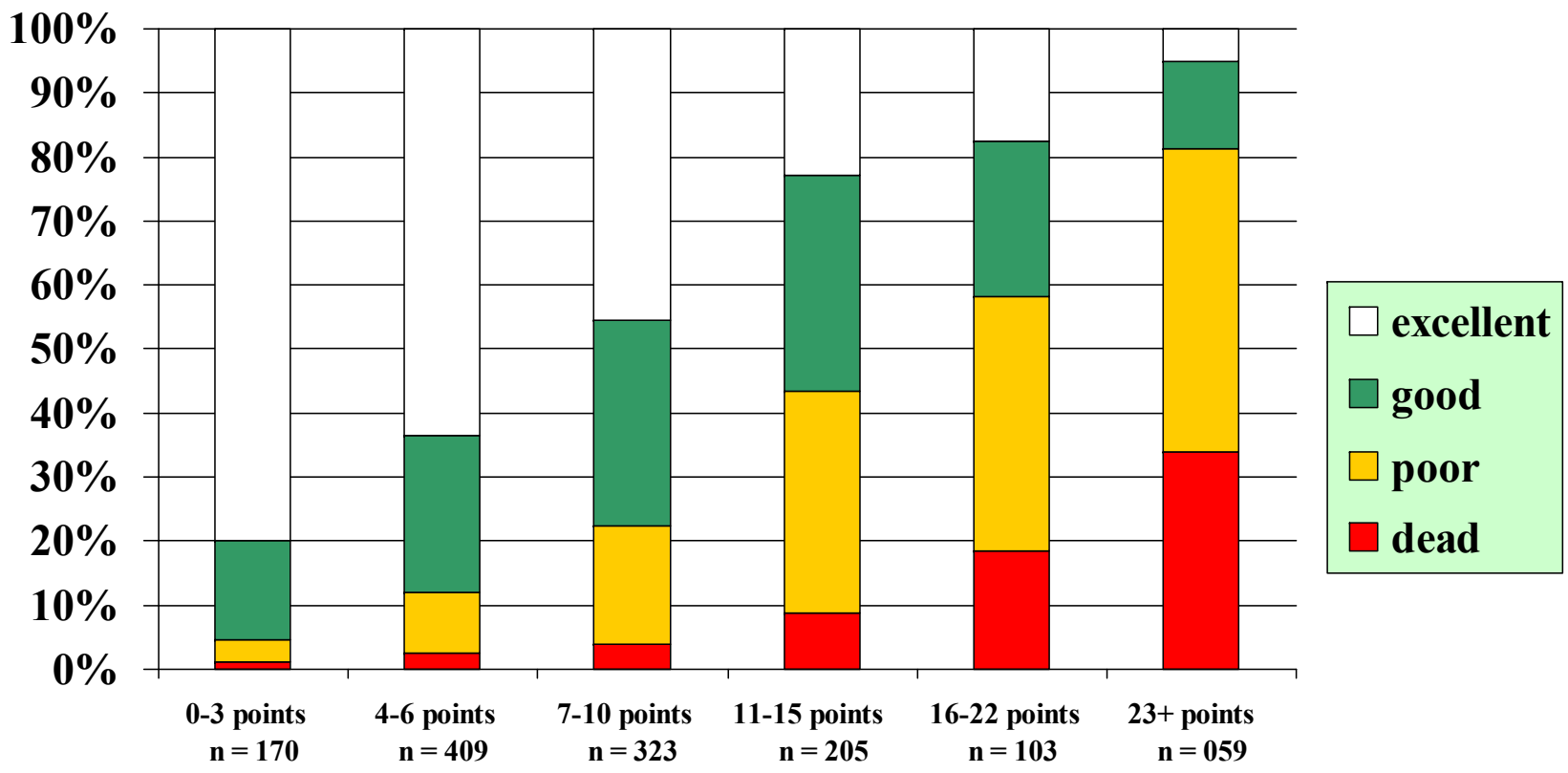
# Correlation of Stroke Severity with Total score on NIH Stroke Scale

- Total score is important in planning for acute care and prognosis
- Range of scores
  - 0 – 3 mild
  - 4 – 8 moderate
  - 9 – 14 mod. severe
  - 15 – 20 severe
  - 21+ very severe



# Prognostic Importance
# NIH Stroke Scale score

UNIVERSITY OF IOWA HEALTH CARE




Adams et al, Neurology, 1999; 53: 126

# Advantages of NIH Stroke Scale

UNIVERSITY OF IOWA HEALTH CARE

- 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

UNIVERSITY OF IOWA HEALTH CARE

- There is a “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

# Reliability of Scores
# NIH Stroke Scale

UNIVERSITY OF IOWA HEALTH CARE

- 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
- National Stroke Association Certification Process
  - 7405 raters (physicians and other health care providers)
  - Range in scoring between raters was considerable
  - Aphasia and facial paresis were the most problematic items
  - Did not improve with taking certification exam several times

    Josephson et al, Cerbrovasc Dis, 2006; 22: 389

# Current Status
# NIH Stroke Scale

UNIVERSITY OF IOWA HEALTH CARE

- 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
  - Score used as an entry criterion for trials and in the selection of interventions
  - Score is now 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 and all physicians dealing with stroke should become proficient in its use

# Global Outcome Scales Recovery After Stroke

UNIVERSITY OF IOWA HEALTH CARE

- Scales widely accepted by medical community, funding authorities, and governmental regulators
  - Broadly differentiate favorable or unfavorable responses
  - Used in both acute and recovery studies
- More useful for testing medications that would have impact in multiple areas of the brain
  - Impact on multiple neurological impairments/disabilities
- Not particularly sensitive to important improvements
  - Discrete areas of neurological disability
  - Over-emphasize some components of recovery
  - May have ceiling- and floor-effects
- Require larger trials

# Modified Rankin Scale

UNIVERSITY OF IOWA HEALTH CARE

- 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

# Modified Rankin Scale

UNIVERSITY OF IOWA HEALTH CARE

- 0 No symptoms at all
- 1 No significant disability despite symptom
  Able to carry out usual duties and activities
- 2 Slight disability
  Unable to carry out all previous activities
  Able to look after own affairs without assistance
- 3 Moderate disability
  Require some help, may walk without assistance
- 4 Moderately severe disability
  Unable to walk without assistance
  Unable to attend to own bodily needs without assistance
- 5 Severe disability
  Bedridden, incontinent, requires constant nursing care
- 6 Death

# Reliability of Scores
# Modified Rankin Scale

UNIVERSITY OF IOWA HEALTH CARE

- 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

# Modality-Specific Scales Recovery After Stroke

UNIVERSITY OF IOWA HEALTH CARE

- 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
  - Diversity of neurological problems after stroke
- Scales may not be well understood by clinicians or the public and results may be widely accepted
  - Inability to compare with other clinical stroke research

# Fugl-Myer Assessment of Motor Recovery after Stroke

UNIVERSITY OF IOWA HEALTH CARE

- 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

# Montreal Cognitive Assessment

UNIVERSITY OF IOWA HEALTH CARE

- Brief screening tool to detect mild cognitive impairments
  - Approximately 10 minutes to assess by direct observation
  - Tests memory, visuospatial functions, executive function, attention and concentration, language, and orientation
  - Similar to what is observed in a clinical setting
  - Incorporates some widely used neuropsychology tests such as the Trail-Making Test
- Has not been extensively tested in stroke trials
  - Educational and certification programs are not available
- May be more sensitive than the Mini-Mental Status Exam but may be less specific

# Barthel Index

UNIVERSITY OF IOWA HEALTH CARE

- Scale to assess disability that is widely used in stroke research
  - 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
- Could be an important secondary outcome measure in stroke recovery trials

# Quality of Life Measures

UNIVERSITY OF IOWA HEALTH CARE

- 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

UNIVERSITY OF IOWA HEALTH CARE

- 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

# Stroke Impact Scale

UNIVERSITY OF IOWA HEALTH CARE

- Assess multiple domains of stroke recovery
  - Reported by patient to reporter
  - Graded 1 (worst) – 5 (best)
- 8 Components
  - Physical activity ( 4 – 20 points)
  - Memory and thinking ( 7 – 35 points)
  - Emotions ( 9 – 45 points)
  - Communication ( 7 – 35 points)
  - Activities of daily living ( 10 – 50 points)
  - Mobility at home and community ( 9 – 45 points)
  - Hand function ( 5 – 25 points)
  - Relationships and activities ( 8 – 40 points)

# Primary Stroke Rating Instruments NINDS Recommendations

UNIVERSITY OF IOWA HEALTH CARE

- 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

UNIVERSITY OF IOWA HEALTH CARE

- 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

UNIVERSITY OF IOWA HEALTH CARE

- 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

UNIVERSITY OF IOWA HEALTH CARE

- 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

UNIVERSITY OF IOWA HEALTH CARE

- 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

# Conclusions V

- Provide a quantitative element to a complex clinical situation
- Foster communication among health care professionals
- 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