Three Ways to Mitigate the Risk of Late-stage Failure in CNS Drug Development

PEER REVIEWED

Central nervous system (CNS) disorders have long been the Bermuda Triangle of drug development. For example, a 2014 study of Alzheimer's disease (AD) drugs tested between 2002 and 2012 showed that 99.6% of them failed one of Phase I, II or III trials (Cummings, Morstorf and Zhong, 2014). Many of these AD drugs failed late, in large Phase III trials, making the failures costlier and more dispiriting for researchers, drug developers, and, certainly, patients. The most recent casualty was a novel tau aggregation inhibitor (tau is a bodily protein that many neuroscientists believe contributes to the brain-destroying effects of AD) that failed to improve cognitive or physical function in the Phase III trial of 891 AD patients (Garde, 2016).

The prevalence of CNS disorders worldwide is staggering. An estimated 350 million people suffer from depression (World Health Organization, 2016a) and over 46 million people live with dementia (Prince, 2015) – a number expected to grow to 131.5 million by 2050 due to aging populations. Approximately 50 million people have epilepsy (World Health Organization, 2016b), and more than 21 million patients have schizophrenia (World Health Organization, 2016c), a disease that places them at high risk for medical comorbidities, higher rates of suicide, and self-harm.

Severe mental disorders cost the world an estimated $2.5 trillion annually, a number expected to increase to $6 trillion by 2030, while the global societal cost of dementia has been pegged at $604 billion per year (Bloom *et al*., 2011).

Despite the urgent need for new CNS therapies, many drug developers have reduced their investment or abandoned the field altogether (RAND Corporation, 2013) due to development times that are 18% longer than those for non-CNS compounds (Tufts Center for the Study of Drug Development, 2014), and the aforementioned low success rates.

High-risk CNS Decisions are Challenged by Incomplete Scientific Knowledge

Late-stage failures in drug development are sadly not uncommon and they involve tens of thousands of participating patients in multiple therapeutic areas, including CNS (Grignolo and Pretorius, 2016). But CNS drug development is particularly treacherous. First of all, the pathophysiology of most CNS disorders is poorly understood. CNS diseases are diverse and range from AD to Parkinson's disease, multiple sclerosis, stroke, epilepsy, depression, addiction, anxiety, pain management, and schizophrenia, to name a few. These conditions may have various etiologies including autosomal dominant conditions, genes associated with increased risk for disease, neurodegenerative conditions, environmentally acquired conditions and many whose etiology is unknown. Despite recognition of heterogeneity within specific diseases, most CNS diseases are diagnosed clinically, particularly for psychiatric conditions.

As Jill Heemskerk, Director of the Office of Research Administration and Acting Deputy Director at the National Institute of Biomedical Imaging and Bioengineering, observed recently, "On average, a marketed psychiatric drug is efficacious in approximately half of the patients who take it. One reason for this low response rate is the artificial grouping of heterogeneous syndromes with different pathophysiological mechanisms into one disorder" (Cedarbaum et al., 2014). At present, however, most CNS diseases lack specific diagnostic biomarkers, as well as other biomarkers (i.e., disease progression, predictors of drug response) that can facilitate separating diseases into more specific syndromes.

Most CNS drug studies test one drug (monotherapy) versus placebo, although some trials use add-on regimens in which an investigational drug is combined with an existing therapy. In CNS, it is quite unusual to combine complementary investigational drugs, even if there is thought to be significant potential for greater efficacy when combined (e.g., a beta secretase inhibitor plus an antibody therapy). The regulatory complexities and hurdles for combination therapies are, understandably, much higher than those for a monotherapy.

Regulatory agencies are receptive to two-drug combinations only when: 1) the treatment is for a serious illness; 2) there is a strong biological rationale for the combination; 3)comprehensive preclinical characterisations have been completed for each drug individually and for the combination product; 4) preclinical and/or clinical data demonstrate a more durable response or a better toxicity profile for the combination product versus each agent; and 5) there is a strong rationale for why the drugs cannot be developed independently. As the FDA notes in its guidance documents, "Codevelopment generally will provide less information about the clinical safety and effectiveness and dose-response of the individual new investigational drugs intended to be used in combination than would be obtained if the individual drugs were developed alone" (U.S. Food and Drug Administration, 2013)

The placebo effect is another significant issue in CNS drug development. Although not unique to CNS, the placebo effect particularly plagues the field due to the lack of objective methods to assess outcomes and the need to rely on patient reports or other subjective clinical assessments that can be susceptible to bias. The placebo effect has been discussed extensively

in the CNS literature and is a notorious issue in studies of depression and schizophrenia. The effect size of the placebo response, which can range from 30% to 60% in depression, has resulted in placebos appearing more effective than investigational agents in some studies of these diseases.

Animal models have not been very successful in predicting efficacy of CNS drugs in man. Metabolic, anatomic, and cellular differences between mice and humans have been blamed for many a failure (Gawrylewski, 2007). Moreover, the traditional heavy reliance on healthy volunteer dose escalation data in early development has not worked well in CNS disorders, generating countless results that could not be reproduced in target patient populations.

An example of misleading results from animal testing and healthy volunteer (HV) studies is Exelon (rivastigmine), an acetylcholinesterase inhibitor which was approved in 2000 for treating mild-to-moderate Alzheimer's dementia, and subsequently also approved for Parkinson's dementia. The drug originally failed in Phase II/III trials because the dose identified as effective in animal models and HV trials was inadequate. Exploratory biomarker readouts from CSF testing that evaluated *ex vivo* acetylcholinesterase/butyryl-cholinesterase inhibition in AD patients who received higher Exelon doses resulted in a plausible 'go ahead' signal for additional testing. Repeat confirmatory studies demonstrated efficacy in AD and the drug was ultimately approved.

Testing a drug's penetration across the blood-brain barrier has not until very recently been routinely performed, even though an understanding of CSF pharmaco-kinetics/pharmacodynamics is essential in evaluating a drug for potential effective doses. In the past, this step was often skipped due to the planning and expense involved, with additional testing often needed in animal models, new imaging ligands, and development of new biomarkers to evaluate target binding and engagement. Resolving chemistry, manufacturing, and control (CMC) issues unique to CNS penetration, and to binding/engaging biologic targets, can increase budgets and elongate development timelines. However, proceeding without this knowledge is more likely to result in inadequate dosing and a diminished understanding of pharmacodynamic effects — which in turn can lead in the long run to more money and time invested or another CNS drug failure.

Most drug developers acknowledge the importance of this early-phase work. They place an increasing emphasis on demonstrating CNS penetration and pharmacological effects of relevance in the target population, typically using a biomarker. Once a drug passes this stage, strategies to de-risk the later-phase clinical development plan become paramount.

A number of clinical trial practices have been adopted to help de-risk late-stage development and to ensure that decision-making around assets is data-driven. These practices include:

1. **Rigorous proof of concept (POC) standards**
2. **More efficient, adaptive study designs**
3. **Close oversight of patient identification and enrolment. Plus continuous monitoring**

Rigorous Proof of Concept (POC) Standards

A high-quality preclinical data package that establishes a solid foundation for a drug's mechanism of action provides the starting point for POC testing in man. In Phase I studies, sensitive measures to evaluate an early signal of efficacy are often included. However, given the limited sample size of Phase I studies, there is typically inadequate power for statistical assessment of these results. Relying too heavily on these data can be misleading for POC studies. Biomarkers may be incorporated into studies for disease diagnosis, progression, and drug response. Preplanning for how the data collected will be used in go/no-go decision-making is essential for ensuring objective decision-making.

Researchers at Pfizer have argued (Morgan *et al*., 2012) that true POC is not achieved — and drug candidate "survival" is not likely — until solid Phase II data demonstrates:

1. Exposure at the target site of action over a desired period (pharmacokinetic/pharmacodynamic principles)
2. Binding to the pharmacological target as expected for the drug's mode of action
3. Expression of pharmacological activity proportional to target exposure and binding

This "three pillars" approach to POC focuses on dose-response evaluation and uses imaging and other biomarkers to measure pharmacodynamic activity. For example, the Pfizer study looked at data from Phase II decisions for 44 internal programmes across the company (covering multiple therapeutic areas) between 2005–2009 and found that of the 14 drugs that met all three pillars, 12 achieved positive POC, and eight advanced to Phase III trials. When none of the three pillars was met, no compounds achieved POC (Morgan *et al*., 2012).

Properly applied, this POC approach can facilitate better compound advancement decisions, allowing companies to terminate failing compounds earlier in development, prior to expensive later-phase studies. Although Pfizer's POC model has concentrated on Phase II data collection and advancement decisions, the process begins in Phase I.

The basis for a go/no-go decision in Phase I is validation (or lack thereof) of a drug's mechanism of action. Does it locate and engage the intended target? Does engagement of that target induce potentially important pharmacological activity?

- Positron emission tomography (PET) and cerebrospinal fluid (CSF) analysis can provide evidence that a drug has crossed the blood-brain barrier and engaged targeted receptor(s). New pharmacodynamic methodologies to quantify a drug's

pharmacological effects include: Functional magnetic resonance imaging (fMRI) – Dynamically evaluates brain regions/networks and their activation. For example, neural activity can be indirectly measured by regional changes in the blood-oxygen level dependent [BOLD] signal in response to a specific task performed by a subject.

- Resting state fMRI (R-fMRI) – A relatively new method for evaluating regional interactions that occur when a subject is not performing an explicit task. R-fMRI potentially reveals the effect of the drug on the connectome, the as-yet unfinished map of neural pathways that underlie human brain function (Neuroscienceblueprint.nih.gov,2016)

- Quantitative electroencephalography (qEEG) and event-related potentials (ERP) – These non-invasive techniques that analyse brain electrical activity can be used to identify biomarkers, document CNS penetration (pre- and post-drug administration), measure adverse events (particularly seizure activity, but also sedation), and evaluate patterns of connectivity.

If a Phase I compound effectively reaches and engages its target, and that engagement correlates with a pharmacological signal, it should advance to Phase II development, assuming safety and tolerability criteria are also met. If Phase I data suggest the target is not engaged, the compound may need to be shelved or further studies may be needed. Too often, compounds advance despite unconvincing evidence from Phase I. Raising the bar for POC in both Phase I and Phase II by utilising the right scientific tools to assess the viability of a drug candidate helps reduces that risk.

More Efficient, Adaptive Study Designs

Many Phase II POC clinical trials provide an efficacy signal when in fact a drug has no effect, i.e., a Type I error. Reducing Type I errors in Phase II would result in fewer compounds failing in Phase III. Thus improved study designs in Phase II help provide a faster, and potentially more accurate, read on a drug candidate. Examples of smarter study designs include:

Seamless Phase I study designs

Traditionally, studies in HVs are used to determine the maximum tolerated dose (MTD) for an experimental drug before it can be administered to patients with active disease. However, in CNS, the effects of a drug on a normally functioning brain can be quite different than on a brain affected by the disease of interest. New study designs combine dose escalation cohorts of healthy volunteers with cohorts of appropriate patients to de-risk drug candidates and support more confident 'go/no go' decisions.

Adaptive designs

In early-phase trials, measuring biomarkers and the magnitude of relevant drug effects can inform dose escalation and allow re-estimation of sample sizes. For example, in AD, the levels of two key proteins (Tau and amyloid beta) in the blood can reveal whether a beta-secretase (BACE) inhibitor is producing sufficient inhibition. Adaptive Phase I and II trials allow dose ranging to be tested more fully and efficiently. Pharmacokinetic and pharmacodynamic test results may show that a planned dose overlaps a prior dose cohort and does not need to be tested, or it may be found that prior modelling was not predictive, and an intermediate dose needs to be evaluated. Planning for flexibility in dose cohorts in both single ascending dose (SAD) and multiple ascending dose (MAD) studies maximises a company's early-phase investment by enabling faster, more informed go/no-go decisions.

A recent analysis demonstrated how an adaptive trial led to early termination of an experimental schizophrenia drug (Shen et al., 2011). The early-stage dose-finding study was meant to identify the optimal dose from among seven doses for use in later trials. Instead, using dose-response modelling, the study found that further testing of the drug would be futile. What was the payoff of the adaptive design? The trial enrolled only 202 patients before reaching a pre-determined threshold for lack of activity, versus the estimated 450 patients that would have been required under a conventional design.

Close Oversight of Patient Identification and Enrolment, Plus Continuous Monitoring

In CNS, reducing the heterogeneity of the patient population is critical as heterogeneity can skew behavioural, biological, and neurochemical response to controls, especially in psychiatric disorders. Enrolling precise subpopulations through the use of stringent inclusion/exclusion criteria and the use of biomarkers should help reduce variability in response to a drug. Conducting placebo run-in phases in which patients who respond to placebo are withdrawn from a study before randomisation begins may help combat the placebo effect. These key strategies come at a cost, however.

For example, slower patient enrolment and randomisation can extend development timelines and increase expenses, potentially endangering a product's success due to later market entry or ROI-driven factors resulting in less competitive pricing.

Various techniques can remotely assess whether or not subjects meet the entry criteria for enrolment into a clinical trial. These remote rater assessments are employed most often in psychiatric studies, especially those in depression and schizophrenia.

Recently, a PAREXEL client used digital audio recordings at screening to increase the appropriateness and homogeneity of a study population. The result was high confidence in the results of this POC study, which evaluated a novel mechanism drug to treat acute schizophrenia. Patients enrolled in the trial consented to the use of digital audio recordings as part of the screening procedure. Screening assessments included recordings of the Brief Psychiatric Rating Scale (BPRS) and the Mini International Neuropsychiatric Interview (MINI), as well as several other tests.

Electronic images of assessment documentation, annotations, and

audio of patient interviews were also collected for remote review. Interviewers/screeners dictated additional clinical information after the interview.

This data was reviewed remotely by a qualified rater to ensure that only appropriate subjects were enrolled. The process helped deliver a more homogeneous sample of subjects and resulted in higher study retention rates than were projected from historical data. The digital audio recordings did not impede study enrolment, which was more rapid than projected. The study demonstrated efficacy and assay sensitivity, supporting a confident test of the drug's mechanism.

Similar data capture of audio recordings or video recordings is important for the assessment of primary and some secondary endpoints in psychiatric studies. These centralised ratings help ensure blinding of the rater, reduce bias and placebo effect and minimise variability in rating scales. The FDA has stated they have a clear preference for the incorporation of centralised ratings for outcome measures in clinical trials for psychiatry.

Recently there has been a fundamental shift in the conceptualisation of neuropsychiatric disorders. Today, psychiatric disorders are seen as the result of dysfunction in several domains of brain function, such as cognition, motivation/reward, and fear/anxiety. Each domain and sub-domain presents symptoms or measurable findings that are linked to brain circuitry networks. A dysfunction in any given domain is reflected in circuitry changes that can be demonstrated in preclinical and clinical studies. This sets the stage for science-based symptomatic (phenotypic) development versus the use of a diagnostic system that rests on traditional, more subjective, and less reliable measures.

The National Institute of Mental Health's Research Domain Criteria (RDoC) initiative, dedicated to creating "a new kind of taxonomy for mental disorders," is at the vanguard of potentially transformative neurocircuitry-based approaches in CNS drug development (National Institute of Mental Health, 2016). RDoC assumes that dysfunction in neural circuits can be measured by electrophysiology (qEEG, ERP), functional neuroimaging (MRI, FDG-PET), neurocognitive/behavioural assessments, and other methods for quantifying connectivity. It connects symptoms and measurable findings in patients with their brain circuitry networks, which could guide the classification of patients for specific research studies.

It has the potential to improve go/no-go decisions by adding data about a drug's effects on a patient's neurocircuitry which goes beyond just measuring its impact on a behavioural clinical endpoint, such as a rating scale.

What could RDoC mean for CNS clinical trials now?

- Early trials could be sized to generate data on objective, more targeted neurocircuitry and behavioural/cognitive measures
- As new mechanisms to assess circuitry/connectivity emerge, accuracy may improve
- Patients can be screened/enrolled on the basis of deficits in a mechanism, rather than on a classic psychiatric Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnosis
- Later-phase trials may be more cost-effective as sites using neurocircuitry techniques may serve as hubs, with other sites referring patients to these sites
- Strong evidence of biomarker/neurocircuitry changes can be gathered throughout development to support label claims to regulators and post-approval gatekeepers (i.e. health technology assessment agencies)

The RDoC approach, however, is in its infancy regarding data collection and validation of this circuitry. It will take time and much work to establish what neurocircuitry changes correlate with desired clinical benefits for this approach to become a standard part of drug development.

In addition to remote rating and RDoC approaches, there is a wide range of tools and strategies that can aid developers in identifying specific enriched patient subpopulations, and enrolling them earlier in the development process in order to enhance signal detection. These include:

- Conducting feasibility studies to obtain clear data on variations in patient availability, medical coverage and the current standard of care across different regions, countries, sites, investigators, and concurrent, competing trials.

- Assessing comorbid illness patterns as well as typical drug prescription profiles in the target population can help to inform impractical inclusion/exclusion criteria. Utilising rating scale training programmes and site surveillance procedures (such as remote monitoring, independent external raters, and the like)

- Promoting study drug compliance by both patients and investigators through targeted training, written or interactive communications and even social media (using new technologies to monitor pill-taking and provide reminders)

- Reducing the incidence of drop-outs through specific retention strategies

Reducing Risk to Fight a Worthy Battle

CNS disorders inflict devastatingly high human and societal costs, but also pose huge risks to drug developers. High profile late-stage trial failures crush the hopes of tens of thousands, even millions, of patients battling disabling diseases, and inflict material financial losses on pharmaceutical companies. The result is a negative feedback loop that diminishes enthusiasm for new drug development in CNS.

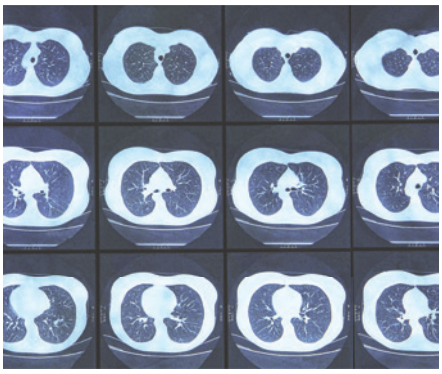
While the techniques and methods described in this article add time and expense to the conduct of trials, they are likely to save developers money in the long run by reducing the risk of late-stage drug failures. If a company has insufficient expertise in-house to pursue all of these measures, it is well worth tapping external resources.

CNS developers can ensure that their POC studies in CNS are rigorous

and adequately powered by employing creative, flexible designs. Doing so allows them to wring better predictive and more definitive data from clinical trials while using state-of-the-art methods for precisely targeting patient populations. This may reduce their risk in a worthy battle that has already seen far too many drug development casualties.

REFERENCES

1. Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z. and Weinstein, C. (2011). The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum. Available at: http://apps.who.int/medicinedocs/documents/s18806en/s18806en.pdf [Accessed 2 Sep. 2016].
2. Cedarbaum, J., Stephenson, D., Rudick, R., Carrillo, M., Stebbins, G., Kerr, D., Heemskerk, J., Galpern, W., Kaufmann, P., Cella, D., Isaac, M. and Walton, M. (2014). Commonalities and Challenges in the Development of Clinical Trial Measures in Neurology. Neurotherapeutics, 12(1), pp.151-169.
3. Cummings, J., Morstorf, T. and Zhong, K. (2014). Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimers Res Ther, 6(4), p.37.
4. Garde, D. (2016). Promising Alzheimer's treatment flops in new trial. [online] STAT. Available at: https://www.statnews.com/2016/07/27/alzheimers-drug-taurx/ [Accessed 2 Sep. 2016].
5. Gawrylewski, A. (2007). The Trouble With Animal Models: Why did human trials fail? The Scientist. [online] Available at: http://www.the-scientist.com/?articles.view/articleNo/25184/title/The-Trouble-with-Animal-Models/ [Accessed 3 Oct. 2016].
6. Grignolo, A. and Pretorius, S. (2016). Phase III Trial Failures: Costly, But Preventable. Applied Clinical Trials, [online] 25(8). Available at: http://www.appliedclinicaltrialsonline.com/phase-iii-trial-failures-costly-preventable [Accessed 3 Oct. 2016].
7. Morgan, P., Van Der Graaf, P., Arrowsmith, J., Feltner, D., Drummond, K., Wegner, C. and Street, S. (2012). Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. Drug Discovery Today, 17(9-10), pp.419-424.
8. National Institute of Mental Health. (2016). NIMH » Research Domain Criteria (RDoC). [online] Available at: https://www.nimh.nih.gov/research-priorities/rdoc/index.shtml [Accessed 3 Oct. 2016].
9. Neuroscienceblueprint.nih.gov. (2016). The Human Connectome Project. [online] Available at: https://neuroscienceblueprint.nih.gov/connectome/ [Accessed 3 Oct. 2016].
10. Prince, M. (2015). World Alzheimer Report 2015: The Global Impact of Dementia. [online] Alzheimer's Disease International. Available at: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf [Accessed 2 Sep. 2016].
11. AND Corporation. (2013). The New Neglected Diseases? Policy Interventions Are Needed to Encourage CNS Drug Development. Perspective. [online] RAND Corporation. Available at: http://www.rand.org/pubs/perspectives/PE117.html [Accessed 3 Oct. 2016].
12. Shen, J., Preskorn, S., Dragalin, V., Slomkowski, M., Padmanabhan, S., Fardipour, P., Sharma, A. and Krams, M. (2011). How Adaptive Trial Designs can Increase Efficiency in Psychiatric Drug Development: A Case Study. Innovations in Clinical Neuroscience, [online] 8(7), pp.26-34. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159542/ [Accessed 3 Oct. 2016].
13. Tufts Center for the Study of Drug Development. (2014). CNS Drugs Take Longer to Develop and Have Lower Success Rates than Other Drugs. [online] Available at: http://csdd.tufts.edu/news/complete_story/pr_ir_nov_dec_ir [Accessed 3 Oct. 2016].
14. U.S. Food and Drug Administration. (2013). Guidance for Industry: Codevelopment of Two or More New Investigational Drugs for Use in Combination. Silver Spring, MD: U.S. Food and Drug Administration. Available at: http://www.fda.gov/downloads/drugs/
15. World Health Organization. (2016a). Depression. [online] Available at: http://www.who.int/mediacentre/factsheets/fs369/en/ [Accessed 2 Sep. 2016].
16. World Health Organization. (2016b). Epilepsy. [online] Available at: http://www.who.int/mediacentre/factsheets/fs999/en/ [Accessed 2 Sep. 2016].
17. World Health Organization. (2016c). Schizophrenia. [online] Available at: http://www.who.int/mediacentre/factsheets/fs397/en/ [Accessed 2 Sep. 2016].



Rebecca M. Evans
MD, MS, Global Therapeutic Area Head for CNS.

Larry Ereshefsky
Pharm.D., BCPP, VP and Global CNS Leader, Early Phase (at the time work was performed).

Alberto Grignolo
Ph.D., Corporate Vice President, Global Strategy.

Sy Pretorius
MD, MMedSc, MBA, MS, FFPM, Senior Vice President & Chief Scientific Officer at PAREXEL International.