

Not all that glitters is gold: A guide to critical appraisal of animal drug trials in epilepsy

*Aristea S. Galanopoulou, and †Wenzhu B. Mowrey

Epilepsia Open, 1(3-4):86–101, 2016
doi: 10.1002/epi4.12021



Aristea S. Galanopoulou is co-Editor in Chief of *Epilepsia Open* and Professor of Neurology and Neuroscience at the Albert Einstein College of Medicine.

SUMMARY

Preclinical studies have produced numerous drugs with antiseizure properties that currently are the standard of care. One third of the human population with epilepsy still continues to have seizures despite the ongoing discoveries. The recognized clinical gaps of care that need to be addressed are the identification of antiepileptogenic and disease-modifying treatments, and treatments for refractory seizures or for seizures and epilepsies with limited or unsatisfactory treatments, such as early life epileptic encephalopathies. In this invited review, we provide a historical summary of the international efforts to reevaluate the strategies adopted in preclinical epilepsy therapy discovery studies. We discuss issues that may affect the quality, interpretation, and validation of preclinical studies and their translation to successful therapies for humans affected with epilepsy. These include the selection of animal models and the study design; research practices that affect rigor (such as appropriate use of statistics and reporting of study methods and results, their validation across models, labs, and preclinical-clinical studies); the need to harmonize research methods and outcome assessment; and the importance of improving translation to clinically appropriate situations. The epilepsy research community is incrementally adopting collaborative research, including consortia or multicenter studies to meet these needs. Improving the infrastructure that can support these efforts will be instrumental in future success.

KEY WORDS: Antiepileptogenesis, Animal model, Seizure, Antiseizure, Preclinical trial, Efficacy endpoint, Drug resistance.

There has been a lot of introspection over the last few years regarding the validity, reproducibility, and translatability of preclinical findings into clinically relevant discoveries that address the current clinical gaps and priorities.¹

Accepted September 19, 2016.

*Saul R. Korey Department of Neurology, Dominick P. Purpura Department of Neuroscience, Laboratory of Developmental Epilepsy, Montefiore/Einstein Epilepsy Center, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York, U.S.A.; and †Division of Biostatistics, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, U.S.A.

Address correspondence to Aristea S. Galanopoulou, Department of Neurology & Neuroscience, Laboratory of Developmental Epilepsy, Albert Einstein College of Medicine, Kennedy Center, 1410 Pelham Parkway South, Rm 306, Bronx, NY 10461, U.S.A. E-mail: aristeia.galanopoulou@einstein.yu.edu

© 2016 The Authors. *Epilepsia Open* published by Wiley Periodicals Inc. on behalf of International League Against Epilepsy.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Concerns that the efficacy of many candidate treatments identified in preclinical studies fails to translate into positive clinical trials or to replicate in animal studies have stirred coordinated efforts from many neurologic research areas to reevaluate the priorities, methodologies, and strategies, and to propose standards for rigorous and clinically translatable practices.

In 2010, the international epilepsy community responded to these concerns by forming a collaboration of investigators involved in therapy development with the task of formulating research priorities and proposing recommendations that would help produce more effective treatments for individuals with epilepsy. Jointly supported by the International League Against Epilepsy (ILAE) and the American Epilepsy Society (AES), this collaborative group met at the First Joint AES/ILAE Translational Workshop in London (2012) to identify the priorities and propose a set of recommendations that could improve epilepsy therapy discovery and validation.² The consensus in this workshop was that,

KEY POINTS

- Animal studies have provided many antiseizure drugs that are currently used in clinical practice
- We still need antiepileptogenic, disease-modifying therapies, and therapies for medically resistant or early life seizures with limited treatments
- Power analysis and appropriate statistical analysis are essential for the correct interpretation of positive or negative results
- To meet the unmet needs, use of appropriate animal models and clinically relevant study designs and endpoints will be necessary
- Harmonization of standards of research practice will be essential to facilitate collaborations and multi-center studies

unlike other research areas, epilepsy has been blessed by the availability of numerous animal models of seizures and epilepsies, which over the course of the years have produced many drugs that are currently approved for use as antiseizure drugs in humans. In the United States, the Anticonvulsant Screening Program (ASP) has been instrumental in this progress.³ In the recent reviews of the ASP program (currently known as Epilepsy Therapy Screening Program [ETSP]), it was estimated that 32,000 compounds had been screened, 10 of which have entered the clinical market, whether based primarily on the ASP/ETSP results or on results that ASP/ETSP screening confirmed or expanded.^{3,4} Independent academic investigators and pharmaceutical companies have contributed additional new therapies that are currently in use. Epileptologists can choose among >30 drugs to treat humans with seizures, balancing efficacy, indications, drug interactions, and safety profiles. Antiseizure drugs are still currently the standard of care.

Several critical reviews have, however, voiced criticisms and concerns that “the current antiseizure drug development has failed to deliver” incrementally over the other drugs. Largely these critiques have been driven by the realization and angst that more needs to be done and new strategies need to be adopted to develop treatments for patient populations who do not benefit from current treatments.^{1,2,5,6} These unmet clinical needs include the following: (1) seizures resistant to current antiseizure drugs; (2) treatments that prevent, stop, or reverse epileptogenesis; (3) epilepsies with no or very limited available treatments, for example, certain early life epileptic encephalopathies and epilepsies; and (4) disease-modifying treatments that prevent or treat epilepsy-related comorbidities. Calls for reevaluation of animal models and screening strategies have led to a first set of proposals to improve technical and methodologic aspects of preclinical studies and the study design of future antiseizure, antiepileptogenic, or disease-modifying therapy discovery studies, develop epilepsy biomarkers, and

improve the reproducibility and translatability of preclinical studies through the implementation of multicenter phase II preclinical studies.^{1,7–11}

These workshop reports, among other publications, have outlined the differences between the preclinical and clinical trials and the different scopes and expectations that these may have. Preclinical studies can better address mechanistic questions and control variables that are heterogeneous in human studies. Animal models are not exact copies or simulations of human disease. There are undeniable genetic, biologic, and habitat differences among species that may affect the pathology, pathogenesis, and outcomes of seizures and epilepsies and drug actions, thereby limiting our ability to predict whether a promising preclinical treatment will be effective or harmful in humans. The landscape is also different in preclinical versus clinical research centers: level of funding, conflicts of interest, number of centers involved, standardization of methods used for data collection and outcome assessment, and research design. How realistic is it to expect that preclinical studies can drastically improve to give us the next-generation treatments and cures for epilepsies and associated comorbidities?

In this invited review, we were asked to critically discuss the current state of preclinical trials that aim to develop new drug treatments for epilepsy and comment on the methodology, interpretation, and relevance of findings to human epilepsies and clinical trials. We focus primarily on issues that enhance the translatability of preclinical data to clinical studies. We briefly discuss issues on the selection and utilization of animal models in therapy screening, the methodologic issues that affect rigor, relevance, and translation of preclinical data to the human studies, including statistical analyses. We comment on practices that could transform and accelerate preclinical research and produce new classes of therapeutic compounds and strategies to guide future clinical trials in relevant target populations.

WHAT HAVE PRECLINICAL DRUG TRIALS IN EPILEPSY DELIVERED? UNMET NEEDS AND WORKS IN PROGRESS

Preclinical therapy trials for seizures in adults

The landscape of epilepsy therapy discovery is undergoing significant changes in response to the reevaluation of strategies, methods, and goals. For example, the new ETSP program has shifted from the model of relying on mostly acutely induced or stimulus-induced models of seizures (maximal electroshock [MES], 6 Hz test, audiogenic seizures, and hippocampal or corneal kindling) and now includes chronic models of drug-resistant seizures or of epilepsy with spontaneous seizures. Currently, identification of promising compounds in the ETSP for drug-resistant

seizures is done in mice using the 6 Hz 44 mA, MES, subcutaneous Metrazol model (pentylenetetrazole), corneal kindling model, and post-kainate status epilepticus (SE) spontaneous bursting in vitro model. Differentiation of compounds that are relevant to spontaneously occurring resistant seizures includes testing in the intrahippocampal kainate model in mice (model of mesial temporal lobe epilepsy [mTLE]), the lamotrigine-resistant amygdala kindled model in rats that produces seizures resistant to sodium blockers, and video-electroencephalography (EEG)-recorded rats with epilepsy induced by kainate-induced SE. Testing in special populations is done in the pilocarpine-induced SE to identify compounds effective in benzodiazepine-resistant SE and the Theiler's murine encephalitis virus (TMEV) epilepsy model. Behavioral toxicity is tested with rotarod, minimal motor impairment assessment of ataxia, gait impairment, and automated locomotor activity assessment. A 2015 update of these protocols and algorithms is available at the ETSP website.⁴ These changes comprise substantial progress toward enabling the identification of compounds with different mechanisms and targets with potential to benefit certain drug-resistant seizures and epilepsies in adults and possibly demonstrate antiepileptogenic effect. Independent academic or industry labs and multicenter collaborations add up to these efforts by identifying new etiologies, new compounds, their targets and chemical properties, testing efficacy and tolerability in a larger variety of animal models and experimental scenarios to define pharmacokinetics-pharmacodynamics (PK-PDs), relevance to specific stages of isogenesis or epileptogenesis, specific target populations, or etiologies. The evolution of concepts and priorities has been reflected in the re-formulation of the *2014 NINDS Benchmarks for Epilepsy Research*,¹² which now prominently recognize the priority in "preventing epilepsy and its progression" (Benchmark II) and "improving treatment options for controlling seizures and epilepsy-related conditions without side effects" (Benchmark III).

Preclinical therapy trials for early life seizures

Seizures and epilepsies are highly prevalent early in life, with newborns, infants, and children exhibiting among the highest age-adjusted incidence rates, whereas in a fifth of individuals with epilepsy, seizures start early in life.^{13–15} Medically resistant epilepsy is at least as common in children as in adults, with almost a third of affected individuals being resistant to current treatments.¹⁶ Extrapolation studies for antiseizure medications have been considered to be appropriate for testing efficacy of drugs on focal-onset seizures in children older than 2 years.^{17–19} However, these studies excluded children with epileptic encephalopathies, SE, progressive cerebral or neurodegenerative diseases, or patients with very frequent seizures that are difficult to count.

Extrapolation does not extend to safety issues or to certain early life epilepsy syndromes that show distinctive

responses to treatments.¹⁷ For example, infantile spasms may respond to hormonal therapy (adrenocorticotrophic hormone [ACTH] or high-dose corticosteroids) or vigabatrin but not to most of the classical antiseizure drugs. Epileptic encephalopathies (Landau-Kleffner syndrome, continuous spike wave in slow wave sleep) may respond to hormonal therapy, whereas seizures in Dravet syndrome can be exacerbated by sodium channel blockers.^{20–23} The neurobiology of the very young brain is very different from that in older age groups; consequently, age- and sex-specific effects, efficacy, and tolerability of tested drugs cannot be predicted by the effects in older subjects.^{24–26} There are also age- and sex-specific mechanisms involved in ictogenesis or epileptogenesis of early life seizures and epilepsies involving different underlying pathologies, network functions, intracellular and neuronal signaling, and communication patterns that are peculiar to certain stages of the disease or developmental stages.^{24,25,27}

Utilization of new animal models for early life medically resistant seizures or epilepsies and consideration of these age- and sex-specific factors involved in a drug's effect, therapeutic or adverse, and in disease pathogenesis, will therefore be of paramount importance in our efforts to address these gaps of care. Recently there has been an emergence of animal models of early life seizures and chronic models of early life epilepsies, genetic or induced, which have been used for therapy discovery.²⁷ A few drugs emerging from such animal models have already acquired orphan drug status for pediatric epilepsy syndromes, like carisbamate and the vigabatrin analog CPP-155.^{28,29} Rapamycin's efficacy in epilepsy due to tuberous sclerosis was first demonstrated in animal models, before being tested in clinical trials.³⁰ The armamentarium of models to investigate early life seizures and epilepsies has grown to include high-throughput screens in a zebrafish Dravet syndrome model, or inducible pluripotent stem cells (iPSCs) used to investigate target mechanisms in certain genetic diseases.^{31,32} The changing landscape of preclinical pediatric therapy development is probably one of the biggest advances and sources of optimism for new therapies and cure for pediatric diseases, which are inherently difficult to test in humans.

INTERPRETING PRECLINICAL THERAPY TRIALS

Animal models: predictive power, validity, and clinical relevance

A common critique in preclinical trials is that the animal model used is not clinically validated or clinically relevant. A validated animal model for therapy discovery for a specific indication would be expected to correctly identify a treatment that is also clinically effective for the same indication in humans. Indeed, several animal models have been useful in identifying such antiseizure treatments, including

acute models (e.g., MES or subcutaneous pentylenetetrazole) or models such as amygdala kindling, although false-positive or false-negative data have been obtained. There are limitations, however, in selecting animal models for therapy discovery strictly based on their perceived validity in this process. First, there is no uniformly accepted definition of clinical validation of an animal model or consensus on how to handle false-positive and false-negative data from a model in this process. Second, for conditions for which there are no clinically available treatments, for example, for drug-resistant seizures or antiepileptogenesis indications, it is not meaningful to discuss validated animal models. In such situations, screening strategies have to rely on the potential of a model to simulate the human condition (phenotypically and/or mechanistically) and to offer itself for testing the specific indications.

An animal model cannot be expected to model humans and cannot model a human disease in all aspects of its primary phenotype and its variations. By the same token, a human epilepsy syndrome is not a perfect model of every individual that is diagnosed with the same syndrome. Building true expectations of what an animal study can model is paramount to both appreciate what science and medicine have accomplished so far, and allow the reevaluation of pre-clinical therapy screening strategies through healthy criticism.

Animal models may model candidate mechanisms or known etiologies (e.g., genetic), may have been found by serendipity (e.g., inbred models of epilepsies), or generated by nonspecific induction methods that induce specific desired phenotypes. For preclinical therapy trials, animal models that exhibit pathologies and seizure phenotypes similar to those of a human syndrome or seizure type have an advantage in that they can best identify with the human condition and define the target population. Models of epilepsy with spontaneous seizures have an advantage over the models of acutely induced seizures in that they allow testing for antiseizure and antiepileptogenic effects with readout the spontaneous seizures. Features that improve the reliability of preclinical data from animal models include having a reproducible phenotype, sufficient prior characterization to permit power analyses and definition of the optimal sample sizes for specific outcome measures and readouts, and to justify the study design. The option to perform video-EEG studies is a strong advantage.

Testing a drug effect across models may inform on whether a treatment may be applicable to similar seizures and/or epilepsies with various mechanisms, yet it may be complicated by the age, strain, species, or phenotypic differences that may occur across models making such comparisons uncertain. We defer the discussion about predictive value and validity of models to systematic reviews of the vast existing literature. Herein we discuss study design issues, including outcome assessments that can affect the

interpretation of drug effects in different types of animal models and may influence comparisons with future clinical trials.

Animal models in preclinical epilepsy drug trials: successes and failures to translate

Models of seizures or epilepsies can be acute or chronic. In acute models, seizures occur only for a few seconds or hours after induction. In chronic models, seizures occur spontaneously for days to months after induction, with or without a latent period without detected seizures.

Testing the antiseizure effects of a drug in an acute model of seizures is typically done by administering the drug to a naive animal prior to seizure induction. A clear advantage is the higher throughput and faster turnover of results compared to the chronic models. Outcomes are then measuring the prevention of or delay of induced seizure occurrence or reduction of the severity of the induced seizures. In essence, this study design would only be expected to predict ability of a drug to prevent or ameliorate seizures that share mechanisms with the induced seizures. A nice example of possible translation is the Metrazol test, which has been used in both rodents and humans. This animal model of acute seizures used extensively for screening for antiseizure drugs, made a transient appearance in the clinical literature as an attempt to differentiate people with epilepsy from those without. Quoting Cure et al., “susceptibility to ‘Metrazol’ seemed related to patient’s convulsive threshold, since it was decreased . . . in those receiving anticonvulsant medication at the time of the examination.”³³ The demonstration that antiseizure drugs with efficacy in acute models of seizures can be useful for human seizures suggests that some of these mechanisms can be at least partially shared. However, it is also not unexpected that acute seizure models may not recapitulate all the mechanisms implicated in human seizures, which have more complex or diverse, often unknown, pathogenic (inducing) mechanisms. A caution in interpreting antiseizure effects shown in acute models is when the tested drug antagonizes the exact mechanisms employed by the inducing drug (e.g., testing a γ -aminobutyric acid [GABA]_A receptor agonist in a model induced by GABA_A receptor antagonist). In such cases, antiseizure effect could be secondary to “insult modification,” and comparisons of its efficacy with other drugs that do not share this mechanism will be difficult without dose–response experiments or testing in models with distinct mechanisms. In addition, the efficacy of a drug may change depending on when it is administered (prior to a seizure, immediately after a brief seizure, or after established status epilepticus),^{34–36} and can be altered by multiple factors including vehicle, species, age and sex factors, brain permeability, and pharmacokinetics. Drug effects in disease-naïve conditions may be very different from those during an established disease process, due to changes in the expression or function of key drug targets during the course of a disease.^{37–43} Testing a drug with

clinically relevant vehicles, testing for target relevance of engagement, using more than one model of seizures including a relevant model of epilepsy or targeted disease or models with different induction methods, and testing across species, ages, and in both sexes may abrogate some of these issues. Clinical scenarios that could fit this study design of drug preadministration include the prevention of possible seizures in the setting of well-recognized insults, for example, anticipated exposure to toxic agents.

Testing the antiseizure effects of a drug in acutely induced seizures with the drug given after seizure onset (“posttreatment”), when feasible, can also inform about the therapeutic window of intervention. This strategy carries a lower probability that “insult modification” may underlie the drug’s effect, and the ability to assess the predrug severity of seizures and normalize results according to these, if needed. In such cases, outcomes look at probability to stop or suppress seizures, duration of antiseizure effect, time to effect, or degree of seizure suppression. The advantage over the pretreatment option is that this is a more clinically applicable scenario, since seizure onset can be identifiable, even when its trigger cannot. A decision that needs to be discussed, however, is whether to adapt the study design to match the clinical setting likely to be utilized in the first in-human trial. For example, comparing drug versus placebo in such cases may be unethical in humans for an experimental drug. On the other hand, designing studies that compare experimental drug versus a comparator drug or administering drug as an add-on to established medical therapy could be better-suited options. In both cases, it is advisable to control for possible pharmacokinetic or drug-drug interactions that could complicate interpretation.

Investigating antiseizure effects in chronic models of seizures and epilepsies has the advantage that (1) testing is done in a “diseased/epileptogenic” brain with spontaneous seizures, (2) concomitant assessment for anticonvulsant and antiepileptogenic effects may be possible, and (3) one can monitor in parallel efficacy and tolerability of the drug, since animals are expected to be freely active between seizures. Such studies are cumbersome, however, due to time requirements, the need to administer the drugs chronically in a nonobtrusive manner (orally or through minipumps), and the care of ensuring proper intake and steady levels of the drug to correctly interpret drug successes, failures, or toxicity. Evidence for target relevance and engagement are also important. Failure to demonstrate an effect in a model where the target is irrelevant cannot contradict a positive effect in a model where target relevance and engagement is demonstrated. The availability of pharmacokinetics-pharmacodynamics (PK-PD) data is necessary to establish drug exposure but also the washout of the drug, which will help interpret potential antiepileptogenic effects.

False-positive effects or failure to translate a therapeutic effect can occur because drugs (1) are too sedative or impair motor abilities of an animal, thereby preventing the motor

manifestations of a seizure; (2) cause side effects in humans, at doses lower than the effective ones, which cannot be predicted in animals; (3) exhibit interactions and altered metabolism in humans who receive other medications for seizures or other conditions; (4) treat animal seizures that are not relevant to the clinical population tested in clinical trials; and (5) target mechanisms not clinically relevant in the context of the clinical trial.

False-negative effects or failure to identify a therapeutic effect can occur because (1) the animal model or experimental study design of the preclinical trial may not be appropriate for the clinical trial context, (2) incomplete dose-searching missed the effective doses, (3) drug targets are irrelevant for the selected animal models or experimental conditions, (4) drugs do not engage targets, and (5) monitoring of drug efficacy is outside the PK-PD informed therapeutic window for this drug.

A common explanation of disparate results in the animal literature is the presence of model, age, sex, contextual, and species differences. Although these may certainly be true and anticipated, studies that provide solid evidence of target relevance and engagement, are informed by PK-PD studies, and interrogate such possible confounders provide more robust evidence for the efficacy or lack of efficacy of a drug. However, the existence of such confounders should not serve as an excuse to neglect aspects of the study that could impair the rigor and robustness of conclusions. Maintaining high standards of rigor and transparency in the manner one reports a study, preliminary or complete, is essential to evaluate the strength of findings and avoid rush decisions to discard or move forward a drug. It is in this light and with the goal to improve the comparability and interpretation of studies, that we discuss in the following section some basic considerations of the statistical design and analyses of preclinical trials.

THE ROLE OF STATISTICS AND POWER ANALYSIS IN THE DESIGN AND ANALYSIS OF PRECLINICAL TRIALS

After selecting the experimental paradigm, appropriate use of statistics is essential to effectively test study hypotheses. It is recommended to pre-specify the following in the design stage: primary endpoint, secondary endpoints, statistical plan for each endpoint, power analysis, interim analyses (optional), and early stopping rules (optional).

Power is defined as the probability of detecting a true effect. Power analysis is most commonly used to determine the sample size needed to detect a given effect size. Sample size has to be properly estimated. When sample size is too small, a study may lack the power to detect a true treatment effect; on the other hand, it may be unethical to run experiments on too many animals. It is conventionally recommended that a study has at least 80% power at 0.05 significance level to balance the tradeoff between type I

(i.e., significance level) and type II errors (equal to $1 - \text{power}$) from statistical hypothesis testing. A two-tailed test is preferred when one is interested in any change, regardless of direction, whereas a one-tailed test is used when the direction of the effect is certain; a two-tailed test usually requires a bigger sample size compared to one-tailed test. Effect size can be estimated from either preliminary data or based on the most relevant literature including effect size from individual studies and the mean effect size from meta-analysis. Mean effect size may be the best because it is the most conservative estimate at the population level; the mean effect is usually computed in a way to account for the effects of the various sample sizes in the individual studies included in the meta-analysis. Power calculations and significance thresholds need to be adjusted for the number of endpoints being tested. It is common in many preclinical studies to examine a large number of endpoints and utilize $p < 0.05$ for each. This practice has high probability in producing a false-positive “significant” result for one of the tested variables. It is also advisable to set the endpoints a priori. Table 1 presents an example that illustrates sample size estimation where the sample size increases with higher study power and smaller effect sizes.

Statistical analysis

Data need to be inspected carefully for correct application of inclusion or exclusion rules, missing data, and outliers. Any animal that had an adverse event, dropped out of the experiment, or died before the completion of the experiment needs to be documented and assessed to see whether it is related to the study condition. Even in the context of preclinical studies where there are relatively homogeneous experimental units and more stringent controls, an animal may not complete the treatment. For example, it may die during a

treatment regimen of repeated daily doses over a few days. In this case, if animals are randomly assigned into treatment groups, intention to treat analysis should be used, where the analysis group is based on the original group assignment.

Baseline characteristics are usually summarized to assess whether groups are comparable at the beginning of the trial. Continuous variables are compared using two-sample *t*-test (for two groups) or analysis of variance (ANOVA; for ≥ 3 groups) if the data follow normal distribution; otherwise, they are compared using Mann-Whitney-Wilcoxon test (for two groups) and Kruskal-Wallis test (for ≥ 3 groups). For the endpoints, time of survival can be analyzed using Kaplan-Meier curves, log-rank tests, and then Cox proportional hazard models for covariate adjustment. Linear mixed models that can account for the repeated observations from the same animal can be used to compare weight gain across groups. Seizure frequencies are not often normally distributed, so the data should be transformed (e.g., logarithmic transformation) before it is analyzed using linear regression models. Responders versus nonresponders can be analyzed using Pearson’s chi-square test (or Fisher’s exact test) for unadjusted analyses and using logistic regression for adjusted analyses. Time-dependent variables such as seizure-free days can be analyzed similarly to time of survival.

In addition to the aforementioned frequentist approach, Bayesian approach can also be used in trials. For both approaches, the statistical plan should be prespecified. A study designed with the frequentist approach cannot be analyzed using Bayesian approach in the end, or vice versa. In the frequentist approach, the unknown parameter, for example, mean of a distribution, is considered to be constant, and any information from previous studies can only be used in the design stage. The frequentist approach has been well established for clinical trials of all kinds including drug

Table 1. Sample size per group to detect a 50% reduction in the proportion of animals with epilepsy

Power	p_1 vs. p_2			
	10% vs. 5%	30% vs. 15%	50% vs. 25%	70% vs. 35%
80%	435	122	59	32
90%	583	163	79	43

The table demonstrates examples for the estimation of sample size. Suppose in a trial, animals are randomized into two groups with equal sample size of n , where one group receives the treatment drug and the other group receives the placebo. We hypothesize that there will be a 50% reduction when comparing the proportion of epilepsy in the treatment group, p_2 , to the proportion of animals with epilepsy in the placebo group, p_1 . Assuming \bar{p} is the average of the two proportions, β is the type II error (power = $1 - \beta$), α is the significance level or type I error, and $Z_{1-\beta}$ and $Z_{1-(\alpha/2)}$ are the quantiles corresponding to the probabilities of $1 - \beta$ and $1 - (\alpha/2)$ in the standardized normal distribution, respectively, then the sample size required in each group is computed by the following formula, where

$$n = \frac{2(\bar{p})(1 - \bar{p}) \left(Z_{1-\beta} + Z_{1-(\alpha/2)} \right)^2}{(p_1 - p_2)^2}.$$

For example, if using a two-sided test at a significance level of $\alpha = 0.05$ ($Z_{1-(\alpha/2)} = Z_{0.975} = 1.96$) and assuming 80% power ($Z_{1-\beta} = Z_{0.80} = 0.84$), $p_1 = 0.10$, $p_2 = 0.05$, and $\bar{p} = 0.075$, then the sample size required in each group is computed as 435. The above table shows the sample size needed in each group when p_1 is 10%, 30%, 50%, or 70%, respectively, at 80% or 90% power ($Z_{0.90} = 1.28$) at a significance level of 0.05 using two-sided tests. The sample size increases with higher power and smaller effect sizes. When the sample size is too small, a study may lack the power to detect a true treatment effect; on the other hand, it may be unethical to run experiments on too many animals in a trial if the sample size is too large.

trials. In the Bayesian approach, we estimate the distribution of the parameter (posterior distribution) based on its prior distribution and the observed data. Prior information can be used in both design and data analysis. Bayesian statistics is computationally intensive, so it was only recently applied in medical device clinical trials thanks to the advancements in computing technology.⁴⁴ Compared to the frequentist approach, incorporation of prior information in the Bayesian approach sometimes may result in smaller sample size or shorter duration for a trial, both of which are very appealing. However, the Bayesian approach faces a few challenges. First, it requires extensive preplanning on the prior distribution, on the data collection plan, and on the model that combines the prior and the data. In fact, the U.S. Food and Drug Administration (FDA) recommends discussing the choice on the prior distribution with them before initiating a trial. Second, it requires extensive model building. Third, it requires involvement of statisticians who specialize in Bayesian methods. Finally, Bayesian and frequentist approaches may differ in their conclusions based on the same data but a different prespecified analysis plan (one Bayesian and the other frequentist), although this gap is currently being resolved by statistical methodologists.⁴⁴ In practice, the frequentist approach continues to be the mainstream.

Interim analyses

Interim analyses are usually performed by an external independent group, preferably consisting of experts in statistics and clinical and preclinical trials. The role of an Independent Data Monitoring Board (IDMB) is quite common in clinical trials but not quite yet in preclinical studies. Unblinded data should be kept confidential from all staff involved in the trial. If interim analysis demonstrates that the effect size is not as anticipated, especially when the original sample size is based on preliminary data, sample size adjustment may be necessary and can be done with proper documentation. A clinical trial may be stopped early if the interim analyses show that there are too many adverse events or superiority of the treatment is clear, or if it is unlikely to show a significant treatment effect.⁴⁵ However, there may be situations when the latter may not necessarily warrant cessation of the studies showing comparative efficacy, but this needs to be pre-set at the beginning of the trial.⁴⁶ It is necessary, however, to report when and why interim analyses prompted changes in the study design and performance, for transparency.

Reporting statistical results

It is recommended that the endpoints and the original statistical plan are prespecified.^{47,48} All analyses should be reported. This includes the prespecified analyses proposed in the original plan. Any additional analyses should be reported and clearly marked as being beyond the original

plan along with the reasons for adding these analyses. Non-significant or negative results should also be reported to prevent other research groups from pursuing the same experiment. In the final results, it is recommended that direction and magnitude of the effect size, the statistical test and actual p-value, and the 95% confidence intervals are reported.^{47–49}

OPPORTUNITIES TO NARROW THE PRECLINICAL-CLINICAL TRIAL GAP AND MEET CLINICAL NEEDS

Even if a preclinical study is well powered and appropriately analyzed and reported, its results do not necessarily foretell the results of a clinical trial. Animal issues (species/strain, biologic, genetic, habitat, environmental) complicate such extrapolations; however, these concerns have not precluded success in bringing a drug to the clinic in the past.⁵⁰ Replication or confirmation of efficacy in similar models of seizures across species may strengthen the confidence of a finding, yet there is insufficient evidence to recommend testing efficacy in more than two species prior to clinical testing.⁵⁰ Regulatory guidelines, however, do require toxicity testing in more than two species. Which among the remediable practices could be optimized to improve translatability to clinical studies but also address the clinical gaps in care?

Homogeneous versus heterogeneous target population: a preclinical–clinical gap in drug trials

A significant difference between the preclinical and clinical trials is the considerably lower heterogeneity in the animal studies compared to humans, even when outbred models are used. This not only refers to variability in the underlying pathology or epilepsy phenotype but also to environmental, comorbid, and other factors that may affect seizure outcomes and drug effects. Although the more controlled conditions of an animal study offer an advantage for the discovery phase of new therapy targets because they help minimize the confounders and therefore the animal use, they may create a challenge when comparing with the clinical trials. Multicenter “phase II” preclinical trials have been proposed to address some of these issues in translation,¹¹ by testing therapies in animals in a manner similar to that done in a clinical phase II trial prior to introducing them to a clinical trial.

Animal models for drug development: beyond seizures or limbic epilepsies

There are numerous models of seizures and epilepsies.⁵¹ Previous publications have addressed the predictability of key models used in therapy development, their appropriateness for certain types of seizures or epilepsies, and suggested changes in the way these are used in therapy development.^{5–11,50–53}

One of the greatest contributions and examples of the old ASP program has been the possibility to create a comparative efficacy/tolerability profile of tested compounds using phenotypic screening with the same protocols, models, and strategies. This systematic and standardized approach generated simple algorithms and rationales for recommending the lead compounds for future clinical testing in human epilepsies. When a plateau was reached, it allowed for constructive criticisms, re-evaluation, and re-structuring so that some of the modern unmet needs and priorities can be addressed in the new ETSP.^{3,4} However, the focus still remains on either acute models of focal or generalized seizures or chronic models with limbic seizures in adult rodents, and attempts to differentiate efficacy in medically resistant seizures or antiepileptogenic potential.

As the field shifts toward the idea of targeting medically resistant epilepsies, epileptogenesis, or disease modification, it may become relevant to study etiology or pathology or age-appropriate models of epilepsy. Currently, therapy development in other models that are common (e.g., pediatric, poststroke, or posttraumatic brain injury models) depends on independent laboratories, many of which have special expertise in specific models and typically perform target-oriented approach. However, these efforts are not streamlined and organized in a manner that would permit comparisons across models and labs, or assessment of comparative efficacy across compounds. Also important is the opportune funding of these labs and funding priorities, which may not guarantee a continuum in therapy screening. Furthermore, it is important to recognize that each model may deliver findings that may not be generalizable to other models or to all human epilepsy syndromes. Developing infrastructure and tools to allow the identification of target populations likely to benefit from discoveries done in certain models, for example, through use of validated biomarkers for therapy implementation, will be essential for the good translation of preclinical findings into the clinics.

It is not realistic to expect that all therapy development be done through the same centralized institution. There is, instead, a recent trend toward team research/consortia of expert clinical and preclinical research groups working on a specific research focus, which is another option that has been initiated with the CURE Infantile Spasms Initiative⁵⁴ and the National Institute of Neurological Disorders and Stroke (NINDS)-funded Centers without Walls. Such initiatives offer unique opportunities to realize that therapy development is a preclinical-clinical continuum and collaboration, and that at the end of the day preclinical testing has to target issues that may transform clinical care. A strong multidisciplinary collaboration is needed of clinical and preclinical trialists, pharmacologists with expertise in PK-PDs and optimal formulations, chemists for lead optimization when necessary, statisticians, researchers involved in biomarker development, and experts on regulatory aspects. Challenges remain, such as the need to agree on common

practices and strategies, share expertise and reconcile data from different models and labs.

Developing drugs for drug-resistant seizures

The definition of drug-resistant epilepsy according to the ILAE is “failure of adequate trials of two tolerated, appropriately chosen and used antiepileptic drug schedules to achieve drug freedom.”⁵⁵ The definition of drug resistance does not depend on etiology but rather differentiates seizures based on response to certain drugs considered appropriate for the seizure type. The etiology of drug resistance in humans is unknown, and probably diverse, although several theories have emerged: target theory, the drug transporter theory, methylation hypothesis, and intrinsic severity factors whereby factors that augment seizure severity (e.g., inflammation, network reorganization, specific pathologies or etiologies, and alterations in synaptic transmission) may contribute to resistance. The models used currently to develop drugs for drug-resistant seizures⁷ employ a more-severe inducing trigger (6 Hz, 44 mA model), preexposure of animals to drugs that result in a resistant phenotype (e.g., lamotrigine-resistant kindled rat), are selected by their resistance to one at least antiseizure drug (e.g., phenobarbital-resistant post-pilocarpine rat with epilepsy), or are generated with pathologies that predispose them to drug resistance (e.g., multiple-hit model of infantile spasms). In most of these models in which extensive testing with multiple antiseizure drugs has been done, there are drugs to which the model is sensitive and drugs to which it is resistant. Therefore, many are models with relative resistance to some of the existing drugs, rather than true models of drug resistance. The expectation is that drugs identified by screening in such models might offer a therapeutic alternative in conditions that may share some of the elements of drug resistance shared with the model, possibly because new drugs employ a different mechanism of action. The challenge is that the etiology of drug resistance is not known in every human with drug-resistant epilepsy to allow validation of these models. Therefore, in our opinion, such models of drug resistance will be useful to identify different drugs that might yield new therapeutic options with probably different mechanisms, which could be worth studying in human drug-resistant seizures. Whether such models will prove to be useful tools for providing viable therapeutic solutions for drug-resistant epilepsies remains to be seen.

There are some additional challenges in studying drug-resistant models. The presence or absence of drug resistance may not be a permanent feature of an epilepsy syndrome. It is not uncommon for patients with human temporal lobe epilepsy to start with a drug-sensitive course, go on to remission, and then re-emerge with a drug-resistance course. Conversely, Sillanpaa and Schmidt⁵⁶ reported that 51% of children with initial drug-resistant childhood-onset epilepsy reached 5-year terminal remission. It is not known if such temporary or delayed drug resistance exists in animal

models, although this is more difficult to document, given the short half-lives and briefer observation periods of these models. For many models their pharmacosensitivity profile has been an inherent criterion for judging their predictive value for new therapies. Is a model that does not respond to the anticipated drugs an invalid model of this epileptic syndrome or a model of drug-resistant epilepsy? An example is the multiple-hit rat model of infantile spasms that does not respond to adrenocorticotrophic hormone and only partially responds to vigabatrin.⁵⁷ In our opinion, response to drugs should be only a modifier (drug-sensitive vs. drug-resistant), not an essential criterion for the validity of the models. This is even more important for certain early life epilepsies that are drug-resistant from the start. The underlying pathology/etiology and observed epilepsy and comorbidity phenotype should be more important for the classification and validity of the models. A final comment is that according to the clinical ILAE definition of drug resistance, one anticipates “failure to achieve seizure freedom.” This may require a significant change in the way preclinical trials are designed in that drug-sensitivity in animal studies may imply some reduction of seizure frequencies, amelioration of certain types of seizures, seizure reduction or disappearance for the brief period of monitoring (rather than terminal remission), whereas dose searching may not always be sufficient to explore the full potential of the drug. We discuss in subsequent sections the issue of seizure freedom definition in animal studies.

Harmonizing preclinical practices in a manner that may improve translation to clinical trials

Progress in clinical epileptology has been made possible due to the existence of structured guidelines of clinical practice and accepted classifications of seizures and epilepsies and standardized protocols for electroencephalography (EEG) tests and interpretation. A recent classification of seizures and epilepsies that applies to companion animals, similar to those in humans,⁵⁸ has been proposed by the International Veterinary Epilepsy Task Force (IVETF).⁵⁹ Yet, there is no such classification for experimental animals and there is no agreement on EEG practices and interpretation. The lack of standards has spurred a lot of discussion and debates among investigators utilizing rodent EEG recordings. Seizure outcomes in epilepsy models are often expressed without differentiating the effects of drugs or intervention on different seizure types. Outcomes (epilepsy or other) are generally dictated by the type of tests used, often differ among labs, and may not translate well to clinical endpoints. Similar issues exist in studies assessing behavior, pharmacology, or physiologic endpoints.

The AES/ILAE Translational Task Force of the ILAE has assembled an international volunteer group of experts (TASKs1–4) to address some of the gaps in the way preclinical and clinical trials are performed.^{1,2,50,60} The creation of a uniform system of interpreting and classifying EEG and

rodent seizures and seizure-like events is one of the key goals of the TASK1 group of the Task Force. The TASK2 group attempts a systematic review of animal literature to evaluate outcome assessment and utilization and predictability of animal models of seizures and epilepsies. TASK3 is currently formulating preclinical common data elements (CDEs) and case report forms (CRFs) for data collection, in alignment with the clinical epilepsy CDEs⁶¹ that have been created recently. TASK4 works towards improving infrastructure supporting multicenter preclinical trials.

Testing clinically relevant treatments

Formulations and routes of delivery that are not appropriate for clinical use or are associated with toxicity are not likely to enter clinical trials. Collaboration with chemists or pharmacology experts early in drug development can help resolve these issues at the early stages when efficacy testing is being conducted. Dose–response experiments guided by tolerability assays help define the maximal tolerated doses (MTDs) in seizure/epilepsy models. The *protective index* (PI), estimated as the ratio median toxic dose (TD₅₀) over the median effective dose (ED₅₀) in 50% of the animals, gauges the safety window of a drug and allows advancement of drugs with high PIs.

The timing of drug administration and monitoring needs to consider clinical relevance and PK-PDs. Pre-treatment is usually done in acutely induced models of seizures, but may be implemented only in conditions when seizure triggers can be predicted (i.e., toxic exposure) or seizures need to be prevented in a high-risk population (e.g., after severe traumatic injury). One can argue that chronic use of antiseizure drugs currently aims to prevent random spontaneous seizures. However, pre-treating naive animals may not necessarily have the same effects as treating an animal with epilepsy to prevent subsequent seizures. Target relevance and engagement may be different in later stages of ictogenesis or epileptogenesis.

PK-PDs and drug levels are critical for the study design and interpretation of the results, positive or negative. Plasma levels are more clinical practice friendly, but brain levels, brain-to-plasma concentration ratio and PK-PDs may ascertain the effective levels of the drug, need for improving the chemical structure, or explain a negative result. PK-PDs, time to peak effect (TPE), and duration of drug exposure and drug washout time help determine when to monitor for efficacy and how to interpret results.

Selecting appropriate endpoints: statistical significance versus biologic importance

Predefining the primary and secondary endpoints of the study and powering the study to test the expected effects at least on the primary endpoints is critical for the success of the study. Many of the parameters and endpoints used in preclinical (Table 2) and clinical (Table 3) epilepsy studies can be extrapolated between the two settings. Others only

Table 2. Endpoints of preclinical therapy development in in vivo seizure/epilepsy models

Endpoints	Preclinical studies	PANACHE database ⁷⁰
Toxicity		
TD50	Yes ⁷¹ (rotarod, observational)	Yes (rotarod, observational)
Minimal motor impairment (MMI)	Yes (observational)	Yes (observational)
Adverse effects of the drug	Yes (model/study-specific: mortality, motor impairment, cognitive and other behavioral tests)	Yes (automated locomotor activity, rotarod, tremors/ataxia, mortality)
Timing of drug delivery		
Effect of timing of treatment on efficacy	Yes (model-specific)	Yes (pilocarpine SE)
Duration of protection by drug	Yes (model/study-specific)	Yes (MES, 6 Hz)
Time to peak effect (TPE)	Yes (model/study-specific)	Yes (MES 6 Hz, bicuculline, picrotoxin, iv Metrazol)
Antiseizure efficacy		
ED50	Yes ⁷¹ (study-specific)	Yes (sc Bicuculline, sc Picrotoxin, Pilocarpine SE, 6 Hz model, Frings audiogenic seizure model)
Protected subjects/total tested	Yes (study-specific; definitions of protected/responders vary)	Yes (pilocarpine SE)
Seizure score	Yes (model/study-specific)	Yes (kindling)
Afterdischarge duration (kindling)	Yes (kindling)	Yes (kindling)
Kindling rate	Yes	
Time to first seizure	Yes (model-specific)	Yes (iv Metrazol test)
Dose of chemoconvulsant to first seizure	Yes ⁷¹ (model-specific)	Yes (iv Metrazol)
Seizure frequency	Yes (vs. vehicle treated, vs. baseline)	
Seizure freedom	Yes (model/study-specific)	
Seizure remission	Yes (model/study-specific)	
Seizure recurrence	Yes (model/study-specific)	
Duration of drug effect	Yes (model/study-specific)	
Effect on mortality from seizures	Yes (study-specific)	Yes
Effects on SE		
SE cessation	Yes (study-specific)	
Time to SE cessation	Yes (study-specific)	
Termination of SE spikes within 30 min	Yes (variable EEG endpoints have been used)	Yes (pilocarpine SE)
50% suppression of electrographic SE	Yes (study-specific)	Yes (pilocarpine SE)
Mean effect of drug vs. vehicle-treated	Yes (study-specific)	Yes (pilocarpine SE)
Power (μV^2) time course	Yes ⁷² (kainic acid SE)	
Seizure severity/burden	Yes (variable measures/scales have been used)	
Neuroprotection	Yes (Fluoro Jade B, silver staining, apoptosis, injury scores, etc.)	Yes (Fluoro Jade in Pilocarpine SE)
Effect on mortality	Yes	Yes (pilocarpine SE)
Effects on epilepsy		
Incidence of epilepsy	Yes	
Seizure burden		Yes (epilepsy from low dose kainic acid)
Seizure frequency	Yes ^{29,73} (vs. baseline of same animal or vs. vehicle treated animals at corresponding observation and treatment periods)	Yes (epilepsy from low dose kainic acid)
Inter-seizure intervals	Yes	
Seizure duration	Yes ⁷³	
Seizure freedom after drug exposure	Yes (variable observation duration)	
Cumulative seizure duration	Yes ⁷³	
Distribution of Racine scales	Yes	Yes (epilepsy from low dose kainic acid)
Frequency of hippocampal paroxysmal discharges (HPDs) (baseline vs. TPE)	Yes	Yes (intrahippocampal kainic acid)
Reduction of HPDs from baseline	Yes	Yes (intrahippocampal kainic acid)

provide evidence of a biologic effect (i.e., kindling rate, dose of chemoconvulsant to first seizure). There is no evidence that certain criteria are more predictive of clinical success, but one would like to believe that those that demonstrate a stronger effect (i.e., seizure freedom vs. 50% reduction) when tested in scenarios similar to the clinical ones, resemble clinical endpoints and have higher PIs could give more confidence in clinical success.

Most of the studies in chronic models of epilepsy report effects on seizures in general and only rarely attempt to differentiate effects on different seizure types. Partly this is due to power analysis issues and partly to the lack of uniform system of classifying and diagnosing seizures in experimental animals. Of note, when carisbamate was tested in the lithium-pilocarpine model, it prevented motor seizures but increased the incidence of spike-wave discharges in

Table 3. Study design and endpoints in clinical epilepsy trials

Study design	European Medicines Agency (EMA) guideline ¹⁷	
Target population		
Seizure type	Inclusion can be given based on a seizure type Seizure types to be evaluated separately (e.g., focal vs. secondarily generalized)	
Epilepsy syndrome	Certain epilepsy syndromes to be evaluated separately Analysis to be done separately by seizure type	
Pediatric populations	Efficacy can be extrapolated from adults in children over 4 years and for focal or genetic generalized epilepsies Testing in certain age-specific epilepsy syndromes (West, Dravet, Lennox-Gastaut, continuous spike wave in slow wave sleep syndromes) needs to be done separately in these populations	
Populations over 65 years	Separate studies are indicated for efficacy, PK-PDs, safety	
Monitoring periods	Separate studies are indicated for efficacy, PK-PDs, safety	
Monitoring period	Pre-defined period	
Baseline	Baseline seizure frequency and duration of observation should be high and low enough to permit detection of changes (increase or decrease) in seizures Patients in whom baseline frequency of seizures differs from their usual frequency should be excluded	
Treatment retention time	Recommended Exit criteria should be predefined	
Endpoints	Clinical studies ⁷⁴⁻⁷⁶	European Medicines Agency (EMA) guideline ¹⁷
Efficacy and toxicity		
Seizure frequency	Change in seizure frequency (primary) Response ratio (primary) <i>Response ratio = difference in seizure frequency between treatment and baseline divided by their sum</i>	Recommended
Seizure freedom	Proportion of seizure-free patients (secondary) Seizure free days (secondary)	Recommended (proportion of patients, distribution) Pre-define duration of seizure freedom period
Time to nth seizure	Secondary endpoint	
Distinction between responders vs. nonresponders	Proportion of responders (primary, >50% reduction in seizure frequency)	Recommended The degree of response should be pre-defined: (i.e., more than 50% reduction, days without seizures, etc.) Potential exacerbation of seizures or new seizures appearing should be also captured and reported
Seizure severity, seizure duration	Secondary endpoint	Recommended
Dose-efficacy studies	Done	Recommended (with plasma level monitoring)
Composite score (seizure frequency, seizure types, adverse effects)		Recommended (secondary endpoint)
EEG pattern for specific syndromes	Done in specific situations	Recommended
Adverse effects	Reported	Recommended to capture and report (including exacerbation of seizures)
Effects on SE		
Cessation of seizures	Primary endpoint May be defined as seizure cessation within X min from drug administration total seizure cessation (for at least X hours after last seizure)	
Time to treatment	Done	
Time from drug delivery to seizure cessation	Done	
In hospital mortality	Primary endpoint	
Length of intensive care stay	Primary endpoint	

Continued

Table 3. Continued.

Study design	European Medicines Agency (EMA) guideline ¹⁷
Adverse events	Secondary endpoint
Duration of mechanical ventilator support	Secondary endpoint
Duration of hospital stay	Secondary endpoint
Cognitive deficits	Secondary endpoint
Long-term outcomes	Secondary endpoint

rats.⁶² As in humans, animal models of epilepsy do not necessarily have one type of seizures. In certain models, epilepsy progresses from predominantly nonconvulsive to predominantly convulsive.⁶³ In others, epilepsy evolves, as in many pediatric animal models.²⁷ Differentiating the effects of a drug on seizure types within the same model, rather than across models of different seizures, may subtract the impact of the underlying pathology and etiology and may help prioritize the effects on different seizure types. Furthermore, the drug effects on seizures versus seizure clusters may be of interest, since clusters may have different outcomes and possible different underlying mechanisms.

Justification for the selection of the baseline and monitoring periods (time points, duration) would be important to report, particularly for chronic treatments in models that demonstrate progression, evolution, or seizure clustering with relatively long intercluster periods.^{27,63,64}

As we advance toward the idea of disease modification and antiepileptogenesis, it becomes important to capture treatment effects, adverse or beneficial, in the population with the specific seizure or epilepsy syndrome rather than in a naive population. Seizures and epilepsies and their coexistent comorbidities alter brain or systemic biology in a manner that may create unexpected effects. For example, mammalian target of rapamycin (mTOR) inhibition in a model of infantile spasms or in Pten mutant mice may improve performance in learning or social interaction, respectively, but tends to deteriorate them if given in naive controls.^{38,43} Regulatory testing for toxicities is typically done in naive animals. Early incorporation of such tolerability testing in the animal models of seizures, epilepsy or comorbidities may better profile the drug's effect.

Endpoints usually have to be adapted for the specific models, seizure types and syndromes, age, and sex groups, and discussion in the context of the specific seizure or epilepsy clinical syndrome would be an important element of the publications. For example, the Racine seizure scale for limbic seizures is occasionally used regardless of model and seizure type or age groups, instead of selecting the age and model appropriate seizure scales. Analysis of seizure scores can be variable, expressed as % of animals protected from more severe seizures (e.g., Racine scale 3 or 5) or compared parametrically, for example, "a drug reduces the mean seizure score from 5 to 4.5." Although this may denote some biologic effect, discussing whether and why this effect may be clinically meaningful would be useful for guiding the

clinical trialist. It is important to differentiate *statistical significance* from *biological importance*. Adopting comparable preclinical endpoints in CRFs and CDEs, an effort done by the AES/ILAE Translational Task Force in partnership with NINDS, could help resolve some of these issues and allow for better validation and translation of preclinical data.

Duration of monitoring for the pretreatment baseline and the treatment response

Clinical trials have been using various methods in assessing seizure freedom, a practice that has been criticized as complicating the comparison of efficacies across trials. The durations of observation periods for outcome assessment have been variable, with a minimum of 6 months. The monitoring period may include the entire stable dose period only or both titration and stable dose periods, whereas divergence exists on whether patients withdrawing are counted as responders or nonresponders or excluded.⁶⁵

In the latest ILAE definition of medically resistant seizures, it was proposed that "the rule of three" for calculating confidence intervals for zero events be followed to monitor seizure freedom or response to treatment in clinical studies.^{55,66} To document a clinically meaningful response, it was also proposed to test for seizure freedom for at least 1 year.⁵⁵ Therefore, a monitoring period at least equal to three times the interevent interval for seizures or 1 year, whichever is longer, has been recommended.⁶⁷

These parameters, of course, are not always easy to incorporate in animal studies. The lifespan of rodents is 2 years, making it unclear how the 1-year observation period in humans would translate to rodents. It would also not be realistic for studies on early life or age-specific seizures, since brain maturation to adult state proceeds over the first 2 months of rodent life. Most importantly, monitoring for seizures in rodents is classically done with use of video-EEG studies, rather than seizure diaries or reports by patients, which renders it cumbersome, expensive, and time and effort demanding to routinely do therapy screening trials with such long-term follow-ups.

In reality, most of the preclinical therapy trials are done over a short period, minutes to weeks for antiseizure effects and up to a few months for antiepileptogenic effects, often implementing a staggered approach of monitoring. When comparing the pre- and posttreatment periods, the duration of these sessions and inclusion/exclusion criteria should be

done based on prior model optimization and the known seizure incidence, frequencies, and interseizure or intercluster intervals. In vehicle/placebo-controlled studies, good randomization and similar handling of all groups would be important. A caveat is that in many models, seizures cluster over a day or two followed by long periods without seizures. Setting the monitoring duration to be long enough and appropriate for the intercluster intervals is particularly important when assessing seizure freedom or seizure remission.

The change of seizure frequency from the baseline can provide useful information about individual response to the drug, which can help differentiate responders from nonresponders or adversely affected subjects who experience worsening of their seizures. The Epilepsy Medicines Agency (EMA) guideline for clinical studies¹⁷ advises inclusion of study individuals for whom baseline frequency is not uncharacteristically different from their usual frequency and exclusion of subjects with slow seizure baseline counts so as to increase the possibility of seeing an effect. Such approaches may or may not be possible in all models of epilepsy, but they should be considered and discussed, depending on the model and scope of study. Unbiased randomization and evaluation of the effect trends on subcohorts with high or low seizure frequencies could be a solution.

The definition of seizure freedom or seizure remission in animal models is not officially defined as in humans. The heterogeneity of approaches used for defining seizures, monitoring for seizures, and diagnosing seizures, and the variability in methods to induce seizures in animals, the variable severity and seizure phenotype across models, and the time, age-, sex-, and strain- dependent differences in natural history of epilepsy, even in the same model, make it difficult to adopt a universal definition for seizure freedom. Therefore, it is important that animal models utilized in a lab set to perform preclinical epilepsy therapy trials are well characterized, at least in the strains, age, and sex groups that will be studied, to derive a detailed description of age of seizure onset, seizure/epilepsy incidence, seizure frequency per seizure type, clustering, circadian variations (important if monitoring is not done continuous over 24 h), seizure remission, and mortality. When seizure frequencies vary considerably in a trial, it may be safe, if feasible, to design the monitoring duration based on the least frequent seizure prior to the treatment. If this is not possible, the methods of outcome assessment should be selected carefully to account for such situations, that is, looking at incidence rates of epilepsy, separating the effects on cohorts with frequent (primary endpoints) versus rare (secondary endpoint) seizures.

Preclinical discoveries lost in translation: reasons and potential remedies

Failure to translate can be due to lack of rigor issues, personal bias, and conflicts of interests or to biologic or strategic issues. Rigor and transparency issues have been covered

extensively in other previous publications.^{5–11,50–53} Assuming that appropriate rigor has been applied in the studies, preclinical findings may not be confirmed in clinical trials either because they have no true clinical relevance or because clinical trials do not test the appropriate scenarios. False-positive preclinical data could be due to the unavoidable species differences, pathologies present in the experimental models (induced or genetic) but not in humans, or preclinical testing strategies that have poor relevance to clinical practice. For example, the old practice of testing drugs as pre-treatment (i.e., given prior to seizure induction) will have little relevance to clinical situations when treatment can be initiated only after the onset of seizures, for example, in a patient on SE due to causes that cannot be predicted.

In contrast, failure to translate a valid preclinical finding could be because clinical trials are missing important elements for treatment success. First, the target population in clinical trials may be heterogeneous and the “likely to benefit” individuals is only poorly represented. “Likely to benefit” individuals may be those with the relevant drug-targeted pathology (appropriate etiology or pathology, therapeutic window, or age or sex groups). Preclinical studies have an advantage over clinical studies because the treatment window can be tailored around specific models and stages of the seizure or epilepsy progression, and specific age and sex groups, unlike studies in humans which test treatments at various time points, wide age ranges and both genders. Second, coexistent comorbidities and their treatments in a subgroup of the treated population may interfere with the PK-PDs of the tested drug and therefore efficacy. Third, unanticipated toxicities in humans may preclude reaching the effective drug levels. Fourth, compliance may be variable in humans and seizure reporting in humans may not be as accurate if it depends on patient or family reports rather than the usual video-EEG studies done in animals.⁶⁸ Fifth, control groups in preclinical and clinical studies may be quite different. In animal studies, the use of vehicle-treated controls may yield higher chance of a demonstrated effect in monotherapy studies. In contrast, testing a drug as an add-on against controls that are either a standard comparator or standard of practice treated group narrows the probability of demonstrating a substantial therapeutic advantage, and studies may therefore be designed as noninferiority trials. Furthermore, the placebo effect may create an important therapeutic artifact in clinical trials that cannot be reproduced in experimental animals bred in our labs. Differences in formulations used in animals and in humans may also contribute to the differential efficacy.

Practices that can improve translation include the following: (1) identification of target mechanisms modified by the treatments that can be used to select target populations and appropriate treatment windows or monitor treatment response; (2) utilization of biomarkers or other tests to monitor target relevance and engagement so that one can

interpret positive or negative results;¹⁰ (3) early discussions among preclinical and clinical investigators so that preclinical and clinical trial study designs are concordant and comparable; (4) testing in clinically relevant formulations, doses, and strategies; and (5) incorporation of tolerability tests in the early preclinical efficacy studies to exclude treatments likely to manifest toxicity.

A proposal to implement multicenter preclinical phase II trials to test the efficacy and tolerability of drugs prior to advancing to full-scale clinical trials has been proposed as a way of improving the translatability of and providing more rigorous screening model for new epilepsy therapies.¹¹ This is currently being done in other fields, like stroke,⁶⁹ and efforts to create the infrastructure for this in epilepsy research is currently the goal of TASK4 or the AES/ILAE Translational Task Force.

PRECLINICAL DRUG TRIALS: PRAGMATIC VERSUS UTOPIC EXPECTATIONS

If the past can predict the future, there is a lot of optimism that the ongoing reevaluation and re-structuring of the preclinical epilepsy research will eventually deliver new treatments that will transform clinical practice. By nature, preclinical research is uniquely placed to identify target-based therapies and hopefully develop biomarkers to validate, implement, and monitor treatments in future clinical trials. If successful, this may promise to de-risk and individualize future treatments of epilepsies. Phenotypic screening using models and strategies that are more appropriate for specific epilepsy syndromes, medically resistant seizures, or epileptogenic processes and epilepsy-related comorbidities may also help identify compounds or new therapy targets that could address the existing unmet needs. It is not realistic to expect almost complete convergence in the results of preclinical and clinical studies. Yet, we should be striving for incremental progress through a healthy criticism, rather than skepticism, and reevaluation of strategies, infrastructure and needs, continuing support of investigators who are committed to deliver better therapies, and fostering collaborations and multi-disciplinary consortia of experts with complementary expertise who can help materialize these goals.

ACKNOWLEDGMENTS

ASG is funded by grants from the National Institutes of Health (NIH) NS091170, Department of Defense (W81XWH-13-1-0180), and the Infantile Spasms Initiative from CURE (Citizens United for Research in Epilepsy), and acknowledges also research funding from the Heffer Family and the Segal Family Foundations and the Abbe Goldstein/Joshua Lurie and Laurie Marsh/Dan Levitz families. ASG is co-chairing the AES/ILAE Translational Task Force of the ILAE and is co-Editor-in-Chief of *Epilepsia Open*. WBM is funded by grants from the Department of Defense (W81XWH-13-1-0180), the Infantile

Spasms Initiative from CURE, and Rett Syndrome Research Trust. We would like to thank her co-chairs, Drs Jacqueline French, Terence O'Brien, and Michele Simonato, as well as the members and volunteers of the Task Force and its working groups for valuable discussions on the topics addressed in this review. We also wish to thank Solomon (Nico) Moshé for critical review of this manuscript.

DISCLOSURE

We have no conflicts of interest in regards to this manuscript. 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

- Galanopoulou AS, Buckmaster PS, Staley KJ, et al. Identification of new epilepsy treatments: issues in preclinical methodology. *Epilepsia* 2012;53:571–582.
- Galanopoulou AS, Simonato M, French JA, et al. Joint AES/ILAE translational workshop to optimize preclinical epilepsy research. *Epilepsia* 2013;54(Suppl. 4):1–2.
- The NIH/NINDS Anticonvulsant Screening Program (ASP): recommendations from the working group's 2012 review of the Program. *Epilepsia* 2012;53:1837–1839.
- NINDS Anticonvulsant Screening Program Working Group. Anticonvulsant screening program report – May 29, 2015 [online]. 2015. Available at: http://www.ninds.nih.gov/research/asp/asp_working_group_report_052915.htm. Accessed May 8, 2016.
- French JA, White HS, Klitgaard H, et al. Development of new treatment approaches for epilepsy: unmet needs and opportunities. *Epilepsia* 2013;54(Suppl. 4):3–12.
- Loscher W, Schmidt D. Modern antiepileptic drug development has failed to deliver: ways out of the current dilemma. *Epilepsia* 2011;52:657–678.
- Wilcox KS, Dixon-Salazar T, Sills GJ, et al. Issues related to development of new antiseizure treatments. *Epilepsia* 2013;54(Suppl. 4):24–34.
- Brooks-Kayal AR, Bath KG, Berg AT, et al. Issues related to symptomatic and disease-modifying treatments affecting cognitive and neuropsychiatric comorbidities of epilepsy. *Epilepsia* 2013;54(Suppl. 4):44–60.
- Pitkanen A, Nehlig A, Brooks-Kayal AR, et al. Issues related to development of antiepileptogenic therapies. *Epilepsia* 2013;54(Suppl. 4):35–43.
- Engel J Jr, Pitkanen A, Loeb JA, et al. Epilepsy biomarkers. *Epilepsia* 2013;54(Suppl. 4):61–69.
- O'Brien TJ, Ben-Menachem E, Bertram EH III, et al. Proposal for a "phase II" multicenter trial model for preclinical new antiepilepsy therapy development. *Epilepsia* 2013;54(Suppl. 4):70–74.
- 2014 NINDS benchmarks for epilepsy research [online]. 2013. Available at: <http://www.ninds.nih.gov/research/epilepsyweb/2014Bechmar-ks-Final-PDF.pdf>. Accessed May 10, 2016.
- Hauser WA, Annegers JF, Kurland LT. Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935–1984. *Epilepsia* 1993;34:453–468.
- Kotsopoulos IA, van Merode T, Kessels FG, et al. Systematic review and meta-analysis of incidence studies of epilepsy and unprovoked seizures. *Epilepsia* 2002;43:1402–1409.
- Benn EK, Hauser WA, Shih T, et al. Estimating the incidence of first unprovoked seizure and newly diagnosed epilepsy in the low-income urban community of Northern Manhattan, New York City. *Epilepsia* 2008;49:1431–1439.
- Sillanpaa M, Schmidt D. Natural history of treated childhood-onset epilepsy: prospective, long-term population-based study. *Brain* 2006;129:617–624.
- Committee for Medicinal Products for Human Use. *Guideline on clinical investigation of medicinal products in the treatment of epileptic disorders* [online]. London, UK: European Medicines Agency; 2010. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070043.pdf. Accessed May 15, 2016.

18. Pellock JM, Carman WJ, Thyagarajan V, et al. Efficacy of antiepileptic drugs in adults predicts efficacy in children: a systematic review. *Neurology* 2012;79:1482–1489.
19. Dunne J, Rodriguez WJ, Murphy MD, et al. Extrapolation of adult data and other data in pediatric drug-development programs. *Pediatrics* 2011;128:e1242–e1249.
20. Galanopoulou AS, Bojko A, Lado F, et al. The spectrum of neuropsychiatric abnormalities associated with electrical status epilepticus in sleep. *Brain Dev* 2000;22:279–295.
21. Pellock JM, Hrachovy R, Shinnar S, et al. Infantile spasms: a U.S. consensus report. *Epilepsia* 2010;51:2175–2189.
22. Guerrini R, Belmonte A, Genton P. Antiepileptic drug-induced worsening of seizures in children. *Epilepsia* 1998;39(Suppl. 3):S2–S10.
23. Chiron C. Current therapeutic procedures in Dravet syndrome. *Dev Med Child Neurol* 2011;53(Suppl. 2):16–18.
24. Giorgi FS, Galanopoulou AS, Moshe SL. Sex dimorphism in seizure-controlling networks. *Neurobiol Dis* 2014;72(Pt B):144–152.
25. Akman O, Moshe SL, Galanopoulou AS. Sex-specific consequences of early life seizures. *Neurobiol Dis* 2014;72(Pt B):153–166.
26. Turski CA, Ikonomidou C. Neuropathological sequelae of developmental exposure to antiepileptic and anesthetic drugs. *Front Neurol* 2012;3:120.
27. Galanopoulou AS, Moshe SL. Pathogenesis and new candidate treatments for infantile spasms and early life epileptic encephalopathies: a view from preclinical studies. *Neurobiol Dis* 2015;79:135–149.
28. Ono T, Moshe SL, Galanopoulou AS. Carisbamate acutely suppresses spasms in a rat model of symptomatic infantile spasms. *Epilepsia* 2011;52:1678–1684.
29. Briggs SW, Mowrey W, Hall CB, et al. CPP-115, a vigabatrin analogue, decreases spasms in the multiple-hit rat model of infantile spasms. *Epilepsia* 2014;55:94–102.
30. Wong M. A critical review of mTOR inhibitors and epilepsy: from basic science to clinical trials. *Expert Rev Neurother* 2013;13:657–669.
31. Baraban SC, Dinday MT, Hortopan GA. Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. *Nat Commun* 2013;4:2410.
32. Tidball AM, Parent JM. Concise review: exciting cells: modeling genetic epilepsies with patient-derived induced pluripotent stem cells. *Stem Cells* 2016;34:27–33.
33. Cure C, Rasmussen T, Jasper H. Activation of seizures and electroencephalographic disturbances in epileptic and in control subjects with metrazol. *Arch Neurol Psychiatry* 1948;59:691–717.
34. Hasson H, Kim M, Moshe SL. Effective treatments of prolonged status epilepticus in developing rats. *Epilepsy Behav* 2008;13:62–69.
35. Goodkin HP, Kapur J. The impact of diazepam's discovery on the treatment and understanding of status epilepticus. *Epilepsia* 2009;50:2011–2018.
36. Morrisett RA, Jope RS, Snead OC III. Effects of drugs on the initiation and maintenance of status epilepticus induced by administration of pilocarpine to lithium-pretreated rats. *Exp Neurol* 1987;97:193–200.
37. Blanco MM, dos Santos JG Jr, Perez-Mendes P, et al. Assessment of seizure susceptibility in pilocarpine epileptic and nonepileptic Wistar rats and of seizure reinduction with pentylenetetrazole and electroshock models. *Epilepsia* 2009;50:824–831.
38. Raffo E, Coppola A, Ono T, et al. A pulse rapamycin therapy for infantile spasms and associated cognitive decline. *Neurobiol Dis* 2011;43:322–329.
39. Pontes JC, Lima TZ, Queiroz CM, et al. Seizures triggered by pentylenetetrazol in marmosets made chronically epileptic with pilocarpine show greater refractoriness to treatment. *Epilepsy Res* 2016;126:16–25.
40. Loscher W. The search for new screening models of pharmacoresistant epilepsy: is induction of acute seizures in epileptic rodents a suitable approach? *Neurochem Res* 2016; in press.
41. Khalilov I, Holmes GL, Ben-Ari Y. In vitro formation of a secondary epileptogenic mirror focus by interhippocampal propagation of seizures. *Nat Neurosci* 2003;6:1079–1085.
42. Cohen I, Navarro V, Clemenceau S, et al. On the origin of interictal activity in human temporal lobe epilepsy in vitro. *Science* 2002;298:1418–1421.
43. Zhou J, Blundell J, Ogawa S, et al. Pharmacological inhibition of mTORC1 suppresses anatomical, cellular, and behavioral abnormalities in neural-specific Pten knock-out mice. *J Neurosci* 2009;29:1773–1783.
44. U.S. Department of Health and Human Services Food and Drug Administration Center for Devices and Radiological Health & Biometrics DoBOoSa. Guidance for the use of bayesian statistics in medical device clinical trials [online]. 2010. Available at: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf>. Accessed May 16, 2016.
45. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER) & (CBER) CfBEaR. Guidance for industry. E9 statistical principles for clinical trials [online]. 1998. Available at: <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf>. Accessed May 16, 2016.
46. Tyson JE, Pedroza C, Wallace D, et al. Stopping guidelines for an effectiveness trial: what should the protocol specify? *Trials* 2016;17:240.
47. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365–376.
48. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
49. Button KS, Ioannidis JP, Mokrysz C, et al. Confidence and precision increase with high statistical power. *Nat Rev Neurosci* 2013;14:585–586.
50. Galanopoulou AS, Kokaia M, Loeb JA, et al. Epilepsy therapy development: technical and methodologic issues in studies with animal models. *Epilepsia* 2013;54(Suppl. 4):13–23.
51. Pitkanen A, Schwartzkroin PA, Moshé SL. *Models of seizures and epilepsy*. San Diego, CA: Elsevier; 2006.
52. Loscher W, Klitgaard H, Twyman RE, et al. New avenues for anti-epileptic drug discovery and development. *Nat Rev Drug Discov* 2013;12:757–776.
53. Barker-Haliski ML, Friedman D, French JA, et al. Disease modification in epilepsy: from animal models to clinical applications. *Drugs* 2015;75:749–767.
54. CURE. Infantile spasms research initiative [online]. Available at: <http://www.cureepilepsy.org/research/is.asp>. Accessed May 15, 2015.
55. Kwan P, Arzimanoglou A, Berg AT, et al. Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia* 2010;51:1069–1077.
56. Sillanpaa M, Schmidt D. Is incident drug-resistance of childhood-onset epilepsy reversible? A long-term follow-up study. *Brain* 2012;135:2256–2262.
57. Scantlebury MH, Galanopoulou AS, Chudomelova L, et al. A model of symptomatic infantile spasms syndrome. *Neurobiol Dis* 2010;37:604–612.
58. Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia* 2010;51:676–685.
59. Berendt M, Farquhar RG, Mandigers PJ, et al. International veterinary epilepsy task force consensus report on epilepsy definition, classification and terminology in companion animals. *BMC Vet Res* 2015;11:182.
60. Barker-Haliski M, Friedman D, White HS, et al. How clinical development can, and should, inform translational science. *Neuron* 2014;84:582–593.
61. NINDS common data elements [online]. Available at: https://common-dataelements.ninds.nih.gov/Epilepsy.aspx-tab=Data_Standards. Accessed May 14, 2016.
62. Francois J, Germe K, Ferrandon A, et al. Carisbamate has powerful disease-modifying effects in the lithium-pilocarpine model of temporal lobe epilepsy. *Neuropharmacology* 2011;61:313–328.
63. Dudek FE, Staley KJ. The time course and circuit mechanisms of acquired epileptogenesis. In Noebels JL, Avoli M, Rogawski MA, et al. (Eds) *Jasper's basic mechanisms of the epilepsies*. 4th Ed. Bethesda, MD: National Center for Biotechnology Information (US). Available from: <https://www.ncbi.nlm.nih.gov/books/NBK98152/>
64. Kharatishvili I, Nissinen JP, McIntosh TK, et al. A model of posttraumatic epilepsy induced by lateral fluid-percussion brain injury in rats. *Neuroscience* 2006;140:685–697.
65. Leppik I, De Rue K, Edrich P, et al. Measurement of seizure freedom in adjunctive therapy studies in refractory partial epilepsy: the levetiracetam experience. *Epileptic Disord* 2006;8:118–130.

66. Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA* 1983;249:1743–1745.
67. French JA. Is the epilepsy responsive or resistant? Only time will tell. *Ann Neurol* 2009;65:489–490.
68. Cook MJ, O'Brien TJ, Berkovic SF, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *Lancet Neurol* 2013;12:563–571.
69. Multi-PART. About Multi-PART. Multicentre Preclinical Animal Research Team is an international collaborative approach to overcoming the translational roadblock in neuroprotection and neuroregeneration research [online]. Available at:<http://www.dcn.ed.ac.uk/multipart/about.html>. Accessed May 15, 2016.
70. Public access to neuroactive chemical evaluations (PANACHE) [online]. Available at:<https://panache.ninds.nih.gov/>. Accessed May 15, 2016.
71. Bialer M, Johannessen SI, Levy RH, et al. Progress report on new antiepileptic drugs: a summary of the Twelfth Eilat Conference (EILAT XII). *Epilepsy Res* 2015;111:85–141.
72. Dzhala VI, Talos DM, Sdrulla DA, et al. NKCC1 transporter facilitates seizures in the developing brain. *Nat Med* 2005;11:1205–1213.
73. Glien M, Brandt C, Potschka H, et al. Effects of the novel antiepileptic drug levetiracetam on spontaneous recurrent seizures in the rat pilocarpine model of temporal lobe epilepsy. *Epilepsia* 2002;43:350–357.
74. French JA. Proof of efficacy trials: endpoints. *Epilepsy Res* 2001;45:53–56; discussion 57–9.
75. Prabhakar H, Bindra A, Singh GP, et al. Propofol versus thiopental sodium for the treatment of refractory status epilepticus. *Cochrane Database Syst Rev* 2012;8:CD009202.
76. Brigo F, Nardone R, Tezzon F, et al. A common reference-based indirect comparison meta-analysis of buccal versus intranasal midazolam for early status epilepticus. *CNS Drugs* 2015;29:741–757.