



Published in final edited form as:

Stroke. 2023 February ; 54(2): 620–631. doi:10.1161/STROKEAHA.122.040638.

Embracing Heterogeneity in The Multicenter Stroke Preclinical Assessment Network (SPAN) Trial

Andreia Morais, PhD¹, Joseph J. Locascio, PhD², Lauren H. Sansing, MD^{3,4}, Jessica Lamb, BS⁵, Karisma Nagarkatti, MS⁵, Takahiko Imai, PharmD, PhD¹, Klaus van Leyen, PhD¹, Jaroslaw Aronowski, MD, PhD⁷, James I. Koenig, PhD⁸, Francesca Bosetti, PhD⁸, Patrick Lyden, MD^{5,6}, Cenk Ayata, MD, PhD^{1,9,*},
on behalf of the SPAN Investigators[#]

¹Department of Radiology, Harvard Medical School, Massachusetts General Hospital, Charlestown, MA, USA.

²Department of Biostatistics, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA.

³Department of Neurology, Yale University School of Medicine, New Haven, CT USA.

⁴Department of Immunobiology, Yale University School of Medicine, New Haven, CT USA.

⁵Department of Physiology and Neuroscience, Zilkha Neurogenetic Institute, Los Angeles, CA USA.

⁶Department of Neurology, Keck School of Medicine at USC, Los Angeles, CA USA.

⁷Department of Neurology, McGovern Medical School, University of Texas HSC, Houston, TX, USA.

⁸National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD USA.

⁹Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA.

Abstract

The Stroke Preclinical Assessment Network (SPAN) is a multicenter preclinical trial platform using rodent models of transient focal cerebral ischemia to address translational failure in experimental stroke. In addition to centralized randomization and blinding and large samples, SPAN aimed to introduce heterogeneity to simulate the heterogeneity embodied in clinical trials for robust conclusions. Here, we report the heterogeneity introduced by allowing the six SPAN laboratories to vary most of the biological and experimental model variables and the impact of

*Corresponding author. cayata@mgh.harvard.edu.

[#] Table S1

DISCLOSURES

Dr Lyden reports compensation from NIH Clinical Center for other services, and compensation from Apex Innovations for consultant services. Dr Ayata reports compensation from Neurelis, Inc. for other services.

SUPPLEMENTAL MATERIAL

Online Figures S1–S4, Tables S1–S6

this heterogeneity on middle cerebral artery occlusion (MCAo) performance. We included the modified intention-to-treat population of the control mouse cohort of the first SPAN trial (n=421) and examined the biological and procedural independent variables and their covariance. We then determined their impact on the dependent variables cerebral blood flow (CBF) drop during MCAo, time to achieve MCAo, and total anesthesia duration using multivariable analyses. We found heterogeneity in biological and procedural independent variables introduced mainly by the site. Consequently, all dependent variables also showed heterogeneity among the sites. Multivariable analyses with the site as a random effect variable revealed filament choice as an independent predictor of CBF drop after MCAo. Comorbidity, sex, use of LDF to monitor CBF, days after trial onset, and maintaining anesthesia throughout the MCAo emerged as independent predictors of time to MCAo. Total anesthesia duration was predicted by most independent variables. / We present with high granularity the heterogeneity introduced by the biological and model selections by the testing sites in the first trial of cerebroprotection in rodent transient filament MCAo by SPAN. Rather than trying to homogenize all variables across all sites, we embraced the heterogeneity to better approximate clinical trials. Awareness of the heterogeneity, its sources, and how it impacts the study performance may further improve the study design and statistical modeling for future multicenter preclinical trials.

INTRODUCTION

The translational failure of experimental acute stroke therapies from animal models to clinical trials has been widely recognized.¹⁻⁶ While the causes of translational failure are still debated, independent and unbiased confirmation of exploratory discoveries in large samples across multiple laboratories has been a critical step missing in the traditional therapeutic development pipelines.⁷⁻⁹ The NIH-funded Stroke Preclinical Assessment Network (SPAN), a large, multicenter preclinical trial platform in a rodent model of transient focal cerebral ischemia, was conceived and launched to address this unmet need.¹⁰

The overarching aim of SPAN is to model a clinical trial in a preclinical network and determine whether cerebroprotective interventions given at the time of or shortly after reperfusion can further improve functional outcomes compared to reperfusion alone. Besides eliminating systematic bias with centralized random allocation and blinding, and the enrollment of unprecedented sample sizes nearly two orders of magnitude higher than most preclinical stroke studies to date,¹¹⁻¹³ SPAN intends to simulate the degree of heterogeneity embodied in human trials in the hopes that an intervention robust enough to demonstrate efficacy in the context of heterogeneity would have a higher likelihood of showing benefit in human trials. Therefore, SPAN experimental protocols permitted several biological and experimental model variables to be varied by each site based on their preferences and prior experience. The ultimate consequences of such heterogeneity remain unknown in this first preclinical trial by SPAN, yet the large sample size allows for *post hoc* analyses of these decisions.¹⁰ Embracing the heterogeneity in biological and model variables could have far-reaching benefits on experimental performance and outcomes.¹⁴

We here aimed to present SPAN biological and model parameters with granularity and transparency, and statistically examine how such heterogeneity in biological and model

variables adopted by SPAN influenced the performance of the surgical middle cerebral artery occlusion (MCAo) procedure across different sites. We hope that these data and insight will help preclinical stroke researchers make informed decisions in their exploratory studies and guide future preclinical trial networks in stroke and other neurological and non-neurological diseases.

STUDY DESIGN AND METHODS

This manuscript adheres to the AHA Journals' implementation of the Transparency and Openness Promotion (TOP) Guidelines. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Study population

The SPAN is a randomized, controlled multicenter preclinical trial funded by the National Institutes of Neurological Disorders and Stroke (NINDS) to address critical rigor, transparency, and reproducibility issues in testing therapeutic cerebroprotective interventions in experimental focal cerebral ischemia.¹⁰ The trial included 4 stages. In Stage 1, only normal young (NY) animals were used. In Stages 2 and 3, aged mice (AG), mice with diet-induced obesity and hyperglycemia (DO), and spontaneously hypertensive rats were used, whereas stage 4 included only young healthy rats. Our study population included the modified intention to treat (mITT) population of male and female NY, AG and DO mice in the control cohort of SPAN (n=421). The mITT population was defined as subjects that received at least partial treatment with the planned intervention. Most mice were purchased from a vendor (n=397, 94%; C57B6), and only 24 (6%) were bred in-house, which was permitted by SPAN protocols. The vendors included The Jackson Lab (n=307, 77%), National Institute on Aging (NIA, n=88, 22%) and Charles River Lab (n=2, 1%).

Experimental sites and animal housing

Six SPAN research laboratories (i.e., sites; Augusta University, Johns Hopkins University, Massachusetts General Hospital, University of Iowa, University of Texas Health Science Center, Yale University; labeled as sites 1–6 in random order) and a coordinating center (CC, University of Southern California), selected through competitive NIH peer review, designed and implemented an endovascular filament MCAo protocol. Subjects were enrolled weekly and labeled using MRI-compatible bar-coded ear tags (RapID Tag, San Francisco). Sites submitted an intention to treat (ITT) form to the SPAN Database (Research Electronic Data Capture; REDCap). Subjects were then randomized to one of 6 intervention arms or their controls, stratified by laboratory, comorbidity, and sex. All sites maintained the subjects in an animal facility with a 12h/12h light/dark cycle, although the lights-on time varied among the sites (see below). All procedures were approved by the local Institutional Animal Care and Use Committee (IACUC) and reported in accordance with the ARRIVE guidelines. Full details of the SPAN protocols have been published, published and all standard operating procedures (SOPs) are available on the SPAN website.¹⁰

Experimental protocols and timeline

The first surgery day of the entire trial was considered day 0, and the dates of subsequent surgeries were expressed relative to day 0 (i.e., trial days 0–539). On the surgery day, the age and weight of the subjects were recorded. Subjects were anesthetized with isoflurane (4% induction, 2–2.5% maintenance in 70:30 N₂O:O₂), and body temperature was kept at 37±0.1°C. Bupivacaine (maximum 5 mg/kg) was injected subcutaneously along the anticipated ventral neck and scalp incision lines.

The decision to monitor the cerebral blood flow (CBF) during the MCAo procedure was left to the sites. Therefore, CBF was monitored during the MCAo procedure in only 298 subjects (71%) by placing a laser Doppler flow (LDF) probe on the skull over the MCA (1 mm posterior and 5 mm lateral to bregma). One site measured CBF before and after MCAo using full-field laser speckle flowmetry but did not use this information to guide their procedure; CBF data from this site was not included in the analyses because their measurement method was different, and the region of interest did not match the LDF location at other sites.

A silicone-coated filament (Doccol, MA) was inserted into the external carotid artery (ECA) and advanced through the internal cerebral artery (ICA) until MCAo. The filament choice was left to the sites. Across the trial, four filament silicone coating thicknesses were used (0.21, 0.22, 0.23, and 0.24 mm), and the silicone coating length was either 1–2, 2–3, or 4–6 mm (Figure S1). If CBF was monitored, the time of MCAo onset was recorded as the time of CBF drop (local time of the day). In the absence of CBF monitoring, the time of MCAo was based on the surgeon's judgment of successful filament placement. The circadian stage of the animal at the time of MCAo was recorded based on when lights were turned on at each site (0:00 zeitgeber time, ZT), either as inactive (0:00–12:00 ZT) or active (12:00–24:00 ZT). The total procedure duration from the anesthesia induction to MCAo (i.e., time to MCAo) was calculated for each subject. Target MCAo duration was 60 minutes. During the occlusion period, sites had the option to keep the subjects under anesthesia or awaken them and then re-anesthetize at reperfusion time.

Each subject was pre-allocated by the CC to one of three control interventions to receive an intraperitoneal injection (IP; 0.2 ml of 0.9% NaCl solution 5 minutes before reperfusion), intravenous infusion (IV; 0.2 ml of reconstituted 0.9% NaCl solution over 20 minutes via the jugular vein starting 5 minutes before reperfusion), or sham remote ischemic conditioning (RIC; non-inflated pressure cuffs placed on the hindlimbs for 40 minutes starting after reperfusion under continued anesthesia). The total anesthesia duration for each subject was also calculated, which varied based on the assigned control intervention. Warm physiologic saline or Ringer's lactate (1ml/100g subcutaneously) was administered at the end of the procedure to prevent dehydration. Post-MCAo care was per local IACUC regulations and sites' preferences and included daily checks and weight measurements for the first two days. Figure 1 shows the experimental timeline. Stratified randomization was used to enroll males and females equally, and enrollment of a particular comorbid model was restricted to 2-week blocks assigned to the sites by the CC. All other variables were left to the sites to choose from based on their preferences.

Statistical Analyses

All data were prospectively entered to the Research Electronic Data Capture platform (REDCap; <https://www.project-redcap.org>)^{15,16} by the experimenters. We captured 12 *independent variables* for each subject determined prior to the MCAo procedure either by the CC or by the site (Figure 1, in red). Among these independent variables, *biological* variables included sex, age, weight, comorbidity (normal young, NY; diet-induced obese, DO; aged, AG), and circadian stage at the time of MCAo. *Procedural* variables included control intervention, endovascular filament silicone tip coating thickness and length, use of CBF monitoring to guide the MCAo procedure, maintaining animals under anesthesia until after reperfusion versus allowing them to awaken after the insertion of endovascular filament, the date on which MCAo surgery was performed, and time of MCAo surgery. In addition, we captured three *observed (i.e., dependent) outcome variables*, including time to MCAo, total anesthesia duration, and cerebral blood flow (CBF) drop during MCAo (Figure 1, in blue).

All statistical analyses were performed using SAS Studio (Version 9.0401M6, SAS Institute Inc., North Carolina). After generating descriptive statistics, we examined bivariate associations among all variables using the χ^2 test of association for two categorical variables, parametric tests such as general linear models (GLM) for a categorical vs. numeric variable or numeric vs. numeric variables, with a linear and quadratic term for the predictor variable to determine the best fit. Reported p values were adjusted for multiplicity by false discovery rate (FDR).

Given the high degree of covariance among many independent variables, we employed multivariable analyses for each dependent variable using a mixed model. The initial model included all independent variables as fixed-effect predictors and the site as a random variable. We then used stepwise backward elimination (cutoff $p < 0.05$) to remove variables without significant relations after adjusting for other predictors. The p values in the final multivariate models were adjusted for multiplicity using stepdown Sidak.

RESULTS

Basic characteristics of the study population and the biological and procedural variables

A total of 436 mice were randomized to the control arms throughout the network. Only fifteen animals were excluded due to procedural failures before receiving the intended intervention. The remaining animals (n=421) formed the mITT population used for this study (Tables S2 and S3, Figure S2). Enrollment was relatively even across the sites (Table S2) and steady throughout the 539 days of the SPAN trial during which mice were enrolled (Figure S2), except for national holidays and two planned enrollment breaks during the trial for interim analyses. Stratified randomization ensured equal numbers of males and females (Table S2). Age showed a bimodal distribution given due to enrollment of a separate aged mouse cohort (range 2.0–18.6 months), whereas body weight (range 16.7–57.5 grams) showed a relatively normal and wide distribution (Table S3, Figure S2). The control cohort was evenly split among IP vehicle, IV vehicle, and the sham intervention for remote ischemic conditioning (Table S2). Surgery was performed between 5:40 and 19:47

hours local time (Table S3, Figure S2). A third of the animals were maintained in reverse light-dark cycle rooms for at least two weeks prior to the MCAo procedure to reverse the circadian rhythm. Hence, one-quarter of the cohort had their MCAo procedure during their active (i.e., dark) circadian stage (Table S2, Figure S2). Three sites did not use LDF during MCAo to guide the procedure in a subset of their animals, amounting to 29% of the study cohort (Table S2). Only 21% of the animals were kept under anesthesia during 60-minute ischemia throughout the network (Table S2). Although the use of short (1–2 mm) versus long (4–6 mm) silicone length of the endovascular filament was relatively even (55% vs. 38%), silicone thickness of 0.22 mm (50%) was favored over 0.21 and 0.23 mm across the network (Table S2). The observed (i.e., dependent) variables time to MCAo and anesthesia duration showed a normal distribution ranging from 7–50 and 11–160 minutes, respectively (Table S3, Figure S2). In contrast, CBF changes during the MCAo procedure showed a skewed distribution, with most animals showing at least a 70% reduction (Table S3, Figure S2). Altogether, these data showed that the network achieved a fair degree of biological and procedural heterogeneity.

Biological and procedural heterogeneity among study sites

Because investigators were allowed to vary many biological and procedural variables, the site covaried with most variables with the exception of sex, which was stratified during randomization, and control intervention, which was assigned randomly by the CC (Figures 2 and 3, Table S4). Notable site heterogeneities included age, time of day MCAo was performed, choice of filament, keeping the animal under anesthesia during the occlusion period, CBF monitoring to guide the MCAo, and circadian stage of the animal at the time of MCAo. As a result, the time it takes to achieve MCAo after anesthesia onset (i.e., time to MCAo), total time subjects spent under anesthesia (i.e., anesthesia duration), and perhaps more importantly, the CBF drop after MCAo significantly varied among the sites (Figure 2A).

Temporal trends in biological and procedural variables

Several biological and procedural variables were unevenly distributed throughout the course of the trial in part due to modifications made to the trial design by the steering committee and in part due to temporal tendencies (Figure S3). For example, the diet-induced obesity (DO) and aged (AG) animal models were adopted instead of normal young (NY) mice, the RIC control intervention arm was eliminated, and the practice of continuous anesthesia until the end of reperfusion was abandoned. As the trial progressed, the body weight and age distribution reflected the adoption of DO and AG models, which also prompted a change in filament choice across the sites. Lastly, CBF monitoring gained slight favor over time. These temporal trends further added to the biological and procedural heterogeneity in SPAN.

Covariance among the biological and procedural variables

Further bivariate analyses showed numerous associations among most other independent variables (Figures 4 and 5, Table S5). Male, AG and DO mice were significantly heavier, consistent with known growth curves (Figure 3A). Thicker filaments were preferred by the surgeons in heavier animals (Figure 3A), as recommended by the manufacturer and in the SPAN standard operating procedures, and therefore, more often used in males, older

animals, and DO mice (Figures 4B and 5A). Body weight and age covaried with each other (Figure 3C) as well as with the choice of maintaining the animals under anesthesia during MCAo (Figure 3A, B), likely reflecting the decision to simultaneously abandon the latter practice and adopt AG and DO during the later stages of the trial (Figure 4D). Control interventions were unevenly distributed among the comorbidity models because RIC was dropped after Stage 1 (Figure 4C). The filament length of 2–3 mm was only used in AG mice, whereas the other lengths were used relatively evenly in all three models (Figures 4B and 5B). The AG mice were less likely to have CBF monitoring during MCAo (Figures 4B and 5E). Filament thickness and length were also associated with the use of CBF monitoring to guide surgery, continuing anesthesia throughout MCAo, and the circadian stage at MCAo (Figure 4A, B). Time of MCAo during the day also covaried with many other variables (Figure 3D). These apparent associations likely reflected the site's biological and procedural choices, as well as changes in the standard operating procedures in successive stages of the trial, which created strong covariance among the independent variables (Figure S4).

Multivariable analyses

We next sought to examine the effect of the heterogeneity in the independent variables on the dependent variables CBF reduction, surgery duration (i.e., time to MCAo), and total anesthesia duration. Strong covariances among the independent biological and procedural variables obligated us to develop a multivariable model for each dependent variable by including fixed predictors of sex, circadian stage, age, and weight at the time of MCAo, control intervention, endovascular filament silicone tip coating thickness and length, use of CBF monitoring using LDF signal drop to guide the MCAo procedure, and the number of days from the start of the trial at the time of MCAo. Two-way interactions among the comorbidity model, circadian stage, age, and weight at the time of MCAo were added as fixed effects. Quadratic terms of continuous numeric variables were also included. The site was introduced in the model as a random effect variable. After stepwise backward elimination of the least significant variables from the initial multivariate model until all remaining variables were significant, several independent predictors emerged for each dependent variable in the final model (Figure 5, Table 1, Table S6).

The use of thicker and longer filaments independently predicted more severe CBF drops during MCAo (Figure 5A, Table 1). The 0.23 mm filament thickness resulted in 12% and 8% more CBF drop compared with 0.21 and 0.22 mm, respectively, whereas the 1–2 mm filament length resulted in 20% and 16% less CBF drop compared with the 2–3 and 4–6 mm, respectively. No other variable independently predicted the CBF drop. The fixed effects of the initial and final multivariable models explained 29% and 18% of the variation in CBF drop, respectively, which increased to about 45% when the site was included (Table S6), suggesting a strong influence by other site-associated factors (e.g., surgeon).

Surgery time to MCAo was predicted by several variables (Figure 5B, Table 1), including sex and comorbidity. Males required 3 minutes longer surgery time than females to achieve MCAo, whereas comorbid animal models AG and DO increased the surgery time by about 9 minutes compared with NY. The decision to use LDF to guide the MCAo added almost 5 minutes, and the decision to keep the animals under anesthesia during MCAo added 8

minutes to surgery time until MCAo was achieved. In addition, surgery time to MCAo decreased as the trial progressed and was several minutes shorter between the beginning and the end of the trial. The fixed effects of the initial and final multivariate models explained approximately 37% and 16% of the variation in time to MCAO, which increased to more than 50% when the site was included (Table S6), once again suggesting a strong influence by the site.

The total duration of exposure to anesthesia had the highest number of predictors (Figure 5C, Table 1). As expected, keeping the animals under anesthesia until reperfusion added 51 minutes more anesthesia exposure. Interestingly, the use of 4–6 mm filament length independently predicted longer anesthesia exposure by 18–21 minutes. The decision to use LDF to monitor CBF added 5 minutes, and IV and RIC control interventions prolonged the anesthesia duration by 16 and 38 minutes, respectively, compared with IP, by the design of the experiment. Anesthesia duration was longer by about 3 minutes in females compared with males. The circadian stage appeared to have a modest effect as well; the active stage was associated with 5 minutes shorter anesthesia duration than the inactive stage. As with surgery duration, anesthesia duration slightly decreased as the trial progressed, and was several minutes shorter between the beginning and the end of the trial. The multivariate model explained more than 90% of the variation in anesthesia duration (Table S6).

DISCUSSION

The data presented here address several crucial issues that impact the pre-clinical assessment of putative treatments. We studied stroke, but our results inform the design of any pre-clinical assessment of candidate interventions. While clinical trial design and performance have been debated and improved over decades,^{17,18} the concept of multi-center, centrally randomized, and blinded pre-clinical trials is relatively new, and design principles are not well established.^{12,13,19} Although it is generally agreed that preclinical trials should try to follow as closely as possible clinical trial methodology,²⁰ it is unclear whether and how preclinical trials can simulate the heterogeneity present in clinical trials, such as age, comorbidities, and stroke severity at onset, as well as unavoidable variations in local clinical practice.^{21–23} Moreover, whether and how introducing this degree of heterogeneity in a preclinical trial might affect the performance and analyses is not known. Therefore, our primary aim in this manuscript was to lay bare the heterogeneity in the independent biological and procedural variables in the first SPAN trial, to allow the readers of future SPAN manuscripts put this heterogeneity in context.

Heterogeneity is a natural consequence of a multicenter preclinical trial network since it would be virtually impossible to achieve identical personnel, equipment, housing, and environmental factors in multiple laboratories. The SPAN investigators embraced further heterogeneity by allowing sites to vary several biological and model variables (Figure 1). As a result, the site (i.e., variable representing the 6 research laboratories) significantly predicted 9 out of 12 independent variables recorded in our study, and more than half of pairwise associations among all biological and model variables were statistically significant, underscoring the degree of covariance introduced by site heterogeneity (Figure S4). The latter obligated us to account for the contribution of the site as a random effect to

determine the independent effects of biological and model variables on the dependent model performance variables CBF drop, time to MCAo, and total anesthesia duration.

Our secondary aim was to run a preliminary analysis of how this heterogeneity might affect the technical performance of the stroke model. We chose the performance metrics CBF drop, time to MCAo, and total anesthesia duration as dependent outcome variables that were captured as part of the SPAN database of common data elements and have been reported to affect functional and morphological outcomes.^{24–26} Multivariable analyses, with site introduced as a random effect variable, revealed several independent predictors of these model performance metrics. For example, longer and thicker filaments led to a larger CBF drop as well, presumably due to more complete occlusions.²⁷ Longer filaments are more likely to have occluded the posterior cerebral artery origin as well, eliminating the variability due to the presence or absence of a posterior communicating artery in the C57BL6 mouse strain²⁸ used in SPAN.¹⁰ The *a priori* decision to monitor CBF by placing an LDF probe to guide the surgery naturally prolonged the procedure, increasing the time to MCAo and anesthesia duration. Comorbidities AG and DO further increased the time to MCAo, possibly reflecting a slow and deliberate approach by the surgeon in these valuable animal models. Conversely, both the surgery duration (i.e., time to MCAo) and anesthesia duration decreased as the trial went on, probably reflecting the surgeons' improving facility with the procedure (i.e., learning curve). Further validating the multivariable model, the *a priori* decision to keep the animals under anesthesia until reperfusion, and the type of control intervention, strongly predicted total anesthesia duration, as expected by the study design. Although not our primary intent, these independent associations provided a deeper practical and procedural insight to help with future study designs.

Of course, the sources of heterogeneity in SPAN do not fully recapitulate those in clinical trials. Nevertheless, SPAN introduced the biological variables sex, age, weight, comorbidities, and circadian stage at MCAo, which (with the exception of sex) greatly differed among the sites (Figure 2, Table S4). Moreover, the use of various lengths and thicknesses of endovascular filaments among the sites (Table S4) might have introduced anatomical differences in infarct location and volume (e.g., longer filaments are more likely to compromise the posterior cerebral artery, leading to hippocampal and/or thalamic infarction). We might also argue that the differences among the sites (and even within a single site) in the time of the day that the MCAo was performed, and anesthesia during MCAo, could to some extent mimic the heterogeneity in clinical trials. Lastly, reperfusion success was not monitored or used as an inclusion criterion; therefore, a subset of animals might not have achieved successful recanalization, mimicking another source of heterogeneity in clinical trials. Besides these biological sources of heterogeneity, we do believe the heterogeneity in local resources and practices (i.e., performance variability) also contributes to the heterogeneity in clinical trials, as previously reported [ref]. Therefore, we believe the sources of heterogeneity in SPAN reasonably approximated those in clinical trials. Importantly, heterogeneity introduced in SPAN was not entirely intentional. In fact, it would be virtually impossible to standardize every possible variable (e.g., surgeons, equipment, environment) that might influence the outcomes.

It is also important to note that whether such heterogeneity in a preclinical trial indeed improves its translational predictive value is not yet known. The question of whether a trial platform should standardize all technical and procedural variables, as opposed to allowing the sites to choose what they are most comfortable with, is up for debate and will be so until a good number of such trials adopt one or the other approach and their translational predictive value are compared. However, given the relatively low sensitivity of clinical outcome assessment tools, it is likely that large effect sizes (e.g., endovascular therapy)^{29,30} are needed to overcome the inherent noise in clinical trials, and that any experimental therapy with a smaller effect size will not succeed even if preclinical data from a single site are promising. Indeed, strong standardization in preclinical studies may underlie poor reproducibility in a multicenter network.^{31,32} The efficacy of an intervention should not depend on the standard operating procedures of the platform they were tested in. An intervention that only works in one testing platform and fails in another would be considered less robust. If the treatment effects are real, they should be reproducible by other multi-center consortia, regardless of the magnitude and sources of heterogeneity, obviating the need for strict standardization.

Therefore, embracing biological and methodological heterogeneity might better mimic clinical trials and improve the predictive value of preclinical testing.^{7,31,33} This has been convincingly demonstrated in simulations using data from single or multiple laboratory studies, which suggested that highly standardized single laboratory studies led to poor reproducibility and predictive value and that introducing heterogeneity in study samples improved reproducibility.³⁴ However, until SPAN, this focus on heterogeneity has not been put into practice. SPAN trial included biological (e.g., sex, age, comorbidities) and procedural (e.g., filament characteristics, anesthesia, CBF monitoring) heterogeneity. Allowing biological and procedural heterogeneity in a pre-clinical trial requires large sample sizes, which can only be accomplished in a multicenter network such as SPAN³³ and Multi-PART.¹³

In summary, this is the first detailed report on the predicted as well as unintended biological and model heterogeneity in the filament MCAo model and their impact on performance metrics in a multicenter preclinical trial of cerebroprotection in ischemic stroke. Our data suggest that the heterogeneity in the SPAN stems mainly from biological and model variations by the testing sites and must be accounted for in statistical models designed for therapeutic testing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was conducted with support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award UL1TR002541) and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University, and its affiliated academic healthcare centers, or the National Institutes of Health.

SOURCES OF FUNDING

National Institutes of Health funding to Massachusetts General Hospital (U01NS113443, Dr. Ayata), Yale University (U01NS113445, Dr. Sansing), University of Texas (U01NS113451, Dr. Aronowski), the University of Southern California (U24NS113452, Dr. Lyden), and The Laboratory of Neuro Imaging Resource (LONIR) at the University of Southern California (P41EB015922).

APPENDIX: SPAN Investigators

Francesca Bosetti, PhD, James I. Koenig, PhD, Patrick D. Lyden, MD, Jessica Lamb, BS, Karisma Nagarkatti, MS, David C. Hess, MD, Pradip K. Kamat, PhD, Mohammad Badruzzaman Khan, PhD, Krishnan Dhandapani, PhD, Ali S. Arbab, MD, PhD, Shahneela Siddiqui, MS, Cameron Smith, MS, Mohammad Nisar, BS, Enrique C. Leira, MD, MS, Anil K. Chauhan, PhD, Nirav Dhanesha, PhD, Rakesh B. Patel, PhD, Mariia Kumskova, PhD, Daniel Thedens, PhD, Kai Wang, PhD, Cenk Ayata, MD, PhD, Andreia Morais, PhD, Takahiko Imai, PharmD, PhD, Tao Qin, Xuyan Jin, PhD, Taylan Denis Erdogan, BS, Lili Yu, MD, Joseph B. Mandeville, PhD, William Taylor Kimberly, MD, PhD, Jonah Patrick Weigand Whittier, BS, Eng Lo, PhD, Ken Arai, PhD, Klaus Van Leyen, PhD, Lauren H. Sansing, MD, Fahmeed Hyder, PhD, Jelena M. Mihailovic, PhD, Basavaraju G. Sanganahalli, PhD, Sebastian Diaz-Perez, BS, Sofia E. Velazquez, BS, Hannah E. Beatty, BS, Conor Johnson, BS, Alison L. Herman, BA, Ligia S. B. Boisserand, PhD, Emma Immakavar, BS, Raymond C. Koehler, PhD, Ted Dawson, PhD, Valina Dawson, PhD, Yanrong Shi, MD, MS, Brooklyn Avery, BS, Steven Lannon, BS, Adnan Bibic, PhD, Kazi Akhter, PhD, Senthilkumar S. Karuppagounder, PhD, Jaroslaw Aronowski MD, PhD, Louise D. McCullough, MD, PhD, Lidiya Obertas, PhD, Andrew Goh, MS, Shuning Huang, PhD, Anjali Chauhan, PhD.

NON-STANDARD ABBREVIATIONS AND ACRONYMS

AG	aged mice
CBF	cerebral blood flow
CC	coordinating center
DO	diet-induced obesity and hyperglycemia
FDR	false discovery rate
IP	intraperitoneal injection
IV	intravenous infusion
LDF	laser Doppler flowmetry
MCAo	middle cerebral artery occlusion
mITT	modified intention to treat
NY	normal young mice
RIC	remote ischemic conditioning

SPAN The Stroke Preclinical Assessment Network

REFERENCES

1. Savitz SI, Fisher M. Future of neuroprotection for acute stroke: in the aftermath of the SAINT trials. *Ann Neurol.* 2007;61:396–402. [PubMed: 17420989]
2. Hill MD, Goyal M, Menon BK, Nogueira RG, McTaggart RA, Demchuk AM, Poppe AY, Buck BH, Field TS, Dowlatshahi D, et al. Efficacy and safety of nerinetide for the treatment of acute ischaemic stroke (ESCAPE-NA1): a multicentre, double-blind, randomised controlled trial. *Lancet.* 2020;395:878–887. doi: 10.1016/S0140-6736(20)30258-0 [PubMed: 32087818]
3. Lyden P, Pryor KE, Coffey CS, Cudkowicz M, Conwit R, Jadhav A, Sawyer RN Jr., Claassen J, Adeoye O, Song S, et al. Final Results of the RHAPSODY trial: A multi-center, Phase 2 trial using a continual reassessment method to determine the safety and tolerability of 3K3A-APC, a Recombinant Variant of Human Activated Protein C, in combination with tissue plasminogen activator, mechanical thrombectomy or both in moderate to severe acute ischemic stroke. *Ann Neurol.* 2018. doi: 10.1002/ana.25383
4. van der Worp HB, Macleod MR, Bath PM, Bathula R, Christensen H, Colam B, Cordonnier C, Demotes-Mainard J, Durand-Zaleski I, Glud C, et al. Therapeutic hypothermia for acute ischaemic stroke. Results of a European multicentre, randomised, phase III clinical trial. *Eur Stroke J.* 2019;4:254–262. doi: 10.1177/2396987319844690 [PubMed: 31984233]
5. Chamorro A, Amaro S, Castellanos M, Segura T, Arenillas J, Marti-Fabregas J, Gallego J, Krupinski J, Gomis M, Canovas D, et al. Safety and efficacy of uric acid in patients with acute stroke (URICO-ICTUS): a randomised, double-blind phase 2b/3 trial. *Lancet Neurol.* 2014;13:453–460. doi: 10.1016/S1474-4422(14)70054-7 [PubMed: 24703208]
6. Lyden P, Hemmen T, Grotta J, Rapp K, Ernstrom K, Rzesiewicz T, Parker S, Concha M, Hussain S, Agarwal S, et al. Results of the ICTuS 2 Trial (Intravascular Cooling in the Treatment of Stroke 2). *Stroke.* 2016;47:2888–2895. doi: 10.1161/STROKEAHA.116.014200 [PubMed: 27834742]
7. Kent TA, Mandava P. Embracing Biological and Methodological Variance in a New Approach to Pre-Clinical Stroke Testing. *Transl Stroke Res.* 2016;7:274–283. doi: 10.1007/s12975-016-0463-9 [PubMed: 27018014]
8. Bosetti F, Koenig JI, Ayata C, Back SA, Becker K, Broderick JP, Carmichael ST, Cho S, Cipolla MJ, Corbett D, et al. Translational Stroke Research: Vision and Opportunities. *Stroke.* 2017;48:2632–2637. doi: 10.1161/STROKEAHA.117.017112 [PubMed: 28751554]
9. Bix GJ, Fraser JF, Mack WJ, Carmichael ST, Perez-Pinzon M, Offner H, Sansing L, Bosetti F, Ayata C, Pennypacker KR. Uncovering the Rosetta Stone: Report from the First Annual Conference on Key Elements in Translating Stroke Therapeutics from Pre-Clinical to Clinical. *Transl Stroke Res.* 2018;9:258–266. doi: 10.1007/s12975-018-0628-9 [PubMed: 29633156]
10. Lyden PD, Bosetti F, Diniz MA, Rogatko A, Koenig JI, Lamb J, Nagarkatti KA, Cabeen RP, Hess DC, Kamat PK, et al. The Stroke Preclinical Assessment Network: Rationale, Design, Feasibility, and Stage 1 Results. *Stroke.* 2022;101161STROKEAHA121038047. doi: 10.1161/STROKEAHA.121.038047
11. Maysami S, Wong R, Pradillo JM, Denes A, Dhungana H, Malm T, Koistinaho J, Orset C, Rahman M, Rubio M, et al. A cross-laboratory preclinical study on the effectiveness of interleukin-1 receptor antagonist in stroke. *Journal of Cerebral Blood Flow & Metabolism.* 2016;36:596–605. doi: 10.1177/0271678x15606714 [PubMed: 26661169]
12. Kleikers PW, Hooijmans C, Gob E, Langhauser F, Rewell SS, Radermacher K, Ritskes-Hoitinga M, Howells DW, Kleinschnitz C, Schmidt HH. A combined pre-clinical meta-analysis and randomized confirmatory trial approach to improve data validity for therapeutic target validation. *Sci Rep.* 2015;5:13428. doi: 10.1038/srep13428 [PubMed: 26310318]
13. Llovera G, Hofmann K, Roth S, Salas-Perdomo A, Ferrer-Ferrer M, Perego C, Zanier ER, Mamrak U, Rex A, Party H, et al. Results of a preclinical randomized controlled multicenter trial (pRCT): Anti-CD49d treatment for acute brain ischemia. *Science translational medicine.* 2015;7:299ra121. doi: 10.1126/scitranslmed.aaa9853

14. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, Macleod MR. Can animal models of disease reliably inform human studies? *PLoS medicine*. 2010;7:e1000245. doi: 10.1371/journal.pmed.1000245 [PubMed: 20361020]
15. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J. The REDCap consortium: Building an international community of software platform partners. *Journal of biomedical informatics*. 2019;95:103208. [PubMed: 31078660]
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*. 2009;42:377–381. [PubMed: 18929686]
17. Whitehead J The design and analysis of sequential clinical trials. John Wiley & Sons; 1997.
18. Bath PMW, Macleod MR, Green AR. Emulating multicentre clinical stroke trials a new paradigm for studying novel. *International Journal of Stroke*. 2009;4:471–479. [PubMed: 19930059]
19. Rewell SS, Churilov L, Sidon TK, Aleksoska E, Cox SF, Macleod MR, Howells DW. Evolution of ischemic damage and behavioural deficit over 6 months after MCAo in the rat: Selecting the optimal outcomes and statistical power for multi-centre preclinical trials. *PLoS ONE*. 2017;12:e0171688. doi: 10.1371/journal.pone.0171688 [PubMed: 28182727]
20. Kochanek PM, Dixon CE, Mondello S, Wang KKK, Lafrenaye A, Bramlett HM, Dietrich WD, Hayes RL, Shear DA, Gilsdorf JS, et al. Multi-Center Pre-clinical Consortia to Enhance Translation of Therapies and Biomarkers for Traumatic Brain Injury: Operation Brain Trauma Therapy and Beyond. *Front Neurol*. 2018;9:640. doi: 10.3389/fneur.2018.00640 [PubMed: 30131759]
21. Tornes M, McLernon D, Bachmann MO, Musgrave SD, Day DJ, Warburton EA, Potter JF, Myint PK. Variations in Rates of Discharges to Nursing Homes after Acute Hospitalization for Stroke and the Influence of Service Heterogeneity: An Anglia Stroke Clinical Network Evaluation Study. *Healthcare (Basel)*. 2020;8. doi: 10.3390/healthcare8040390
22. Seenan P, Long M, Langhorne P. Stroke units in their natural habitat: systematic review of observational studies. *Stroke*. 2007;38:1886–1892. doi: 10.1161/STROKEAHA.106.480871 [PubMed: 17463308]
23. Tornes M, McLernon D, Bachmann M, Musgrave S, Warburton EA, Potter JF, Myint PK, Anglia Stroke Clinical Network Evaluation Study G. Does service heterogeneity have an impact on acute hospital length of stay in stroke? A UK-based multicentre prospective cohort study. *BMJ Open*. 2019;9:e024506. doi: 10.1136/bmjopen-2018-024506
24. Bleilevens C, Roehl AB, Goetzenich A, Zoremba N, Kipp M, Dang J, Tolba R, Rossaint R, Hein M. Effect of anesthesia and cerebral blood flow on neuronal injury in a rat middle cerebral artery occlusion (MCAO) model. *Exp Brain Res*. 2013;224:155–164. doi: 10.1007/s00221-012-3296-0 [PubMed: 23064810]
25. Gaidhani N, Sun F, Schreihofner D, Uteshev VV. Duration of isoflurane-based surgical anesthesia determines severity of brain injury and neurological deficits after a transient focal ischemia in young adult rats. *Brain Res Bull*. 2017;134:168–176. doi: 10.1016/j.brainresbull.2017.07.018 [PubMed: 28755978]
26. Slupe AM, Kirsch JR. Effects of anesthesia on cerebral blood flow, metabolism, and neuroprotection. *J Cereb Blood Flow Metab*. 2018;38:2192–2208. doi: 10.1177/0271678X18789273 [PubMed: 30009645]
27. Gubskiy IL, Namestnikova DD, Cherkashova EA, Chekhonin VP, Baklaushev VP, Gubsky LV, Yarygin KN. MRI Guiding of the Middle Cerebral Artery Occlusion in Rats Aimed to Improve Stroke Modeling. *Transl Stroke Res*. 2018;9:417–425. doi: 10.1007/s12975-017-0590-y [PubMed: 29178027]
28. Yuan F, Tang Y, Lin X, Xi Y, Guan Y, Xiao T, Chen J, Zhang Z, Yang GY, Wang Y. Optimizing suture middle cerebral artery occlusion model in C57BL/6 mice circumvents posterior communicating artery dysplasia. *J Neurotrauma*. 2012;29:1499–1505. doi: 10.1089/neu.2011.2105 [PubMed: 22054164]
29. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, Roy D, Jovin TG, Willinsky RA, Sapkota BL, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. 2015;372:1019–1030. doi: 10.1056/NEJMoa1414905 [PubMed: 25671798]

30. Campbell BC, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, Yan B, Dowling RJ, Parsons MW, Oxley TJ, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*. 2015;372:1009–1018. doi: 10.1056/NEJMoa1414792 [PubMed: 25671797]
31. Usui T, Macleod MR, McCann SK, Senior AM, Nakagawa S. Meta-analysis of variation suggests that embracing variability improves both replicability and generalizability in preclinical research. *PLoS Biol*. 2021;19:e3001009. doi: 10.1371/journal.pbio.3001009 [PubMed: 34010281]
32. Dirnagl U, Group MotM-SOP. Standard operating procedures (SOP) in experimental stroke research: SOP for middle cerebral artery occlusion in the mouse. *Nature Precedings*. 2012. doi: 10.1038/npre.2012.3492.3
33. Dirnagl U, Fisher M. International, multicenter randomized preclinical trials in translational stroke research: it's time to act. *J Cereb Blood Flow Metab*. 2012;32:933–935. doi: 10.1038/jcbfm.2012.51 [PubMed: 22510602]
34. Voelkl B, Vogt L, Sena ES, Wurbel H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS biology*. 2018;16:e2003693. doi: 10.1371/journal.pbio.2003693 [PubMed: 29470495]

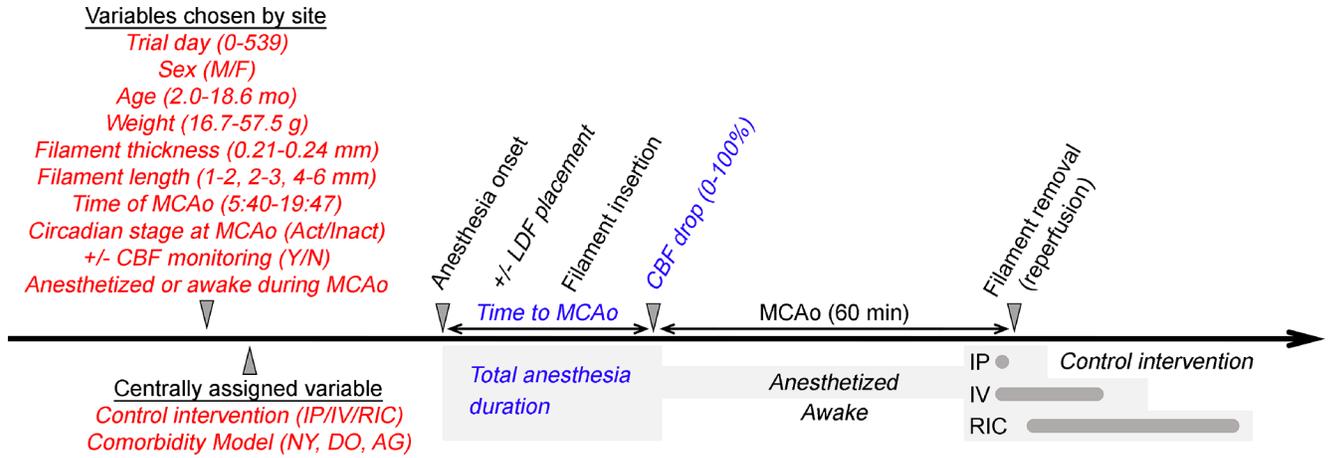


Figure 1. Study variables and timeline.

The timeline shows the independent (red) and dependent (blue) variables in relation to the experimental protocol. Two independent variables were assigned by the CC for each subject, while the rest were left to each site’s preference. Time to MCAo was calculated from anesthesia onset to onset of MCAo. Total anesthesia duration (gray shaded areas) included the duration of surgery, the duration of MCAo (60 minutes) in animals maintained under anesthesia until reperfusion, the reperfusion procedure, and the time required to administer the centrally allocated control intervention (IP, intraperitoneal vehicle; IV, intravenous vehicle; RIC, sham remote ischemic conditioning).

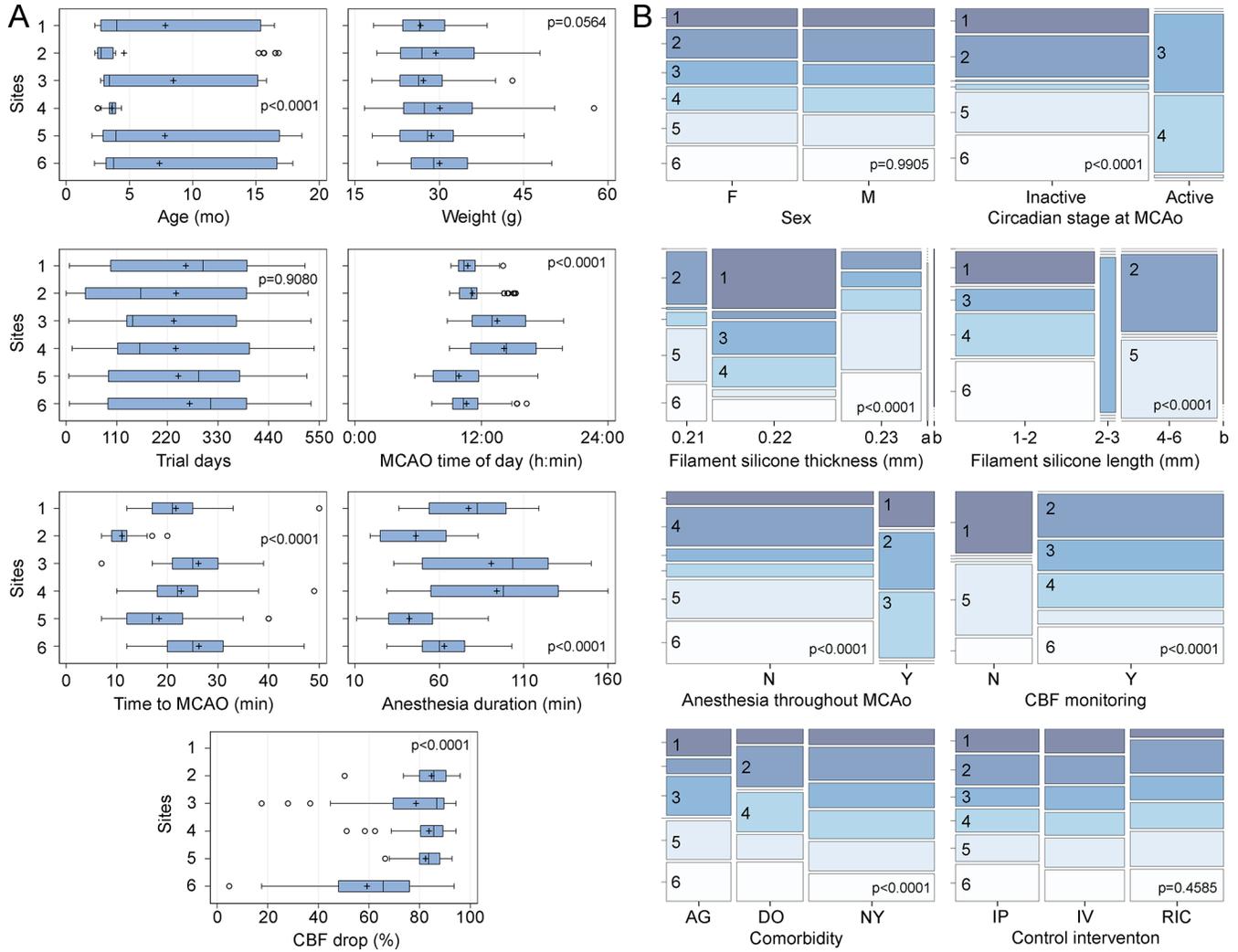


Figure 2. Heterogeneity of numerical and categorical variables across sites.

A: Whisker-box plots highlight the heterogeneity among the study sites. All numerical variables significantly varied among the six sites, except for weight at the time of MCAo (one-way ANOVA). Circles indicate data points outside the 1.5x interquartile range (IQR). All p-values are adjusted by FDR. **B:** Mosaic plots show the distribution of categorical variables among the study sites. All categorical variables significantly varied among the six sites, except for sex and control intervention type (χ^2). All p-values are adjusted by FDR. a, 0.24 mm; b, missing value; AG, aged mice; DO, diet-induced obese mice; F, female; IP, intraperitoneal vehicle; IV, intravenous vehicle; M, male; N, no; NY, normal young mice; RIC, sham remote ischemic conditioning; Y, yes.

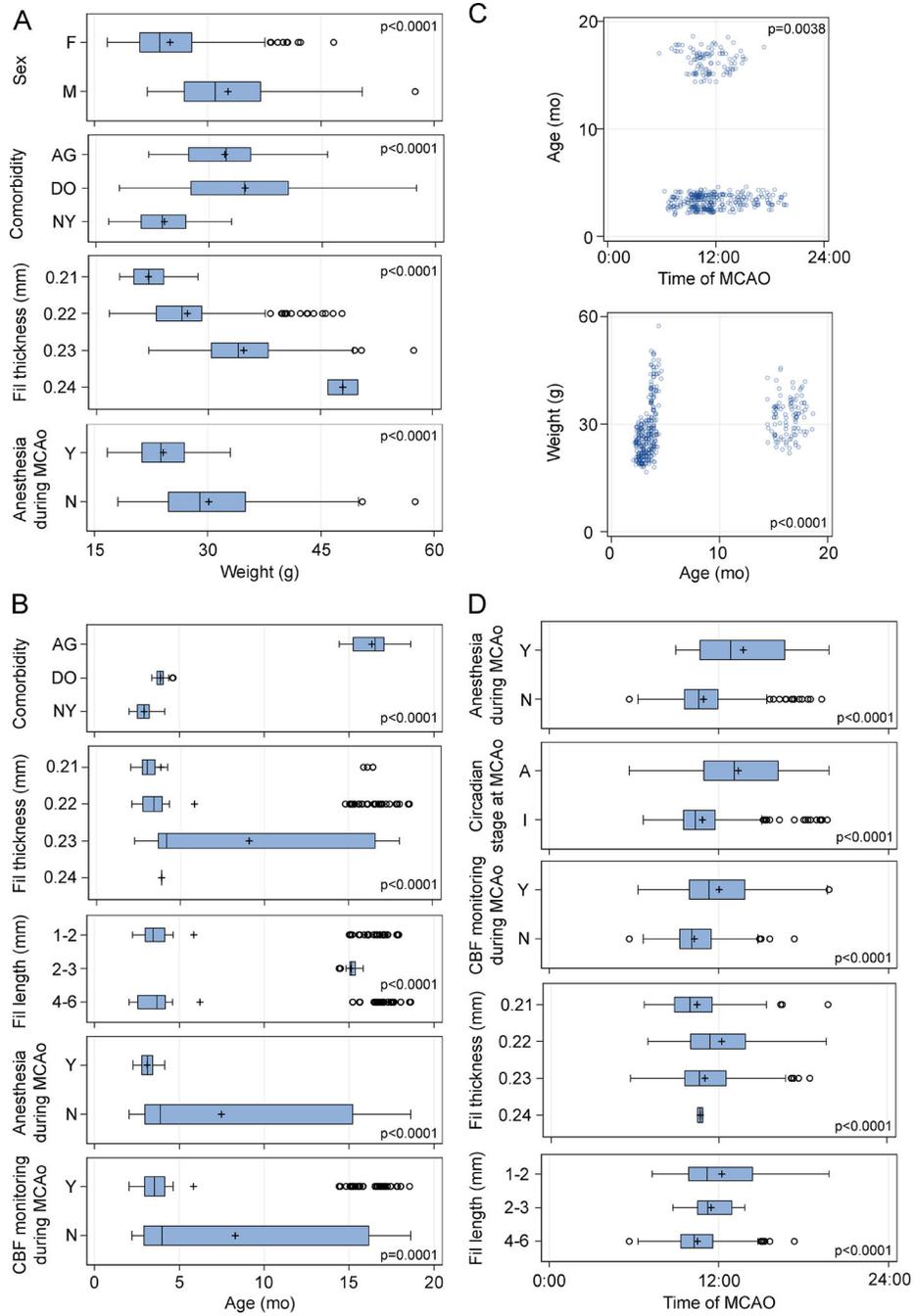


Figure 3. Bivariate associations of numerical independent variables.

Whisker-box and scatter plots show statistically significant associations (i.e., covariance) of independent variables with weight (A), age (B and C), and time of day at MCAo (D), analyzed using t-test, one-way ANOVA, or general linear model. Circles in the box-whisker plots indicate data points outside the 1.5x interquartile range (IQR). All p-values are adjusted by FDR. A, active; AG, aged mice; DO, diet-induced obese mice; F, female; I, inactive; M, male; N, no; NY, normal young mice; Y, yes.

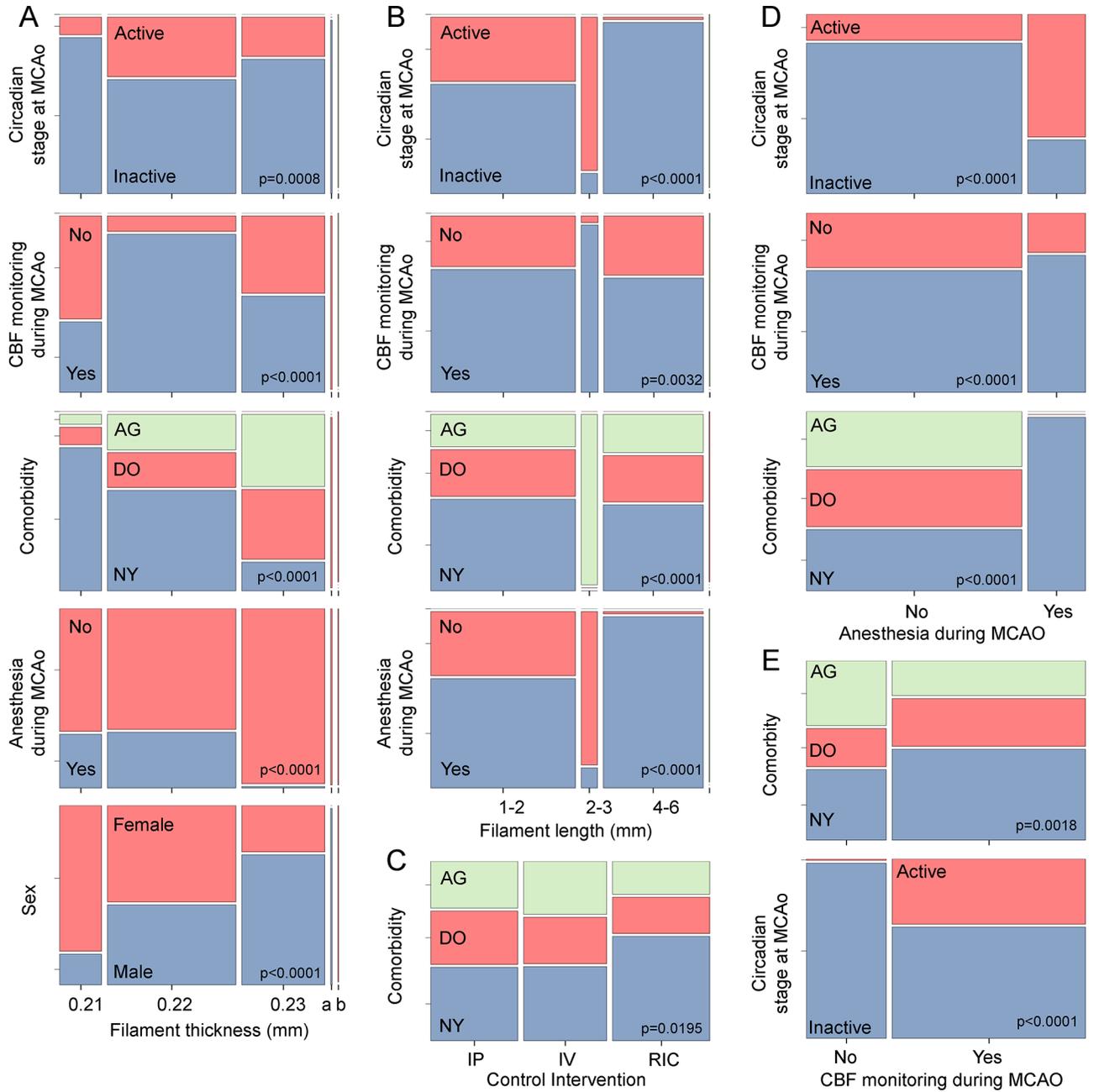


Figure 4. Bivariate associations of categorical independent variables.

Mosaic plots show statistically significant associations (i.e., covariance) among categorical independent variables analyzed using χ^2 . Detailed data are shown in Table S5. All p-values are adjusted by FDR. AG, aged mice; DO, diet-induced obese mice; NY, normal young mice

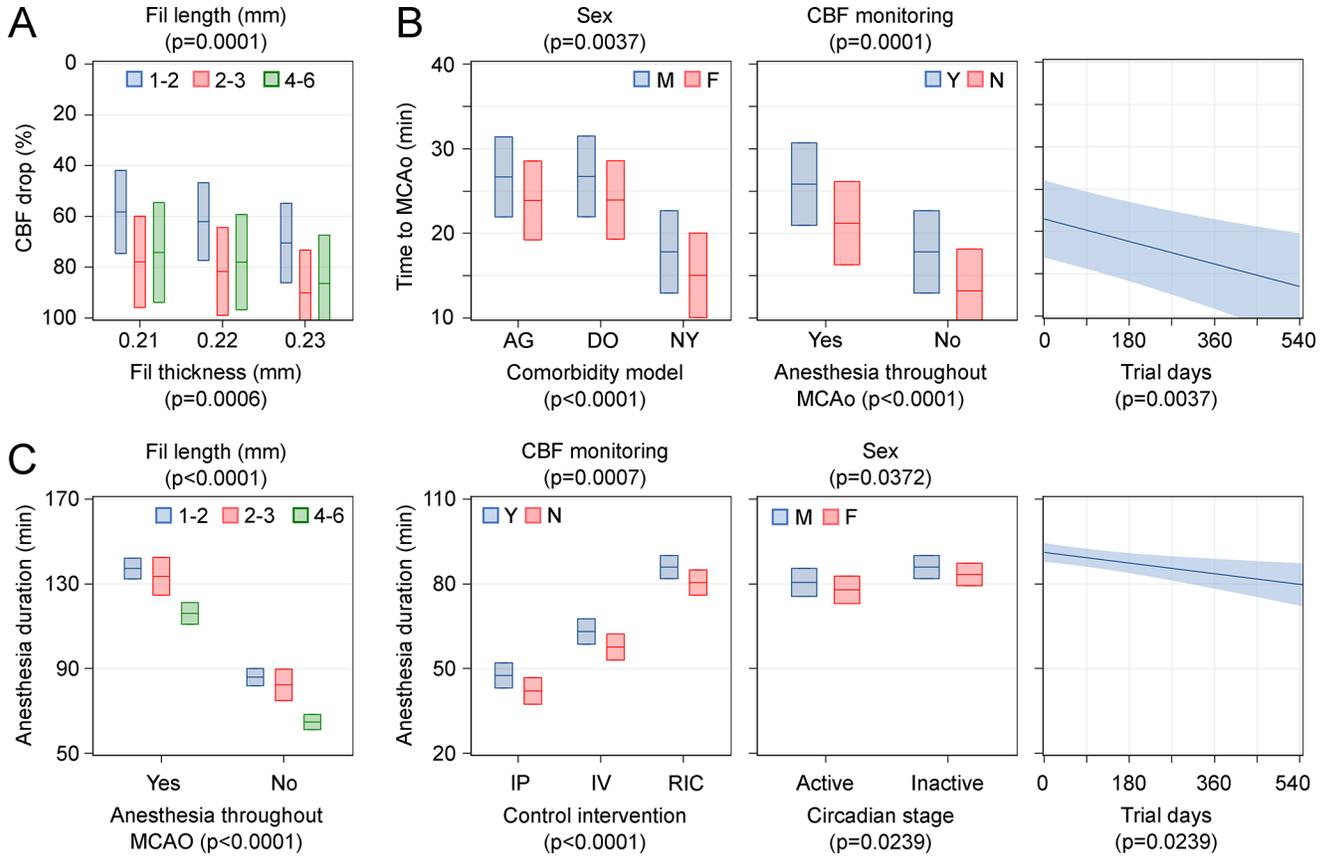


Figure 5. Predictors of dependent variables CBF drop, time to MCAO, and total anesthesia duration.

Independent predictors of CBF drop (A), time to MCAO (B), and total anesthesia duration (C) in multivariable models after stepwise backward elimination are shown (see Table 1 and Table S6 for details). Effect plots display the behavior of the predicted fitted model while holding other covariates fixed. Plots show predicted effects with 95% confidence intervals (shaded area). When a variable is not represented in the plot, it is fixed at the mean or mode of the entire cohort. When more than one independent variable is represented on a graph, their p-values are shown next to the variable they belong to. All p-values are adjusted for multiple comparisons by stepdown Sidak. F, female; IP, intraperitoneal vehicle; IV, intravenous vehicle; M, male; N, no; RIC, sham remote ischemic conditioning; Y, yes.

Table 1.

Effect sizes in the final multivariable model

		Estimate	SE	Lower 95% CI	Upper 95% CI	DF	t	Sig
CBF drop (%)								
Fil thickness (mm)	0.21 vs. 0.22	-3.8	3.7	-11.1	3.5	289	-1.0	n.s
	0.21 vs. 0.23	-12.2	3.7	-19.5	-4.9	289	-3.3	**
	0.22 vs. 0.23	-8.4	2.7	-13.7	-3.1	289	-3.1	**
Fil length (mm)	1 to 2 vs. 2 to 3	-19.6	4.5	-28.5	-10.8	289	-4.4	****
	1 to 2 vs. 4 to 6	-15.9	12.2	-40.0	8.1	289	-1.3	n.s.
	2 to 3 vs. 4 to 6	3.7	12.8	-21.5	28.9	289	0.3	n.s.
Time to MCAO (min)								
Sex	F vs. M	-2.78	0.82	-4.38	-1.17	405	-3.40	***
Comorbidity model	NY vs. DO	-9.07	1.87	-12.76	-5.39	405	-4.84	****
	NY vs. AG	-8.98	1.87	-12.67	-5.30	405	-4.80	****
	DO vs. AG	0.091	0.89	-1.66	1.84	405	0.10	n.s.
CBF monitoring	No vs. Yes	-4.60	1.04	-6.64	-2.57	405	-4.44	****
Anesthesia during MCAO	No vs. Yes	-8.03	1.2	-10.43	-5.64	405	-6.59	****
Trial days [†]	(min/day)	-0.02	0.01	-0.02	-0.006	405	-3.33	***
Anesthesia duration (min)								
Anesthesia during MCAO	No vs. Yes	-51.29	2.13	-55.49	-47.09	342	-24.03	****
Fil length (mm)	1 to 2 vs. 2 to 3	3.68	3.15	-2.51	9.87	342	1.17	n.s.
	1 to 2 vs. 4 to 6	21.20	1.34	18.57	23.83	342	15.86	****
	2 to 3 vs. 4 to 6	17.52	3.26	11.12	23.93	342	5.38	****
CBF monitoring	No vs. Yes	-5.49	1.41	-8.26	-2.73	342	-3.91	***
Control Intervention	IP vs. IV Control	-15.56	1.21	-17.95	-13.18	342	12.84	****
	IP vs. RIC Control	-38.39	1.54	-41.41	-35.37	342	-25.00	****
	IV vs. RIC Sham	-22.83	1.56	-25.89	-19.77	342	-14.66	****
Sex	F vs. M	-2.62	1.04	-4.67	-0.57	342	-2.51	*
Circadian stage	Active vs. Inactive	-5.45	1.94	-9.26	-1.64	342	-2.81	**
Trial days [†]	(min/day)	-0.021	0.01	-0.035	-0.01	342	-2.84	**

Estimates are the partial unstandardized regression coefficient for each given numeric predictor (†), or the unstandardized covariate-adjusted mean difference between the indicated categories for the categorical predictors.

*
p < 0.05

**
p < 0.01

p < 0.001

p < 0.0001

n.s., nonsignificant. F, female; M, male; AG, Aged; DO, diet-induced obesity; NY, normal young; IP, intraperitoneal; IV, intravenous; RIC, remote ischemic conditioning; CBF, cerebral blood flow.