Surrogate Outcomes
A Cautionary Note

Joanna M. Wardlaw, MD

Drug development for acute stroke treatments is complex, very expensive, very time-consuming, and frequently disappointing. Surrogate outcome theory states that new therapies for stroke could be evaluated with smaller sample sizes than are required when traditional functional outcome measures are used by using surrogate outcomes in clinical trials. This would enable new therapies to be discarded or confirmed as candidates for larger definitive clinical trials more rapidly than at present.

This theory presupposes that surrogate outcomes work by removing the “noise” inherent in all clinically based assessments so that the therapeutic “signal” can be detected more clearly, thus requiring fewer patients to achieve a positive result. It also assumes that the effect of the treatment on the surrogate outcome profiles the therapeutic signal equally and uniformly across all key clinical characteristics (like stroke severity or age). Thus, it assumes that the surrogate mirrors both the beneficial effects and risks so that no systematic bias is inadvertently introduced that might skew the results away from the true therapeutic effect. Smaller samples should reduce costs, speed up drug assessments, and hence enable effective treatments to be found more quickly. It is therefore understandable why surrogates appear attractive, particularly for future trials in acute ischemic stroke, an area in which the discovery of effective new treatments has proved elusive.

Imaging has the potential to provide various surrogate outcome markers for acute ischemic stroke trials. Evidence of ischemic tissue injury is easy to see and appears simple to measure on MR diffusion imaging performed very early after onset; on fluid-attenuated inversion recovery at subacute times; and on T2-weighted imaging at later times. Recent thinking proposes that the amount of potentially salvageable penumbral tissue may be estimated by the difference between the perfusion and diffusion imaging lesions at presentation. Thus, the amount of potentially salvageable penumbral tissue may be estimated by the difference between the initial ischemic lesion at presentation, the amount of reduction in the extent of the perfusion lesion. Thus, the size of the initial ischemic lesion at presentation, the amount of growth by later time points, the proportion with, and the extent of, any penumbra (by cross-sectional area or volume) could all be determined in patients allocated to placebo and compared with patients allocated to active treatment.

However, this presupposes that all patients will have the required imaging outcome and that there is no interaction between survival (ie, being able to obtain the imaging outcome variable of interest, eg, final infarct size) and the treatment under evaluation. It depends on there being no interaction between the 2 main variables of interest (like final infarct volume and treatment allocation) and any other factor such as initial infarct size. It also requires measurement error for key outcome measures to be small and less than those of clinical outcome measures and, finally, that each imaging variable has been standardized. The measurement error for diffusion lesions is not trivial, and may actually inflate the sample size. The use of thresholds may improve reproducibility but not validity. The choice of perfusion metric or threshold still requires standardization because different perfusion metrics yield substantially different estimates of lesion volume and hence differing proportions of patients with penumbra. All that aside, is there evidence that imaging surrogate outcomes could be useful?

The Echoplanar Imaging Thrombolysis Evaluation Trial (EPITHET) aimed to determine whether the presence and extent of the ischemic penumbra in acute stroke using diffusion and perfusion imaging would predict patients most likely to respond to thrombolytic therapy. Using the EPITHET definition, 86% of patients had penumbra. The 90-day MR scan was required to assess final infarct extent, but because approximately 30% of patients did not have the 90-day scan, the missing scan results were imputed from the “last value carried forward.” The primary analysis showed no difference in geometric mean infarct growth between baseline diffusion-weighted imaging and the 90-day final T2-weighted imaging lesion (exponential of the mean log of relative growth, 1.24 with alteplase and 1.78 with placebo, Student t test, \( P = 0.239 \)) or in the median relative infarct growth (1.18 with alteplase, 1.79 with placebo, Wilcoxon’s test, \( P = 0.054 \)), or in functional outcome between patients allocated to alteplase and patients allocated to placebo. However, reperfusion was more common with alteplase than with placebo and was associated with less infarct growth (\( P = 0.001 \)), better neurological outcome (\( P < 0.0001 \)), and better functional outcome (\( P = 0.010 \)) than no reperfusion. There were too few patients without penumbral tissue to determine whether the effect of alteplase was materially different in patients who had, versus who did not have, penumbra.

In this edition of Stroke, the EPITHET authors further explore whether an earlier time point than 90 days could be used to assess the final infarct because this would help to overcome the problem of increasing numbers of missing
scans at later times. Of 101 patients originally included in EPITHET, 91 were scanned at Days 3 to 5 and only 72 at 90 days.\textsuperscript{7} The authors found little difference in lesion size between 3- to 5-day and 90-day scans; and similar correlations among acute, 3- to 5- and 90-day lesion volume, and the 90-day National Institutes of Health Stroke Scale score. They also found that alteplase reduced infarct growth (change in infarct volume from baseline to 3- to 5-day diffusion-weighted imaging) at 3 to 5 days ($P=0.03$) in contrast to the negative results of the primary analysis when 90-day final infarct volume was used.\textsuperscript{6} They suggest that sample size for future stroke treatment trials could be reduced by using growth in MR diffusion lesion volume between baseline and 3 to 5 days as the primary outcome rather than wait and scan at 90 days and risk loss of outcome data through missing scans. Removing the requirement for 90-day assessment would also reduce costs and speed up trial completion. So what are the bear traps?

One danger is that by focusing on imaging surrogate outcomes, to which only patients who survive and can be scanned can contribute, important adverse clinical effects may be missed. For example, in the case of alteplase in EPITHET, almost twice as many patients who received alteplase did not have the 90-day scan (18 alteplase versus 10 placebo) and this was mostly because they had died (13 alteplase versus 7 placebo). Even at 3 to 5 days, the majority of the 10 patients who were not scanned were from the alteplase group and were unable to have the scan because they had died (6 alteplase versus one placebo). Thus, although the imaging surrogate picked up a positive signal for alteplase (less infarct growth by 3 to 5 days), it completely missed the negative clinical signal of increased death. Perhaps we need imaging surrogates for adverse effects, not just for the beneficial effects. One would hope that if in the future, new drugs for stroke are tested in trials with an imaging surrogate, this would be quickly followed by trials based on relevant clinical outcomes that were large enough to identify important benefits and exclude major hazards. Perhaps use of an imaging surrogate would be less likely to risk missing an important adverse effect in the case of drugs with few hazards, but drugs with few side effects tend to have less striking beneficial effects as well and therefore may be of limited value in acute ischemic stroke.

Are there other interactions that need to be considered regarding use of infarct growth or infarct volume as an imaging surrogate? The imbalance in lesion volume at baseline in EPITHET is a problem, although it was not conventionally significant. Increase in infarct volume is not simply due to an increase in the extent of tissue involved, but is also due to swelling of the affected tissue. The degree of swelling is variable, but all acute ischemic lesions swell up to some extent.\textsuperscript{8} Swelling is maximal at approximately 3 to 5 days.\textsuperscript{9} Simply measuring volume does not differentiate between an increase in infarct size due to true increase in extent from that due to increase in swelling without any increase in extent. The major determinant of infarct swelling is initial infarct size. Large infarcts swell more than small ones.\textsuperscript{9} Therefore, any imbalance in lesion size at baseline between treatment groups, even one which is not conventionally significant, could introduce a systematic bias between treatment groups that is difficult to correct even with appropriate statistical adjustment. The group with larger baseline lesions will have more swelling and therefore apparently larger 3- to 5-day lesion volumes without there necessarily having been any increase in the actual lesion extent. On a similar note, lesions that are smaller to start with are likely to be smaller at 3 to 5 days simply because they have further to grow than larger lesions.

In EPITHET, the baseline lesions were almost 50% larger in the placebo group (20.51 mL) than the alteplase group (14.37 mL, $P=\text{nonsignificant}$, equivalent to diameters of 3.4 cm and 3.0 cm, respectively. This difference might not seem like much, possibly quite difficult to detect by eye, but remember that the majority of the volume of a lesion is in its outer layer. A small proportionate increase in diameter (eg, from 9 to 10 cm [10%]) equates to a much larger proportionate increase in volume (in this example, from 382 to 524 mL [30%]). Thus, the lesions in the EPITHET placebo group at 3 to 5 days would be expected to be larger than those in the alteplase group because, being larger to start with, they were going to swell up more, regardless of whether they actually grew in extent. The greater swelling would proportionately increase their volume even further.

Are imaging surrogates likely to be better outcome predictors than clinical variables? The nonlinear association between the National Institutes of Health Stroke Scale and functional outcome (very steep in mild stroke, much less steep in severe stroke)\textsuperscript{10} explains why, in the case of diffusion imaging lesions, there is a weak independent association between lesion volume and functional outcome in severe strokes, but not in moderate to mild strokes in which the association is swamped by the overwhelmingly powerful association between National Institutes of Health Stroke Scale and functional outcome.\textsuperscript{11,12} Therefore, how close diffusion lesion volume comes to being a true surrogate outcome measure for functional outcome will depend on the case mix of patients included in the trial.

Thrombolysis increases the rate of reperfusion and reperfusion is highly beneficial,\textsuperscript{11} whether spontaneous (as occurs in approximately 20% of patients by 24 hours) or therapeutically induced.\textsuperscript{14} For a given size of infarct, reperfusion within the first 24 hours is also associated with less infarct swelling at 3 to 5 days than if there had been no reperfusion,\textsuperscript{9} a further reason why alteplase-allocated patients would have smaller infarcts at 3 to 5 days. Perhaps, in a situation such as this in which the main mechanism of action of the drug is known, measuring reperfusion would be a more direct surrogate for alteplase. Infarct volume has been widely used as the primary outcome measure in experimental studies of new stroke treatments, yet so many new pharmaceuticals coming through this route have failed to realize their promise in clinical trials, often because an unexpected side effect emerged. Surrogate outcomes may be useful if used cautiously and judiciously, but if so, we must not forget to look “outside the box” to see what else the treatment under assessment might be doing.
Disclosures
None.

References

Key Words: outcome □ diffusion-weighted imaging □ acute ischemic stroke □ surrogate outcome □ thrombolysis
Surrogate Outcomes: A Cautionary Note
Joanna M. Wardlaw

Stroke. 2009;40:1029-1031; originally published online February 26, 2009;
doi: 10.1161/STROKEAHA.108.540641
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/40/4/1029

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published
in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office.
Once the online version of the published article for which permission is being requested is located, click
Request Permissions in the middle column of the Web page under Services. Further information about this
process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/