Published in final edited form as:

Transl Stroke Res. 2016 August ; 7(4): 274–283. doi:10.1007/s12975-016-0463-9.

Embracing Biological and Methodological Variance in a New Approach to Pre-Clinical Stroke Testing

Thomas A. Kent^{1,2} and Pitchaiah Mandava^{1,2}

¹Stroke Outcomes Laboratory, Department of Neurology, Baylor College of Medicine, McNair Campus, 7200 Cambridge St. 9th Floor, MS: BCM609, Houston, TX 77030, USA

²Michael E. DeBakey VA Medical Center Stroke Program and Center for Translational Research on Inflammatory Diseases, Houston, TX, USA

Abstract

High-profile failures in stroke clinical trials have discouraged clinical translation of neuroprotectants. While there are several plausible explanations for these failures, we believe that the fundamental problem is the way clinical and pre-clinical studies are designed and analyzed for heterogeneous disorders such as stroke due to innate biological and methodological variability that current methods cannot capture. Recent efforts to address pre-clinical rigor and design, while important, are unable to account for variability present even in genetically homogenous rodents. Indeed, efforts to minimize variability may lessen the clinical relevance of preclinical models. We propose a new approach that recognizes the important role of baseline stroke severity and other factors in influencing outcome. Analogous to clinical trials, we propose reporting baseline factors that influence outcome and then adapting for the pre-clinical setting a method developed for clinical trial analysis where the influence of baseline factors is mathematically modeled and the variance quantified. A new therapy's effectiveness is then evaluated relative to the pooled outcome variance at its own baseline conditions. In this way, an objective threshold for robustness can be established that must be overcome to suggest its effectiveness when expanded to broader populations outside of the controlled environment of the PI's laboratory. The method is model neutral and subsumes sources of variance as reflected in baseline factors such as initial stroke severity. We propose that this new approach deserves consideration for providing an objective method to select agents worthy of the commitment of time and resources in translation to clinical trials.

Keywords

Pre-clinical; Animal model; Outcome modeling; Hyperglycemia; Infarct size; Nano anti-oxidants

Correspondence to: Thomas A. Kent.

Conflict of Interest TAK and PM are copyright holders of pPREDICTS. TAK is a co-inventor in a patent application related to PEG-HCCs and is an owner in shares of Acelerox, Inc, to pursue commercialization of nano-antioxidants.

Ethical Approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. This article does not contain any studies with human participants.

Introduction

Despite decades of research and tremendous progress in the understanding of the mechanisms of cell death and identification of several neuroprotectant drugs that thwart those mechanisms in pre-clinical stroke models, to date only treatments targeted at enhancing recanalization have shown clinical success [1]. With numerous neuroprotectant compounds tested and failed [2, 3], the inability to predict successful clinical outcomes from apparently positive pre-clinical and early-phase clinical studies is all too common. While there are likely several explanations for these failures, including lack of power [4, 5], rigor in design and execution, and a disconnection between the pre-clinical design and human trial [6, 7], we believe there is a fundamental problem in the identification of effective agents due to the way clinical and pre-clinical studies are designed and analyzed for disorders heterogeneous with respect to the myriad of factors that influence outcome, including innate biological, environmental, and methodological variance and experimental noise [8].

Failures in clinical translation are not unique to stroke, but have occurred in major trials in disorders such as Alzheimer's disease, amyotrophic lateral sclerosis, and traumatic brain injury [9–11]. A common theme of these failures is that treatments appeared to be effective in pre-clinical models and showed some apparent signals of efficacy in early-phase trials, yet go on to fail in larger phase 3 trials. These conditions also share in common that baseline factors, such as initial disease severity and complex subtypes, influence outcome but are frequently unevenly distributed between patient arms and in different settings.

Recognizing that imbalances in baseline factors can influence trial results, investigators frequently employ statistical methods to adjust or control for these factors, especially when investigating smaller subgroups. Most of these statistical methods are variations on multi-variate regression approaches. For their valid application, all these methods require overlapping, normally distributed factors that are related to each other in a linear fashion. However, our prior work indicated that these assumptions are almost never verified, and indeed even in the landmark National Institute of Neurological Disorders and Stroke (NINDS) recombinant tissue plasminogen activator (rt-PA) trial, the distribution of baseline factors did not meet these stringent assumptions [8]. Non-overlapping and non-linear relationship between baseline factors essentially invalidates the use of traditional multi-variate correction methods [12, 13]. In the case of rt-PA, we have shown using alternative analytical methods that the strength of effect of rt-PA overcame the influence of imbalances [14], but this has not been the case for the majority of neuroprotective agents. Indeed, we recently reported a contribution from the analytical approach to errors in the conclusions in 90 % of early-phase trials leading up to failed phase 3 [8]. Furthermore, methodological variation inherent in measurement of functional outcome can contribute as much as 25 % error to outcome depending on what method is used to quantitate [15]. It is highly unlikely that variability, noise, and the errors they impose will be evenly distributed in smaller pre-pivotal phase trials. As a result, if the uneven distributions of factors that influence outcome and errors in measurement by chance favor a better response in a treatment group, a false positive result emerges. A false negative result can occur if those errors favored better outcomes in the placebo arm. Statistical tests that compare the means and distributions of outcomes do not reflect underlying imbalances of the distribution of these factors.

Here, we contend that there are analogous issues in the current approach to pre-clinical stroke studies that hamper the ability to predict which agents will be reproducibly effective when tested outside of the controlled environment of the PI's laboratory or when expanded to different stroke models or patients. Efforts primarily focused on increasing experimental rigor, as in the 2009 Stroke Therapy Academic Industry Roundtable (STAIR) meeting and publication [16] and the Multi-*PART* consortium that mimics a phase 3 clinical trial scheme but in a pre-clinical setting [17–20], while clearly advances will, in our view, not be sufficient to identify an effective therapy when the treatment is extended to a larger, more heterogeneous population. There are too many sources of variability to anticipate and control. Also, a focus on homogeneity may make pre-clinical studies less translationally relevant [21]. Instead, we suggest here a dramatically different alternative with a new approach that, rather than attempt to eliminate variability, incorporates biological variability and noise into decision making.

pPREDICTS® as a Method to Quantify Variability

In prior work, the authors developed a novel analytical method for clinical trial decision making that generates an objective threshold that any new therapeutic must exceed in order to predict efficacy when extended from an early phase to a broader, more representative population in late-stage trials. This method addresses the additional sources of errors and non-random variation as the study extends outside of the controlled environment of the early-phase trial to a broader population [8]. Our clinical method, pPREDICTS (**p**ooled **P**lacebo **R**esponse **D**ICTates **T**reatment **S**uccess [22]), is applicable when baseline factors are predictive of outcome. For example, in stroke, initial severity explains approximately 80 % of functional outcome variance [22, 23]. There are additional contributions from the patient's age [24, 25] and whether the patient underwent recanalization therapy [1]. Baseline glucose is also highly associated with outcome, particularly when the patient has undergone recanalization [26], consistent between human and animal models [27]. Collaterals, some genetically determined, are also known to play a role in stroke outcomes in both humans and animals and can be altered by acquired conditions such as cardiovascular risk factors [28–32]. Uneven distribution of these factors among treatment arms can independently influence outcome beyond that achieved by the treatment under study.

The net result of the uneven distribution of these sources of biological, experimental, and methodological variance is the need for a larger strength of effect than would be anticipated from individual laboratory results [8, 33]. At present, granting agencies and other entities interested in translating a new treatment can only guess at the robustness of treatment effect that needs to be achieved to increase confidence that a treatment will be positive when its use is extended to other populations.

Notably, biological, methodological, and measurement variance may be a more general phenomenon in clinical studies. For example, a recent report indicated that only 39 % of 100 social psychology studies selected for their importance in the field were replicated despite adopting identical methods to the original study [34]. Indeed, several results were the opposite direction [34]. The only factor predictive of replication was robustness of the initial effect size (and not, for example, reputation of the original authors). We developed our

method specifically to generate an objective threshold for robustness that a treatment would need to overcome to raise confidence that the treatment can improve outcomes even when tested in a heterogeneous population.

pPREDICTS consists of a pooled outcome model generated from the placebo arms of randomized clinical trials (RCTs). We add the additional novel feature of multidimensional statistical or “prediction” intervals around this function providing a threshold by which a new therapy can be evaluated at its own baseline characteristics. Different than confidence intervals, prediction intervals take into account noise in a future observation as well as noise in generation of the minimized function and the variability of the data used to model the function [35, 36]. Confidence intervals pertain to function generated from available data and prediction intervals pertain to a future observation.

Our current pPREDICTS model is based on the control arms from 47 studies representing 11,000 patients. As the number of trials published increased, we were able to develop newer versions based on more specific subgroups of patients, such as differences in extent of rt-PA use [37] and intracranial hemorrhage (ICH) [38]. pPREDICTS is thus far successful in identifying that those early-phase clinical trials would go on to fail in phase 3 both retrospectively and prospectively and would have correctly predicted the two major positive stroke trials, intravenous rt-PA and endovascular stent retrievers [8]. The halting for futility of ATACH2 involving aggressive treatment of hypertension in acute ICH was also predicted by our models, as the prior ICH studies of aggressive hypertension management did not show any signal of efficacy in our model [38–40].

Clinical use of pPREDICTS is illustrated in Fig. 1a. The pPREDICTS middle surface shows the inverse relationship between the chances for a good functional outcome (z -axis: modified Rankin Scale of 0–2), baseline NIHSS (x -axis), and age (y -axis), with $p = .05$ prediction surfaces around the function that a treatment must overcome to be considered positive. To assess a study’s potential, outcomes are plotted at their respective baseline conditions, in this case, median NIHSS and age, and then visually inspected with respect to the outcome function (middle surface) and whether treatment outcome exceeds the $p = .05$ surfaces (Fig. 1b). Here, the NINDS trial phases 1 and 2 are superimposed on a pooled control model of patients who had not received rt-PA. The outcome of the treatment arms of the NINDS rt-PA trial [41] is superimposed on the model surfaces plotted at their baseline NIHSS and age (phase 1 and 2 involved different treatment time windows). Both treatment arm results exceed the $p = .05$ threshold indicating their effectiveness. Interestingly, the control arm results lie near to the pooled control arm outcomes, indicating that outcomes achieved in this trial are still relevant to modern stroke trials as all studies published since the NINDS trial are included in this model. As shown in Fig. 1c, both treatment phases of the failed SAINT trial of the spin-trap antioxidant NXY059 [42, 43] show no hint of improved outcome as treatment arm outcomes lie on the pooled control arm function at their baseline NIHSS and age [22]. This analysis suggests that SAINT-I was erroneously viewed as positive based on small differences between the treatment and control arms, but as shown here, those differences are well within the variance expected when a trial is extended to a broader population.

As conceived and applied here, this method, based on cumulative sources of variability around measurable factors that contribute to outcome, provides an objective threshold that a new therapy would need to exceed if it is expected to be effective when tested in a broader population. Had SAINT-I been analyzed in this method, it would have predicted futility of the agent, as ultimately shown in the larger, better balanced SAINT-II trial. In this paper, we suggest the development of an analogous method utilizing baseline data from pre-clinical stroke models to generate an expected outcome and generating quantitative thresholds for assessing the likelihood that a treatment is truly positive. If successful, this method could potentially predict the ultimate effectiveness of therapies, taking innate biological and methodological variance into account, and thus contribute greatly to improved decision-making and allocation of scarce pharmaceutical resources to develop new therapies.

Clinical Trial Failures as a Paradigm to Understand Pre-clinical Failures

While pre-clinical researchers have been discouraged at their failure to identify successful clinical therapeutics, they generally do not appreciate that most pilot clinical trials also have not predicted a positive result in late-phase/pivotal trials [8]. This analogy between failures in early clinical and pre-clinical phases extends to potential sources of variance. Table 1 indicates potential analogous sources of biological and methodological variability in the clinical and pre-clinical setting. Despite genetic homogeneity of rodent models, most preclinical investigators have already identified factors associated with outcomes, such as extent of cerebral blood flow (CBF) reduction during occlusion [45], temperature [46], occurrence of seizures [47], and others. Investigators recognize that even with permanent middle cerebral artery occlusion (MCAO), rats demonstrate considerable variance. To try to reduce variance in outcome, investigators commonly monitor CBF and exclude rats who do not achieve a specific threshold of reduction. Animals are fasted overnight to reduce peri-procedure glucose and the incidence of seizures. Temperature is regulated to the extent possible. However, the overall contribution of these maneuvers to reducing variability has not been quantified to our knowledge. By limiting heterogeneity, such as restricting the range of baseline glucose due to fasting, the important contribution of hyperglycemia to stroke outcome is ignored [48–53] and the effectiveness of a therapy in a hyperglycemic setting will not be known prior to translation to humans, where as many as one third of stroke patients are hyperglycemic on admission [54].

Initial post-procedure stroke severity, while often reported, is not explicitly associated with outcome other than to screen for success of the stroke model or stratify subjects. Initial stroke severity in humans is the most important predictor of outcome, and imbalances in this key factor has caused tremendous controversy in the clinical setting [14, 55–57]. With regards to functional testing, most investigators recognize that rodents must be “pre-tested” prior to stroke to see if they are motivated to participate in behavioral and motor testing or trained to perform complex tasks [58, 59]. While this pre-training likely reduces variability of functional measures, there is no definitive threshold at which all rodents have the same motivation and it can be expected that any functional differences are due to the stroke and or/treatment. Similarly in humans, many factors outside of stroke severity can influence outcome including depression [60], access and quality of rehabilitation, and adherence to secondary prevention [61–63]. With respect to outcome assessment, inter- and intra-rater

reliability for these measures has not been rigorously tested. In human studies, variability remains despite extensive training or central adjudication [64] and contributes considerable errors to outcome measures that require an increase in sample size to improve the chances that errors will be evenly distributed between study arms. To our knowledge, it is not known if there is similar variability in pre-clinical functional testing.

Looking at pre-clinical testing from this perspective leads us to conclude that it is assumed, but as far as we can tell, not tested formally, that incorporating methods such as overnight fasting and pre-testing for behavioral effort will be sufficient to generate homogeneous strokes so that any outcome differences observed among controls and treated animals will be due mostly to the treatment under study. We believe that such an assumption is not validated by the failure of these models to predict clinical efficacy. Moreover, emphasis on homogeneity may make these models even less relevant to the human condition. Lack of an objective tool to quantify the effect size hampers rational decision making regarding selection of agents to proceed through the pre-and clinical trial process.

A Proposal for a Pre-clinical Methodology that Incorporates Major Sources of Variability: pPREDICTS Pre-clinical

We developed pPREDICTS for the clinical setting as a new approach that embraces variances while requiring no adjustment for imbalances. By generating a continuous function relating outcome to baseline factors from a pooled population and adding “prediction” intervals, the effect size necessary for a reproducible result is obtained. We propose here consideration of a similar approach in the pre-clinical setting. One example (Fig. 2) of how baseline data not typically reported can be used to generate a threshold for outcome in the preclinical stroke setting can be seen from data generated in our laboratory in a 90-min transient suture occlusion of the MCAO in male adult rats [65]. To generate a spectrum of pre-stroke glucose levels and mimic the typical clinical trial situation, the pancreatic toxin streptozotocin was administered to some rats 2 days prior to increase of blood glucose levels [27], and then an effort was made to maintain glucose under 300 mg/dL post-operatively; otherwise, all pre- and post-operative care was identical. Baseline glucose is plotted vs infarct size (controlled for edema [66]), and a pPREDICTS function fitted to this data.

Detailed methods to generate the pPREDICTS function and predication intervals are published [22]. The model was developed by the following: (1) fitting a function to outcome proportions in multiple dimensions using a Matlab routine (nlinfit) and Freeman-Tukey modification of the arc sine square root of a proportion; (2) eliminating large outliers based on Studentized residuals; and (3) generating multidimensional prediction interval surfaces at the $p < .05$ level using a Jacobian matrix. Note that there are many potential scenarios by which baseline factors can be associated with outcome, including logistic/sigmoid, exponential, or linear functions.

By plotting baseline blood glucose vs infarct size (Fig. 2), we can see a highly correlated, sigmoid relationship that indicates that glucose explains considerable infarct size variance ($r^2 = .92$). Generating $p = .05$ prediction intervals around this function indicates the

threshold that an agent would need to exceed to overcome the variability not explained by this factor. Superimposed on this function is the mean $\pm$ confidence interval of infarct size in a group of streptozotocin-treated rats treated at the time of reperfusion with a novel high-capacity and broadly active carbon nanoparticle antioxidant, pegylated hydrophilic carbon clusters (PEG-HCCs) under investigation in the author's laboratory [68]. Infarct size at the treatment group's baseline glucose is smaller than predicted by the control function, suggesting that this treatment effectively reduces stroke size in a setting of hyperglycemia ($p = 0.003$, comparing PEG-HCC mean $\pm$ confidence intervals vs function mean $\pm$ prediction intervals at the same baseline glucose). We propose that this outcome model can be expanded to include other baseline factors such as initial stroke severity as determined either by post-operative neurological function or noninvasive imaging and a multi-dimensional model developed analogous to that seen in clinical pPREDICTS (Fig. 1). Different outcomes, such as percent that achieve a certain follow-up neurological endpoint [69, 70], analogous to the modified Rankin Scale could be employed to provide a range of potential outcomes.

A suggested paradigm for pre-clinical testing that could be adapted to this analytical approach is shown in Fig. 3. In bold are factors we consider the minimal to utilizing this method as an initial screen for a therapeutic agent. Because this approach in theory would be model-neutral, the other potential conditions and factors such as alternative stroke models as shown in phase 1 could serve to broaden the test of any agent to mimic what it might encounter in a clinical situation. This approach would be feasible as long as baseline factors such as post-stroke neurological deficit were associated with outcome. We propose that both pre-stroke and post-stroke measures, as illustrated in phases 2 and 3, could be subjected to modeling a multi-dimensional relationship between baseline factors and outcome, as well as prediction intervals generated to incorporate the cumulative sources of errors discussed above.

Testing of validity of the models

Due to paucity of work in this area, it is unclear if the models as conceived here express causal relationships between factors, although independent relationship is not a requirement as long as the model is useful in predicting outcomes [33]. Of considerable importance is the internal and external validity of this approach when broadened in its application [4, 71–73]. Poor internal validity has been identified as a major contributor to false positive results in pre-clinical studies [4, 5]. In our case, internal validity could be improved by including several laboratories in the initial model development and using our iterative pPREDICTS method to exclude extreme outliers based on objective a priori criteria [74]. External validity could be tested by developing the model on an initial cohort of data and then comparing this to a subsequent set. In the clinical stroke setting, pPREDICTS is continuously adapting to newer treatment approaches, and we envision in the preclinical setting that addition of different stroke models and conditions would contribute to a more broadly externally valid model. Of course, basic tenets of rigor as derived from STAIR, Multi-PART, and others would be essential to both types of validity.

If adopted by the larger stroke community, we anticipate that the approach proposed here could take a direction similar to what transpired for our clinical method [33]. Our first study employed a published outcome model based on a few sites, yet was able to correctly predict lack of efficacy of pre-stent retriever endovascular interventions [75] from early case series. As more studies were published and investigators provided us additional data, the model was expanded to stroke subtypes, populations, and treatments [1, 22, 37, 76]. We will encourage our colleagues to contribute similar data points obtained by them. Over time, a comprehensive pre-clinical pPREDICTS model may emerge that addresses a wider range of stroke models.

Potential uses of the pPREDICTS pre-clinical model

Individual investigators and granting agencies could employ the thresholds generated in this growing sample to identify the strength of effect a new therapy needs to achieve in order to suggest its benefit when extended to a clinical population. Investigators can superimpose outcomes of their study along with confidence intervals onto the pPREDICTS model. Whether their control group is representative of a broader population can be tested by superimposing the outcomes of their control group onto the pooled model (analogous to Fig. 1b). If outcome $\pm$ confidence intervals of their treatment arm are outside the 95 % prediction intervals of the pooled control arm, there is compelling reason to proceed to the next phase of investigation. Similar to any screening methods, there is a chance of missing a potentially effective therapy (type II error) whose robustness may not be great enough to exceed other sources of variance. However, this may be a more general phenomenon related to the complexity of the disease being studied. Methods such as ours could target more narrow conditions and encourage use of more objective, less error-prone outcomes, all with the goal of narrowing the variance that needs to be exceeded.

Since errors may also result from a non-representative control group outcome, investigators can check their control results compared to this pooled population which suggests that a larger control group's outcomes may drift toward the pooled outcomes. Funnel plots, substituting the pPREDICTS outcome function as the expected outcomes, can be generated to detect publication bias [5, 75].

This method could be used to obtain a sample size. If an investigator has conducted a single trial and knows the variance of their study, then variance of the pooled model at the same baseline independent variable can be utilized to obtain the sample size requirements for the next phase [35, 77]. This approach will obviate the need for making an assumption of equal variance. Power of a study can be determined in a similar manner [77].

Summary

High-profile failures in stroke clinical trials such as SAINT-II [43] discouraged major pharmaceutical and governmental investments in clinical translation of cellular protective agents in stroke. Explanations for failure focused on lack of power [4, 5], rigor in design and execution, and a disconnection between the pre-clinical design and human trial [6, 7]. An effort to address this problem included the (STAIR) meeting and subsequent publication [16]. STAIR has improved pre-clinical reproducibility but has not yet demonstrated

improved prediction. While we applaud these efforts, we do not believe them to be sufficient to ensure reproducible results when extended outside a few laboratories and models. We propose here to incorporate contributions from STAIR—e.g. blinding, randomization, careful record keeping, and reporting, from Multi-*PART* [4, 5, 78]—engaging multiple laboratories in investigating the same agent and extending these methods to include a consideration of the innate biological and methodological variability that this agent will ultimately encounter when translated to humans. We propose that this new approach deserves consideration as providing an objective measure of the robustness of treatment effect to aid in the selection of an agent worthy of the commitment of time and resources in definitive clinical trials.

Acknowledgments

Funding TAK received funding from NIH R21NS084290-02 and R01NS094535-01.

Computing resources of the Computational and Integrative Biomedical Research Center at Baylor College of Medicine were utilized for this study (PM).

References

1. Mandava P, Shah SD, Sarma AK, Kent TA. An outcome model for intravenous rt-PA in acute ischemic stroke. *Translat Stroke Res*. 2015; 6(6):451–7. DOI: 10.1007/s12975-015-0427-5
2. O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 1,026 experimental treatments in acute stroke. *Ann Neurol*. 2006; 59(3):467–77. DOI: 10.1002/ana.20741 [PubMed: 16453316]
3. Minnerup J, Wersching H, Schilling M, Schabitz WR. Analysis of early phase and subsequent phase III stroke studies of neuroprotectants: outcomes and predictors for success. *Exp Translat Stroke Med*. 2014; 6(1):2.doi: 10.1186/2040-7378-6-2
4. Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, et al. Where have all the rodents gone? The effects of attrition in experimental research on cancer and stroke. *PLoS Biol*. 2016; 14(1):e1002331.doi: 10.1371/journal.pbio.1002331 [PubMed: 26726833]
5. Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol*. 2010; 8(3):e1000344.doi: 10.1371/journal.pbio.1000344 [PubMed: 20361022]
6. Lapchak PA, Zhang JH, Noble-Haeusslein LJ. RIGOR guidelines: escalating STAIR and STEPS for effective translational research. *Translat Stroke Res*. 2013; 4(3):279–85. DOI: 10.1007/s12975-012-0209-2
7. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke*. 2008; 39(10):2824–9. DOI: 10.1161/STROKEAHA.108.515957 [PubMed: 18635842]
8. Kent TA, Shah SD, Mandava P. Improving early clinical trial phase identification of promising therapeutics. *Neurology*. 2015; 85(3):274–83. DOI: 10.1212/WNL.0000000000001757 [PubMed: 26109712]
9. Cudkovic ME, van den Berg LH, Shefner JM, Mitsumoto H, Mora JS, Ludolph A, et al. Dexamipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): a randomised, double-blind, phase 3 trial. *Lancet Neurol*. 2013; 12(11):1059–67. DOI: 10.1016/S1474-4422(13)70221-7 [PubMed: 24067398]
10. Robertson CS, Hannay HJ, Yamal JM, Gopinath S, Goodman JC, Tilley BC, et al. Effect of erythropoietin and transfusion threshold on neurological recovery after traumatic brain injury: a randomized clinical trial. *JAMA*. 2014; 312(1):36–47. DOI: 10.1001/jama.2014.6490 [PubMed: 25058216]

11. Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014; 370(4):322–33. DOI: 10.1056/NEJMoa1304839 [PubMed: 24450891]
12. Crager MR. Analysis of covariance in parallel-group clinical trials with pretreatment baselines. *Biometrics*. 1987; 43(4):895–901. [PubMed: 3427174]
13. Hair, JF., Black, WC., Babin, BJ., Anderson, RE., Tatham, RL. *Multivariate data analysis*. upper Saddle River, NJ: Pearson/Prentice-Hall; 2006. *Multivariate analysis of variance*.
14. Mandava P, Kalkonde YV, Rochat RH, Kent TA. A matching algorithm to address imbalances in study populations: application to the National Institute of Neurological Diseases and Stroke Recombinant Tissue Plasminogen Activator acute stroke trial. *Stroke*. 2010; 41(4):765–70. DOI: 10.1161/STROKEAHA.109.574103 [PubMed: 20167911]
15. Mandava P, Krumpelman CS, Shah JN, White DL, Kent TA. Quantification of errors in ordinal outcome scales using shannon entropy: effect on sample size calculations. *PLoS One*. 2013; 8(7):e67754.doi: 10.1371/journal.pone.0067754 [PubMed: 23861800]
16. Fisher M, Feuerstein G, Howells DW, Hurn PD, Kent TA, Savitz SI, et al. Update of the stroke therapy academic industry roundtable preclinical recommendations. *Stroke*. 2009; 40(6):2244–50. DOI: 10.1161/STROKEAHA.108.541128 [PubMed: 19246690]
17. Llovera G, Hofmann K, Roth S, Salas-Perdomo A, Ferrer-Ferrer M, Perego C, et al. Results of a preclinical randomized controlled multicenter trial (pRCT): Anti-CD49d treatment for acute brain ischemia. *Sci Transl Med*. 2015; 7(299):299ra121c.doi: 10.1126/scitranslmed.aaa9853
18. Multi-PART. A new paradigm for translational medicine. <http://www.dcn.ed.ac.uk/multipart/>. Accessed February 19 2016
19. Dirnagl, U. Thomas Willis Lecture. International Stroke Conference. 2016. https://my.americanheart.org/idc/groups/ahamam-public/@wcm/@sop/@scon/documents/downloadable/ucm_480867.pdf
20. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. *Lancet*. 2014; 383(9912):101–4. DOI: 10.1016/S0140-6736(13)62329-6 [PubMed: 24411643]
21. Richter SH, Garner JP, Wurbel H. Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nat Methods*. 2009; 6(4):257–61. DOI: 10.1038/nmeth.1312 [PubMed: 19333241]
22. Mandava P, Kent TA. A method to determine stroke trial success using multidimensional pooled control functions. *Stroke*. 2009; 40(5):1803–10. DOI: 10.1161/STROKEAHA.108.532820 [PubMed: 19286598]
23. Uchino K, Billheimer D, Cramer SC. Entry criteria and baseline characteristics predict outcome in acute stroke trials. *Stroke*. 2001; 32(4):909–16. DOI: 10.1161/01.STR.32.4.909 [PubMed: 11283391]
24. Nakayama H, Jorgensen HS, Raaschou HO, Olsen TS. The influence of age on stroke outcome. The Copenhagen Stroke Study. *Stroke*. 1994; 25(4):808–13. [PubMed: 8160225]
25. Rosen CL, Dinapoli VA, Nagamine T, Crocco T. Influence of age on stroke outcome following transient focal ischemia. *J Neurosurg*. 2005; 103(4):687–94. DOI: 10.3171/jns.2005.103.4.0687 [PubMed: 16266051]
26. Mandava P, Martini SR, Munoz M, Dalmeida W, Sarma AK, Anderson JA, et al. Hyperglycemia worsens outcome after rt-PA primarily in the large-vessel occlusive stroke subtype. *Translat Stroke Res*. 2014; 5(4):519–25. DOI: 10.1007/s12975-014-0338-x
27. Fabian RH, Kent TA. Hyperglycemia accentuates persistent “functional uncoupling” of cerebral microvascular nitric oxide and superoxide following focal ischemia/reperfusion in rats. *Translat Stroke Res*. 2012; 3(4):482–90. DOI: 10.1007/s12975-012-0210-9
28. Wang S, Zhang H, Wiltshire T, Sealock R, Faber JE. Genetic dissection of the Canq1 locus governing variation in extent of the collateral circulation. *PLoS One*. 2012; 7(3):e31910.doi: 10.1371/journal.pone.0031910 [PubMed: 22412848]
29. Moore SM, Zhang H, Maeda N, Doerschuk CM, Faber JE. Cardiovascular risk factors cause premature rarefaction of the collateral circulation and greater ischemic tissue injury. *Angiogenesis*. 2015; 18(3):265–81. DOI: 10.1007/s10456-015-9465-6 [PubMed: 25862671]

30. Menon BK, Smith EE, Coutts SB, Welsh DG, Faber JE, Goyal M, et al. Leptomeningeal collaterals are associated with modifiable metabolic risk factors. *Ann Neurol*. 2013; 74(2):241–8. DOI: 10.1002/ana.23906 [PubMed: 23536377]
31. Faber JE, Zhang H, Lassance-Soares RM, Prabhakar P, Najafi AH, Burnett MS, et al. Aging causes collateral rarefaction and increased severity of ischemic injury in multiple tissues. *Arterioscler Thromb Vasc Biol*. 2011; 31(8):1748–56. DOI: 10.1161/ATVBAHA.111.227314 [PubMed: 21617137]
32. Faber JE, Chilian WM, Deindl E, van Royen N, Simons M. A brief etymology of the collateral circulation. *Arterioscler Thromb Vasc Biol*. 2014; 34(9):1854–9. DOI: 10.1161/ATVBAHA.114.303929 [PubMed: 25012127]
33. Simard JM. Essentially all models are wrong, but some are useful. *Neurology*. 2015; 85(3):210–1. DOI: 10.1212/WNL.0000000000001769 [PubMed: 26109711]
34. Open SC. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*. 2015; 349(6251):aac4716.doi: 10.1126/science.aac4716 [PubMed: 26315443]
35. Process Modeling. NIST/SEMATECH e-Handbook of Statistical Methods. 2012. <http://www.itl.nist.gov/div898/handbook/toolaids/pff/pmd.pdf>. Accessed March 30 2014
36. Beers, K. Numerical methods for chemical engineering Applications in matlab. Cambridge: Cambridge University Press; 2007.
37. Mandava P, Martini SR, Shah VA, Fabian RH, Kent TA. Factors influencing differential outcomes in stent retriever trials: Comparison to a new outcome model based on percent utilization of rt-PA. *Stroke*. 2016; 47:AWMP1.
38. Mandava P, Samson Y, Kent TA. Modeling outcome after intracerebral hemorrhage indicates no benefit from lowering blood pressure. *Stroke*. 2015; 46:A95.
39. Anderson CS, Heeley E, Huang Y, Wang J, Stapf C, Delcourt C, et al. Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage. *N Engl J Med*. 2013; 368(25):2355–65. DOI: 10.1056/NEJMoa1214609 [PubMed: 23713578]
40. Antihypertensive Treatment of Acute Cerebral Hemorrhage (ATACH-II). 2015. This study has been suspended <https://clinicaltrials.gov/ct2/show/NCT01176565?term=ATACH&rank=2>
41. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med*. 1995; 333(24):1581–7. DOI: 10.1056/NEJM199512143332401 [PubMed: 7477192]
42. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, et al. NXY-059 for acute ischemic stroke. *N Engl J Med*. 2006; 354(6):588–600. DOI: 10.1056/NEJMoa052980 [PubMed: 16467546]
43. Shuaib A, Lees KR, Lyden P, Grotta J, Davalos A, Davis SM, et al. NXY-059 for the treatment of acute ischemic stroke. *N Engl J Med*. 2007; 357(6):562–71. DOI: 10.1056/NEJMoa070240 [PubMed: 17687131]
44. Chu PL, Keum S, Marchuk DA. A novel genetic locus modulates infarct volume independently of the extent of collateral circulation. *Physiol Genomics*. 2013; 45(17):751–63. DOI: 10.1152/physiolgenomics.00063.2013 [PubMed: 23800850]
45. Fan X, Qiu J, Yu Z, Dai H, Singhal AB, Lo EH, et al. A rat model of studying tissue-type plasminogen activator thrombolysis in ischemic stroke with diabetes. *Stroke*. 2012; 43(2):567–70. DOI: 10.1161/STROKEAHA.111.635250 [PubMed: 22052516]
46. Kim Y, Busto R, Dietrich WD, Kraydieh S, Ginsberg MD. Delayed postischemic hyperthermia in awake rats worsens the histopathological outcome of transient focal cerebral ischemia. *Stroke*. 1996; 27(12):2274–80. DOI: 10.1161/01.STR.27.12.2274 [PubMed: 8969793]
47. Uchino H, Smith ML, Bengzon J, Lundgren J, Siesjo BK. Characteristics of postischemic seizures in hyperglycemic rats. *J Neurol Sci*. 1996; 139(1):21–7. [PubMed: 8836968]
48. Hafez S, Coucha M, Bruno A, Fagan SC, Ergul A. Hyperglycemia, acute ischemic stroke, and thrombolytic therapy. *Translat Stroke Res*. 2014; 5(4):442–53. DOI: 10.1007/s12975-014-0336-z
49. Hafez S, Hoda MN, Guo X, Johnson MH, Fagan SC, Ergul A. Comparative analysis of different methods of ischemia/reperfusion in hyperglycemic stroke outcomes: interaction with tPA. *Translat Stroke Res*. 2015; 6(3):171–80. DOI: 10.1007/s12975-015-0391-0

50. Palomares SM, Sweet JG, Cipolla MJ. Acute rosiglitazone treatment during reperfusion after hyperglycemic stroke is neuroprotective not vascular protective. *Translat Stroke Res.* 2012; 3(3): 390–6. DOI: 10.1007/s12975-012-0190-9
51. Cipolla MJ, Godfrey JA. Effect of hyperglycemia on brain penetrating arterioles and cerebral blood flow before and after ischemia/reperfusion. *Translat Stroke Res.* 2010; 1(2):127–34. DOI: 10.1007/s12975-010-0014-8
52. Chen C, Cheng Y, Chen J. Transfection of Noggin in bone marrow stromal cells (BMSCs) enhances BMSC-induced functional outcome after stroke in rats. *J Neurosci Res.* 2011; 89(8): 1194–202. DOI: 10.1002/jnr.22662 [PubMed: 21544853]
53. Cui X, Chopp M, Zacharek A, Ye X, Roberts C, Chen J. Angiopoietin/Tie2 pathway mediates type 2 diabetes induced vascular damage after cerebral stroke. *Neurobiol Dis.* 2011; 43(1):285–92. DOI: 10.1016/j.nbd.2011.04.005 [PubMed: 21515377]
54. Bruno A, Williams LS, Kent TA. How important is hyperglycemia during acute brain infarction? *Neurologist.* 2004; 10(4):195–200. [PubMed: 15245585]
55. Hoffman JR, Schriger DL. A graphic reanalysis of the NINDS Trial. *Ann Emerg Med.* 2009; 54(3): 329–36. DOI: 10.1016/j.annemergmed.2009.03.019 [PubMed: 19464756]
56. Ingall TJ, O'Fallon WM, Asplund K, Goldfrank LR, Hertzberg VS, Louis TA, et al. Findings from the reanalysis of the NINDS tissue plasminogen activator for acute ischemic stroke treatment trial. *Stroke.* 2004; 35(10):2418–24. DOI: 10.1161/01.STR.0000140891.70547.56 [PubMed: 15345796]
57. Mandava P, Dalmeida W, Anderson JA, Thiagarajan P, Fabian RH, Weir RU, et al. A pilot trial of low-dose intravenous Abciximab and unfractionated heparin for acute ischemic stroke: translating GP IIb/IIIa receptor inhibition to clinical practice. *Translat Stroke Res.* 2010; 1(3):170–7. DOI: 10.1007/s12975-010-0026-4
58. Rosell A, Agin V, Rahman M, Moranchio A, Ali C, Koistinaho J, et al. Distal occlusion of the middle cerebral artery in mice: are we ready to assess long-term functional outcome? *Translat Stroke Res.* 2013; 4(3):297–307. DOI: 10.1007/s12975-012-0234-1
59. Schaar KL, Brenneman MM, Savitz SI. Functional assessments in the rodent stroke model. *Exp Translat Stroke Med.* 2010; 2(1):13. doi: 10.1186/2040-7378-2-13
60. Ayerbe L, Ayis SA, Crichton S, Wolfe CD, Rudd AG. Natural history, predictors and associated outcomes of anxiety up to 10 years after stroke: the South London Stroke Register. *Age Ageing.* 2014; 43(4):542–7. DOI: 10.1093/ageing/aft208 [PubMed: 24375225]
61. Micieli G, Cavallini A, Quaglini S, Guideline Application for Decision Making in Ischemic Stroke (GLADIS) Study Group. Guideline compliance improves stroke outcome: a preliminary study in 4 districts in the Italian region of Lombardia. *Stroke.* 2002; 33(5):1341–7. [PubMed: 11988613]
62. Levine DA, Neidecker MV, Kiefe CI, Karve S, Williams LS, Allison JJ. Racial/ethnic disparities in access to physician care and medications among US stroke survivors. *Neurology.* 2011; 76(1):53–61. DOI: 10.1212/WNL.0b013e318203e952 [PubMed: 21084692]
63. Ellis C, Boan AD, Turan TN, Ozark S, Bachman D, Lackland DT. Racial differences in poststroke rehabilitation utilization and functional outcomes. *Arch Phys Med Rehabil.* 2015; 96(1):84–90. DOI: 10.1016/j.apmr.2014.08.018 [PubMed: 25223490]
64. Quinn TJ, Dawson J, Walters MR, Lees KR. Variability in modified Rankin scoring across a large cohort of international observers. *Stroke.* 2008; 39(11):2975–9. DOI: 10.1161/STROKEAHA.108.515262 [PubMed: 18688002]
65. Longa EZ, Weinstein PR, Carlson S, Cummins R. Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke.* 1989; 20(1):84–91. [PubMed: 2643202]
66. McBride DW, Klebe D, Tang J, Zhang JH. Correcting for brain swelling's effects on infarct volume calculation after middle cerebral artery occlusion in rats. *Translat Stroke Res.* 2015; 6(4):323–38. DOI: 10.1007/s12975-015-0400-3
67. Quast MJ, Wei J, Huang NC, Brunder DG, Sell SL, Gonzalez JM, et al. Perfusion deficit parallels exacerbation of cerebral ischemia/reperfusion injury in hyperglycemic rats. *J Cereb Blood Flow Metab.* 1997; 17(5):553–9. DOI: 10.1097/00004647-199705000-00009 [PubMed: 9183293]
68. Samuel EL, Marciano DC, Berka V, Bitner BR, Wu G, Potter A, et al. Highly efficient conversion of superoxide to oxygen using hydrophilic carbon clusters. *Proc Natl Acad Sci U S A.* 2015; 112(8):2343–8. DOI: 10.1073/pnas.1417047112 [PubMed: 25675492]

69. Lapchak PA. Translational stroke research using a rabbit embolic stroke model: a correlative analysis hypothesis for novel therapy development. *Translat Stroke Res.* 2010; 1(2):96–107. DOI: 10.1007/s12975-010-0018-4
70. Belayev L, Khoutorova L, Atkins KD, Eady TN, Hong S, Lu Y, et al. Docosahexaenoic acid therapy of experimental ischemic stroke. *Translat Stroke Res.* 2011; 2(1):33–41. DOI: 10.1007/s12975-010-0046-0
71. Navarro, D. Learning Statistics with R A tutorial for psychology students and other beginners. Adelaide: University of Adelaide; 2014.
72. Slack MK, Draugalis JR. Establishing the internal and external validity of experimental studies. *Am J Health Syst Pharm.* 2001; 58(22):2173–81. [PubMed: 11760921]
73. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? *PLoS Med.* 2010; 7(3):e1000245.doi: 10.1371/journal.pmed.1000245 [PubMed: 20361020]
74. Pedhazur, ESL. Measurement, design, and analysis: an integrated approach. New York: Taylor & Francis; 1991.
75. Mandava P, Kent TA. Intra-arterial therapies for acute ischemic stroke. *Neurology.* 2007; 68(24): 2132–9. DOI: 10.1212/01.wnl.0000264898.55747.80 [PubMed: 17562835]
76. Mandava P, Shah SD, Sarma AK, Dalmeida D, Kent TA. A pooled analysis indicates the promise and pitfalls of improving upon intravenous rt-PA for ischemic stroke. *Stroke.* 2014; 45:AWP60.
77. Rosner, B. Estimation of sample size and power for comparison of two means *Fundamentals of biostatistics.* Boston: Brooks/Cole; 2011. p. 301-4.
78. Dirnagl U, Hakim A, Macleod M, Fisher M, Howells D, Alan SM, et al. A concerted appeal for international cooperation in preclinical stroke research. *Stroke.* 2013; 44(6):1754–60. DOI: 10.1161/STROKEAHA.113.000734 [PubMed: 23598526]

Fig. 1.

Clinical application of the pPREDICTS model. **a** The pPREDICTS model derived from the placebo arms of 47 trials representing >11,000 subjects. The *middle surface* shows the inverse relationship between the chances for a good functional outcome (z-axis: modified Rankin Scale of 0–2), baseline NIHSS (x-axis) and age (y-axis) in the linear portion of the curve. The *surfaces* above and below the *middle function* are $p = .05$ prediction surfaces around the function. These *surfaces* indicate the thresholds that a treatment must overcome to be considered positive at a study's own baseline parameters. **b** The positive 1995 NINDS

rt-PA trial shows that both treatment phases exceed that threshold for good functional outcome despite baseline imbalances (note that baseline NIHSS differs between the placebo and treatment arms, particularly in part 2). **c** Both clinical trial phases of the failed SAINT neuroprotectant trials show no hint of improved outcome as treatment arm results lie on the middle surface indicating no difference than expected from placebo. This result indicates that the slight improvement seen in the treatment arm of SAINT 1 was incorrectly interpreted as potentially effective, but instead was well within the variability expected in a large stroke sample. (Figure redrawn from Kent et al. Neurology 2015 [8])

Fig. 2.

pPREDICTS pre-clinical application to infarct size vs baseline glucose level in 90-min MCA occlusion/recanalization rat model. Data generated utilizing transient 90-min middle cerebral artery occlusion using the Longa suture model [65]. Infarct size (*y-axis*) determined histologically at 48–72 h, with correction for edema [66]. A range of blood glucose was achieved by treating some rats with streptozotocin 60 mg/kg IP, 2 days prior to stroke surgery [67]. Glucose at the time of initiation of the stroke procedure is shown in the *x-axis*. Results indicated a highly significant sigmoid relationship between infarct volume as glucose levels increase (*solid line*, $r^2=.92$; $p < .002$). pPREDICTS prediction intervals were generated around this function (*dashed line*) indicating the threshold that a treatment would have to exceed to show improved or worsened outcome when compared at baseline glucose. Superimposed on this model is the mean infarct volume $\pm$ confidence intervals (*red dot*: 111.9 mm^3 ; CI $58.7\text{--}165.1$) of six rats treated at the time of recanalization with the novel carbon nanoparticle antioxidant, pegylated hydrophilic carbon clusters under investigation in the author's laboratory [68] (p value for comparison vs pPREDICTS <0.003 by t test). Infarct size in this group is well below the $p = .05$ lower threshold of the prediction interval, indicating that this treatment appears effective even with a range of baseline glucose as would be expected in a clinical population

Fig. 3.

Proposed pre-clinical stroke model paradigm. Our proposed modifications to test effectiveness of a therapeutic agent in pre-clinical stroke models. Four phases are shown in *blue shade* and the specific procedures outlined in each section. Phases 2 and 3 illustrate key parameters that will permit the generation of a model to relate baseline conditions to outcome. Phase 1 indicates that, as long as baseline factors contribute to outcome, pPREDICTS could be model-neutral and incorporate the range of conditions likely to be encountered in a clinical population. Procedures in *bold* are suggested by the authors as minimal requirements to provide sufficient data to relate baseline conditions to outcome and define extent of biological and methodological variability. Additional procedures are listed that could enhance the model. Key measures such as baseline glucose and initial stroke severity would likely provide sufficient information to mimic variability expected in human trials and thereby provide an objective threshold that a treatment would need to exceed to be considered likely positive when translated to a larger population

Table 1

Potential sources of variability in stroke: analogies between clinical and pre-clinical

Factor	Clinical stroke studies	Pre-clinical animal models
Initial stroke severity	Ca. 80% of outcome variance explained by initial stroke severity. Frequent imbalances	Often obtained, but rarely quantitatively related to outcome
Stroke subtype	Influences recurrence rate, response to thrombolysis, presence of penumbra, propensity to hemorrhage	Likely similar to humans, but different stroke etiologies rarely considered
Age	Influences rate of recovery	Similar to humans
Sex	Influences stroke severity, subtype, response to therapy	Similar to humans
Physiology	Glucose: hemorrhage, mortality Temperature: stroke severity Blood pressure: hypotension	Similar to humans
Collaterals	Influences outcome from ischemia; modified by cardiovascular risk factors	Similar in humans
Genetic Factors	Neuroplasticity genes (e.g., BDNF) Stroke types are influenced by genetics [44]	Some evidence of influence of genetic factors; several genetic stroke models exist
Motivation for recovery	Depression and other behavioral and psychological factors influence stroke outcome in humans	Rats often pre-tested for motivation or training to complete behavioral testing; no single threshold exists to differentiate pre-stroke ability
Social factors	Access to high-quality acute stroke therapy and post-stroke follow-up Access to rehabilitation	Post-stroke conditions differ in different laboratories, e.g., climate, housing conditions, quality of post-op care
Errors in measures	Subjective outcome measures can add as much as 25 % error to outcome; requires larger sample size Errors in quantifying stroke imaging	Errors likely exist in functional assessment and in quantifying stroke volume, edema, perfusion/diffusion match, but rarely reported for a pre-clinical trial