OUTCOME MEASURES IN NEUROLOGY CLINICAL TRIALS

Jeremy M. Shefner
Kemper and Ethel Marley Professor and Chair of Neurology
Barrow Neurological Institute
Disclosures

• Personal compensation received from Cytokinetics, Biogen, Ionis, MT Pharma, Neuraltus, Denali
• Research funding received from Cytokinetics, Biogen, Synapse, Neuraltus, Biotie, Amylyx, ALS Association, MDA, NINDS
Objectives

• Describe classes of endpoints
• Discuss phases of clinical development with a focus on endpoints
• Discuss strengths and limitations of different endpoints
• Take home message:
  – The choice of outcome measure in a clinical trial is a critical one, and can influence the outcome
  – Future progress in neurology clinical trials depends to a large extent on better markers
General Classification of Endpoints

- Clinically Relevant Endpoints
  - How a patient feels, functions, or survives
    - Phase 3

- Surrogate Endpoints
  - A laboratory value, image, or objective assessment intended to substitute for or predict a clinically relevant outcome
    - When available, commonly are phase 2 endpoints

- Pharmacodynamic Endpoints
  - An endpoint that reflects activity of treatment on a specific pathway, receptor, etc, or reflects an aspect of the disease process itself
  - No implication that this endpoint reflects a clinical change
  - Used in phase 2, increasingly phase 1
What is a Biomarker?

- generally refers to a measurable indicator of some biological state or condition. (Wikipedia)
- a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (NIH)
- Definition is so broad that a biomarker can be any one of above
- A clinically relevant endpoint can also be a biomarker
Clinically Relevant Endpoints

- Clinically relevant endpoints are required for phase 3 trials
  - May be subjective (I feel better) or objective (I can walk across the room better)
  - Survival
  - Time to event

- However:
  - Clinical Relevance is often a fuzzy target
    - Is vital capacity clinically relevant?
    - Is strength clinically relevant?
  - Clinical relevance does not necessarily imply relevance to potential therapeutic mechanism
  - Issues of variability may limit utility
    - Disease related
    - Measure related
Functional Scales

• Functional Scales are considered clinically relevant
  • They directly ask patients about functional capacity, or assess these functions by observation
  • However, size of effect that is important is not always clear
  • The scale properties are critical and often undefined
    • Interval Scaling
    • Continuous vs discrete
Functional Scales

- Can be disease or attribute specific
- Scoring of individual items should have characteristics of an interval scale: i.e., a change of 1 unit should be the same anywhere on the scale
- Often comprised of well defined domains capable of assessing different aspects of function
Limitations of Functional Scales

- Often combine attributes so it is difficult to attribute a change to a specific function
- The minimum clinically significant change is undetermined
- Lack of interval scaling may mask small changes
- Variability of scoring may limit use or increase sample size
- Individual items are usually strikingly non-linear; averaging many items together can create appearance of linearity
Commonly Used Functional Scales

- Kurtzke EDSS
- ALS Functional Rating Scale- Revised (ALSFRS-R)
- Unified Parkinson’s Disease Rating Scale (UPDRS)
- Alzheimer’s Disease Assessment Scale-cognitive subscale (ADAS-cog)
- Modified Rankin Scale
Disability Scores do not linearly decline in MS

From: www.mult-sclerosis.org
ALSFRS-R

- Speech
- Salivation
- Swallowing
- Handwriting
- Cutting food, handling utensils
- Dressing and Hygiene
- Turning in bed and adjusting bed clothes
- Walking
- Climbing stairs
- Dyspnea
- Orthopnea
- Respiratory insufficiency

From: Cedarbaum et al, 1999
### ALSFRS Climbing Stairs

<table>
<thead>
<tr>
<th>Score</th>
<th>ALSFRS-R Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Normal</td>
</tr>
<tr>
<td>3</td>
<td>Slow</td>
</tr>
</tbody>
</table>
| 2     | Mild unsteadiness or fatigue  
Patient needs to rest between steps, or feels unsteady, but does not need rail |
| 1     | Needs assistance  
Patient needs assistance including handrail or caregiver |
| 0     | Cannot do             |
ALSFRS-R Sub-Domains

- Changes in sub-domain scores validated across two studies conducted a decade apart in time.

Respiratory questions are 25% of the scale, but only 13% of the change over time.

Cedarbaum et al. 1999
Decline of ALSFRS-R is often linear

From: Cudkowicz et al., 2006, 2013
Edaravone Phase 3 Trial

Edaravone ALS 19 Study Group, 2017
Unified Parkinsons Disease Rating Scale

- 44 items: some assessed by interview, some from direct evaluation. Ranges from 0 (normal) to 176 (maximum disability)
- 3 domains: mentation, behavior and mood, ADLs, and motor function
- Questions have significant overlap between domains
Characteristics of UPDRS

- Change over time of total score and subscores usually linear
- Overlapping questions make the ADL and motor scores interdependent
- Within individual scales, different attributes are combined (e.g., impairment vs. disability)
- A change in UPDRS may reflect either symptomatic improvement or disease modification
UPDRS and Subscales are linear

From: PSG, Arch Neurol, 2004;61:1044-1053
Binary/Time to Event

• **Advantages**
  – Easy to understand
  – Power calculations are straightforward

• **Disadvantages**
  – Only subjects who reach endpoint are useful
  – Only 1 change of state is deemed important
Time to Event: Survival

- Useful only when events are likely to occur
  - Stroke
  - SAH
  - ALS

- Depending on disease state and target, may not be sensitive to experimental intervention
  - Nuedexta for Emotional Lability
    - Approved for ALS, but unlikely to impact survival
Survival as an outcome measure in ALS

From: Drachmann et al., 2000

From Lacomblez et al., 1996
Time to event is an example of a binary endpoint

- Time to event endpoints
  - Survival
  - Hospital readmission
  - Time to new vascular event
  - Time to initiation of NIV

- Other binary endpoints
  - Achieving functional independence
  - Achieving independent ambulation
A 7 point scale is often dichotomized (0-2 vs 3-6) for primary analysis

### Table 1. Modified Rankin Scale

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No symptoms at all</td>
</tr>
<tr>
<td>1</td>
<td>No significant disability despite symptoms; able to perform all usual duties and activities</td>
</tr>
<tr>
<td>2</td>
<td>Slight disability; unable to perform all previous activities, but able to take care of self without assistance</td>
</tr>
<tr>
<td>3</td>
<td>Moderate disability; requiring some help, but able to walk without assistance</td>
</tr>
<tr>
<td>4</td>
<td>Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance</td>
</tr>
<tr>
<td>5</td>
<td>Severe disability; bedridden, incontinent, and requiring constant nursing care and attention</td>
</tr>
<tr>
<td>6</td>
<td>Dead</td>
</tr>
</tbody>
</table>

Adapted from Saver, 2007
Symptom relief vs. disease progression

• For neurodegenerative disease, ultimate goal is to affect disease progression

• However, symptom relief is often difficult to distinguish from change in disease course
  – Survival
    • Hypothesis: if survival is impacted, disease course is as well
      – Converse is likely true: in neurodegenerative disease, if disease progression rate is slowed, survival is likely to be improved.
      – However, many interventions may impact survival independent of disease progression
  – Functional rating scales
    • Often considered to be assessments of symptom burden, but reflect disease progression as well
Examples of Survival Benefit Without Impacting Disease Course

- Trial of Permanent Assisted Ventilation with Tracheostomy in patients with ALS.
- Trial of PEG in ALS
Pattern of change in an endpoint can suggest symptom relief or change in disease progression.

May suggest disease modification; however, depends on relationship of measure to disease process.

May suggest symptom relief; however, disease burden may still be less at end of study.
Using symptomatic endpoints to assess disease modification

From: Olanow et al., Movement Disorders, 2008.
Effect of Early vs Late Rasagiline on UPDRS

From: Hauser et al., Movement Disorders 24, 2009, 564–573

Caveat: after 6 mo, Active treatment was Open label

**P < 0.05

**P < 0.0001

Placebo-Rasagiline

Rasagiline

Graph showing the mean % change in total UPDRS over time for Placebo-Rasagiline and Rasagiline groups.
Other clinically relevant endpoints

- Diaries
  - Headache
  - Seizure
- Performance
  - 6 minute walk
- Pain Scales
- Rate of word production
Non clinically relevant outcomes

• Safety measures
  – Lab
  – Clinical events

• Pharmacokinetic measures

• Health economic measures

• Pharmacodynamic measures
  – An endpoint that reflects activity of treatment on a specific pathway, receptor, etc, or reflects an aspect of the disease process itself
Pharmacodynamic Endpoints

– Reflect an important aspect of treatment or disease
  • A physiological or imaging measure that changes with disease progression
  • Presence at the intended site of action
  • Engagement of intended target
  • Alteration of potentially relevant pathway
– PD endpoint hopefully shows a signal more quickly or more sensitively than a clinically relevant phase 3 endpoint
– The distinction between PD and clinical relevance may be quite blurred
Methods of assessment can be very important

• Strength is a functional marker that may be important in studying many diseases

• However, how it is measured affects its utility
  ▪ Single muscle group
    ▪ Vital capacity
    ▪ Handgrip
  ▪ Global Assessment
    ▪ MRC manual muscle testing
      ▪ Any number of muscle can be tested on a 0-5 ordinal scale
    ▪ Quantitative muscle testing
      ▪ TQNE
      ▪ HHD
Uneven Steps Between MRC Grades

MMT scale compared with actual dynametric force measurement of the biceps brachii

(modified from van der Ploeg: J Neurol, 1984)
Quantitative Muscle Testing: Standardized Training and Validation

- Standardized positions
- Video and hands on training
- Requirement for demonstration of adequate training
- Test-retest reliability criterion
Decline in individual muscle groups

Right

Left

Months

Biogen Empower Study
Decline in UE Megascore, 1 site, Ceftriaxone
Proportion of zero force per muscle

Bulbar Onset

Extremity Onset
Sensitivity of time to first zero muscle compared to survival
Electrical Impedance Myography (EIM)

- Pioneered by Seward Rutkove
- Technique based on the application of high-frequency electrical current to localized areas of muscle with measurement of resulting voltages.
  - Painless
  - Non-invasive
  - Can apply to virtually any superficial muscle
    - Tongue, paraspinals, proximal muscles all possible
- Sensitive to alterations in muscle composition, structure, atrophy
A. Healthy Muscle

- Applied current
- Surface voltages result as current flows through resistance and capacitance in the tissue. The capacitance also causes a phase shift.
- Measured Voltage

B. Diseased Muscle

- Applied current
- Increased tissue resistance in diseased muscle results in a larger voltage, and reduced capacitance results in less phase shift.
- Measured Voltage
EIM has been studied in several NM diseases

- ALSA-funded Longitudinal Study in ALS
- Ongoing SBIR
- Neuralstem study of stem cells in ALS
- SMA
- Animal models
- A variety of muscle diseases
EIM vs other measures

**Coefficient of Variation:**
- **ALSFRS (Relative):** 0.81
- **Mean HHD (Relative):** 0.93
- **EIM Phase (Relative):** 0.62
- **MUNE:** 0.72

From: Rutkove et al., 2012
Shefner et al., 2011
EIM in SMA

From Rutkove et al., 2010
EIM vs Strength in SMA

From: Rutkove et al., 2010
ALS Rat Data: Measuring Disease Progression
16 animals followed from pre-symptomatic to death

Wang et al, Clin Neurophys 2011
Summary

• The choice of outcome measures is critical in the design of clinical trials
• Outcomes should be reliable, meaningful, and sensitive to disease modification
• An appropriate choice of outcome measure should increase the probability of correctly concluding the presence of therapeutic effect
• The currently available toolbox of measures is not adequate to meaningfully shorten trials or reduce sample size for most neurological diseases
Efficacy and safety of oral fumarate in patients with relapsing-remitting multiple sclerosis: a multicentre, randomised, double-blind, placebo-controlled phase IIb study

Ludwig Kappos, Ralf Gold, David H Miller, David G MacManus, Eva Havrdova, Volker Limmroth, Chris H Polman, Klaus Schmierer, Tarek A Yousry, Minhua Yang, Mefkûre Eraksoy, Eva Meluzinova, Ivan Rektor, Katherine T Dawson, Alfred W Sandrock, Gilmore N O’Neill, for the BG-12 Phase IIb Study Investigators

• 257 patients, 3 doses vs placebo for 24 weeks
• Primary endpoint: new GdE lesions
  – Clear dose response; lesions reduced by 69% at highest dose
• Secondary endpoint: relapse rate
  – No dose response; overall, relapse rate declined by 32% (p=0.27)

Kappos et al
Lancet 2008
RRMS: Gd+ lesions
A marker of disease activity

Kappos et al
Lancet 2008
Placebo-Controlled Phase 3 Study of Oral BG-12 for Relapsing Multiple Sclerosis

Ralf Gold, M.D., Ludwig Kappos, M.D., Douglas L. Arnold, M.D., Amit Bar-Or, M.D., Gavin Giovannoni, M.D., Krzysztof Selmaj, M.D., Carlo Tornatore, M.D., Marianne T. Sweetser, M.D., Ph.D., Minhua Yang, M.S., Sarah I. Sheikh, M.D., and Katherine T. Dawson, M.D., for the DEFINE Study Investigators*

---

**Hazard ratio vs placebo (95% CI)**
- BG-12 bid: 0.51 (0.40 – 0.66); p < 0.001
- BG-12 tid: 0.50 (0.39 – 0.65); p < 0.001

**Estimated proportion with relapse at 2 years**
- BG-12 bid: 27%
- BG-12 tid: 26%
- placebo: 46%

---

**Patients with relapse (%)**

---

**No. at risk**
- Placebo: 408 356 321 282 243 224 205 190 115
- BG-12 bid: 410 353 324 303 286 267 255 243 154
- BG-12 tid: 416 346 322 301 286 270 251 244 166

---

**Weeks on study**
Summary

• The choice of outcome measures is critical in the design of clinical trials
• Outcomes should be reliable, meaningful, and sensitive to disease modification
• An appropriate choice of outcome measure should increase the probability of correctly concluding the presence of therapeutic effect
• The currently available toolbox of measures is not adequate to meaningfully shorten trials or reduce sample size for most neurological diseases
In Rats, EIM correlates strongly to MUNE and survival

Wang et al, *Clin Neurophys* 2011
Progression rate of ALSRS at diagnosis predicts survival

$$\Delta FS = \frac{48 - \text{initial ALSFRS-R}}{\text{duration from onset}} \quad n=115$$
CoV for different measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Dexpanetopenole</th>
<th>Ceftriaxone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope ALSFRS-R</td>
<td>-1.14 0.97 -0.84</td>
<td>-1.18 0.85 -0.72</td>
</tr>
<tr>
<td>SVC</td>
<td>-3.11 2.83 -1.09</td>
<td>FVC</td>
</tr>
<tr>
<td>Slope Z megascore total</td>
<td>-0.09 0.08 -0.97</td>
<td>-0.08 0.06 -0.85</td>
</tr>
<tr>
<td>Slope Z megascore upper limbs</td>
<td>-0.1 0.09 -0.95</td>
<td>-0.08 0.08 -0.95</td>
</tr>
<tr>
<td>Slope Z megascore lower limbs</td>
<td>-0.07 0.09 -1.28</td>
<td>-0.08 0.08 -1.01</td>
</tr>
</tbody>
</table>

Table 1: Slopes and coefficient of variation (CoV) for rate of change estimates based on regression analysis. Slope ALSFRS-R points, % predicted, vital capacity, and Z score change for Amyotrophic Lateral Sclerosis Functional Rating Scale-revised (ALSFRS-R), vital capacity, and blymescores, respectively. Data are expressed as change/mo.
Effect of Tirasemtiv on SVC

Least Square Mean Change from Baseline (percentage points)

Effect of Tirasemtiv on SVC

Shefner et al., 2016