Evidence from Biomarkers and Surrogate Endpoints

Andrew Feigin

Center for Neurosciences, North Shore–Long Island Jewish Research Institute, Manhasset, New York 11030

Summary: The use of physiological, anatomical, and other biological tests is commonplace in the practice of medicine. In neurology, objectively measured tests termed biomarkers (BMs) are playing an increasing role in diagnosis and management of disease, both in clinical practice and in experimental therapeutics. This article will discuss the various applications of BMs to the assessment of therapies for neurological diseases and will use examples from neurological diseases to elucidate the strengths and potential weaknesses of BMs. As the understanding of the pathophysiology of many neurological diseases has improved, new BMs have been developed, and efforts have been made to use these as proxies for clinical endpoints. A BM used in this manner is referred to as a surrogate endpoint (SE). There are many potential advantages and disadvantages of using SEs in the evaluation of new therapies, and these will be reviewed as well. Furthermore, the evidence required for the development of an SE and the nature of the evidence that can be derived from the use of BMs and SEs will be discussed. **Key Words:** Biomarker, surrogate, endpoint, therapeutics, evidence.

INTRODUCTION

Simple, inexpensive, reliable, and rapidly obtainable measures of a disease process may have great utility both in clinical care and in clinical trials. These types of measures, referred to as biomarkers (BMs) and surrogate endpoints (SEs), may be useful for diagnosis, prognosis, therapy, and drug development. Neurological disorders provide examples of many important issues related to the development and use of these measures. In addition, some neurological disorders, especially those that are neurodegenerative, may present unique challenges to the acceptance of these measures. This article will review the definitions, applications, successes, and failures of biomarkers and surrogate endpoints in neurology. Examples used in this paper will focus on multiple sclerosis (MS), stroke, and neurodegenerative diseases including Alzheimer's disease (AD), Parkinson's disease (PD), and Huntington's disease (HD), although similar examples could be found for many other neurological disorders.

The terms "marker" and "endpoint" imply certain applications. Markers in general can be understood as measures potentially applicable in clinical care as well as in research, whereas endpoints are by definition used in clinical research. A "clinical endpoint" is "a characteristic or variable that reflects how a patient feels, functions or survives."[1] Clinical endpoints are considered the gold standard for evaluating the efficacy of a new therapy. The term "surrogate marker" is sometimes used in common parlance and in the medical literature,[2–4] but it is a poorly defined term, and its use has been discouraged.[1] This paper will propose a definition for this term that may have a unique application to neurodegenerative disorders.

BIOMARKERS

The Biomarkers Definitions Working Group (BDWG) of the National Institutes of Health defines a BM as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention."[1]

The development of BMs relies on the understanding of the pathological basis for the disease under study. This understanding could be at any level, from molecular genetic, to biochemical, to anatomical. An example of a molecular genetic BM is the cytosine–adenosine–guanosine (CAG) repeat length in HD. This test can serve as a diagnostic test in symptomatic individuals[5–7] or as a prognostic test for presymptomatic individuals who know they are at risk for HD.[8,9] This BM would be unlikely to be useful as a proxy for a clinical endpoint in

a clinical trial, because it would be unlikely to change with a therapeutic intervention and can be abnormal in individuals with no clinical signs of disease (presymptomatic HD). Biochemical BMs used in clinical practice and clinical trials include measuring CSF oligoclonal bands to confirm the diagnosis of MS,[10–12] and functional imaging of dopaminergic neurons in PD with dopamine transporter ligands, measures of dopamine metabolism (fluorodopa), or others.[13,14] Anatomical BMs include magnetic resonance imaging (MRI) of lesion burden or cerebral atrophy in MS[10,15–17] and confirmation of the clinical diagnosis of stroke with computed tomography (CT) or MRI.[18] In fact, in the absence of clinical signs or symptoms, these imaging methods can be used in isolation to diagnose stroke (silent stroke).[19–21]

APPLICATIONS OF BIOMARKERS

There are many potential valuable applications of BMs in medicine in general and in neurological diseases in particular. These applications may include the following: 1) diagnosis and differential diagnosis, 2) screening potential new therapies either *in vitro* or *in vivo*, 3) measuring severity, progression of disease, and responses to therapy, 4) predicting prognosis, and 5) measuring toxicity.

Diagnosis and differential diagnosis

There are multiple examples of BMs used in clinical neurology to aid with diagnosis. In most instances, the diagnosis of the condition remains predominantly a clinical one, and the BMs serve an auxiliary role. In MS, for example, although the diagnosis remains predominantly a clinical one,[22] the presence of white matter plaques on MRI has become a major contributor to the diagnosis.[11,15,17,23,24] In fact, white matter lesions that appear and disappear in different locations on MRI may be viewed with the same diagnostic value as clinical signs and symptoms that vary over time and space. In AD, CSF measures of tau protein, phosphorylated tau protein, and the 42 amino-acid residue form of amyloid-β have been proposed as helpful for distinguishing AD from mild cognitive impairment (MCI).[25–27] In early PD, given the difficulty of making an accurate diagnosis before exposure to levodopa, there has been great interest in developing a BM that would help in distinguishing idiopathic PD from atypical parkinsonism.[14,28–31] Of these, the most promising appear to be imaging measures of either brain metabolism with ^{18}F-fluoro-2-deoxy-D-glucose and positron emission tomography (PET) or of cerebral blood flow as measured with SPECT.[32,33] Studies examining the diagnostic accuracy of these and other methods have been difficult, because they rely on comparisons with the clinical diagnosis (the gold standard),

which is known to be inaccurate ~15-20% of the time.[34] Ideally, a new diagnostic test for PD should be evaluated by comparing the results with a known postmortem pathological result, but no large-scale study using this method has been reported. Another approach, although imperfect, could compare imaging results early in the course of disease to the best clinical diagnosis after many years of follow-up; such studies are underway. In summary, there are many examples of BMs used to assist with diagnosis in neurology, and new diagnostic BMs will likely alter the practice of neurology in the future.

Preclinical drug discovery

Using BMs to screen new therapies for neurological diseases has proven to be difficult. Several potential problems arise when attempting to develop a BM as a predictor of therapeutic efficacy. One issue is whether the biological process used as a screen is relevant to the intended disease. For example, in HD there is much evidence that glutamate-mediated excitotoxicity is involved in the pathogenesis of the disease,[35–38] so drugs have been screened *in vitro* and in animal models to assess their ability to block this effect. Drugs that are effective in the screening process (i.e., block glutamate-mediated excitotoxicity) have failed in clinical trials, suggesting that the screening model may not be relevant to the disease.[39] This is not to say that glutamate is not involved in the pathogenesis of HD; there are multiple possible reasons (some of which may be difficult to foresee) that a biologically relevant screening test might not accurately predict therapeutic efficacy.[40] Some potential confounding factors include the following: 1) a drug might have additional biological effects which counter the known effects, 2) *in vitro* screening tests may not account for relevant *in vivo* biological processes, e.g., a drug that is useful in a screening test but does not cross the blood brain barrier would not be effective in the human disease, 3) the screening test might test for only one of several important mechanisms underlying the disease, and could represent only a small amount of the variance in the disease process, and 4) side effects could result in poor outcomes despite a favorable effect on the disease mechanism (see discussion of surrogate endpoints).

Assessments of disease severity, progression, and responses to therapy

There are many examples in neurology of BMs that are used for assessing severity of disease. One example already mentioned is the use of MRI for assessing lesion volume or burden in MS. The main use of objective BMs of disease severity is to document changes in response to therapy. In MS, MRI is used for this purpose both in clinical practice and in clinical trials.[11] In PD, BMs are generally not used in the assessment of symptomatic therapies, because clinical endpoints are relatively inex-

pensive and sensitive for short-term efficacy trials. BMs have, however, been used extensively in clinical trials aimed at altering the progression of disease.[4,13,41–44] Specifically, imaging measures of dopaminergic nigrostriatal integrity have been used to examine potential neuroprotective effects of drugs. In some of these studies, the BMs identified relative preservation of the dopaminergic system in subjects treated with dopamine agonists,[42,43] suggesting slowing of disease progression. This occurred in the setting of better clinical function in the group not treated with dopamine agonists (but treated with levodopa). The results of these trials have raised significant and difficult questions regarding the use of BMs in neurodegenerative diseases, including the following: Can drugs affect the BMs independent of their effects on the disease process, and if so, are the BMs any better than clinical measures? Can side effects of drugs outweigh benefits in slowing disease progression? How much benefit is sufficient to warrant a given degree of toxicity? Conversely, how should symptomatic benefits in the absence of disease-modifying effects be weighed against modest neuroprotective effects in the absence of symptomatic benefits? Despite these concerns, BMs will likely continue to gain acceptance in the assessment of neurological disease severity and progression. As this occurs, they will be used more often in the assessments of therapies, although they will likely continue to be used in conjunction with clinical endpoints, and their utility will be dependent on careful assessments of reproducibility and validity.

Predicting prognosis

BMs may have utility in predicting prognosis or risk. In stroke, measuring low-density lipoprotein (LDL) and high-density lipoprotein cholesterol or electrocardiogram changes can provide important information regarding risk.[45–50] Changes in these measures in response to therapy can reliably alter the risk. For example, converting atrial fibrillation to normal sinus rhythm will reduce the risk of stroke.[50] Similarly, angiography of the carotid arteries can be viewed as a BM that has utility in directing the mode of therapy for stroke prevention (i.e., carotid endarterectomy).[51–54] BMs such as these will likely have an increasing role in the future in tailoring therapies for individuals within a given diagnosis as well as in designing clinical trials to optimize the likelihood of finding a benefit. For example, certain genetic markers have been identified that appear to predict responses to antidepressants[55,56]; perhaps in the future, PD patients with depression could be tested for specific genetic markers that would direct the choice of antidepressant.

Measuring toxicity

Safety and tolerability assessments in both clinical practice and clinical trials rely on BMs in addition to clinical evaluations. In clinical practice, for example, the safety of clozapine is monitored by weekly measures of white blood cell count.[57] Similarly, the use of tolcapone for the treatment of PD requires regular monitoring of liver function tests (LFTs).[58] Clinical trials use these measures and other standard laboratory tests routinely to monitor safety. Of all the uses and potential uses of BMs, this application to assessments of toxicity is probably the most widely accepted as having value independent of clinical assessments, i.e., the use of BMs as outcome measures in phase II clinical trials is not controversial.[40,59] For example, if an individual has elevated LFTs, even in the setting of no clinical signs of liver dysfunction, the drug under study will likely be either withheld or reduced.

SURROGATE ENDPOINTS

The preceding discussion focused on BMs in general. SEs are unique and well characterized BMs that fit the following definition: "a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence."[1] It is worth noting that although a SE is by definition a BM, not all (and, in fact, a very small minority) of BMs will meet the standard of a SE.

Surrogate endpoint validation

Strict criteria for the validation of a SE do not exist. Nonetheless, general guidelines for the interpretation of clinical trials using SEs have been proposed.[60] These guidelines provide a framework for understanding the nature of the evidence that is provided by SEs, and include the following: 1) "Is there a strong, independent, consistent association between the surrogate endpoint and the clinical endpoint?", which is a necessary but not sufficient condition, 2) "Is there evidence from randomized trials in other drug classes that improvement in the surrogate endpoint has consistently led to improvement in the target outcome?" 3) "Is there evidence from randomized trials in the same drug class that improvement in the surrogate endpoint has consistently led to improvement in the target outcome?" 4) Is there a "large, precise, and lasting...treatment effect?" and 5) "Are the likely treatment benefits worth the potential harms and costs?"[60]

For a BM to be regarded as an SE, there must be a causal relationship between the BM and the clinical outcome. That is, establishing a connection between a BM and a disease process is not sufficient. Simply identifying a physiological defect in a given disease does not provide evidence that the defect is related to the clinical signs of the disease. A BM must be linked to a clinical outcome through a randomized controlled clinical trial (RCCT) to be characterized as an SE (FIG. 1). For example, many
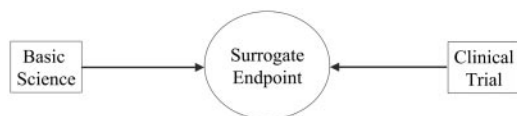
FIG. 1. Surrogate endpoints. The validation of a surrogate endpoint requires data from both basic science (mechanism) and clinical trials.

studies *in vitro* and *in vivo* have established that there is abnormal oxidative stress in PD,[61–65] suggesting that BMs that measure oxidative stress might be useful for screening and assessing therapies for PD. The link between this particular mechanism and the clinical symptoms of PD, however, is weak, so that at this point, such a BM could not be considered a SE. Stated another way, "If there is no clearly established causal link between (biomarkers of) a given pharmacologic action and clinical outcome, then no amount of evidence confirming that the pharmacologic action is linked to treatment will increase confidence that the clinical outcome is also linked to treatment."[66]

The US Food and Drug Administration (FDA) utilizes the definition of a SE to define how SEs may be used to achieve drug approval in their Fast Track program [described by Katz in the present issue of *NeuroRx*® (1:307–316, 2004)], stating that a BM can be characterized as a SE if it is "...reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit."[1] The problem is that the term "reasonably likely" remains poorly defined. Until 1997, the FDA required the results of several RCCTs to approve a drug. Specifically, the guidelines required "adequate and well controlled investigations," and this was interpreted to mean a minimum of two such trials.[66] It is worth noting that the original intent of this regulation has often been subverted by the common practice of performing two identical RCCTs.[66] That is, the purpose of requiring more than one investigation was to insure that the findings could be replicated in different populations at varying times and places; simply repeating the same trial does not necessarily achieve this goal, and may undermine the value of replication. In 1997 [see Katz in the present issue of *NeuroRx*® (1:307–316, 2004)], the FDA altered the requirement to "data from one adequate and well controlled investigation and confirmatory evidence."[67]

The term "confirmatory evidence" has not been adequately defined, but clearly has relevance to BMs in general and SEs in particular. Although BMs are identified through basic science research, their causal relationship to the clinical manifestations of a disease can only be confirmed through clinical trials, i.e., a RCCT of a drug that works through a specific known mechanism that demonstrates an improvement in clinical symptoms provides evidence that the mechanism is relevant to the disease. When other drugs with the same mechanism of

action also are shown to have clinical benefit, a BM reflecting this mechanism may be accepted as a SE. Lowering of LDL cholesterol provides an example of this. Measurements of LDL, although a BM, could not be considered an SE until RCCTs demonstrated a link between lowered LDL and reduced risk of stroke and cardiovascular disease. As this discussion suggests, it is almost impossible to imagine an SE being accepted for the evaluation of a "first-in-class" drug.[66] That is, evidence may suggest that a particular pathomechanism underlies a given disease, and drugs that affect this mechanism might be developed; nonetheless, a BM reflecting the proposed mechanism could not be used as an SE in a RCCT until at least one drug with the relevant mechanism of action has been demonstrated to produce a clinical benefit using standard clinical endpoints.

An ideal SE will reflect all relevant clinical information, not just information relating to efficacy.[68] This must be the case if a SE is to be used in the absence of standard clinical endpoints. The reason for this is that unexpected side effects of a therapy could produce negative outcomes even if a well characterized SE predicted clinical benefit. The most widely cited example of this is the Cardiac Arrhythmia Suppression Trial (CAST). High rates of ventricular premature contractions (VPCs) following myocardial infarction are associated with increased risk of sudden death; in the CAST study, however, successful suppression of VPCs resulted in a threefold increase in mortality.[69] In this case, the SE of frequency of VPCs did not predict clinical benefit. A subtler version of this problem has been raised by the recent trial comparing different cholesterol-lowering agents for the reduction of atherosclerosis (presented at the American Heart Association meeting in November, 2003). In this trial, Lipitor was compared to Pravochol in a blinded manner. When matched for degree of cholesterol lowering, Lipitor produced a statistically significantly greater benefit in reversing atherosclerotic disease, suggesting that measuring cholesterol does not capture all of the relevant clinical information. That is, an SE might be useful for predicting benefit, but might not be satisfactory for determining degree of benefit. Cholesterol measures are perhaps the most widely cited and accepted SEs, yet they are imperfect. These observations suggest that even relatively well characterized SEs should be used in combination with clinical endpoints.

SURROGATE MARKERS

The BDWG states that the term surrogate marker (SM) should not be used because it implies that the "substitution is for a marker rather than for a clinical endpoint."[1] In neurodegenerative disorders such as PD, HD, and AD, however, the goals of experimental therapies are twofold—better symptomatic therapies, and

treatments to slow disease progression or delay disease onset.[70] With regard to the latter, clinical endpoints are used as biomarkers (clinical measures such as blood pressure have been used as BMs previously), and in fact probably should be referred to as clinical markers, i.e., they are not measured for the purpose of detecting clinical benefit but for their reflection of the underlying pathological neurodegenerative process. In fact, in a neuroprotective trial for PD, for example, the goal of the trial is to detect no change in clinical status; even a worsening in clinical status could be considered a success if the rate of worsening is slower than expected. If an improvement in clinical status occurs, this is considered a potential confound. For this very reason, clinical scales are quite poor measures of disease progression, because they may be affected by symptomatic therapies and can therefore lead to misleading interpretations of results. The Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) study provides an example.[71] This study hypothesized that deprenyl, a monoamine oxidase B inhibitor, would slow the progression of PD, and earlier small-scale studies had demonstrated no symptomatic benefit from deprenyl in PD. In this study, 800 PD subjects were randomized to receive deprenyl and/or tocopherol alone or in combination in a 2×2 factorial design. The primary outcome measure was need for dopaminergic therapy. The subjects randomized to deprenyl were found to be less likely to require dopaminergic therapy over time, and this was initially interpreted as evidence that deprenyl slowed the progression of PD. Subsequent analysis, however, found that the reason for the difference was that deprenyl produced a small but statistically significant symptomatic benefit (detected because of the increased power in this large-scale trial). A commentary on this trial at the time stated: "The null hypothesis of this protocol is: the investigational drugs do not retard the pathologic process...The null hypothesis may validly be rejected only if there is a significant difference between control and treated groups in the absence of symptomatic effects...This restrictive stipulation is essential because the protocol's only significant measure of pathologic process is symptom severity."[72] This sentiment reflects the widely held belief that measuring clinical severity alone in degenerative diseases is not sufficient for proving neuroprotection.

The DATATOP study demonstrates that the goal of biological substitute measures for clinical measures of progression is to improve on clinical measures, not to equal or approximate the clinical measures. This is conceptually quite different from a surrogate endpoint; the BDWG states that "the assessment of clinical endpoint in controlled clinical trials provides the most convincing evidence for the benefit of an intervention,"[1] but this is not necessarily the case for neuroprotection trials, as demonstrated by the DATATOP study and its subsequent critique. In this regard, clinical measures cannot be considered the gold standard for measuring disease progression. Though fraught with potential problems (e.g., distinguishing symptomatic from neuroprotective effects), this is in fact how clinical measures are currently used in neuroprotection clinical trials. Therefore, in this sense, a biomarker substituting for a clinical measure of progression (clinical marker) should be considered a SM.

This distinction between a clinical endpoint and a clinical marker is important and counter-intuitive. In medicine in general, useful therapies are understood to have measurable benefits for an individual patient. That is, for example, levodopa improves the symptoms of PD, statins lower cholesterol (something measurable) thereby reducing the risk of stroke, or antihypertensives lower blood pressure (again, something measurable) resulting in lower risk of stroke. Neuroprotective therapies, by contrast, might be expected to produce no measurable change (clinical or biological) in a given individual, and can only be shown to be effective over time in an RCCT. For example, if 2-β-carbomethoxy-3-β-(4-[^{125}I]iodophenyl)tropane SPECT were used as a biomarker in the assessment of PD progression, an individual subject would demonstrate a progression of their disease over time even if a therapeutic intervention were successful in slowing progression; there would be no way of knowing on the basis of this BM or on clinical grounds if the individual was benefitting from the drug, and the only way to demonstrate benefit is through a group comparison in an RCCT. This concept of a therapy producing a benefit that cannot be measured in an individual is counter-intuitive and not often encountered by patients or physicians.

Presymptomatic HD provides another example of how the term surrogate marker may have a role. Studies aimed at slowing progression in presymptomatic carriers of the HD gene mutation cannot use clinical measures because the subjects are by definition asymptomatic and without clinical signs of HD. If a group of investigators chooses to study progression in this population using an imaging modality such as MRI or PET, these biomarkers cannot be said to be acting as surrogate endpoints because they are not substituting for a clinical endpoint (see definitions above). They are, however, acting as proxies for the measurement of pathological disease progression and therefore can be considered surrogate markers. One might counter that presymptomatic HD could be studied using onset of clinical symptoms as a clinical endpoint, but although this is true, it is not a requirement. In this latter situation, the biomarkers would be substituting for a clinical endpoint (analogous to lowering cholesterol), reducing the risk over time of developing HD, and therefore could be considered surrogate endpoints. The distinction is not simply academic; at this point, neurologists consider the slowing of disease progression in
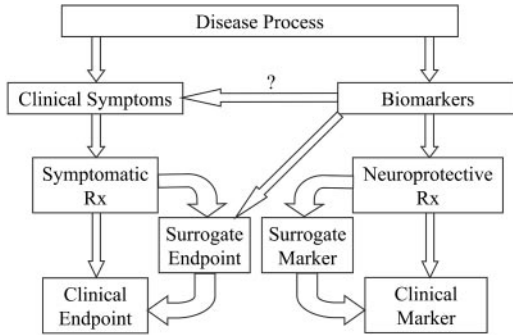
FIG. 2. Proposed terminology for neurodegenerative disorders. Clinical measures are used in neurodegenerative disorders to assess both symptomatic therapies (clinical endpoints) and potential neuroprotective therapies (clinical markers). Surrogate endpoints may substitute for clinical endpoints, whereas surrogate markers may substitute for clinical markers (see text). Surrogate endpoints and surrogate markers are both biomarkers.

neurodegenerative disorders a goal in and of itself, independent of any demonstrable clinical benefit, because it is assumed that slower pathological progression will mean improved clinical outcome. This assumption may not be valid, and this view may change in the future, but for now, the distinction between a clinical endpoint and clinical marker, and the resultant distinction between a surrogate endpoint and a surrogate marker (FIG. 2) seem appropriate.

SEs and SMs could be viewed as part of a spectrum of surrogate outcomes, SEs being validated by a causal relationship to clinical outcomes, and SMs not being validated in this way [see discussion by Katz in the present issue of *NeuroRx*® (1:307–316, 2004) of validated *vs* unvalidated surrogates]. The use of the separate terms SE and SM is favored here because they reflect the duality of the purpose of the surrogate, i.e., long-term outcome in neurodegenerative disorders is a function of both short-term symptomatic effects and pathological progression, and clinical measures alone cannot be used to tease apart these aspects of disease. In fact, neurodegenerative disorders may be unique with regard to the relationship between pathological progression and clinical status. For example, early PD patients may appear clinically normal on levodopa despite the loss of 50-70% of their nigral dopaminergic neurons and the presence of Lewy bodies; wide clinical variations (improvement or worsening) may not be associated with pathological changes. Again, the purpose of using imaging measures in neurodegenerative disorders is to provide information regarding the pathological status of the disease, not to substitute for the clinical measures. A BM that was 100% linked to clinical status in PD, for example, would be less useful for this purpose, so it would seem inappropriate to refer to it as "validated." Similarly, a BM used for measuring pathological progression that has not been linked to clinical outcome might still have significant value, which would not be reflected in the term "unvalidated."

The terms SE and SM capture the dual intent of these measures.

There are several potential risks of using surrogates in clinical trials regardless of whether they are validated. Some of these risks, specifically with regard to SEs, have already been discussed. SMs, or unvalidated surrogates, carry additional risks, including the following: 1) results contradicting clinical outcomes may be difficult or impossible to interpret, 2) unexpected acute effects of the therapy on the SM could confound results analogous to the problems with using clinical measures of progression (see DATATOP discussion above), and 3) positive results from SMs in the absence of clinical benefit could be erroneously used to promote the use of a drug. These problems can be mitigated to some extent by always using clinical outcomes in conjunction with an SM.

Given that SMs, as defined in this paper, are not substituting for clinical endpoints, the evidentiary standard for their use should be quite different than that of SEs. Clinical endpoints are the gold standard for evaluating symptomatic therapies, but clinical markers may not be the gold standard for evaluating neuroprotection. As mentioned above, surrogate markers should always be used in combination with clinical endpoints in neuroprotective clinical trials, because information regarding the short-term symptomatic effects of a proposed neuroprotective agent will need to be known to fully understand any disease modifying effects. In addition, suggesting a neuroprotective effect in the absence of any clinical measure of benefit will remain a controversial result until a given SM is accepted as truly reflecting the disease process. How this will be achieved remains uncertain.

CONCLUSION

BMs have become commonplace in clinical neurology, and evidence from BMs in clinical trials is gaining in acceptance. BMs can be used for many purposes in clinical neurology and in experimental therapeutics, including but not limited to diagnosis, prognosis, and evaluation of therapeutic efficacy and toxicity. BMs may substitute for clinical measures of efficacy (symptomatic therapies) or for measures of pathological disease progression (neuroprotective therapies). When a BM is used to substitute for a traditional clinical endpoint, it is referred to as an SE. The standards for validation of an SE require data from both basic science and clinical trials, and the use of SEs in subsequent trials must be undertaken with great caution, especially when not used in conjunction with clinical endpoints. This paper proposes that when a BM is used to substitute for a clinical measure of pathological disease progression (clinical marker) it should be referred to as a surrogate marker. This type of BM will need to be used with traditional clinical

endpoints in future clinical trials aimed at slowing disease progression in neurodegenerative disorders.

REFERENCES

1. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69:89–95, 2001.
2. De Milito A, Titanji K, Zazzi M. Surrogate markers as a guide to evaluate response to antiretroviral therapy. *Curr Med Chem* 10: 349–365, 2003.
3. Fleming TR. Surrogate markers in AIDS and cancer trials. *Stat Med* 13:1423–1435 [discussion 1437-1440], 1994.
4. Morrish PK. How valid is dopamine transporter imaging as a surrogate marker in research trials in Parkinson's disease? *Mov Disord* 18 [Suppl 7]:S63–S70, 2003.
5. Claes S, Van Zand K, Legius E, Dom R, Malfroid M, Baro F et al. Correlations between triplet repeat expansion and clinical features in Huntington's disease. *Arch Neurol* 52:749–753, 1995.
6. MacMillan JC, Davies P, Harper PS. Molecular diagnostic analysis for Huntington's disease: a prospective evaluation. *J Neurol Neurosurg Psychiatry* 58:496–498, 1995.
7. Wang V, Yeh TP, Chen CM, Yan SH, Soong BW. Usefulness of molecular testing in Huntington's disease. *Zhonghua Yi Xue Za Zhi* 62:586–590, 1999.
8. Rubinsztein DC, Barton DE, Ferguson-Smith MA. Issues in Huntington's disease testing. *Q J Med* 87:71–73, 1994.
9. Sanchez A, Castellvi-Bel S, Mila M, Genis D, Calopa M, Jimenez D et al. Huntington's disease: confirmation of diagnosis and presymptomatic testing in Spanish families by genetic analysis. *J Neurol Neurosurg Psychiatry* 61:625–627, 1996.
10. Pirttila T, Nurmikko T. CSF oligoclonal bands, MRI, and the diagnosis of multiple sclerosis. *Acta Neurol Scand* 92:468–471, 1995.
11. Michalowska-Wender G, Losy J, Wender M. Biological markers to confirm diagnosis and monitor the therapy in multiple sclerosis patients. *Folia Neuropathol* 39:1–6, 2001.
12. Fortini AS, Sanders EL, Weinshenker BG, Katzmann JA. Cerebrospinal fluid oligoclonal bands in the diagnosis of multiple sclerosis. Isoelectric focusing with IgG immunoblotting compared with high-resolution agarose gel electrophoresis and cerebrospinal fluid IgG index. *Am J Clin Pathol* 120:672–675, 2003.
13. Marek K, Innis R, van Dyck C, Fussell B, Early M, Eberly S et al. [123I]beta-CIT SPECT imaging assessment of the rate of Parkinson's disease progression. *Neurology* 57:2089–2094, 2001.
14. Burn DJ, Sawle GV, Brooks DJ. Differential diagnosis of Parkinson's disease, multiple system atrophy, and Steele-Richardson-Olszewski syndrome: discriminant analysis of striatal 18F-dopa PET data. *J Neurol Neurosurg Psychiatry* 57:278–284, 1994.
15. Arnold DL, Matthews PM. MRI in the diagnosis and management of multiple sclerosis. *Neurology* 58 [Suppl 4]:S23–S31, 2002.
16. Bylesjo II, Brekke OL, Prytz J, Skjeflo T, Salvesen R. Brain magnetic resonance imaging white-matter lesions and cerebrospinal fluid findings in patients with acute intermittent porphyria. *Eur Neurol* 51:1–5, 2004.
17. Debaene A, Lavieille J, Vion-Dury J, Stanoyevitch JF, Saint-Jean JC, Legre J. Contribution of MRI to the diagnosis of multiple sclerosis. *J Neuroradiol* 13:1–10, 1986.
18. Corea F, Tambasco N, Luccioli R, Ciorba E, Parnetti L, Gallai V. Brain CT-scan in acute stroke patients: silent infarcts and relation to outcome. *Clin Exp Hypertens* 24:669–676, 2002
19. Corea F, Henon H, Pasquier F, Leys D. Silent infarcts in stroke patients: patient characteristics and effect on 2-year outcome. *J Neurol* 248:271–278, 2001.
20. Bernick C, Kuller L, Dulberg C, Longstreth WT Jr, Manolio T, Beauchamp N et al. Silent MRI infarcts and the risk of future stroke: the cardiovascular health study. *Neurology* 57:1222–1229, 2001.
21. Leary MC, Saver JL. Annual incidence of first silent stroke in the United States: a preliminary estimate. *Cerebrovasc Dis* 16:280–285, 2003.
22. Simon JH, Thompson AJ. Is multiple sclerosis still a clinical diagnosis? *Neurology* 61:596–597, 2003.
23. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 50:121–127, 2001.
24. Mitchell LA. Radiology quiz. MRI in the diagnosis of multiple sclerosis. *Aust Fam Physician* 27:631–632, 1998.
25. Hampel H, Buerger K, Kohnken R, Teipel SJ, Zinkowski R, Moeller HJ et al. Tracking of Alzheimer's disease progression with cerebrospinal fluid tau protein phosphorylated at threonine 231. *Ann Neurol* 49:545–546, 2001.
26. Hampel H, Goernitz A, Buerger K. Advances in the development of biomarkers for Alzheimer's disease: from CSF total tau and Abeta(1-42) proteins to phosphorylated tau protein. *Brain Res Bull* 61:243–253, 2003.
27. Morishima-Kawashima M, Ihara Y. Alzheimer's disease: beta-amyloid protein and tau. *J Neurosci Res* 70:392–401, 2002.
28. Antonini A, Leenders KL, Vontobel P, Maguire RP, Missimer J, Psylla M et al. Complementary PET studies of striatal neuronal function in the differential diagnosis between multiple system atrophy and Parkinson's disease. *Brain* 120:2187–2195, 1997.
29. Acton PD, Mozley PD, Kung HF. Logistic discriminant parametric mapping: a novel method for the pixel-based differential diagnosis of Parkinson's disease. *Eur J Nucl Med* 26:1413–1423, 1999.
30. Brooks DJ. PET studies on the early and differential diagnosis of Parkinson's disease. *Neurology* 43 [Suppl 6]:S6–S16, 1993.
31. Tzen KY, Lu CS, Yen TC, Wey SP, Ting G. Differential diagnosis of Parkinson's disease and vascular parkinsonism by (99m)Tc-TRODAT-1. *J Nucl Med* 42:408–413, 2001.
32. Eidelberg D, Moeller JR, Ishikawa T, Dhawan V, Spetsieris P, Chaly T et al. Early differential diagnosis of Parkinson's disease with 18F-fluorodeoxyglucose and positron emission tomography. *Neurology* 45:1995–2004, 1995.
33. Feigin A, Antonini A, Fukuda M, De Notaris R, Benti R, Pezzoli G et al. Tc-99m ethylene cysteinate dimer SPECT in the differential diagnosis of parkinsonism. *Mov Disord* 17:1265–1270, 2002.
34. Stacy M, Jankovic J. Differential diagnosis of Parkinson's disease and the parkinsonism plus syndromes. *Neurol Clin* 10:341–359, 1992.
35. Beal MF. Huntington's disease, energy, and excitotoxicity. *Neurobiol Aging* 15:275–276, 1994.
36. Kieburtz K. Antiglutamate therapies in Huntington's disease. *J Neural Transm Suppl* 55:97–102, 1999.
37. Reynolds GP, Pearson SJ. Glutamate in Huntington's disease. *Lancet* 344:189–190, 1994.
38. Sanberg PR, Johnston GA. Glutamate and Huntington's disease. *Med J Aust* 2:460–465, 1981.
39. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. *Neurology* 57:397–404, 2001.
40. Ellenberg SS. Surrogate endpoints: the debate goes on. *Pharmacoepidemiol Drug Saf* 10:493–496, 2001.
41. Morrish PK. The harsh realities facing the use of SPECT imaging in monitoring disease progression in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 74:1447, 2003.
42. Whone AL, Watts RL, Stoessl AJ, Davis M, Reske S, Nahmias C et al. Slower progression of Parkinson's disease with ropinirole versus levodopa: the REAL-PET study. *Ann Neurol* 54:93–101, 2003.
43. Dopamine transporter brain imaging to assess the effects of pramipexole vs levodopa on Parkinson disease progression. *JAMA* 287:1653-1661, 2002.
44. Biglan KM, Holloway RG. Surrogate endpoints in Parkinson's disease research. *Curr Neurol Neurosci Rep* 3:314–320, 2003.
45. Ansell BJ. Cholesterol, stroke risk, and stroke prevention. *Curr Atheroscler Rep* 2:92–96, 2000.
46. Bots ML, Elwood PC, Nikitin Y, Salonen JT, Freire de Concalves A, Inzitari D et al. Total and HDL cholesterol and risk of stroke. EUROSTROKE: a collaborative study among research centres in Europe. *J Epidemiol Community Health* 56 [Suppl 1]:i19–24, 2002.
47. Bowman TS, Sesso HD, Ma J, Kurth T, Kase CS, Stampfer MJ et al. Cholesterol and the risk of ischemic stroke. *Stroke* 34:2930–2934, 2003.

48. Moroney JT, Tang MX, Berglund L, Small S, Merchant C, Bell K et al. Low-density lipoprotein cholesterol and the risk of dementia with stroke. *JAMA* 282:254–260, 1999.
49. Sheikh K. High-density lipoprotein cholesterol and risk of stroke. *JAMA* 286:1573–1574, 2001.
50. Schuchert A, Behrens G, Meinertz T. Impact of long-term ECG recording on the detection of paroxysmal atrial fibrillation in patients after an acute ischemic stroke. *Pacing Clin Electrophysiol* 22:1082–1084, 1999.
51. Yu D, Schaefer PW, Rordorf G, Gonzalez RG. Magnetic resonance angiography in acute stroke. *Semin Roentgenol* 37:212–218, 2002.
52. Bunt TJ, Cropper L. The role of complete cerebral angiography in the evaluation of patients with prior stroke. *Am Surg* 53:77–79, 1987.
53. Ballantyne CM. Surrogate endpoints and newer risk markers in atherosclerosis management. *Am J Manag Care* 7 [Suppl 5]:S144–S147, 2001.
54. Amoli SR, Turski PA. The role of MR angiography in the evaluation of acute stroke. *Neuroimaging Clin N Am* 9:423–438, 1999.
55. Wu WH, Huo SJ, Cheng CY, Hong CJ, Tsai SJ. Association study of the 5-HT(6) receptor polymorphism (C267T) and symptomatology and antidepressant response in major depressive disorders. *Neuropsychobiology* 44:172–175, 2001.
56. Yu YW, Chen TJ, Hong CJ, Chen HM, Tsai SJ. Association study of the interleukin-1 beta (C-511T) genetic polymorphism with major depressive disorder, associated symptomatology, and antidepressant response. *Neuropsychopharmacology* 28:1182–1185, 2003.
57. FDA approves clozapine for treatment of schizophrenia; careful monitoring required. *Hosp Community Psychiatry* 40:1310, 1989.
58. Olanow CW. Tolcapone and hepatotoxic effects. Tasmar Advisory Panel. *Arch Neurol* 57:263–267, 2000.
59. De Gruttola VG, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedman L et al. Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop. *Control Clin Trials* 22:485–502, 2001.
60. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature. XIX. Applying clinical trial results. A. How to use an article measuring the effect of an inter-
vention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA* 282:771–778, 1999.
61. Zhang Y, Dawson VL, Dawson TM. Oxidative stress and genetics in the pathogenesis of Parkinson's disease. *Neurobiol Dis* 7:240–250, 2000.
62. Youdim MB, Grunblatt E, Levites-Royak Y, Mandel S. Drugs to prevent cell death in Parkinson's disease. Neuroprotection against oxidative stress and inflammatory gene expression. *Adv Neurol* 86:115–124, 2001.
63. Jenner P. Oxidative stress in Parkinson's disease. *Ann Neurol* 53 [Suppl 3]:S26–S38, 2003.
64. Jenner P, Olanow CW. Oxidative stress and the pathogenesis of Parkinson's disease. *Neurology* 47 [Suppl 3]:S161–S170, 1996.
65. Ebadi M, Govitrapong P, Sharma S, Muralikrishnan D, Shavali S, Pellett L et al. Ubiquinone (coenzyme q10) and mitochondria in oxidative stress of Parkinson's disease. *Biol Signals Recept* 10:224–253, 2001.
66. Peck CC, Rubin DB, Sheiner LB. Hypothesis: a single clinical trial plus causal evidence of effectiveness is sufficient for drug approval. *Clin Pharmacol Ther* 73:481–490, 2003.
67. Food and Drug Administration Modernization Act of 1997. Public Law 105-115, Page 111, STAT 2295, 1997.
68. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 8:431–440, 1989.
69. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH et al. Mortal and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 324:781–8, 1991.
70. Ravina BM, Fagan SC, Hart RG, Hovinga CA, Murphy DD, Dawson TM et al. Neuroprotective agents for clinical trials in Parkinson's disease: a systematic assessment. *Neurology* 60:1234–1240, 2003.
71. DATATOP: a multicenter controlled clinical trial in early Parkinson's disease. Parkinson Study Group. *Arch Neurol* 46:1052–1060, 1989.
72. Landau WM. Clinical neuromythology IX. Pyramid sale in the bucket shop: DATATOP bottoms out. *Neurology* 40:1337–1339, 1990.