Embracing failure: What the Phase III progesterone studies can teach about TBI clinical trials

Donald Stein, *Emory University*

Journal Title: Brain Injury
Volume: Volume 29, Number 11
Publisher: Taylor & Francis: STM, Behavioural Science and Public Health Titles | 2015-01-01, Pages 1259-1272
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.3109/02699052.2015.1065344
Permanent URL: https://pid.emory.edu/ark:/25593/rgj8g

Final published version: http://dx.doi.org/10.3109/02699052.2015.1065344

BRAIN INJURY

http://tandfonline.com/ibij
ISSN: 0269-9052 (print), 1362-301X (electronic)

Brain Inj, 2015; 29(11): 1259–1272
Published with license by Taylor and Francis. DOI: 10.3109/02699052.2015.1065344

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

Embracing failure: What the Phase III progesterone studies can teach about TBI clinical trials

Donald G. Stein

Department of Emergency Medicine, Emory University, Atlanta, GA, USA

Abstract

Background: Despite positive preclinical studies and two positive Phase II clinical trials, two large Phase III clinical trials of progesterone treatment of acute traumatic brain injury (TBI) recently ended with negative results, so a 100% failure rate continues to plague the field of TBI trials.

Methods: This paper reviews and analyses the trial structures and outcomes and discusses the implications of these failures for future drug and clinical trial development. Persistently negative trial outcomes have led to disinvestment in new drug research by companies and policy-makers and disappointment for patients and their families, failures which represent a major public health concern. The problem is not limited to TBI. Failure rates are high for trials in stroke, sepsis, cardiology, cancer and orthopaedics, among others.

Results: This paper discusses some of the reasons why the Phase III trials have failed. These reasons may include faulty extrapolation from pre-clinical data in designing clinical trials and the use of subjective outcome measures that accurately reflect neither the nature of the deficits nor long-term quantitative recovery.

Conclusions: Better definitions of injury and healing and better outcome measures are essential to change the embrace of failure that has dominated the field for over 30 years. This review offers suggestions to improve the situation.

Keywords

Clinical trials, progesterone, ProTECT III, SyNAPSe, traumatic brain injury

History

Received 23 March 2015
Revised 19 June 2015
Accepted 19 June 2015
Published online 14 August 2015

The problem

Traumatic brain injury (TBI) is a serious public health problem across the globe, yet after decades of clinical trials, there has been a 100% failure to identify a drug that works in the acute stage of the injury when neuroprotection is most critical. Comprehensive reviews and suggestions from expert consensus panels addressing the problem began to appear over a dozen years ago (see Narayan et al. [1] for an excellent earlier discussion of the issues), but little has changed since then in the conduct and outcomes of ensuing clinical trials. Society is still facing the same questions, despite very specific recommendations that could have been followed in subsequent trials. While the focus of this review is on the problem with recent trials for acute TBI, the high Phase III negative outcome rate problem is endemic, affecting sepsis, stroke, cancer, cardiology and orthopaedics research, to name just a few. These persistent failures have had a chilling effect on pharmaceutical industry investments in new drug development and the costs to pharma and government are staggering. It is no wonder that funding agencies and policy-makers in both sectors are deeply concerned and realize that continuing to do the same things in the same way cannot go on. Recent reviews of the literature support this contention and the titles are revealing:

- Why do phase III clinical trials in oncology fail so often? [2].
- Uncertainty in the translation of pre-clinical experiments to clinical trials. Why do most phase III clinical trials fail? [3].
- Why do Phase III trials of promising heart failure drugs often fail? The contribution of regression to the truth [4].
- Why are there no good treatments for diabetic neuropathy? [5].
- Understanding history and not repeating it. Neuroprotection for acute ischaemic stroke: From review to preview [6].
- Animal models of sepsis: why does pre-clinical efficacy fail to translate to the clinical setting? [7].

Despite much recent discussion, little has been done to change the situation for the better. This paralysis could be due in part to the lack of any broad consensus about where the most basic problems lie. One faction argues that the fault lies with pre-clinical animal studies that fail to translate to the

Correspondence: Donald G. Stein, 1365 B Clifton Rd NE, Suite 5100, Atlanta, GA, 30322, USA. E-mail: dstei04@emory.edu

human condition. Among the many papers on this issue, a recent article by Jickling and Sharp [8] argues that this is the problem in the field of stroke, while Begley and Ellis [9] make the same claim for pre-clinical cancer research. This is a substantive issue that deserves attention—in fact, animal studies have often *not* translated easily to clinical trial methodology or outcomes, but this is the arena in which most, if not all, mechanistic and exploratory drug studies are currently done.

Other factions cite issues with the design, execution and analysis of clinical trials themselves. Does the problem lie with the laboratory scientists or with the clinical trials? Yes! In a recent comprehensive review, Goodman and Gerson [10] evaluate the strength of evidence of pre-clinical drug development data used to support clinical trials and it is clear from their annotated bibliography (well worth a read) that there are substantive problems. The ratio of screened compounds to marketed drugs is ~1:10 000. Some drugs show strong biological and functional signals in animal studies, but fail to work in the clinic. However, using Bayesian analyses, the authors show that, although pre-clinical evidence lacks precision, it still yields quantitatively more evidence than the clinical phase of testing can do on its own. In fact, Goodman and Gerson estimate that basic research leading up to a clinical trial increases the odds of success 12-fold over the clinical research process by itself. In addition, the US Food and Drug Administration (FDA) requires animal tests before patients can be exposed to any new molecular entities for the simple reason of ensuring safety. So, while the problems facing pre-clinical investigations are serious, they're far from the only reason for the many failures in clinical trials. What can be done to better the situation in both areas of inquiry?

Why are TBI clinical trials so problematic?

This review will focus mostly on the failures of clinical trials testing pharmacological interventions for acute TBI and in particular on the failure of two major Phase III trials of progesterone in 2014. Progesterone is a natural steroid that can be synthesized by endocrine glands as well as by cells in the central nervous system (CNS), where it can act locally as a hormone [11,12]. After several decades of pre-clinical work supporting the neuroprotective effects of progesterone, the failure of the clinical trials is a significant setback for the future of research using this hormone as a treatment for brain injury. Understanding why these trials failed to demonstrate treatment effectiveness following moderate-to-severe TBI is critically important and may serve as a useful training paradigm for the design and analysis of future clinical studies if researchers are to ever move beyond the 100% failure rate in this field.

The background: Progesterone treatment showed promise in pre-clinical research

In many animal models of CNS injury, progesterone shows multi-factorial benefits in the repair of the damaged brain. Over 300 pre-clinical studies in both male and female subjects report that, given in the early stages of injury, progesterone reduces the expression of inflammatory cytokines [13], levels of glutamate excitotoxicity [14,15] and vasogenic and intracellular cerebral oedema [16,17]; prevents apoptosis and necrosis [18,19]; restores blood–brain-barrier integrity [20,21]; and enhances functional recovery on sensory, cognitive and motor tasks [22–26]. At the morphological level, acute administration of progesterone after neural injury can stimulate glial cells to increase myelin formation [27–29] and restore metabolic function through its effects on the mitochondrial transition pore and in calcium channel modulation after injury [14,30,31]. The hormone has growth-promoting properties and, in animal models of injury, stimulates the expression and release of brain-derived neurotrophic factor, nerve growth factor and insulin-like growth factor, which help to repair the damaged brain [32–34] by stimulating neurogenesis and synaptogenesis [35–39]. There is increasing evidence that progesterone regulates many genes involved in the expression of trophic factors and the inhibition of inflammatory cytokines and that it works through multiple nuclear and membrane-bound receptor mechanisms [40] to regulate growth-promoting and anti-inflammatory genes involved in CNS and other tissue repair [41]. At the pre-clinical level, progesterone and its metabolites produce all these beneficial effects in the brain and spinal cord after contusion injuries, nerve crush injuries [42], diffuse axonal injury [43], stroke [44,45], haemorrhage [46,47], cytotoxic injury [48] and even in certain models of neurodegenerative diseases [49–51].

Phase II trials seemed promising

While there are a few reports in the pre-clinical literature showing no benefits of progesterone treatment [52–54], the vast majority of studies supported the idea of testing it in patients with TBI and indicated that it had no toxic effects in the doses needed for neuroprotection after brain injury. Accordingly, beginning early in the new century, two single-centre clinical trials were conducted, one in Atlanta, GA with 100 patients with moderate-to-severe TBI and one in Hangzhou, China with 159 patients with severe TBI.

ProTECT II [55] was a randomized, double-blind, placebo-controlled trial conducted at a Level 1 trauma centre in Atlanta, GA. It enrolled 100 male and female adult patients, with consent, within 11 hours after their injuries, who had either moderate or severe Glasgow Coma Scores (GCS) of 4–12. This scale goes from 3–15, where the lower number represents the worst level of consciousness and 15 represents the best outcome. There were four patients in the 3-day intravenous progesterone treatment group for every one patient in the placebo control group. Seventy-seven patients received progesterone in an Intralipid emulsion and 23 received just the Intralipid. The loading dose of progesterone was 0.71 mg kg^{-1} h^{-1} at 14 mL h^{-1} for the first hour and then 10 mL h^{-1} of 0.5 mg kg^{-1} h^{-1} for 11 hours. Five additional doses were given at 10 mL h^{-1} for a total of 3 days of treatment at 12 mg kg^{-1} day^{-1}; the average treatment delay was 6.3 hours. The frequency of serious adverse events (SAEs) and mortality at 30 days post-injury were the measures of drug safety. The primary measure of functional benefit, also measured at 30 days post-TBI, was the dichotomized Glasgow Outcome Scale-extended (GOS-E). The Disability Rating Scale (DRS), another quality-of-life measure, was also used.

No SAEs were observed for the treatment group compared to controls, who received current standard of care. At 30 days post-injury, treated patients with severe TBI (GCS 4–8) remained in coma longer, but had a significantly lower mortality rate compared to controls, yet had slightly worse GOS-E and DRS scores compared to controls, possibly due to survival of a badly-injured treated sub-population that would have died had they been members of the control group. The patients diagnosed with moderate TBI (GCS 9–12) and given progesterone had better 30-day outcomes on the GOS-E and DRS than those who received placebo. Patients who had been discharged from the hospital were assessed by telephone. The authors of the study were careful to note that this was a small, single-centre study with a 4:1 ratio of experimental treatment to controls examining outcomes *only* at 30 days post-injury using measures that an National Institutes of Health (NIH) expert consensus panel noted 'can miss clinically important findings that may be detectable by more sophisticated neuro-psychological tests' [56]. They also noted that their primary purpose was to assess safety, not efficacy. No biomarkers, dose–response, duration of treatment or timing of treatment initiation parameters were evaluated.

At about the same time, a Phase II, single-centre, 1:1 randomized trial conducted in Hangzhou, China [57] focused on 159 male and female adult patients with severe TBI (GCS 3–8) only. Blinded treatment with progesterone or vehicle ($1\ mg\ kg^{-1}$) was given by intramuscular injection in camellia oil vehicle in patients enrolled within 8 hours after injury and then once every 12 hours for 5 days at $2\ mg\ kg^{-1}\ day^{-1}$. No dose–response, duration of treatment or timing of treatment initiation was evaluated. The average treatment delay was 3.7 hours. Primary neurological outcomes were measured by the GOS dichotomized into favourable or unfavourable outcomes. For secondary efficacy measures the trial used the Functional Independence Measure (FIM) and mortality at 3 and 6 months post-injury.

In this study, no SAEs during hospitalization were reported for the treatment group. With 84% availability of patients at 6-month follow-up, mortality was also significantly lower for the progesterone group (18% vs 32%). At both 3- and 6-month follow-ups, the patients treated with progesterone also reportedly had significantly better dichotomized GOS and FIM scores than the controls.

Although the protocols were significantly different and each of the two studies had serious limitations that the authors were careful to note in their publications, the results of the Phase II trials were interpreted to suggest that at both the 1-month end-point of the ProTECT II trial and the 3- and 6-month end-points of the Chinese trial, severely injured TBI patients given progesterone treatments had lower mortality than controls. On the other hand, in ProTECT II GOS scores were improved at 1 month, but only in the moderately injured patients, while such improvements were seen in the severely injured patients in the Chinese trial at both 3 and 6 months. Both trials reported no SAEs due to treatment.

If at first you don't fail: Phase III protocols and results

At the time of these reports there were no other approved neuroprotective treatments for acute TBI, so the results were considered encouraging enough to go forward to Phase III testing. Two independent, FDA-approved, NIH- and industry-sponsored Phase III trials were conducted almost simultaneously and both failed to support the Phase II studies [58,59]. Once again, the field faces a 100% failure rate in finding a safe and effective acute-stage, neuroprotective treatment for TBI.

ProTECT III, the NIH/NINDS trial, was a double-blinded, two-arm, 1:1, 49-centre trial that intended to enrol ~1200 patients with moderate-to-severe acute TBI, based primarily on the evaluation of patients with GCS scores of 4–12 [58]. Over 17 600 patients with TBI were screened to obtain the ~1200 intended for enrolment. The goal was to obtain a 10% absolute difference in outcome between the treatment and control groups. The median age of the patients was 35 years, although some were in the range of 70–94 years. Seventy-four per cent were males. Patients ranged from moderate to severe: 29% were classified as moderate, 54% as moderate-to-severe and ~18% as severe at the time of randomization. Patients received i.v. progesterone in Intralipid within 4 hours after the injury or Intralipid alone. The progesterone was first dissolved in ethanol by the Emory Investigational Drug Service and then shipped to the individual centres where it was mixed at each site with a 20% Intralipid emulsion according to the weight of the patient. The actual preparation of the drug ($0.05\ mg\ kg^{-1}$ body weight per millilitre of infusate) or vehicle was coded and blinded to the staff. A new intravenous drip bag was provided every 24 hours. The treatment started within 4 hours after injury and began with a 1-hour loading dose followed by 71 hours of continuous infusion at $12\ mg\ kg^{-1}\ day^{-1}$ and then 24 hours of tapering towards the end of the treatment for a total infusion of 96 hours. Thus, patients received $14\ ml\ h^{-1}$ for 1 hour and then $10\ ml\ h^{-1}$ for the duration of treatment. Patients were followed for ~6 months (±30 days) post-injury using the stratified, dichotomized GOS-E as the primary outcome measure to assess favourable or unfavourable outcomes based on initial evaluation of injury severity. Patients with a less severe initial injury had to have a 'better' recovery than those with a more severe injury. Secondary outcomes at 5–7 months were incidence of mortality, DRS scores and frequency and types of SAEs.

After enrolment of 882 patients and a second interim analysis of the blinded data, ProTECT III was stopped for futility when no statistically significant differences between the progesterone-treated and placebo groups were found in mortality or on the primary functional GOS-E outcome at 6 months post-injury. At this stage the only positive finding was that the hormone had 'an acceptable safety profile' in which there were no SAEs due to the treatment. There were no data to suggest that progesterone had any benefit over placebo.

The SyNAPSe Phase III trial was supported by BHR Pharma, a privately held company with a long history of making progesterone products, and conducted in ~180 centres across 21 countries in Asia, Europe and North and South America [59]. Over 10 500 patients were screened for eligibility and 9000 excluded from further enrolment. The SyNAPSe double-blind, randomized, two-arm, 1:1 study completed its enrolment of 1195 patients, all with an initial

diagnosis of severe TBI defined as GCS 3–8. The patients were 16–70 years of age and received either i.v. progesterone in a proprietary lipid emulsion of 6% soybean oil and 1.2% egg lecithin or just the emulsion within 8 hours after injury and continued for 120 hours. The patients in the treatment group received a 1-hour loading dose of $0.71\,mg\,kg^{-1}$ followed by $0.50\,mg\,kg^{-1}\,h^{-1}$ for 119 hours at $12\,mg\,kg^{-1}\,day^{-1}$ without a final taper. Blood levels of progesterone were analysed 2 days after initial dosing and showed a median level of $335\,ngml^{-1}$, similar to the levels reported in the ProTECT II trial.

For SyNAPSe, the primary outcome measure was the GOS and then the GOS-E at 6 months post-injury. The slightly different GOS-E was employed later after the initial injury because it uses several more categories of disability (and recovery) classification and was thought to have more potential to find an effect. A 36-item short form health survey administered at 3 and 6 months was also used to assess patients' quality-of-life—if they were able to complete the questionnaire at all. To meet FDA requirements for fast-track approval of their progesterone formulation, the study investigators set out to find a 10% improvement in outcome in patients with severe TBI at the relatively stringent two-tailed significance level of $p<0.01$. The trial investigators also examined the possibility that there would be geographic differences in outcome as well as differences in outcomes based on estimates of best, worst and intermediate recovery, as determined by the GOS scores.

On the 6-month GOS-E, there were no statistically significant differences between the placebo and progesterone groups. The proportion of progesterone-treated patients with good recovery was the same as the placebo group (50.4% vs 50.5%). Even with a sliding dichotomy analysis, there were no significant differences on the GOS or GOS-E at either 3 or 6 months. As with ProTECT III, there were no more adverse events in the progesterone group than in the placebo.

What went wrong?

Both the ProTECT and the SyNAPSe authors propose a number of factors that could have led to the negative results of the trials. Skolnick et al. [59] cite the probably inappropriate characterization of TBI as a uni-dimensional disorder based primarily on GCS and Marshall classification scores of initial injury that are based on categorizing CT scan abnormalities; also the insensitivity of the measures and the lack of any mechanistic early outcome end-points and biomarkers as possible 'major obstacles' to obtaining positive findings. Many of these issues were discussed and evaluated while the trials were still in progress (see Maas et al. [60] for an excellent review). Additional factors that deserve attention are: possible issues with the dose levels and/or durations of treatment selected for the Phase II studies; not optimizing those parameters in additional Phase II studies prior to planning Phase III studies; errors of execution of approved protocols; effects of rehabilitation and other therapies on outcome measures; and even time pressure from patent and other economic factors pushing to complete the project before the possibility of late recovery with or without treatment could be assessed.

In the case of progesterone for TBI, clinicians had access to hundreds of published animal studies in a variety of central and peripheral neural injury models with molecular, physiological markers and multiple functional/behavioural outcomes supporting the neuroprotective effects of the hormone in CNS injuries. These pre-clinical reports pointed to a number of critical parameters such as dosing levels, duration of treatment, window of treatment, route of administration, sex and age differences as variables needing attention in clinical trial designs. Yet, for reasons that are not yet evident, the two very similar Phase III trials went forward without incorporating any drug optimization studies into their design.

In a recent editorial in *Nature Reviews Neurology*, Menon and Maas [61] conclude that better trial design, better patient selection procedures, better outcome measures and better options for when to take them and how often are needed. The subject needs better everything! Indeed, Menon and Maas raise the question of whether TBI is just 'intrinsically unmodifiable' or the outcome tests are 'intrinsically too insensitive'. The ProTECT III investigators [58] ask whether 'It is possible that the heterogeneity of the injury, confounding pre-existing conditions, and characteristics of individual patients (e.g., resilience), which can well be controlled in animal models, play too large a role to overcome in human disease' (p. 9). They propose that what is needed are rigorous multi-centre trials in animals that better simulate human trials if the field is to advance, despite the fact that the pre-clinical experiments that informed the Phase II progesterone trials met all of the Stroke Therapy Academic Industry Roundtable recommendations, with the exception of testing in non-human primates. (It may be worth noting that primates—human patients—were in fact tested, with more subjects than would likely have been used if the subjects were monkeys.) The SyNAPSe group [59], also using just the GOS-E, point to the long history of failures, citing the insensitivity of current outcome measures among other 'major obstacles' to the development of successful treatments and the limitations in the ability to translate experimental data to the context of TBI in humans.

Defining TBI

It appears that for many diseases, including and perhaps especially TBI, the outcome measures and end-points selected for evaluation do not clearly reflect the course of the disease being studied and, in the case of TBI and stroke, are not sensitive enough to quantitatively measure short- and long-term deficits and their gradual recovery over time, particularly in cases where recovery is only modest. The same problems may hold for clinical trials conducted in the post-acute stage. In a recent consensus paper, Maas et al. [60] highlighted the problems with both acute and post-acute clinical trial research, noting that,

> In particular, there is little continuity in research between acute and post-acute care studies. Nevertheless, disparities in access to post-acute care may influence the recovery process and confound interpretation of outcomes. A major challenge in the post-acute care phase is posed by the highly variable time periods at which data are recorded,

> confounding comparability of studies and interpretation of their results. Thus, a great need exists for more prospective longitudinal studies bridging the gap between acute and post-acute research in TBI (p. 38).

Furthermore, there is still no clear consensus on how TBI should be characterized and defined. If the disease is hard to define and if the measures are not sensitive or consistently representative of the extent of the injury, sophisticated statistics techniques (dichotomized scaling, logistic regressions, probabilistic odds ratios and so on) will not avail. Again, as emphasized by Maas et al. [60],

> Traditional clinical trials and studies have relied upon a hypothesis-driven, model-based approach. While this reductionistic approach has been very successful in developing treatments for infectious diseases and cancer, where single organisms or cell types are responsible for the pathology, there has been only limited success using this approach for more heterogeneous complex diseases, such as inflammatory disease, diabetes and cardiovascular disease. In these complex disorders there is likely no single factor that is responsible for the disease. This is particularly true for disorders of the central nervous system, such as traumatic brain injury, for which there is significant heterogeneity in the aetiology, pathology, mechanisms and outcome (p. 39).

Heterogeneity of the injuries, the potential for 'spontaneous recovery' (where some patients, despite massive physical damage to the brain, show surprisingly extensive recovery) and the lack of any correlative measures indicating locus and extent of the damage could have led to problems in defining the TBI itself. In some cases, for example, depending on the locus a circumscribed injury could produce low GCS and GOS scores while a larger injury in another area might result in a much better outcome score. In closed head injuries where diffuse axonal damage may occur throughout the brain, imaging technologies are not yet sensitive enough to determine what is damaged and how widespread the damage is. In both Phase III trials, thousands of potential patients had to be screened to find those that met the criteria for enrolment. ProTECT III patients had to be enrolled within 4 hours after their injuries, so detailed screenings to eliminate fluctuations in GCS scores were very limited. Patients enrolled with co-morbidities could make it more difficult to assess whether their functional status was due to the injury itself or to pre-existing, hard-to-screen disease conditions which could affect the response to treatment, such as a history of alcohol or substance abuse. Both the wide variety of medical treatments for such co-morbid conditions and additional treatments for TBI-related pathologies (e.g. anti-epileptics and analgesics) present potential confounds. As just one example, in a small trial of post-acute amantadine administration after TBI which failed to find a difference between treatment and placebo, 'approximately one-third of the patients received potentially confounding medications' ([62], p. 823). The problem of heterogeneity of the patient samples enrolled in trials has plagued the field, as was noted in the failure of the COBRIT Phase III trial testing the effects of oral, long-term dosing with citicoline to obtain any differences in the outcome of treatment and placebo groups. This trial admitted a new category of patients with 'mild-complicated', as well as patients with moderate and severe TBI [63]. It is clear that how TBI is characterized needs to be changed.

Dosing

In ProTECT III and SyNAPSe, with just one dose level and two very similar durations of treatment, it's possible, even likely, that a sub-optimal dose and/or schedule was used in both trials. Phase III trials are normally used to confirm efficacy in a large population after a particular dose and schedule have already been optimized in previous Phase II trials [64]. Unfortunately, in this case no attempt was made to optimize either dose or schedule prior to Phase III initiation. The two Phase II trials that were done differed significantly in several ways: they used 6-fold different dose levels (12 vs $2\,mg\,kg^{-1}\,day^{-1}$), different routes (i.v. vs i.m.) and vehicles (Intralipid vs camellia oil) in the ProTECT II and Chinese trials, respectively. On the other hand, two similar, relatively short, dosing durations (3 and 5 days) were used in the two studies and it's entirely possible that neither duration approached what would be necessary for optimal treatment in humans. Instead of attempting to reconcile and optimize the above conditions in an additional Phase II trial, the investigators chose perhaps a more risky strategy of employing the high dose level used in ProTECT II for both ProTECT III and SyNAPSe and 3-day (plus a 1-day taper) and 5-day dosing durations, respectively, also similar to those used in Phase II. It is important to note that, in pre-clinical studies in both TBI and stroke models, an inverted U-shaped dose–response curve was observed, with higher doses of progesterone showing no efficacy. However, these data were not available to the clinical investigators until well after the trials were underway [45,65,66]. This is one reason why having a dosing optimization study as part of the Phase III (or Phase II) trials might have been worthwhile.

Over-valuing false positives

The two Phase III trial reports in the *NEJM* were accompanied by an editorial [67], attempting to put the issue of constant Phase III trial failure into perspective. Schwamm [67] notes that both trials had much in common with other failed neuroprotection trials and attributes these failures, in part, to the 'lack of informative pre-clinical models and biomarkers' relevant to TBI in humans. He concurs with a widely cited article by Ioannidis [68], arguing that clinical research in general typically fails because the laboratory studies on which trials are based often do not replicate one another and have too many false-positive findings. According to Schwamm [67] and Ioannidis [68], these published papers then lead to unwarranted and over-enthusiastic estimations of effect sizes and are uncritically reported in the peer-review literature and accepted and used by the clinical community to go forward with clinical trials (see also Mak et al. [69] for further discussion of this issue as related to cancer clinical trials).

Schwamm [67] suggests that the over-optimistic interpretation of pre-clinical and Phase II data and the uncritical

reliance on a system that tends to generate false positives could be one of the stanzas in the 'Siren Song' that leads clinicians astray in setting up clinical trials. Schwamm correctly observes that there was only modest improvement in functional outcomes at 1 month for ProTECT II and at 6 months for the Chinese trial, where (he asserts) re-assignment of just one patient in the placebo group from unfavourable to favourable outcome would have changed the overall trial results to favour no effect between the treatment and control groups. Schwamm also notes that, for the ProTECT II study, there was an expectation that 50% of the patients in the placebo group would have good outcomes and, indeed, in the Phase III trial 55% of the patients in the placebo group did have favourable outcomes—perhaps indicating that just being enrolled in a clinical trial can lead to better outcomes, possibly because all the patients are receiving more attention and better overall care.

The claims made by Schwamm and other critics do have merit. There *is* a lack of translation from the pre-clinical animal research to effective Phase III trial outcomes. Nor does every published paper meet the rigorous standards operationally defined by some researchers using meta-analyses to determine what passes for good work and what does not [10,68,70]. If the problem is as widespread as claimed and if most pre-clinical research cannot be trusted, this could very well turn out to be an existential question for basic research in neuroscience. If it does not translate, what is its purpose? Are the monitors and critics right in saying that pre-clinical studies have nothing to offer in informing clinical trial research?

Whose criteria should be used to decide what constitutes 'solid' pre-clinical data? What determines a 'well-vetted hypothesis'? How much corroboration and replication is 'required'? How do outcome measures for different disease conditions become standardized? How will it be decided when there is enough supportive evidence to go forward with a clinical trial? Again, some progress is being made in this area and perhaps in future trials this issue will be handled more appropriately [71–73].

Outcome measures: Blunt instruments

Schwamm [67] suggests that future problems can be avoided by pooling pre-clinical data and requiring more co-ordinated and sequential Phase II trials using *standardized outcomes* to replicate potential findings. However, while important to consider as a valid strategy for planning future trials, this approach may actually be one of the key reasons that clinical trials for TBI (and stroke) continue to fail: the great majority of them use the same standardized, short, blunt instruments evaluating quality-of-life that do not approach the detail and precision of the measures used to evaluate morphological and functional outcomes required of animal studies. Subjective rating measures in pre-clinical studies would not be publishable if bench scientists used non-quantitative 'quality-of-life' outcome measures applied to animals analogous to a 7–10-minute questionnaire about how patients think they are doing at 30 days or 6 months after moderate-to-severe brain injury. This is a very substantive difference between pre-clinical research and clinical trials.

Surprisingly, the issue of faulty outcome measurement is well known to clinicians leading clinical trials. Both mainstays of TBI outcome, the GCS and the GOS, have been taken to task as tools for evaluating brain damage. An opinion article by Green [74] finds that 'It is time to abandon the... GCS, as this ubiquitous neurologic scoring system is confusing, unreliable and unnecessarily complex, and its manner of common clinical use is statistically unsound.... To be accurate and useful, a clinical scale must be reproducible. Unfortunately the GCS contains multiple subjective elements and has repeatedly demonstrated surprisingly low interrater reliability in a variety of settings' (p. 427). Yet if this test is the very basis for categorizing patients by the severity of their injury (mild, moderate, severe) and determining treatment protocols, it should be obvious that a serious situation exists.

Others have raised similar issues (see Maas et al. [60] and Retzios [75] for a discussion of problematic neuroprotection end-points in clinical trials). If the GOS has problems similar to the GCS (and it does), then misclassification of patients, moving them arbitrarily into one category or another (regardless of how ordinal the measures may be), inter-rater reliability and levels of competence in scoring across many centres and indeed across continents and countries could produce substantial variability in how the outcome measures are interpreted [76,77].

The subjective nature of the GOS-E

The GOS relies on patient or close caregiver evaluations in a yes/no format. A sample of questions:

- Is the assistance of another person at home essential every day for some activities of daily living?
- Do you need frequent help or someone to be around most of the time?
- Did you need assistance before your injury?
- Are you able to shop without assistance?
- Can you travel locally without assistance?
- Are you currently able to work to previous capacity?
- Do you/the patient have psychological problems?

The subjective nature of the GOS-E questions and the lack of quantification of deficits or recovery from them are obvious. In addition, self-assessments of one's condition can often be flawed and subject to unrealistic optimism or pessimism, which can be influenced by the patient's family, caregivers and economic circumstances (see Dunning et al. [78] for in-depth discussion). As an example, since most patients with moderate-to-severe TBI, at least in the US, often receive disability payments from insurers or the government, their responses to the GOS-E at only 3 or 6 months after their injury might alter the status of their reimbursements. How would this be evaluated in a yes/no survey of quality-of-life? Should this question even be asked? Could this be a confounding variable when the maximum time of testing is at 6 months after a TBI? 'Post-concussion syndrome' resulting from a mild TBI, while often resolving within 3 months, can last for up to a year in some patients [79–81]. If this is the case, would it be likely that patients would respond

well to a quality-of-life questionnaire? There is also evidence to suggest that there are sex differences in how long it takes to recover from a concussion in young patients [82]. So, if a mild concussion can have such long-term effects, when, indeed, is it appropriate to measure the long-term consequences of moderate to severe TBI—especially on quality-of-life indicators?

The weakness of the GOS-E as a primary assessment tool has been re-emphasized in describing the criteria and design of a new TBI clinical trial for the testing of cerebrolysin as a potential neuroprotective agent [83]. As Poon et al. [83] note, 'In a clinical trial, a single outcome measure cannot capture all clinical relevant information from any type of TBI survivor' (p. 572). They further state that the

> commonly used GOS and GOS-E, which measure global functioning after TBI, are insensitive to important and specific deficits in behaviour, executive function, memory and emotion that may produce significant disabilities ... Further, previous trials often dichotomize the GOS and GOS-E to enable logistic regression analysis of multiple predictors as covariates, *a method that has been shown to discard potentially relevant information, limit statistical power, and not correspond well with clinical practice* (pp. 572–573; emphasis added).

Are long(er)-term outcome measures worthwhile?

Because of cost and the difficulty of retaining patients and repeatedly visiting them to monitor outcomes, the longest TBI and stroke trials typically measure outcomes is ~6–12 months after the injury. As noted, in the case of severe TBI, this may simply not be enough time for patients and their caregivers to feel that they have made substantial progress on quality-of-life measures.

Particularly with a *severe* brain injury, how likely is it that quality-of-life would, indeed, substantially improve by only 6 months? In any case, there is substantial evidence from psychological studies that self-assessment often leads to flawed outcomes and inconsistent experimental findings. In addition, if the scores also have only moderate-to-weak correlations with other biomarker measures, the problems with interpretation of outcomes become even more severe and the trials more likely to 'fail'. The extent of recovery may also depend on the level and duration of rehabilitation and counselling therapy that patients receive after injury. The substantial variability that would be contributed by the type, duration and professional level of rehabilitation therapies (or not), treatment centres, states and countries could have profoundly influenced the outcomes of these two trials—especially in the 3–6 months following the injuries. Recently, Lingsma et al. [84] analysed individual data from close to 10 000 patients with moderate-to-severe TBI that were enrolled in 10 randomized, controlled trials using the GOS at 6 months as the primary measure. The investigators reported 'substantial differences' in overall quality of care and outcomes, particularly in Europe (3.8-fold) vs the US, where the differences were 'only' 2.4-fold. Lingsma et al. [84] noted that, in some centres, unfavourable outcomes were more than double the average, while other hospitals had less than half the average of patients with unfavourable scores. Wide-ranging differences like those reported by Lingsma et al. have the potential to obscure any treatment effects in any trial, regardless of the agent used, especially if the only outcome measure is the GOS and GOS-E or other similarly blunt and qualitative measures. Even if this issue could have been addressed in some way, another problem stems from the contextual structure of the GOS-E questions themselves, which can appear to guide the patients and their caregivers into focusing on what is wrong with them and what they cannot do compared to their pre-injury state, rather than on how much they have been able to accomplish despite their injury.

More quantitative measures are available, but they are expensive and time-consuming

To improve assessment of the patients' condition, the recording of more details about their specific disabilities and/or extent and type of recovery from them would seem to be in order. As in animal studies, it is also possible to measure and quantify the rate and extent of recovery of gait, sensory and cognitive abilities over time [66] and such information would help the patient recognize that they are (or are not) improving by a specific amount—even though they may remain in their original category as assessed by the GOS-E over time. Such subtle changes in recovery or worsening are missed when only subjective sliding scales and broad categories of quality-of-life outcomes are used—and, without better monitoring, it is difficult to know whether responders to the GOS questionnaire over- or under-estimate the patient's condition at any given time [78] (see also Menon and Maas [61] for more discussion of this issue). As noted, comprehensive neuropsychological tests require the services of a skilled professional, are time consuming (4 or more hours/session as opposed to the 15-minute GOS-E), fatiguing to patients with disabilities, more intrusive, harder to interpret and certainly more expensive to administer—especially if testing over repeated intervals is part of the clinical trial design. These factors militate against their use and clearly favour simplified, easy-to-use and understand tests like the GOS.

Clearly, the many problems surrounding the GOS or GOS-E (or similar short-form evaluations such as the Galveston Orientation and Amnesia Test (GOAT), which is thought to measure a patient's level of 'disorientation') as a primary outcome measure in TBI trials are known and understood. Roozenbeek et al. [85] recently critiqued the GOS in these terms: '[S]etting an arbitrary threshold which patients must cross to demonstrate clinical improvement is not reflective of the clinical situation and, in fact, substantially reduces chances of showing benefit... Although, perhaps intuitively attractive because it is so simple, the traditional approach to dichotomize the GOS is counterproductive and disregards potentially valuable information contained in the ordinal scale (p. 40). Yet, the authors note, ''In the absence of early mechanistic end-points, TBI investigators and regulatory authorities have both adopted the ... GOS ... as the standard for primary efficacy analysis' (p. 4).

Others have proposed dichotomizing GOS scores rather than using the standard GOS outcomes [86]. These authors conclude that there are statistical problems with the dichotomization approach and that using sliding dichotomies does not automatically provide more power to detect proportional differences between treatment and control groups. The ProTECT III trial stratified the GOS-E scores according to initial injury severity with different criteria for reporting favourable outcomes in each stratum. From the Price et al. report, one can conclude that the issue of creating more subjective categories of good recovery vs bad recovery to find significant differences among patients is far from being resolved.

Ordinal scaling may be better, but is it good enough?

Even with the use of ordinal scales to improve the evaluation (adding a few more subjective categories than just good or bad recovery), the GOS-E measures themselves are still blunt. When there are large differences in outcomes after moderate and severe TBI across centres, the problem of reliability and validity of the outcome measures is of serious concern. In the face of such large between-centre disparities, it seems reasonable to ask why anyone would continue to use the qualitative, judgmental assessment as a primary or only outcome measure. It would have been highly informative to discuss this issue in the SyNAPSe trial publication [59]. Perhaps the temptation to use these blunt instruments is one of the verses of the Siren's song that needs to be kept from the clinician's ears. Indeed, what are the temptations? It may be worthwhile to note that none of the problems with clinical trials would have been unambiguously solved by strengthening the pre-clinical studies—increasing the numbers of animals/group, using more sophisticated and stringent statistics or better blinding and selection techniques or providing more mechanistic analysis of drug effects—yet trial after trial uses the same outcomes and follows the same trial designs. Why?

Are there good reasons to keep the GOS as a primary measure?

Why persist with the use of the GOS-E as a primary outcome measure (or use it at all) in TBI trials? There are many good alternatives, but they are much more time-consuming, much more expensive to administer and perhaps frustrating for patients if they don't understand why such testing might benefit them and their caregivers. Although the NIH Toolbox (free and online: http://www.nihtoolbox.org/Pages/default.aspx) provides 41 normed and validated multi-dimensional tests to evaluate cognitive, sensory, motor and emotional function in people from 3–85 years of age, the GOS and GOS-E are still the most-used primary outcome measures.

There is a rationale for doing this:

- It is widely used because it is widely used. The FDA likes it and, despite the efforts that went into the development of the NIH Toolbox, paradoxically, the GOS has been *recommended* by the NIH (the funding agency) for use as a primary outcome measure [87,88] (but see also [89] addressing additional concerns with this test).
- Compared to quantitative neuropsychological testing, it is easy to use.

Can be given by structured interview, phone or even by mail.
Takes only ~10–15 minutes to administer.
Does not require extensive training or professional certification to administer.
Does not require any high technology to administer.

- It is relatively inexpensive.
- The categories are clinically relevant, address activities of daily living and are understandable by patients or their caregivers.

For practical reasons, then, it might make sense for FDA to encourage use of the GOS as a primary outcome measure, but if so many researchers agree that it has serious problems that in the long run can sink a clinical trial and a promising treatment, is the practicality worth it? How should one weigh the risks and benefits of a pragmatic approach to the assessment of a very complex and heterogeneous disease? The question of the use of GOS-E as a primary outcome in future clinical trials is not likely to be resolved anytime soon and there will be further debate on its value compared to the perspective taken by Poon et al. [83] and others cited above. In a very recent article, Alali et al. [90] argue that improving analysis methods for using the GOS-E under more sophisticated forms of statistical evaluation can lead to better prediction of functional outcomes just as effectively as the use of more multi-faceted testing approaches and this is why the FDA has accepted the GOS-E as 'the single primary outcome measure for TBI treatment trials' (p. 586).

The unspoken challenge: Time is money

One major factor that does not often get discussed in the literature is cost and finance in the broad sense of the term. Everyone knows that large, multi-centre clinical trials take a long time, are extremely expensive and require vast amounts of effort. If sponsored by federal agencies there are usually 3–5 years of funding before another application for continuation is needed and over that interval government budgets can be dramatically altered, rarely for the better. Once approved for Phase III testing, ProTECT III took almost 5 years before it was stopped and the privately-funded SyNAPSe trial took somewhat over 3 years.

There was considerable pressure to complete the trials as quickly as possible, so centres were added on an *ad hoc* basis as the trials proceeded. In the haste to add more centres to recruit more patients, was quality control affected? Analysis of data entry errors and transgressions over time may answer this question. Using simplified outcome measures and short trial intervals (e.g. 3–6 month follow-ups) meant that the trials could be completed more quickly and there was an additional, important economic incentive for doing so. Before a new drug can be used in clinical practice someone has to manufacture it. For any investors or company manufacturing a potential new therapeutic agent, an important key to successful drug development is to have patent protection and exclusive control over the agent if it goes into clinical practice. In the US, new drug patents are issued for 20 years from the application date. Most patent holders do not disclose the details of a new invention (or drug) until they have national and international

patent protection and the drug cannot be used in clinical trial until its details are disclosed to the FDA.

The longer a clinical trial takes to complete, the less time remains to have control over the manufacture and sale of the drug and there are often substantial domestic and international fees to maintain the patents. Patents are not free. If the trials begin with Phase I/II clinical testing and then go to Phase III, much of the time remaining for the patent protection is consumed—so there is every motivation to move as quickly as possible to completion. Long-term follow-ups and extensive testing quickly become very expensive and time-consuming. It is reasonable to speculate that, in addition to the hope of getting a successful treatment to patients as soon as possible, the SynAPSe investigators also sought fast-track approval for their drug to beat the clock before patent protection ran out.

Even if subsequent/secondary analyses of the progesterone trials were to show signs of efficacy and even if the FDA were to approve a smaller, more limited trial based on further pre-clinical and clinical data, progesterone's relatively low cost and now limited 'use' patent protection are negative incentives to any new investment in developing it as a treatment for TBI. Despite all the pre-clinical evidence of strong biological and functional signal, progesterone may never see success in clinical application for TBI. New trials would run out the limited use patents that currently exist and there would be no economic incentive to produce the drug for general clinical use in TBI without some protection. This is likely the reason so many other drugs that once showed promise in early development are never tested again in light of any new data—the patents run out and there is no longer an economic reason to make them. There is a substantial trade-off between taking the time to do all the appropriate clinical trial measures at the optimal time-points and the need to get the drug to market and recoup costs of research and development before the patent protection runs out.

The fallout from the Phase III trial failures: What if a good drug never gets used?

Despite the positive pre-clinical evidence and the design issues surrounding both Phase III TBI clinical trials, there is a growing assumption that any further pre-clinical study of progesterone for TBI or other diseases is now futile. Even if the science of progesterone is good, for some, the negative trial findings make the possibility for translation to clinical use very unlikely. Thus, for example, this could mean that, even if there is substantial laboratory evidence that progesterone may work for stroke, it should never proceed to clinical trial because it didn't work for TBI. Or be tested for any other CNS disorder? This 'progesterone fatigue' becomes a serious dilemma, especially if there are no better treatment alternatives for TBI or stroke patients. It may very well be true that in humans, progesterone is simply not effective in treatment of CNS disorders, but this claim is far from being proven by the TBI trials. Is it rational to assume that, if a drug does not work in one setting, it will never work in any other setting? Many neurological specialists will argue that stroke and TBI are very different disorders, so is it logical to assume that mechanisms of TBI are virtually identical to the mechanisms of all the different kinds of stroke? Even before all the facts are in, many colleagues working with neurosteroids have already reported that, despite good scientific reviews of their research proposals, 'enthusiasm' for funding is diminished because the clinical trials were not positive. How is further progress to be made? Are there better options than a rush to judgement?

Time to change some rules

Growing awareness of how serious the problem has become is demonstrated by the Agency for Healthcare Research and Quality (AHRQ)-mandated white paper [10] and by a recent $17 M DoD grant, the TBI Endpoints Development (TED) Award, to the University of California San Francisco to support a public-private partnership to focus on designing and developing better clinical trials for TBI interventions. The group plans to examine records from thousands of patients to see if they can tease out what needs to be done to improve patient selection and diagnostic and outcome measures. Colonel Dallas Hack, Director of the US Army Combat Casualty Care Research Program reported to have said he initiated the TED programme because '[i]t had been a growing source of frustration to me that we couldn't get anything through the system. *If you can't win the game you have to change the rules*' (emphasis added) [91].

To-do list

Besides careful adherence to institutionally mandated (FDA, NIH, individual centres, etc.) guidelines, here are four suggestions to improve the chances of a successful clinical trial. This list is by no means exhaustive and, as the TED research program goes forward, it is likely that other recommendations will emerge.

(1) Agree on how to define brain injury

A more precise and focused pathoanatomical characterization of TBI must be devised and used consistently to select patients for enrolment in clinical trials. This has been a high priority for the NIH and was extensively discussed in a consensus report back in 2008 [92]. Despite its ease of administration, it may be time to retire the GOS. As in animal studies, subject/patient selection should be based on a combination of imaging, biomarker and functional outcome measures that permit more localization and quantification of the injury as it evolves over time; however, the technology currently available may not yet be completely applicable to patients (see discussion by Wright et al. [58]). This is one area where animal TBI studies and clinical TBI studies show a wide divergence. In virtually all animal studies, a well-defined type and anatomical locus of brain injury is selected for study (e.g. bilateral, frontal cortex injury over the midline, sagittal sinus, striatum, motor cortex, etc.). The drugs are targeted to manage the specific injury and corroboration comes from multiple independent studies demonstrating whether the drug effect is robust or not. In clinical trials, TBI may be defined as 'blunt force to the head', 'penetrating injury', 'complicated mild' [63], etc. Blunt? Penetrating? Any locus? Diffuse axonal injury? [93] Subcortical or cortical bleeding? Bilateral or unilateral? Cortical or subcortical oedema? All of the above in the same

trial? No wonder there are substantial differences between pre-clinical and clinical outcomes.

Large, multi-centre trials based on very broad categories and definitions of brain injury are not working. 'Brain injuries' are not all the same. As Colonel Hack has noted, there are currently 42 different published definitions of concussion. What is the best definition to use? The same problem likely applies to other forms of brain injury, as the neurosurgeon Geoffrey Manley also suggested in a recent interview on how to improve TBI outcomes [91]. The fact that there are a number of upcoming policy-level meetings to build consensus on what to measure and when attests to the importance of this problem for the field of TBI. Manley recently noted that 'TBI lags 40–50 years behind heart disease and cancer in terms of progress and understanding of the actual disease process and its potential aftermath' [91].

Although Manley's points are well taken and important, the issue goes well beyond problems with clinical research on TBI. It is an endemic issue facing clinical trial research in general and how much time, effort and money is spent following false leads based on shaky foundations. In a special issue of *Injury* on outcomes research, one paper argues that

> Fundamental to the reduction of bias in clinical research is the choice of valid and reliable outcomes. The choice of ideal outcome in clinical study has become increasingly complex due to the variety of measures promoted in orthopaedics. The consistency in which outcomes are chosen or applied in the orthopaedic literature is highly variable. Despite decades of fracture research, *we continue to struggle with basic definitions of healing* [94, p. 231] (emphasis added).

No clear consensus on what constitutes healing? How does one develop a drug if one doesn't know what it's supposed to treat?

One alternative is to select fewer patients with more specifically defined brain injuries and study them more intensively with more quantitative measures over longer periods of time. The notion of a 'primary outcome' measure may have to be discarded in favour of evaluating multiple quantitative behavioural and morphological measures indicative of both short- and long-term recovery. This would substantially reduce variability, but would take more time (and money) to accomplish, but in the long-run might be much cheaper than failure. Fewer centres (and only those with an established record of effective clinical trial performance) should be used, at least in the first stages of a Phase III trial. The NIH creation of the NETT is a step in the right direction and, once the program has had time to be evaluated, NIH-supported clinical trials should be limited to centres that meet all the stringent criteria for handling brain-injured patients. All data from the trials, including the data entry errors and transgressions, should be publicly available for further independent analyses.

(2) Improve animal studies

To the extent possible, animal models of brain injury should be designed to reflect aspects of human brain injury that will be examined in clinical trials. Both bench and clinical investigators should first reach consensus on this issue through active collaboration and consultation in pre-clinical and clinical trial designs. This can be done in the early stages of planning by having a steering committee consisting of basic researchers and clinicians and funding agency officials who can take advantage of the vast amount of information available online. The trial parameters should be independently evaluated by review panels of basic and clinical research experts.

While pre-clinical studies can inform clinical trial planning as to dose, duration and window of treatment parameters, they are only suggested guidelines that must be replicated in early-stage Phase I/II clinical trials. Animal studies can only partially replicate issues concerning drug metabolism rate and absorption, patient drug and environmental history and genomics and can only serve as rough guidelines for clinical research. In turn, surprisingly few pre-clinical studies examine the interactions between age, sex, environment, pre-morbid conditions such as vitamin/hormonal deficiencies, injury-induced systemic inflammatory disease, stress and handling and the strain of the animals used in the conduct of the research. This is where consensus between the pre-clinical and clinical communities is essential to determine what is needed. To force the issue, it may be necessary for federal agencies such as the FDA and NIH to establish very specific criteria for defining different categories of TBI and then limiting a trial to only those patients in just one (or two) of those categories. This is essentially what is done in the pre-clinical studies to enable replication and substantiation of results. Complaining about patient heterogeneity and the experimental variance caused by it needs to be replaced with more pro-active management and less ambitious, more carefully focused enrolment criteria for TBI trials.

When clinical trials report disappointing results there is a tendency to claim that, if pre-clinical animal studies had reported their negative as well as positive findings, the problem of going forward with a clinical trial based on limited results would have been avoided. There may be some truth to this and there has been a slight uptick in the reporting of negative outcomes in some journals. However, the pre-clinical study must be closely examined to confirm that it did not fail because it was poorly conducted. In this case journals can supply common data elements as supplementary material (since most are online anyway) and readers can judge for themselves. The same should be done for clinical trials.

(3) Develop better physiological and predictive biomarkers

Although better biomarkers that are sensitive to the acute injury cascade could help with prognoses and even with patient assignment/enrolment (for TBI, early-stage imaging, serum markers of the inflammatory cascade [95], breakdown of blood–brain-barrier and its restoration and the regulation of trophic factors to prevent further cell death), another key area for brain injury and stroke is when and how often to measure functional outcomes. Both pre-clinical and clinical studies need to determine how long after an injury functional

and metabolic markers are needed to assess outcome and when recovery or the lack of it is manifest. Are assessments at 3 and 6 months post-injury sufficient to judge whether quality-of-life is fully stabilized? In a largely forgotten chapter on 'late changes in the nervous system', Geschwind [96], a leading neurologist in his time, wrote that, after brain lesions, 'it may be the rule that one never achieves an equilibrium. There are immediate changes, and changes occurring over seconds, hours, days weeks, months and indeed years' (p. 468). Geschwind chronicles late changes in recovery in peripheral nerve, spinal cord, cerebellum and brain stem, hemiplegia and aphasia. One of his patients who was severely aphasic up to a year after his injury was advised that, with his persistent condition, he would not be able to return to work. At 2 years and with no additional therapy or treatment, he told Geschwind that he had returned to his full-time work as a salesman. Under the current clinical trial outcomes for TBI, based on a single quality-of-life questionnaire, the authors have no idea what patient outcomes will be a year or more after injury and whether there would have been a differential response to earlier pharmacological and rehabilitation treatments. As Geschwind and others (see Luria [97]) knew over 40 years ago, there is no law of neurology or neuroscience (but there may be such a principle for health insurance companies) declaring that if recovery does not occur within 3–6 months, it will never occur. In sum, long(er)-term follow-ups of perhaps 2–3 years after injury, with use of more sensitive and quantitative measures of functional recovery, should be required in all Phase III trials for TBI and stroke.

(4) Report and address problems with data handling

Along with other enrolment criteria, the type, extent and frequency of data entry errors (for each participating institution) and the techniques used to clean up the errors should be reported as supplementary data in all clinical trial reports. It will be important to address the problem of data entry errors in larger-scale clinical trials, often involving hundreds of variably trained staff across dozens or even hundreds of centres and different countries. Even single-centre studies can have problems with data errors and transgressions in clinical trial databases, which can lead to inconsistent reports, transfer of data in appropriate format, adverse events, misinterpretation of information and issues with the specific care of patients in the trial. For example, in one study, Goldberg et al. [98] used two large oncology databases from an academic centre in Boston to evaluate errors that could influence treatment and outcome of oncology patients. The investigators found that errors in data were common and led to as much as a 13.5% error rate in each of the databases they studied at just the one centre. Even if 'cleaned up' prior to analysis, such an error rate could translate into dosing errors and erroneous assignments of patients. In the case of the two clinical trials seeking to find a 10% treatment effect, it would seem that the possibility of even a 13% error rate could do serious damage. Data entry error rates and the methods for clearing up the mistakes should have been reported along with the other data published in the two recent progesterone Phase III trial reports [58,59].

Conclusions

In the final analysis, it is not just about what is lost in translation from pre-clinical animal studies to clinical trials. It's about having precise definitions about how best to diagnose and characterize disease (in this case TBI), precise and quantitative measures of outcomes, how to select the most appropriate patients, how to tailor treatment parameters to suit the patient's condition and several other issues like trial design and execution.

Continuing to perform clinical trials under the same mandated standards makes no sense. One major new initiative is the application of adaptive design to clinical trials, a methodology being considered to replace the traditional RCT with its rigorously fixed sample sizes and equal randomization to study groups. Adaptive design typically uses Bayesian computer modeling to help 'adapt' patient assignment to groups, adjust sample sizes as the data come in and allow for modifications in the dosing and duration of drug treatment protocols as the trial progresses and data accumulate. With proper planning and if the trial hypotheses are clearly stated during the design and planning process, new cohorts of patients can even be added (see Yin [99] for comprehensive discussion on these points). This topic is well beyond the scope of this chapter, but there are now excellent resources available to learn more (as but one example, see He et al. [100]).

Thus, on many fronts, clinical trialists will have to take the lead in pressing for the development of new trial paradigms, but it would be folly to do so without much more collaboration with basic, pre-clinical scientists who need to do the studies that make the most sense to move new drugs and techniques to the patient's bedside. This may require overcoming some of the siloization and often stultifying bureaucracy that currently affects the government/academic/scientific and medical communities.

Declaration of interest

Although the author no longer has any financial gains, royalties or licensing agreements from research on progesterone, he does hold active patents related to the use of progesterone in TBI and certain forms of CNS tumours.

References

1. Narayan RK, Michel ME, Ansell B, Baethmann A, Biegon A, Bracken MB, Bullock MR, Choi SC, Clifton GL, Contant CF, et al. Clinical trials in head injury. Journal of Neurotrauma 2002;19:503–557.
2. Amiri-Kordestani L, Fojo T. Why do phase iii clinical trials in oncology fail so often? Journal of the National Cancer Institute 2012;104:568–569.
3. Lowenstein PR, Castro MG. Uncertainty in the translation of preclinical experiments to clinical trials. Why do most phase III clinical trials fail? Current Gene Therapy 2009;9:368–374.
4. Krum H, Tonkin A. Why do phase III trials of promising heart failure drugs often fail? The contribution of ''regression to the truth". Journal of Cardiac Failure 2003;9:364–367.
5. Malik RA. Why are there no good treatments for diabetic neuropathy? Lancet Diabetes & Endocrinology 2014;2:607–609.
6. Grupke S, Hall J, Dobbs M, Bix GJ, Fraser JF. Understanding history, and not repeating it. Neuroprotection for acute ischemic stroke: From review to preview. Clinical Neurology & Neurosurgery 2015;129C:1–9.

7. Dyson A, Singer M. Animal models of sepsis: Why does preclinical efficacy fail to translate to the clinical setting? Critical Care Medicine 2009;37(1 Suppl):S30–37.
8. Jickling GC, Sharp FR. Improving the translation of animal ischemic stroke studies to humans. Metabolic Brain Disease 2015;30:461–467.
9. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature 2012;483:531–533.
10. Goodman SN, Gerson J, Mechanistic evidence in evidence-based medicine: A conceptual framework. Rockville, MD: Research White Paper Agency for Healthcare Research and Quality; 2013. Report No.: 13-EHC042-EF.
11. Baulieu EE, Robel P, Schumacher M. Neurosteroids: Beginning of the story. International Review of Neurobiology 2001;46:1–32.
12. Schumacher M, Akwa Y, Guennoun R, Robert F, Labombarda F, Desarnaud F, Robel P, De Nicola AF, Baulieu EE. Steroid synthesis and metabolism in the nervous system: Trophic and protective effects. Journal of Neurocytology 2000;29:307–326.
13. Lei B, Mace B, Dawson HN, Warner DS, Laskowitz DT, James ML. Anti-inflammatory effects of progesterone in lipopolysaccharide-stimulated bv-2 microglia. PLoS One 2014;9: e103969.
14. Luoma JI, Kelley BG, Mermelstein PG. Progesterone inhibition of voltage-gated calcium channels is a potential neuroprotective mechanism against excitotoxicity. Steroids 2011;76:845–855.
15. Atif F, Sayeed I, Ishrat T, Stein DG. Progesterone with vitamin d affords better neuroprotection against excitotoxicity in cultured cortical neurons than progesterone alone. Molecular Medicine 2009;15:328–336.
16. He L, Zhang X, Wei X, Li Y. Progesterone attenuates aquaporin-4 expression in an astrocyte model of ischemia/reperfusion. Neurochemical Research 2014;39:2251–2261.
17. Wang X, Zhang J, Yang Y, Dong W, Wang F, Wang L, Li X. Progesterone attenuates cerebral edema in neonatal rats with hypoxic-ischemic brain damage by inhibiting the expression of matrix metalloproteinase-9 and aquaporin-4. Experimental & Therapeutic Medicine 2013;6:263–267.
18. Chen G, Shi JX, Qi M, Wang HX, Hang CH. Effects of progesterone on intestinal inflammatory response, mucosa structure alterations, and apoptosis following traumatic brain injury in male rats. Journal of Surgical Research 2008;147:92–98.
19. Candolfi M, Jaita G, Zaldivar V, Zarate S, Ferrari L, Pisera D, Castro MG, Seilicovich A. Progesterone antagonizes the permissive action of estradiol on tumor necrosis factor-alpha-induced apoptosis of anterior pituitary cells. Endocrinology 2005;146:736–743.
20. Si D, Li J, Liu J, Wang X, Wei Z, Tian Q, Wang H, Liu G. Progesterone protects blood-brain barrier function and improves neurological outcome following traumatic brain injury in rats. Experimental & Therapeutic Medicine 2014;8:1010–1014.
21. Pascual JL, Murcy MA, Li S, Gong W, Eisenstadt R, Kumasaka K, Sims C, Smith DH, Browne K, Allen S, et al. Neuroprotective effects of progesterone in traumatic brain injury: Blunted *in vivo* neutrophil activation at the blood-brain barrier. American Journal of Surgery 2013;206:840–845; discussion 845–846.
22. Sayeed I, Stein DG. Progesterone as a neuroprotective factor in traumatic and ischemic brain injury. Progress in Brain Research 2009;175:219–237.
23. Deutsch ER, Espinoza TR, Atif F, Woodall E, Kaylor J, Wright DW. Progesterone's role in neuroprotection, a review of the evidence. Brain Research 2013;1530:82–105.
24. Garcia-Ovejero D, Gonzalez S, Paniagua-Torija B, Lima A, Molina-Holgado E, De Nicola AF, Labombarda F. Progesterone reduces secondary damage, preserves white matter, and improves locomotor outcome after spinal cord contusion. Journal of Neurotrauma 2014;31:857–871.
25. Santarsieri M, Niyonkuru C, McCullough EH, Dobos JA, Dixon CE, Berga SL, Wagner AK. Cerebrospinal fluid cortisol and progesterone profiles and outcomes prognostication after severe traumatic brain injury. Journal of Neurotrauma 2014;31: 699–712.
26. Yousuf S, Sayeed I, Atif F, Tang H, Wang J, Stein DG. Delayed progesterone treatment reduces brain infarction and improves functional outcomes after ischemic stroke: A time-window study in middle-aged rats. Journal of Cerebral Blood Flow and Metabolism 2014;34:297–306.
27. Kipp M, Amor S, Krauth R, Beyer C. Multiple sclerosis: Neuroprotective alliance of estrogen-progesterone and gender. Frontiers in Neuroendocrinology 2012;33:1–16.
28. Garay L, Gonzalez Deniselle MC, Sitruk-Ware R, Guennoun R, Schumacher M, De Nicola AF. Efficacy of the selective progesterone receptor agonist nestorone for chronic experimental autoimmune encephalomyelitis. Journal of Neuroimmunology 2014; 276:89–97.
29. Hussain R, El-Etr M, Gaci O, Rakotomamonjy J, Macklin WB, Kumar N, Sitruk-Ware R, Schumacher M, Ghoumari AM. Progesterone and nestorone facilitate axon remyelination: A role for progesterone receptors. Endocrinology 2011;152:3820–3831.
30. Kelley BG, Mermelstein PG. Progesterone blocks multiple routes of ion flux. Molecular and Cellular Neurosciences 2011;48: 137–141.
31. Linnertz R, Wurm A, Pannicke T, Krugel K, Hollborn M, Hartig W, Iandiev I, Wiedemann P, Reichenbach A, Bringmann A. Activation of voltage-gated na(+) and ca(2)(+) channels is required for glutamate release from retinal glial cells implicated in cell volume regulation. Neuroscience 2011;188:23–34.
32. Coughlan T, Gibson C, Murphy S. Progesterone, bdnf and neuroprotection in the injured cns. International Journal of Neuroscience 2009;119:1718–1740.
33. Frye CA, Koonce CJ, Walf AA. Progesterone, compared to medroxyprogesterone acetate, to c57bl/6, but not 5alpha-reductase mutant, mice enhances object recognition and placement memory and is associated with higher bdnf levels in the hippocampus and cortex. Neurosci Letters 2013;551:53–57.
34. Gonzalez SL, Labombarda F, Gonzalez Deniselle MC, Mougel A, Guennoun R, Schumacher M, De Nicola AF. Progesterone neuroprotection in spinal cord trauma involves up-regulation of brain-derived neurotrophic factor in motoneurons. Journal of Steroid Biochemistry & Moleccular Biology 2005;94:143–149.
35. Chan M, Chow C, Hamson DK, Lieblich SE, Galea LA. Effects of chronic oestradiol, progesterone and medroxyprogesterone acetate on hippocampal neurogenesis and adrenal mass in adult female rats. Journal of Neuroendocrinology 2014;26:386–399.
36. Mouihate A. Tlr4-mediated brain inflammation halts neurogenesis: Impact of hormonal replacement therapy. Frontiers in Cellular Neuroscience 2014;8:146. doi: 10.3389/fncel.2014.00146.
37. Barha CK, Ishrat T, Epp JR, Galea LA, Stein DG. Progesterone treatment normalizes the levels of cell proliferation and cell death in the dentate gyrus of the hippocampus after traumatic brain injury. Experimental Neurology 2011;231:72–81.
38. Bali N, Arimoto JM, Iwata N, Lin SW, Zhao L, Brinton RD, Morgan TE, Finch CE. Differential responses of progesterone receptor membrane component-1 (pgrmc1) and the classical progesterone receptor (pgr) to 17beta-estradiol and progesterone in hippocampal subregions that support synaptic remodeling and neurogenesis. Endocrinology 2012;153:759–769.
39. Li Z, Wang B, Kan Z, Zhang B, Yang Z, Chen J, Wang D, Wei H, Zhang JN, Jiang R. Progesterone increases circulating endothelial progenitor cells and induces neural regeneration after traumatic brain injury in aged rats. Journal of Neurotrauma 2012; 29:343–353.
40. Anderson GD, Farin FM, Bammler TK, Beyer RP, Swan AA, Wilkerson HW, Kantor ED, Hoane MR. The effect of progesterone dose on gene expression after traumatic brain injury. Journal of Neurotrauma 2011;28:1827–1843.
41. Guennoun R, Labombarda F, Gonzalez Deniselle MC, Liere P, De Nicola AF, Schumacher M. Progesterone and allopregnanolone in the central nervous system: Response to injury and implication for neuroprotection. Journal of Steroid Biochemistry & Moleccular Biology 2015;146:48–61.
42. Roglio I, Bianchi R, Gotti S, Scurati S, Giatti S, Pesaresi M, Caruso D, Panzica GC, Melcangi RC. Neuroprotective effects of dihydroprogesterone and progesterone in an experimental model of nerve crush injury. Neuroscience 2008;155:673–685.
43. O'Connor CA, Cernak I, Johnson F, Vink R. Effects of progesterone on neurologic and morphologic outcome following diffuse traumatic brain injury in rats. Experimental Neurology 2007;205: 145–153.
44. Lee RJ, Kim JK, Chao D, Kuo L, Mally A, McClean ME, Pemberton HE, Wilmington AR, Wong J, Murphy SP. Progesterone and allopregnanolone improves stroke outcome in male mice via

distinct mechanisms but neither promotes neurogenesis. Journal of Neurochemistry 2015;132:32–37.
45. Yousuf S, Atif F, Sayeed I, Tang H, Stein DG. Progesterone in transient ischemic stroke: A dose-response study. Psychopharmacology 2014;231:3313–3323.
46. Chang CM, Su YF, Chang CZ, Chung CL, Tsai YJ, Loh JK, Lin CL. Progesterone attenuates experimental subarachnoid hemorrhage-induced vasospasm by upregulation of endothelial nitric oxide synthase via akt signaling pathway. BioMed Research International 2014;2014:207616.
47. Won S, Lee JH, Wali B, Stein DG, Sayeed I. Progesterone attenuates hemorrhagic transformation after delayed tpa treatment in an experimental model of stroke in rats: Involvement of the vegf-mmp pathway. Journal of Cerebral Blood Flow and Metabolism 2014;34:72–80.
48. Morales T. Recent findings on neuroprotection against excitotoxicity in the hippocampus of female rats. Journal of Neuroendocrinology 2011;23:994–1001.
49. Liu S, Wu H, Xue G, Ma X, Wu J, Qin Y, Hou Y. Metabolic alteration of neuroactive steroids and protective effect of progesterone in alzheimer's disease-like rats. Neural Regeneration Research 2013;8:2800–2810.
50. Noorbakhsh F, Baker GB, Power C. Allopregnanolone and neuroinflammation: A focus on multiple sclerosis. Frontiers in Cellular Neuroscience 2014;8:134. doi: 10.3389/fncel.2014.00134.
51. Grimm A, Schmitt K, Lang UE, Mensah-Nyagan AG, Eckert A. Improvement of neuronal bioenergetics by neurosteroids: Implications for age-related neurodegenerative disorders. Biochimica et Biophysica Acta 2014;1842:2427–2438.
52. Spratt NJ, Tomkins AJ, Pepperall D, McLeod DD, Calford MB. Allopregnanolone and its precursor progesterone do not reduce injury after experimental stroke in hypertensive rats - role of postoperative temperature regulation? PLoS One 2014;9:e107752.
53. Gilmer LK, Roberts KN, Scheff SW. Efficacy of progesterone following a moderate unilateral cortical contusion injury. Journal of Neurotrauma 2008;25:593–602.
54. Wong R, Gibson CL, Kendall DA, Bath PM. Evaluating the translational potential of progesterone treatment following transient cerebral ischaemia in male mice. BMC Neuroscience 2014;15:131.
55. Wright DW, Kellermann AL, Hertzberg VS, Clark PL, Frankel M, Goldstein FC, Salomone JP, Dent LL, Harris OA, Ander DS, et al. Protect: A randomized clinical trial of progesterone for acute traumatic brain injury. Annals of Emergency Medicine 2007;49:391–402.
56. Hannay HJ, Exrachi O, Contant CF, Levin HS. Outcome measures for patients with head injuries: Report of the outcome measure subcommittee. Journal of Head Trauma Rehabilitation 1996:1141–1150.
57. Xiao G, Wei J, Yan W, Wang W, Lu Z. Improved outcomes from the administration of progesterone for patients with acute severe traumatic brain injury: A randomized controlled trial. Critical Care 2008;12:R61.
58. Wright DW, Yeatts SD, Silbergleit R, Palesch YY, Hertzberg VS, Frankel M, Goldstein FC, Caveney AF, Howlett-Smith H, Bengelink EM, et al. Very early administration of progesterone for acute traumatic brain injury. New England Journal of Medicine 2014;371:2457–2466.
59. Skolnick BE, Maas AI, Narayan RK, van der Hoop RG, MacAllister T, Ward JD, Nelson NR, Stocchetti N, the STI. A clinical trial of progesterone for severe traumatic brain injury. New England Journal of Medicine 2014;371:2467–2476.
60. Maas AI, Menon DK, Lingsma HF, Pineda JA, Sandel ME, Manley GT. Re-orientation of clinical research in traumatic brain injury: Report of an international workshop on comparative effectiveness research. Journal of Neurotrauma 2012;29:32–46.
61. Menon DK, Maas AI. Traumatic brain injury in 2014: Progress, failures and new approaches for tbi research. Nature Reviews Neurology 2015;11:71–72.
62. Giacino JT, Whyte J, Bagiella E, Kalmar K, Childs N, Khademi A, Eifert B, Long D, Katz DI, Cho S, et al. Placebo-controlled trial of amantadine for severe traumatic brain injury. New England Journal of Medicine 2012;366:819–826.
63. Zafonte RD, Bagiella E, Ansel BM, Novack TA, Friedewald WT, Hesdorffer DC, Timmons SD, Jallo J, Eisenberg H, Hart T, et al. Effect of citicoline on functional and cognitive status among patients with traumatic brain injury: Citicoline brain injury treatment trial (cobrit). Journal of the American Medical Association 2012;308:1993–2000.
64. Rivera SM, Gilman AG. Drug invention and the pharmaceutical industry. In: Brunton LL, editor. Goodman and Gilman's the pharmacological basis of therapeutics. 12th ed. New York:McGraw Hill; 2001. pp 3–16.
65. Goss CW, Hoffman SW, Stein DG. Behavioral effects and anatomic correlates after brain injury: A progesterone dose-response study. Pharmacology, Biochemistry and Behavior 2003;76:231–242.
66. Wali B, Ishrat T, Won S, Stein DG, Sayeed I. Progesterone in experimental permanent stroke: A dose-response and therapeutic time-window study. Brain 2014;137:486–502.
67. Schwamm LH. Progesterone for traumatic brain injury–resisting the sirens' song. New England Journal of Medicine 2014;371:2522–2523.
68. Ioannidis JP. Why most published research findings are false. PLoS Medicine 2005;2:e124.
69. Mak IW, Evaniew N, Ghert M. Lost in translation: Animal models and clinical trials in cancer treatment. American Journal of Translational Research 2014;6:114–118.
70. Gibson CL, Gray LJ, Bath PM, Murphy SP. Progesterone for the treatment of experimental brain injury; a systematic review. Brain 2008;131:318–328.
71. Hicks R, Giacino J, Harrison-Felix C, Manley G, Valadka A, Wilde EA. Progress in developing common data elements for traumatic brain injury research: Version two–the end of the beginning. Journal of Neurotrauma 2013;30:1852–1861.
72. Bell MJ, Kochanek PM. Pediatric traumatic brain injury in 2012: The year with new guidelines and common data elements. Critical Care Clinics 2013;29:223–238.
73. Stead LG, Bodhit AN, Patel PS, Daneshvar Y, Peters KR, Mazzuoccolo A, Kuchibhotla S, Pulvino C, Hatchitt K, Lottenberg L, et al. TBI surveillance using the common data elements for traumatic brain injury: A population study. International Journal of Emergency Medicine 2013;6:5. doi: 10.1186/1865-1380-6-5.
74. Green SM. Cheerio, laddie! Bidding farewell to the glasgow coma scale. Annals of Emergency Medicine 2011;58:427–430.
75. Retzios AD. Why do so many clinical trials fail? Issues in Clinical Research 2009:1–46. http://adrclinresearch.com/Issues_in_Clinical_Research_links/Why%20Pivotal%20Clinical%20Trials%20Fail%20-%20Part%201_v12L_a.pdf, accessed 12 May 2015..
76. Lu J, Murray GD, Steyerberg EW, Butcher I, McHugh GS, Lingsma H, Mushkudiani N, Choi S, Maas AI, Marmarou A. Effects of glasgow outcome scale misclassification on traumatic brain injury clinical trials. Journal of Neurotrauma 2008;25:641–651.
77. Lu J, Marmarou A, Lapane KL, Investigators I. Impact of gos misclassification on ordinal outcome analysis of traumatic brain injury clinical trials. Journal of Neurotrauma 2012;29:719–726.
78. Dunning D, Heath C, Suls J. Flawed self-assessment: Implications for health, education, and the workplace. Psychological Science in the Public Interest 2004;5:69–106.
79. McKee AC, Daneshvar DH. The neuropathology of traumatic brain injury. Handbook of Clinical Neurology 2015;127:45–66.
80. Katz DI, Cohen SI, Alexander MP. Mild traumatic brain injury. Handbook of Clinical Neurology 2015;127:131–156.
81. Pearce AJ, Hoy K, Rogers MA, Corp DT, Maller JJ, Drury HG, Fitzgerald PB. The long-term effects of sports concussion on retired australian football players: A study using transcranial magnetic stimulation. Journal of Neurotrauma 2014;31:1139–1145.
82. Covassin T, Schatz P, Swanik CB. Sex differences in neuropsychological function and post-concussion symptoms of concussed collegiate athletes. Neurosurgery. 2007;61:345–350; discussion 350–341.
83. Poon W, Vos P, Muresanu D, Vester J, von Wild K, Homberg V, Wang E, Lee TM, Matula C. Cerebrolysin asian pacific trial in acute brain injury and neurorecovery: Design and methods. Journal of Neurotrauma 2015;32:571–580.
84. Lingsma HF, Roozenbeek B, Li B, Lu J, Weir J, Butcher I, Marmarou A, Murray GD, Maas AI, Steyerberg EW. Large

between-center differences in outcome after moderate and severe traumatic brain injury in the international mission on prognosis and clinical trial design in traumatic brain injury (IMPACT) study. Neurosurgery 2011;68:601–607; discussion 607–608.
85. Roozenbeek B, Lingsma HF, Maas AI. New considerations in the design of clinical trials for traumatic brain injury. Clinical Investigations (London) 2012;2:153–162.
86. Price M, Hertzberg V, Wright DW. Does the sliding dichotomy result in higher powered clinical trials for stroke and traumatic brain injury research? Clinical Trials 2013;10:924–934.
87. Nichol AD, Higgins AM, Gabbe BJ, Murray LJ, Cooper DJ, Cameron PA. Measuring functional and quality of life outcomes following major head injury: Common scales and checklists. Injury 2011;42:281–287.
88. Shukla D, Devi BI, Agrawal A. Outcome measures for traumatic brain injury. Clinical Neurology & Neurosurgery 2011;113: 435–441.
89. Bagiella E, Novack TA, Ansel B, Diaz-Arrastia R, Dikmen S, Hart T, Temkin N. Measuring outcome in traumatic brain injury treatment trials: Recommendations from the traumatic brain injury clinical trials network. Journal of Head Trauma Rehabilitation 2010;25:375–382.
90. Alali AS, Vavrek D, Barber J, Dikmen S, Nathens AB, Temkin NR. Comparative study of outcome measures and analysis methods for traumatic brain injury trials. Journal of Neurotrauma 2015;32: 581–589.
91. Norris J. $17m DoD award aims to improve clinical trials for traumatic brain injury. San Francisco: University of California San Francisco; 2014. Available online at: http://www.ucsf.edu/news/2014/10/118596/us-aims-traumatic-brain-injury-clinical-trial-success, accessed 2 February 2014.
92. Saatman KE, Duhaime AC, Bullock R, Maas AI, Valadka A, Manley GT, Workshop Scientific T, Advisory Panel M. Classification of traumatic brain injury for targeted therapies. Journal of Neurotrauma 2008;25:719–738.
93. Meythaler JM, Brunner RC, Johnson A, Novack TA. Amantadine to improve neurorecovery in traumatic brain injury-associated diffuse axonal injury: A pilot double-blind randomized trial. Journal of Head Trauma Rehabilitation 2002;17:300–313.
94. Bhandari M, Giannoudis PV. Assessing patient outcomes: Pearls for clinical practice and research. Injury 2011;42:231.
95. Hong CM, Tosun C, Kurland DB, Gerzanich V, Schreibman D, Simard JM. Biomarkers as outcome predictors in subarachnoid hemorrhage–a systematic review. Biomarkers 2014;19:95–108.
96. Geschwind N. Problems in the anatomical understanding of the aphasias. In: Stein D, Rosen J, Butters N, editors. Plasticity and recovery of function in the central nervous system. New York: Academic Press; 1974. pp 467–508.
97. Luria AR. Higher cortical functions in man. New York: Basic Books; 1966.
98. Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases 2008. MIA Annual Symposium Proceedings 2008;2008:242–246.
99. Yin G. Clinical trial design: Bayesian and frequentist adaptive methods. Hoboken, NJ: Wiley; 2012.
100. He W, Pinero J, Kuznetsova OM, editors. Practical considerations for adpative trial design and implementation. Statistics for biology and health. New York: Springer; 2014.