SUPPORTING INFORMATION (ONLINE ONLY)

Further Details of Standardized Evaluations

Clinical Neurophysiology: Spike detection was performed using spike detection software, and laterality was determined by investigator judgment. A standardized montage was used at all sites, and ictal onset patterns were characterized as well-localized to the ipsilateral temporal lobe, lateralized to the ipsilateral hemisphere but not localized, nondiagnostic, or contradictory, according to a standardized protocol. In order to be randomized, at least three seizures had to be recorded if interictal spikes were unilateral, four seizures had to be recorded if interictal spikes were bilateral or if there were no interictal spikes, and ictal onset had to be lateralized concordant to Class I imaging, or localized to the concordant anterior temporal region for Class II imaging. Participants were excluded if contradictory seizure onsets were recorded, or if extratemporal or purely (> 90%) contralateral interictal spikes or slowing occurred. For seizures that occurred in clusters, defined as four hours or less apart, only the first seizure was counted. Nondiagnostic ictal onsets did not count toward the total required.

Neuroimaging: The MRI acquisition protocol required a 1.5-Tesla MR instrument, and the (regular or turbo) spin-echo sequences included T1- and T2-weighted axial and coronal images. Additional images, including gadolinium, were permitted at the discretion of site investigators. PET was carried out with i.v. 18F-fluorodeoxyglucose (FDG), 5-10 mCi, while the participant was awake, with eyes and ears unoccluded in a dim, quiet environment. The tomography protocol included the entire temporal lobes in every scan. No generalized tonic-clonic seizures were permitted within five days before FDG injection and no complex partial seizures were

permitted within six hours before FDG injection. Continuous EEG monitoring was performed from 30 minutes before until 30 minutes after FDG injection.

Intracarotid Amobarbital Procedure: Memory assessment was based on yes/no recognition of six-to eight objects presented during hemiparesis, interspersed among 16 similar foils. Participants were considered to have failed if they recognized < 50% of objects following the ipsilateral injection. Up to two repeated injections were permitted to adjust for obtundation and other factors interfering with performance. If a participant failed the IAP, the participant continued in the study in the surgical arm, in accordance with the intention-to-treat principle, but did not receive surgery.

Statistical Analysis

The analysis of the primary outcome variable, freedom from disabling seizures during the second year of follow-up, will involve fitting a logistic regression model with treatment group as the factor of interest and center, age group, and side of ictal onset as stratification factors. The treatment effect will be summarized using the adjusted treatment group odds ratio and its associated 95% confidence interval. For continuous outcome variables such as the QOLIE-89 overall raw score, repeated measures analysis of covariance models will be used to estimate treatment effects over time, with effects at two years being of primary interest. Again, center, age group, and side of ictal onset will be included in the model as stratification factors and the baseline value of the outcome variable will be included as a covariate. Treatment effects will be summarized using differences between adjusted group mean responses and their associated 95% confidence intervals.

Two important issues regarding plans for statistical analysis were 1) how to deal with missing data and 2) how to deal with treatment "cross-overs", i.e., participants assigned to the surgical arm who ultimately failed the IAP or declined surgical treatment and participants assigned to the medical arm who ultimately received surgery prior to the end of the scheduled two-year follow-up period. For the primary outcome variable, participants who prematurely withdrew from the trial were considered to have had a disabling seizure during the second year of follow-up. For analyses of continuous secondary outcome variables (e.g., HRQOL), the repeated measures analysis of covariance model used maximum likelihood to estimate the parameters of interest (treatment effects) using available data from all participants. A key assumption underlying this analysis is that the missing data were "missing at random" (MAR), i.e., the probability that the responses were missing for a participant depended only on the set of observed data for that participant and not on the specific missing values that were not obtained (Little and Rubin, 2002). Sensitivity analyses using other methods of dealing with missing data (e.g., multiple imputation (Little and Rubin, 2002) and pattern-mixture models (Verbeke and Molenberghs, 2000)) were planned.

The primary analysis will be performed in accordance with the intention-to-treat principle and will include all available data from all randomized participants. All participants will be counted in their originally assigned treatment groups regardless of the treatment actually received during the course of the trial. Secondary sensitivity analyses will be performed that exclude data from treatment "cross-overs" that were obtained after the time that the alternate treatment was received.

Sample Size Considerations

An initial sample size of 65 participants per group was chosen to provide > 90% power to detect a group difference of 80% (surgical arm) vs. 50% (medical arm) in the percentage of participants free of disabling seizures in the second year of follow-up using a χ^2-test and a significance level of 5%. It was anticipated that the actual percentage of seizure-free participants would be less than 30% in the medical arm, however (Wiebe et al., 2001). Therefore, the trial was believed to have more than adequate power to detect the anticipated group difference in seizure-free outcome. The sample size was instead chosen largely to address the important secondary aim of examining treatment group differences in mean change in HRQOL.

For purposes of sample size planning, the change from baseline to two years in the QOLIE-89 overall raw score was the outcome variable considered. The difference in mean response between the surgical and medical arms that was thought to be of minimal clinical importance to try to detect was 11-12 points (Wiebe et al., 2002). Also, the standard deviation of the two-year change in QOLIE-89 overall raw score was conservatively estimated to be approximately 20 points (McLachlan et al., 1997; Wiebe et al., 2001). Because of the absence of a valid quality of life instrument that can be applied to both adults and adolescents, the analyses of the HRQOL outcomes would have to be performed separately for these two age groups. It was anticipated that 85% of the trial participants would be adults, so the sample size considerations focused on detecting treatment effects on the QOLIE-89 overall raw score in the adult subset of the trial cohort.

A sample size of 55 adult participants per group (110 total) was chosen to provide power ranging between 80% and 90% to detect group differences in mean response of 11-12 points, using a t-test and a significance level of 5% (two-tailed). This does not take into account treatment "cross-overs", however. Let π_1 be the proportion of participants in the surgical arm who never receive surgery, and let π_2 be the proportion of participants in the medical arm who receive surgery within two years of randomization. Furthermore, suppose that among participants who do not "cross over", the mean response in the surgical arm is μ_1 and the mean response in the medical arm is μ_2. Finally, it will be assumed that participants assigned to the surgical arm who do not receive surgery have a mean response of μ_2 (the same as that for participants in the medical arm who do not "cross over"), and participants assigned to the medical arm who receive surgery have a mean response of θ. If surgical therapy is indeed superior to medical therapy (i.e., if $\mu_1 > \mu_2$), then it would be expected that $\mu_2 < \theta < \mu_1$ since many of the participants who "cross-over" to surgery would not receive surgery immediately after randomization, and the benefit of surgery may not be fully realized at the end of the two-year follow-up period.

If all participants were fully compliant with their assigned treatment (i.e., if there were no treatment "cross-overs"), the treatment effect (surgical – medical) would be $\Delta = \mu_1 - \mu_2$. After taking into account treatment "cross-overs", the treatment effect would be

$$\begin{aligned}\Delta' &= \left[\left(1-\pi_1\right)\mu_1 + \pi_1\mu_2\right] - \left[\left(1-\pi_2\right)\mu_2 + \pi_2\theta\right] \\ &= \left(\mu_1-\mu_2\right) - \pi_1\left(\mu_1-\mu_2\right) - \pi_2\left(\theta-\mu_2\right).\end{aligned}$$

Note that this reduces to $\mu_1 - \mu_2$ if $\pi_1 = \pi_2 = 0$ (no treatment "cross-overs"). Holding all other factors (power, significance level, standard deviation) constant, the sample size required for a

clinical trial is inversely proportional to the square of the treatment effect, $\Delta^2 = (\mu_1 - \mu_2)^2$. Therefore, the inflation factor for sample size due to treatment "cross-over" is given by:

$$C = \left(\frac{\Delta}{\Delta'}\right)^2 = \frac{1}{\left[1 - \pi_1 + \pi_2\left(\frac{\theta - \mu_2}{\mu_1 - \mu_2}\right)\right]^2}.$$

The worst-case scenario in terms of sample size occurs if $\theta = \mu_1$, which will happen if participants assigned to the medical arm who receive surgery realize the full benefit of surgery. In this case,

$$C = \frac{1}{\left(1 - \pi_1 - \pi_2\right)^2}.$$

On the other hand, if participants assigned to the medical arm who receive surgery realize no benefit of surgery, then if $\theta = \mu_2$ and

$$C = \frac{1}{\left(1 - \pi_1\right)^2}.$$

For planning purposes, the conservative assumption was made that 5% of the participants assigned to the surgical arm would refuse surgical treatment or fail the IAP (π_1 = 5%) and 10% of participants assigned to the medical arm would ultimately receive surgical treatment during the trial (π_2 = 10%). In the worst-case scenario described above ($\theta = \mu_1$), the required sample size was multiplied by a factor of C = 1.3841, resulting in a total sample size of 110 × 1.3841 =

152 adult participants, or 76 adult participants per group. Since adult participants are expected to comprise 85% of the total sample, a total of 180 participants (152 adults and 28 children) are required for enrollment.

The above considerations assume that data on the QOLIE-89 are collected for every trial participant. The required sample size was increased from 180 to 200 participants to account for an anticipated 10% rate of loss-to-follow-up.

Interim Monitoring

A DSMB appointed by NIH provided independent monitoring of study integrity, data quality, and participant safety. They received periodic reports from the CTCC and Biostatistics Center and met at least twice per year to review the accumulating trial data. Of primary concern were the data on adverse events and worsening in neuropsychological test results. A single formal interim analysis for efficacy was planned after 50% of the participants had completed two years of follow-up. This analysis, however, was not to be based on the primary outcome variable of freedom from disabling seizures during the second year of follow-up because it was thought to be important to prematurely halt the trial only if there was definitive evidence of superiority of surgical treatment in terms of HRQOL after two years of follow-up. Therefore, the interim analysis was to be based on the outcome of change from baseline to two years in QOLIE-89 overall raw score. An O'Brien-Fleming (O'Brien and Fleming, 1979) adjusted significance level of 0.0052 (two-tailed) was chosen for the interim analysis so that the significance level of the final analysis (0.048) was maintained close to the originally-planned 0.05.

REFERENCES

Little RJA, Rubin DB. (2002) *Statistical Analysis with Missing Data, Second Edition*. John Wiley and Sons, Hoboken, NJ.

McLachlan RS, Rose KJ, Derry PA, Bonnar C, Blume WT, Girvin JP. (1997) Health-related quality of life and seizure control in temporal lobe epilepsy. *Ann Neurol* 41:482-489.

O'Brien PC, Fleming TR. (1979) A multiple testing procedure for clinical trials. *Biometrics* 35:549-556.

Verbeke G, Molenberghs G. (2000) *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

Wiebe S, Blume WT, Girvin JP, Eliasziw M. (2001) A randomized, controlled trial of surgery for temporal lobe epilepsy. *N Engl J Med* 345:311-318.

Wiebe S, Matijevic S, Eliasziw M, Derry PA. (2002) Clinically important change in quality of life in epilepsy. *J Neurol Neurosurg Psychiatry* 73:116-120.

Table S1. Quality of Life and Psychosocial Measures

Domain	Instrument(s) / Responder	
	Adolescents	Adults
Self Reported QOL	QOLIE-AD-48 / patient (Cramer et al., 1999) CHQ-98-PF / parent (Landgraf and Abetz, 1996)	QOLIE-89 / patient (Devinsky et al., 1995) ESI-55 / patient (Vickrey et al., 1992)
Psychopathology	CBCL / parent (Achenbach and Edelbrock, 1991) Structured interview K-SADS (Kaufman et al., 1997)	MINI-SCID / patient (Sheehan et al., 1994) Structured interview
Behavior	CBCL / parent	
Socialization	CBCL / parent	
Locus of control		Rotter's I-E scale / patient (Rotter, 1966)
Neuroticism	PANAS / patient (Peters and Derry, 2001)	PANAS / patient
Family Function	FAD / patient (Byles et al., 1988) FEICS / patient (Shields et al., 1992)	FAD / patient FEICS / patient
Stressful Life Events	Coddington's scale / parent (Coddington, 1972)	LES / patient (Sarasan et al., 1978)
Academic Achievement	Grade point average Educational stream	
Employment		Employment status

CBCL: Child Behavior Check List
CHQ: Child Health Questionnaire
ESI: Epilepsy Surgery Index
FAD: Family Assessment Device
FEICS: Family Emotional Involvement and Criterion Scale
LES: Life Experience Survey
MINI-SCID: Mini International Neuropsychiatric Interview, Structured Clinical Interview for DSM-III-R
PANAS: Positive Affect Negative Affect Schedule
QOLIE: Quality of Life in Epilepsy
QOLIE-AD: Quality of Life in Epilepsy for Adolescents

Table S2. Schedule of Follow-Up Evaluations

Seizure log and report forms for each seizure
Site coordinator to call biweekly for the first 12 weeks

Clinic Visits	Tests Administered
Baseline	FAD, FEICS-PC, PANAS, BSI, MINI or K-SADS, QOLIE-89/ESI-55 or QOLIE-48-AD, CHQ and CBCL, neuropsychological tests, ancillary outcomes, MRI, PET, adverse events, seizure logs
3 months	MRI (surgical patients only), adverse events, seizure logs
6 months	PANAS, BSI, MINI or KSADS, QOLIE-89/ESI-55 or QOLIE-48-AD, CHQ and CBCL, adverse events, seizure logs
9 months	Adverse events, seizure logs
12 months	FAD, FEICS-PC, PANAS, BSI, MINI or K-SADS, QOLIE-89/ESI-55 or QOLIE-48-AD, CHQ and CBCL, neuropsychological tests, ancillary outcomes, adverse events, seizure logs
15 months	Adverse events, seizure logs
18 months	PANAS, BSI, MINI or KSADS, QOLIE-89/ESI-55 or QOLIE-48-AD, CHQ and CBCL, adverse events, seizure logs
21 months	Adverse events, seizure logs
24 months	FAD, FEICS-PC, PANAS, BSI, MINI or K-SADS, QOLIE-89/ESI-55 or QOLIE-48-AD, CHQ and CBCL, neuropsychological tests, ancillary outcomes, MRI, PET, adverse events, seizure logs

BSI: Brief Symptom Inventory
K-SADS: Kiddie Schedule of Affective Disorders and Schizophrenia
MRI: Magnetic Resonance Imaging
PET: Positron Emission Tomography

Others as in Table 1