Epilepsia, 52(9):1750–1756, 2011
doi: 10.1111/j.1528-1167.2011.03155.x

SPECIAL REPORT

Epilepsy imaging study guideline criteria: Commentary on diagnostic testing study guidelines and practice parameters

***†William D. Gaillard, ‡J. Helen Cross, §John S. Duncan, ¶Hermann Stefan, †William H. Theodore, and Task Force on Practice Parameter Imaging Guidelines for the International League Against Epilepsy, Commission for Diagnostics**

***Center for Neuroscience, Children's National Medical Center, George Washington University, Washington, District of Columbia, U.S.A.; †National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, U.S.A.; ‡Neurosciences Unit, University College of London Institute of Child Health, and Great Ormond Street Hospital for Children, London, United Kingdom; §Department of Clinical and Experimental Epilepsy, University College of London Institute of Neurology, London, United Kingdom; and ¶Epilepsy Center – Neurological Clinic, University Hospital Erlangen, Germany**

Summary

Recognition of limited economic resources, as well as potential adverse effects of "over testing," has increased interest in "evidence-based" assessment of new medical technology. This creates a particular problem for evaluation and treatment of epilepsy, which are increasingly dependent on advanced imaging and electrophysiology, since there is a marked paucity of epilepsy diagnostic and prognostic studies that meet rigorous standards for evidence classification. The lack of high quality data reflects fundamental weaknesses in many imaging studies but also limitations in the assumptions underlying evidence classification schemes as they relate to epilepsy, and to the practicalities of conducting adequately powered studies of rapidly evolving technologies. We review the limitations of current guidelines and propose elements for imaging studies that can contribute meaningfully to the epilepsy literature.

KEY WORDS: Epilepsy, Imaging, Guidelines.

Motivation and Needs

There are now a bewildering variety of ''guidelines'' and practice parameters, and a burgeoning literature devoted to them. Current trends emphasize applying strict criteria to diagnostic and therapeutic studies in order to assess the strength of evidence presented. Typically a series of studies is reviewed to determine to what extent available evidence may address specific practice questions, the ''quality'' of evidence is rated, and conclusions of varying ''strength'' drawn, often quite weak (for a recent epilepsy imaging example see (Harden et al., 2007), and often followed by recommendations for further research to fill gaps in knowledge.

The American Academy of Neurology (AAN) has a formal guidelines procedure that allows studies to be considered in broad distinct categories: therapeutics; diagnosis; and prognosis, screening, and causation. Each of these is relevant to the role of technology in the evaluation and care of patients with epilepsy. Therapeutic guidelines are the most clear, and are based on a long history of medication and intervention trials (see Table 1 for the AAN classification of evidence). Other examples include the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group (Atkins et al., 2004) employed by the U.K. National Institute for Health and Clinical Excellence (NICE) and by the World Health Organization (WHO). Some but not all guidelines processes use prospective, double-blind, randomized, controlled trials (RCTs) as the highest standard of evidence for therapeutic or diagnostic efficacy (French, 2009). A recent antiepileptic drug (AED) monotherapy guideline published by the International League Against Epilepsy (ILAE) Commission on Therapeutic Strategies adopted stricter criteria than the AAN, adding duration and power criteria for study classification (Glauser et al., 2006).

Others, such as the GRADE method, appear more open to evidence from ''well-designed observational studies'' or cohort studies than does the AAN process. For example, observational studies may be considered to have the same level of evidence as RCTs if there is high relative risk in two or more studies, and no plausible confounders (Atkins et al., 2004). The NICE explicitly considers social and economic

Accepted May 5, 2011; Early View publication July 8, 2011.
Address correspondence to William Davis Gaillard, M.D., Department of Neuroscience, Children's National Medical Center, 111 Michigan Ave NW, Washington, DC 20010, U.S.A. E-mail: wgaillar@cnmc.org

Table 1. American Academy of Neurology (AAN) classification schemes

Therapeutic
- Class 1. Prospective, randomized, controlled; masked to outcome, representative populations
  - (1) Outcome defined
  - (2) Inclusion/exclusion defined
  - (3) Account for dropouts
  - (4) Baseline characterization
- Class 2. Prospective, matched group cohort; masked to outcome, representative populations with above OR randomized clinical trial lacking one of four criteria in class 1 study above
- Class 3. Controlled trial representative population, outcome independently assessed, objective outcome measures (control population, natural history or patients)
- Class 4. Uncontrolled, case series, opinion, and so on.

Diagnostic
- Class 1. Prospective, broad population, defined by GOLD standard for case definition, test applied in blinded evaluation, (can assess test for diagnostic accuracy), all patients are + or – for disease determined
- Class 2. Prospective narrow population OR retrospective broad population with condition defined by GOLD standard compared to broad controls; test applied in blinded fashion
- Class 3. Retrospective, patients and controls narrow spectrum; references standard measured (if not objective, performed by outside person who did not perform test)
- Class 4. No independent evaluation; case series without controls

Prognostic
- Class 1. Prospective, broad population, predictive ability using GOLD standard for case definition, Predictor (test) masked to presentation; outcome measured masked to presence of predictor/test. All have test and outcome measured
- Class 2. Prospective narrow population OR retrospective broad population with condition compared to broad controls; measures prognostic accuracy of factor/test using GOLD standard for case definition applied in blinded fashion, test measured masked to presence of outcome
- Class 3. Retrospective, patients and controls narrow spectrum predictive ability using GOLD standard for case definition. Outcome measured (if not objective, performed by outside person independent of person who measured test/predictor)
- Class 4. No independent evaluation; Case series without controls

criteria, and includes a wide range of ''stakeholders,'' such as patient groups, in the process (National Institute for Health and Clinical Excellence, 2009). It uses ''expert consensus to make decisions if evidence is poor or lacking.'' And note that although therapeutic studies, especially of medication, lend themselves well to a prospective double-blind study design, this is not the case for diagnostic studies.

Despite clear advantages, including standardization, reduced bias, and reasonable objectivity, applying rigorous approaches to technology-based studies aimed at diagnosis and prognosis may lead to difficulties: The classification criteria have important limitations (see Tables 1 and 2). Technologic approaches, including imaging [computed tomography (CT), magnetic resonance imaging (MRI), and radioisotope based], and neurophysiologic [electroencephalography (EEG) and magnetoencephalography (MEG) now routinely are coregistered with MRI data, making the term ''imaging'' a useful shorthand] studies have several applications, including diagnosing etiology, syndrome classification for clinical trials, prognosis in long-term outcome studies (see population based, epidemiologic (Harvey et al., 1997; Spooner et al., 2006; Shinnar et al., 1994), the proposed new ILAE classification (Berg, 2010) and, perhaps most prominently, focus localization, to plan epilepsy surgery and predict surgical outcome.

Epilepsy imaging investigators may add to the problem by failing to report complete data, as well as understand the criteria and processes for assessing the strength of clinical

Table 2. Essential elements of a quality imaging or neurophysiologic studies

No.	Item
1	Clear study question and clearly stated study design
2	Clearly defined study population, and where relevant note study base, based on agreed diagnostic categories
3	Clearly defined control population
4	Prospective data collection; where possible following standardized protocols
5	Test applied to all patients uniformly (unless randomized design)
6	Clearly defined experimental measure and comparison/ outcome measure
7	Data analysis clearly defined, preferably objective measures; if not, skilled visual raters, better if measure of replicability provided
7	Assessments blinded to patient identity from the rater and from caring physicians
8	Assessment of population size and homogeneity/heterogeneity and study power
9	For surgical series, pathologic confirmation
10	For outcome, follow up >1 year ascertained by person without a vested interest in outcomes
11	State how data were (not) considered in decision making process
12	Provide data in tabular form for external assessment
13	Data analysis with appropriate statistical test (validity, sensitivity, specificity): comparison with other method
14	State practicalities and limitations, including sources of selection bias, or insurmountable factors modifying above statistical measures (e.g., known incomplete resection)
15	For surgical outcome, list seizure freedom and degree of seizure reduction in those not seizure-free

evidence. However, existing guidelines may in part be inappropriate for imaging and neurophysiologic studies. This review examines controversies and challenges that confront investigators in study design and conduct. The review also provides suggestions for how best to organize and conduct a study that will provide optimal information and contribute meaningfully to the literature and to improved practice.

Main Challenges and Limitations of Current Literature

Guideline reviews of diagnostic literature—structural imaging, functional imaging, and neurophysiologic studies in epilepsy—seem to raise particular problems leading to ''low'' evidence ratings. Sample sizes are small, and randomization and blinding are uncommon. Most criteria for investigative criteria are designed to assess procedures on fairly narrow ''diagnostic'' criteria, rather than the more fluid localization and prognostic questions important for intractable epilepsy [AAN; Center for Evidence-based Medicine (CEBM); http://www.cebm.net/index.aspx?o=1157 – accessed July 27 2009]. Given the limitations in the available data, and disagreement about the process, it is challenging to develop guidelines based on satisfactory quality of data that would appear generally satisfactory for a number of important questions to help guide clinical practice:

1 What is the best imaging approach for determining the cause(s) and prognosis of epilepsy?
2 What is the best imaging approach for location of the seizure focus?
3 What are the best imaging studies to predict surgical outcome?

Limitations of current imaging guideline criteria

The classification of evidence for diagnostic and outcome studies of technology derives primarily from therapeutic trials (see Table 1), which outline clear study populations, control populations, intervention, measures, and outcomes. Technology does not readily lend itself to classification in this model format. Devices are usually evaluated in terms of their accuracy, reliability, therapeutic potential, and cost effectiveness. In epilepsy studies, devices and techniques usually are directed at diagnosis and prognosis for seizure control. There are several aspects of epilepsy that make application of evidence classification schema problematic.

The course of epilepsy is irregular, with remissions and exacerbations. It may take as long as 10 years after seizure onset for patients to develop persistent ''intractable epilepsy'' (Spooner et al., 2006; Berg, 2009). Imaging modalities used early in prospective studies may be obsolete by the time the data are analyzed, and thus are irrelevant to current practice.

For surgical planning, identifying—or confirming—the area responsible for seizures and, therefore, for surgical resection is considered to be paramount, based on the data showing that patients with focal findings on imaging or neurophysiology do better than those with normal studies (e.g., Mcintosh et al., 2004). These data themselves, however, generally would receive low ratings in the AAN scheme (due to lack of blinding and randomization, among other issues), perhaps doing slightly better in the GRADE classification. To complicate matters, patients may have a restricted zone of epileptogenicity within a structural lesion, a wider zone beyond it, multiple lesions, or a more broadly defined ''epileptogenic network,'' which is not evident on imaging studies.

Imaging studies are predicated on the assumption that a visualized abnormality is linked to cause, pathology, seizure focus, and outcome. MRI evidence of hippocampal sclerosis is usually taken to have pathophysiologic significance. However, this presumption is based on the observation that such MRI findings have been rare in the large number of normal volunteers scanned for neuropsychologic studies. Some investigators suggest that hippocampal sclerosis is not always associated with intractable epilepsy (Stephen et al., 2001; Kobayashi et al., 2002). Moreover, hippocampal sclerosis in the setting of refractory epilepsy may have different significance than when found in new-onset seizure populations (Spooner et al., 2006) or asymptomatic people. The lesion that has been shown to progress over time (Theodore et al., 1999; Mathern et al., 2002) may be a consequence as well as a cause of seizures.

Not all MRI abnormalities—including hippocampal sclerosis, cavernomas, gliomas, and malformations—cause seizures and not all seizures originate from identified structural cerebral abnormalities. It is necessary to establish with clinical and neurophysiologic data whether a given lesion is likely to be responsible for the seizures. Nevertheless, the consensus that identifying clear [hippocampal sclerosis, malformation of cortical development (MCD), tumor; not gliosis or encephalomalacia] imaging abnormalities is associated with good surgical outcome would make it very difficult to perform a prospective study (see also the large scale retrospective ILAE 2004 pediatric surgery outcome data [Harvey et al., 2008]).

Both diagnostic and prognostic classification schemes are based on some variety of a ''final common criterion,'' often referred to, with unintended irony, as a ''gold'' standard. The criterion itself may be elusive or flawed; in some instances there is no standard. The standard for identification of a seizure focus may be based on video–scalp ictal EEG, intracranial ictal EEG, pathology, or postoperative seizure freedom. For diagnostic purposes the standard usually means the seizure focus, initially defined electrophysiologically, with supporting evidence from imaging and sometimes pathology. This approach of course runs the risk of creating circular arguments, although new imaging approaches can be evaluated in comparison to ''established'' ones.

Linking imaging standards to pathology can be difficult as well: changes may be subtle, or missed, due to limited tissue availability and quality for review or insufficient expertise. Moreover, the relation between underlying pathology and clinical seizures is inexact. Pathologic classification schemes are subject to debate and reconsideration; changing pathologic classification schemes, like changing MR technology, can make comparison of new and old data difficult (Palmini et al., 2004; Blümcke et al., 2011).

Many factors may affect clinical outcome. For surgical studies the ideal measure—seizure freedom—is problematic. Surgical outcome depends on the surgeon, the approach, and functional/anatomic constraints. A success rate of <100% may not mean that imaging was incorrect. Sometimes the abnormality or the focus cannot be entirely removed for technical reasons (e.g., vascular), pathologic reasons (e.g., gliomas), or functional reasons (e.g., overlap with eloquent cortex). A reduction in seizures may suggest that the imaging data were correct, but the resection was incomplete. A further difficulty is the variability in time at which postoperative outcome is assessed. Postoperative seizure frequency fluctuates, as may patient compliance with postoperative AED treatment. A patient could be seizure-free for several years, experience one or more seizures, followed by another extended remission, or longer relapse. These confounds will effect sensitivity and specificity measures by underestimating or overestimating the value of diagnostic and prognostic testing.

For language and memory lateralization, the intracarotid amobarbital test (IAT) is often considered a ''gold'' standard. Yet there are clearly flaws: the IAT includes measurable risk, limited time for cognitive assessment of variables of interest, poor validation of memory, inaccurate results of IAT, as well as technical and vascular reasons for failure. Electrocortical stimulation (ECS) is considered the ''gold'' standard for functional localization but is limited in time for assessment, and sampling can only be performed at sites of implanted electrodes. Postoperative cognitive assessment could be considered a standard, but no study will randomize patients to removal of areas where language or memory are thought to reside on the basis of an imaging procedure—one can only examine unintended adverse surgical consequences.

Common shortcoming of current imaging literature

Although there are flaws in current guideline criteria, the current imaging literature commonly lacks study designs necessary to provide meaningful contributions to clinical practice. A limitation that plagues epilepsy surgery investigations is the size of study populations, especially for new or limited availability technology, and in pediatrics. Initial reports on imaging and physiology studies are usually small (15–30 patients) with follow-up studies rarely >100, and smaller when ionizing radiation is involved. With these limited numbers it is often impossible to generalize findings because of the heterogeneity of patient populations and limited statistical power. Imaging technology also changes rapidly, with upgrades annually and major changes of equipment every 5 years common place. Even at the most active epilepsy treatment sites it takes several years to obtain homogenous patient populations, with a minimum of 12 months postoperative follow-up, that have sufficient power to make meaningful conclusions. Meanwhile new positron emission tomography (PET) ligands or MR sequences may have been introduced.

Only a minority of epilepsy imaging studies have control populations. Exceptions include some adult PET studies, functional MRI (fMRI) language studies, diffusion tensor imaging (DTI), and structural MRI–voxel-based morphometry (VBM)–based approaches to data analysis [primarily structural, DTI, magnetization transfer, fluid-attenuated inversion recovery (FLAIR)] (Cook et al., 1992; Rugg-Gunn et al., 2001; Gaillard et al., 2002; Rugg-Gunn et al., 2003; Salmenpera et al., 2007; Focke et al., 2008a,b, 2009). Ionizing radiation used for PET and single-photon emission computed tomography (SPECT) precludes obtaining normal data in children (Chugani et al., 1987; Gaillard et al., 2002). Even when controls are available, the set may not be large enough to ensure that data accurately reflect population age-related norms; the control population must be appropriately powered for experimental comparisons (e.g., MRI-VBM methods require 30 or more subjects (Focke et al., 2009). Defining control populations for imaging studies in epilepsy populations with respect to outcome is also problematic. In therapeutic trials one can more readily randomize patient populations, and then move to open label or cross-over design (see below). The usual approach is to choose a more or less homogeneous sample of subjects with an epilepsy syndrome of interest (usually temporal lobe epilepsy, TLE) and perform an imaging study in order to compare clinical characteristics and surgical outcome between patients with positive and negative imaging findings.

Therapeutic trials are facilitated by an infrastructure for multisite trials and strict government criteria for approval [e.g., the European Medicines Agency (EMEA) and Food and Drug Administration (FDA)]. There is no mechanism for conducting comparable multisite imaging studies that would be the equivalent of ''pivotal'' medication trials. Other impediments to multisite technology studies include expense, limited availability, and expertise. Perceived technical differences in machines and sequences are viewed as impediments to studies although these differences are less than patient heterogeneity.

Diagnostic data are, with rare exception, used in the decision-making process [for the exception, see Theodore 1992, where fluorodeoxyglucose–PET (^{18}FDG-PET) data were obtained but not provided for surgical planning and intervention]. Sometimes imaging data identify an abnormality that leads to intracranial EEG and subsequent resection in a patient previously considered not to be a surgical candidate

(Salmenpera et al., 2007; Focke et al., 2009). It is difficult, in these circumstances, to test the data independently and without compromising good clinical practice in the use of accepted techniques. For example, it would be difficult to do such studies with MRI, SPECT, language fMRI, or MEG; one should be able to do so with new MRI sequences (diffusion/perfusion).

Recent alternative study designs advocate presentation of novel image or neurophysiologic data, after a case conference decision has been made using standard clinical and imaging material, in order to assess how reconsideration with the new information alters decision making (Medina et al., 2005; Knowlton et al., 2008a,b). Here one does not know what would happen with those patients who do not undergo the procedure and effect ultimate outcome. It is not clear how this can be avoided without compromising good clinical practice. The practice introduces a selection bias; TLE patients with normal MRI may be less likely to have surgery, and the effect is greater for extratemporal lobe epilepsy. Studies often do not evaluate how novel imaging changes practice.

There is a general failure to collect data prospectively. Ideally all the imaging analysis should be done before surgery, unless results of analysis may bias study conduct (e.g., preoperative fMRI to predict postoperative memory outcome). Imaging and physiologic data, inherently objective, lend themselves to independent review; but retrospective analysis may introduce several sources of bias. Many studies do not interpret data or assess outcomes blindly. Data need to be analyzed by a person blinded to patient identity and without a vested interest in the outcome. Most centers do not have special expertise in all imaging modalities, thereby complicating multimodal comparisons.

Another major limitation is the continuing and rapid evolution in technology. Although there are no class 1 studies on 1.5 MRI, imaging has moved to 3T and 7T studies are commencing. New MRI sequences and changes in scanner hardware and software are introduced every few years, but their application and proper place in epilepsy evaluation is not well established. In short, the technology does not stand still long enough to enable adequately powered studies with adequate follow-up to be carried out.

There is also an issue of sensitivity and specificity. Subtle focal cortical malformations are considered to be the likely cause of many cases of nonlesional focal epilepsy. With higher resolution scanners and sequences it will be difficult to be certain that increasingly subtle findings are clinically relevant unless adequate numbers of healthy controls are studied. Last, there is the issue of how to pay for new technologic assessment of efficacy; this is most problematic for new PET ligands, and less an issue for new MRI sequences that can be added to a clinical series. Studies of new data do not test whether a given technology is equivalent, and more importantly do not test when a test may be redundant (e.g., FDG-PET when MRI and video-EEG are concordant, or IAT when fMRI language laterality is clear).

For an ethical clinical trial there must be equipoise between the two arms of the study in terms of patient benefit. This may not be possible with many imaging studies. It would not now be considered ethical to withhold fMRI language lateralization results from a surgical team to determine whether the study could predict postoperative dysphasia. This would, however, be feasible at this time for fMRI studies of memory, not yet generally accepted. Here, there is reasonable equipoise as to whether and how the data should influence surgical decision making.

Solutions and Proposals for the Conduct of Quality Imaging Studies

Although all the current systems of evidence classification have flaws, they all emphasize essential features of a study that could contribute meaningfully to evaluation and care of patients with epilepsy. This section outlines items that can, and should, be incorporated in imaging studies (Table 2). STROBE (http://www.strobe-statement.org) and CONSORT (http://www.bmj.com/content/340/bmj.c869.full) are efforts to help standardize and improve presentation of data from observational studies and randomized trials, elements of which may also help inform planning and reporting of imaging studies. It may be possible to conduct a ''class 1'' epidemiologic study on prognosis for developing intractable epilepsy based on standardized imaging if given enough time (Berg, 2009). However, it is not likely that broad population, randomized imaging trials will be conducted with control populations for epilepsy surgery. We propose below study designs and elements that address many of the current difficulties in the epilepsy imaging literature. Studies that contain these essential elements should be strongly considered as meeting best clinical research practice that informs clinical care.

Investigators must clearly define the clinical or pathophysiologic question (e.g., comparison with EEG, pathology, surgical outcome, IAT, other imaging) and design a study to answer it. The patients and data should be prospectively obtained with clearly defined populations and study selection criteria, in agreed diagnostic categories. Because patients in imaging/neurophysiology epilepsy studies are unlikely to be randomized, the imaging modality should be applied to all patients with the caregiver blinded, when equipoise is present, to study result. The image analysis methods and measures should be clearly defined. Preferably the image data should be assessed by objective, quantitative measures, or where not possible, by expert blinded raters, with a separate image set used to assess interpretative reliability. All assessments need to be blinded to patient

identity from the rater, and when equipoise is present, from the caregiver.

Studies need to contain a sufficiently large patient and control population and be powered to accommodate heterogeneity and allow statistically valid subgroup analyses of more homogeneous subpopulations. Where diagnostic considerations are paramount, pathologic confirmation should be provided in surgical series; these data should be analyzed on image findings not pathology findings. Where outcome is paramount, a prolonged (at least 1 year) and complete follow-up should be made; outcomes should be defined and ascertained by a person without a vested interest in the outcome.

Control populations are the hallmark of any clinical study, yet remain problematic for epilepsy. Some studies more readily lend themselves to normal control populations and need to be used whenever possible. Other studies will be conducted only in patient populations, where the next best option is to examine the data between those who undergo a procedure in question or who do not have the procedure. In this setting comment cannot be made, particularly regarding outcomes, on those who did not have the procedure.

Ideally the experimental data will not be used in the decision process. Where ethical restraints prohibit such a design, one can make a decision without the data and then reconsider the clinical decision with the data provided (change in practice model). In these circumstances meticulous documentation of how the information altered decision making would need to be provided. For example, one scenario to establish the utility of a new test is to apply the new test to cases in whom a clinical answer is not clear (e.g., nonlesional) and then to determine if new information is provided that changes the plan (proceed vs. not proceed to surgery) and then whether it leads to a good outcome. Other possible models are to set up a sham committee with the two sets of data, or to set up a study where one center employs the new technology and the other does not in order to see if the new technology influences outcomes presuming comparable patient populations and, where relevant, surgical approach and expertise. In this circumstance, data would need to be examined to assure the patient populations are comparable.

The investigators should provide a data table showing results for each subject explicitly. The presentation of data allows independent assessment, facilitates comparison of data, and facilitates future meta-analysis. The data should be analyzed with the appropriate statistical test, which will usually be some variant of ''validity:'' sensitivity and specificity, and positive predictive value. It is also important to acknowledge limitations including potential sources of referral bias. Methods should be clear, and when possible with standardization protocols, in order to facilitate study replication and pooling of data across specialty centers. A broad range and spectrum of patients necessary for class one diagnostic and outcome studies are unlikely to derive from any single center. If a different method is used, comparison to more common methods should be included with a determination of positive contribution and redundancy made.

Potential conflict of interest needs to be addressed in guideline development. In addition to relationships with industry, it is important to consider that investigators may have substantial clinical income, grant support, or academic publications and prestige related to particular techniques.

What Can Be Achieved?

The U.S. Institute of Medicine recently issued a report (http://www.iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-Systematic-Reviews.aspx) with a set of standards already generally adhered to by most organizations, and in particular designed to evaluate comparative effectiveness data, little of which exists, as yet, for epilepsy imaging. The guideline process is in flux, with a desire to achieve at least some degree of international harmonization. One risk is that guideline processes with the most rigorous evidence classification schemes will be diluted in the interests of compromise. However, objective and rational assessments and procedures are necessary that meet the demands and constraints of what is practicable and achievable.

The care of patients with epilepsy will be improved when those who care for patients with epilepsy have a clear sense of the quality and integrity of data we draw upon to make decisions for our patients. Ideally a standardized approach with standardized assessments will be made. With standardized assessment and collection, large repositories may be established. Such approaches will allow for converging evidence from small studies and facilitate meta-analyses based on good data in absence of large scale studies. Large repositories allow discernment, within a heterogeneous population, based on multiple clinical variables (such as the ILAE pediatric epilepsy surgery outcome project and the National Institute of Neurological Disorders and Stroke common measures initiative). Care in acquisition of image data and clinical variables using these methods proposed will improve the quality of data and clinical care. Moreover, in a field evolving as rapidly as epilepsy imaging, guidelines must be reviewed frequently.

Acknowledgments

We thank Mr Alexander Zeitchick for assistance in manuscript preparation.

Disclosure

William D Gaillard has served on an educational course supported by Lundbeck Inc. His department derives clinical income from the evaluation and management of children with epilepsy, and receives research support from Lundbeck Inc., King Pharmaceuticals, PRA International, Eisai Inc., and Marinus Pharmaceuticals, Inc. He is supported by federal funding from

the NIH [NINDS 1R01NS44280-01 (PI) and NICHD 1P30HD40677-01 (IDDRC, core director), NCRR 1K12RR17613-01 (mentor), NIMH 1 R01 MH065395-01A2 (Co-I), and CDC-APTR R-03 (Paid consultant)].

J Helen Cross Reports no conflict of interest. John S Duncan has received Institutional support from Eisai, GSK, Janssen-Cilag, UCB Pharma, GE Healthcare, Medtronic, and research grants from Wellcome Trust, Medical Research Council, Action Medical Research, European Union. Hermann Stefan has received honoraria and travel support for talks and consultations from Desistin Arzneimittel GmbH, Electa, Eisai GmbH, GlaxoSmithKline, Novartis Pharma, UCB Pharma, and Eisia GmbH. In addition he was supported by federal funding from the German Ministry of Health, DFG, and OTAN. William H Theodore received honoraria from serving as an editor of *Epilepsy Research*, receives research support and salary from NINDS DIR, and holds stock options in GE.

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

References

Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S. (2004) Grading quality of evidence and strength of recommendations. *BMJ* 328:1490–1494.

Berg AT. (2009) Identification of pharmacoresistant epilepsy. *Neurol Clin* 27:1003–1013.

Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, Engel J, French J, Glauser TA, Mathern GW, Moshé SL, Nordli D, Plouin P, Scheffer IE. (2010) Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia* 51:676–685.

Blümcke I, Thom M, Aronica E, Armstrong DD, Vinters HV, Palmini A, Jacques TS, Avanzini G, Barkovich AJ, Battaglia G, Becker A, Cepeda C, Cendes F, Colombo N, Crino P, Cross JH, Delalande O, Dubeau F, Duncan J, Guerrini R, Kahane P, Mathern G, Najm I, Ozkara C, Raybaud C, Represa A, Roper SN, Salamon N, Schulze-Bonhage A, Tassi L, Vezzani A, Spreafico R. (2011) The clinicopathologic spectrum of focal cortical dysplasias: a consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission. *Epilepsia* 52:158–174.

Chugani HT, Phelps ME, Mazziotta JC. (1987) Positron emission tomography study of human brain functional development. *Ann Neurol* 22: 487–497.

Cook MJ, Fish DR, Shorvon SD, Straughan K, Stevens JM. (1992) Hippocampal volumetric and morphometric studies in frontal and temporal lobe epilepsy. *Brain* 115 (Pt 4):1001–1015.

Focke NK, Symms MR, Burdett JL, Duncan JS. (2008a) Voxel-based analysis of whole brain FLAIR at 3T detects focal cortical dysplasia. *Epilepsia* 49:786–793.

Focke NK, Yogarajah M, Bonelli SB, Bartlett PA, Symms MR, Duncan JS. (2008b) Voxel-based diffusion tensor imaging in patients with mesial temporal lobe epilepsy and hippocampal sclerosis. *Neuroimage* 40:728–737.

Focke NK, Bonelli SB, Yogarajah M, Scott C, Symms MR, Duncan JS. (2009) Automated normalized FLAIR imaging in MRI-negative patients with refractory focal epilepsy. *Epilepsia* 50:1484–1490.

French JA. (2009) Is the epilepsy responsive or resistant? Only time will tell. *Ann Neurol* 65:489–490.

Gaillard WD, Kopylev L, Weinstein S, Conry J, Pearl PL, Spanaki MV, Fazilat S, Venzina LG, Dubovsky E, Theodore WH. (2002) Low incidence of abnormal (18) FDG-PET in children with new-onset partial epilepsy: a prospective study. *Neurology* 58:717–722.

Glauser T, Ben-Menachem E, Bourgeois B, Cnaan A, Chadwick D, Guerreiro C, Kalviainen R, Mattson R, Perucca E, Tomson T. (2006) ILAE treatment guidelines: evidence-based analysis of antiepileptic drug efficacy and effectiveness as initial monotherapy for epileptic seizures and syndromes. [see comment]. *Epilepsia* 47:1094–1120.

Harden CL, Huff JS, Schwartz TH, Dubinsky RM, Zimmerman RD, Weinstein S, Foltin JC, Theodore WH; Therapeutics & Technology Assessment Subcommittee of the American Academy of N. (2007) Reassessment: neuroimaging in the emergency patient presenting with seizure (an evidence-based review): report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 69, 1772–1780.

Harvey AS, Berkovic SF, Wrennall JA, Hopkins IJ. (1997) Temporal lobe epilepsy in childhood: clinical, EEG, and neuroimaging findings and syndrome classification in a cohort with new-onset seizures. *Neurology* 49:960–968.

Harvey AS, Cross JH, Shinnar S, Mathern BW. (2008) Defining the spectrum of international practice in pediatric epilepsy surgery patients. *Epilepsia* 49:146–155.

Knowlton RC, Elgavish RA, Bartolucci A, Ojha B, Limdi N, Blount J, Burneo JG, Ver Hoef L, Paige L, Faught E, Kankirawatana P, Riley K, Kuzniecky R. (2008a) Functional imaging: II. Prediction of epilepsy surgery outcome. *Ann Neurol* 64:35–41.

Knowlton RC, Elgavish RA, Limdi N, Bartolucci A, Ojha B, Blount J, Burneo JG, Ver Hoef L, Paige L, Faught E, Kankirawatana P, Riley K, Kuzniecky R. (2008b) Functional imaging: I. Relative predictive value of intracranial electroencephalography. *Ann Neurol* 64:25–34.

Kobayashi E, Li LM, Lopes-Cendes I, Cendes F. (2002) Magnetic resonance imaging evidence of hippocampal sclerosis in asymptomatic, first-degree relatives of patients with familial mesial temporal lobe epilepsy. *Arch Neurol* 59:1891–1894.

Mathern GW, Adelson PD, Cahan LD, Leite JP. (2002) Hippocampal neuron damage in human epilepsy: Meyer's hypothesis revisited. *Prog Brain Res* 135:237–251.

Mcintosh AM, Kalnins RM, Mitchell LA, Fabinyi GC, Briellmann RS, Berkovic SF. (2004) Temporal Lobectomy: long-term seizure outcome, late recurrence and risks for seizure recurrence. *Brain* 127:2018–2030.

Medina LS, Bernal B, Dunoyer C, Cervantes L, Rodriguez M, Pacheco E, Jayakar P, Morrison G, Ragheb J, Altman NR. (2005) Seizure disorders: functional MR imaging for diagnostic evaluation and surgical treatment – prospective study. *Radiology* 236:247–253.

National Institute for Health and Clinical Excellence. (2009) *How NICE clinical guidelines are developed: an overview for stakeholders, the public and the NHS*, 4th ed. National Institute for Health and Clinical Excellence, London.

Palmini A, Najm I, Avanzini G, Babb T, Guerrini R, Foldvary-Schaefer N, Jackson G, Luders HO, Prayson R, Spreafico R, Vinters HV. (2004) Terminology and classification of the cortical dysplasias. *Neurology* 62:S2–S8.

Rugg-Gunn FJ, Eriksson SH, Symms MR, Barker GJ, Duncan JS. (2001) Diffusion tensor imaging of cryptogenic and acquired partial epilepsies. *Brain* 124:627–636.

Rugg-Gunn FJ, Eriksson SH, Boulby PA, Symms MR, Barker GJ, Duncan JS. (2003) Magnetization transfer imaging in focal epilepsy. *Neurology* 60:1638–1645.

Salmenpera TM, Symms MR, Rugg-Gunn FJ, Boulby PA, Free SL, Barker GJ, Yousry TA, Duncan JS. (2007) Evaluation of quantitative magnetic resonance imaging contrasts in MRI-negative refractory focal epilepsy. *Epilepsia* 48:229–237.

Shinnar S, Kang H, Berg AT, Goldensohn ES, Hauser WA, Moshe SL. (1994) EEG abnormalities in children with a first unprovoked seizure. *Epilepsia* 35:471–476.

Spooner CG, Berkovic SF, Mitchell LA, Wrennall JA, Harvey AS. (2006) New-onset temporal lobe epilepsy in children: lesion on MRI predicts poor seizure outcome. [see comment]. *Neurology* 67:2147–2153.

Stephen LJ, Kwan P, Brodie MJ. (2001) Does the cause of localisation-related epilepsy influence the response to antiepileptic drug treatment? *Epilepsia* 42:357–362.

Theodore WH, Sato S, Kufta C, Balish MB, Bromfield EB, Leiderman DB. (1992) Temporal lobectomy for uncontrolled seizures: the role of positron emission tomography. *Ann Neurol* 32:789–794.

Theodore WH, Bhatia S, Hatta J. (1999) Progressive hippocampal atrophy in patients with complex partial seizures: the effect of epilepsy duration. *Neurology* 52:132–136.