JVIN | Journal of Vascular and Interventional Neurology

OFFICIAL JOURNAL OF ZEENAT QURESHI STROKE INSTITUTE

Treatment-arm count, Phase 2 information value, and complexity as predictors of Phase 3 success in neurological drug development

Adnan I. Qureshi[1] MD, Kamran Zahoor[1,2,*] MD, Jonathan Beall[3] PhD, Byron Gajewski[4] PhD, Renee H. Martin[3] PhD, Jose I. Suarez[5] MD

[1] Zeenat Qureshi Stroke Institute, Columbia, MO, USA.

[2] University of Missouri, Columbia, MO, USA.

[3] Medical University of South Carolina (MUSC), Charleston, SC, USA.

[4] University of Kansas Medica Center, Kansas City, KS, USA.

[5] John Hopkis University, Baltimore, MD, USA.

Abstract

Background— Phase 2 neurological trials often include multiple treatment arms to select doses, characterize exposure-response, evaluate regimens or subgroups, and make earlier go/no-go decisions. The possibility that multi-arm Phase 2 trials may be more informative and improve Phase 3 trial translation may be offset by added complexity that may impair translation by reducing per-arm precision, increasing multiplicity and optimistic selection, lengthening development timelines, and creating operational or interpretive problems.

Objective— To test two linked hypotheses in neurological drug development: first, that increasing Phase 2 trial treatment-arm count would be associated with higher Phase 3 trial primary-endpoint success because of greater information value; and second, that high Phase 2 trial design complexity could reduce translation into Phase 3 trial success.

Methods— We extracted 61 linked Phase 2 trial to Phase 3 trial pharmacological or biologic neurological trial pairs across major neurological disease areas. Exposures were Phase 2 trial total-arm count, active-arm count, multi-arm/randomized design, high total-arm complexity (>=4 total arms), high active-arm complexity (>=3 active arms), Phase 2 trial signal, biomarker/genetic enrichment, and target/mechanism validation. The primary outcome was whether the later Phase 3 trial or pivotal readout met it's prespecified primary endpoint. Analyses were descriptive and hypothesis-generating.

Results— Phase 2 trial treatment-arm count did not show a monotonic relationship with Phase 3 trial success (Spearman rho=0.06, p=0.66). Phase 3 trial primary success occurred in 4/5 single-arm pairs, 8/19 two-arm pairs, 7/10 three-arm pairs, 9/15 four-arm pairs, 3/8 five-arm pairs, and 4/4 six-arm pairs. High total-arm complexity (>=4 arms) was not associated with lower or higher Phase 3 trial success (16/27 (59.3%) vs 19/34 (55.9%) for <4 arms; OR 1.15, 95% CI 0.41-3.20; Fisher p=1.00). Clear Phase 2 trial positivity and strong target/mechanism validation were stronger predictors.

Conclusions— Multiple treatment arms may increase Phase 2 trial information value and higher arm count alone does not reduce successful Phase 3 trial translation. Complexity appeared beneficial when paired with strong biology and clear efficacy signals, and unhelpful when used to explore weakly validated mechanisms.

Keywords— neurology, Phase 2 trial, Phase 3 trial, treatment arms, clinical trial complexity, adaptive design, drug development, trial translation.

INTRODUCTION

Phase 2 trials are intended to provide an evidentiary bridge between early safety/pharmacology studies and large confirmatory Phase 3 trials. In neurology, this bridge is difficult because disease biology, outcome measures, time horizons, placebo response, and trial feasibility differ substantially across disorders. These challenges are consistent with broader evidence that clinical development success varies substantially by therapeutic area and that late-stage failures remain common despite earlier clinical testing.[1,2] Programs in multiple sclerosis, migraine, spinal muscular atrophy, and myasthenia gravis often target clinically or genetically

Vol.16, No. 1, pp. 84-91, Published July, 2026.

All Rights Reserved by JVIN. Unauthorized reproduction of this article is prohibited.

M. Fareed K. Suri MD served as Guest Editor for this manuscript to ensure independent review process.

*Corresponding Author: Kamran Zahoor, University of Missouri, Columbia, MO. Kzwyc@heath.missouri.edu.

validated pathways, whereas programs in Alzheimer disease, Parkinson disease, amyotrophic lateral sclerosis, Huntington disease, stroke, and traumatic brain injury have often advanced with less reliable proof-of-concept evidence.

Recent neurology evidence has shown that many Phase 3 trials begin without a clearly positive preceding Phase 2 trial efficacy result, supporting the importance of Phase 2 trial evidence quality for late-stage development decisions.[3] However, whether the structure of Phase 2 trial evidence - particularly the number of treatment arms - effect Phase 3 trial success is less clear. Arm count is not merely a formatting feature: it can represent dose-ranging, regimen exploration, common-control efficiency, biomarker subgroup evaluation, or selection among competing interventions.

Theoretical rationale: why more arms might improve Phase 3 trial translation

A multi-arm Phase 2 trial may be more informative than a simple two-arm or single-arm trial for several reasons. First, multiple active doses can define dose-response and exposure-response relationships, reducing the risk that Phase 3 trial tests a subtherapeutic or poorly tolerated dose. Second, a common control arm can allow several doses, regimens, or candidate interventions to be compared within one protocol, improving consistency of eligibility criteria, endpoint assessment, and calendar time. Multi-arm multi-stage designs can improve efficiency by sharing control groups, allowing interim dropping of ineffective arms, and reducing expected sample size while maintaining operating characteristics.[4,5] Third, multi-arm designs can support biomarker or population selection by evaluating whether treatment effects are concentrated in a defined subgroup. Fourth, adaptive or multi-stage approaches can drop ineffective arms early and allocate subsequent development resources toward more promising candidates. Food and Drug Administration (FDA) adaptive-design guidance describes advantages including statistical efficiency, ethical advantages from early stopping, and improved understanding of drug effects, including adaptive dose selection and enrichment.[6] Multi-arm multi-stage frameworks similarly allow multiple treatments to be tested under a single protocol, with poorly performing arms dropped and new arms added as appropriate.[7]

Theoretical rationale: why complexity might impair Phase 3 trial translation

The opposite hypothesis is also plausible. Multi-arm trials can dilute sample size per arm if total enrollment does not increase, producing unstable estimates and increasing the probability that an apparently promising arm is selected by chance. Low per-arm precision is especially important in neuroscience, where underpowered studies can exaggerate effect sizes and reduce reproducibility.[8] Multiple comparisons, adaptive selection, subgroup exploration, and dose choice can inflate false-positive risk or treatment-effect optimism unless prespecified and statistically controlled. Adaptive designs also require careful attention to bias in treatment-effect estimation and multiplicity control, particularly when interim decisions influence which arms continue.[6,9,10] FDA guidance explicitly notes that adaptive and complex designs can introduce Type I error inflation, biased treatment-effect estimates, longer planning lead times, logistical challenges, and more complex documentation requirements.[6] More generally, trial-complexity analyses have identified number of arms, number of endpoints, eligibility criteria, countries, sites, and enrollment size as features contributing to trial duration and complexity.[11] In neurology, longer trials may be especially vulnerable to changing standards of care, endpoint drift, recruitment bottlenecks, and operational variability. Complexity may also be a marker of biological uncertainty: programs with many exploratory arms may be trying to compensate for uncertain target validity rather than clarifying a mature mechanism.

Hypotheses tested

We therefore reframed the arm-count question as a bidirectional empirical test. The information hypothesis predicted that a larger number of Phase 2 trial arms would increase Phase 3 trial success by improving doses, population, regimen, and go/no-go decisions. The complexity hypothesis predicted that highly complex Phase 2 trial designs could reduce successful translation into Phase 3 trial by increasing statistical, operational, and interpretive friction. We tested these hypotheses at the individual trial-pair level rather than by summarizing previous reviews.

METHODS

Design and analytic unit

This was a Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA)-informed rapid systematic review and individual study-level evidence synthesis. The analytic unit was a linked trial pair: a Phase 2 trial, Phase 2b trial, or Phase 2/3 trial proof-of-concept or dose-finding study of a pharmacological or biologic neurological intervention, paired with a later pivotal or confirmatory Phase 3 trial/readout in a comparable disease setting. When several Phase 3 trials existed for the same program, the row was coded according to the pivotal program-level primary-endpoint interpretation.

Eligibility criteria

- Population: patients with neurological diseases, including Alzheimer/dementia, multiple sclerosis, Parkinson disease, amyotrophic lateral sclerosis, migraine, epilepsy, Huntington disease, spinal muscular atrophy, myasthenia gravis, stroke, and traumatic brain injury.
- Intervention: pharmacological or biologic drug interventions. Devices, surgery, rehabilitation-only interventions, and behavioral interventions were excluded.
- Evidence linkage: a public Phase 2 trial or Phase 2 trial-like proof-of-concept/dose-finding signal had to be linkable to a later Phase 3 trial readout in a similar indication.
- Evidence rows: individual trials, trial registries, or sponsor primary result announcements. Previous systematic reviews and narrative reviews were used only for context, not as data rows.

TABLE 1: Phase 3 trial primary-endpoint success by disease area.

Disease area	Trial pairs	Phase 3 trial primary successes	Success %
Multiple sclerosis	15	12	80.0%
Alzheimer/dementia	8	2	25.0%
Migraine	8	8	100.0%
Amyotrophic lateral sclerosis	7	1	14.3%
Epilepsy	7	6	85.7%
Myasthenia gravis	4	3	75.0%
Parkinson disease	4	1	25.0%
Huntington disease	3	0	0.0%
Spinal muscular atrophy	2	2	100.0%
Stroke	2	0	0.0%
Traumatic brain injury	1	0	0.0%

TABLE 2: Phase 3 trial primary success by Phase 2 trial total-arm count.

Phase 2 trial total arms	Trial pairs	Phase 3 trial primary successes	Success %
1 arm	5	4	80.0%
2 arms	19	8	42.1%
3 arms	10	7	70.0%
4 arms	15	9	60.0%
5 arms	8	3	37.5%
6 arms	4	4	100.0%

TABLE 3: Arm-count complexity proxies and Phase 3 trial primary success.

Exposure	Phase 3 trialsuc-cess in exposed	Phase 3 tiral success in reference	OR (95% CI)	Fisher p
Multi-arm/randomized (>1 total arm)	31/56 (55.4%)	4/5 (80.0%)	0.31 (0.03-2.95)	0.38
High total-arm complexity (>=4 total arms)	16/27 (59.3%)	19/34 (55.9%)	1.15 (0.41-3.20)	1.00
High active-arm complexity (>=3 active arms)	16/27 (59.3%)	19/34 (55.9%)	1.15 (0.41-3.20)	1.00

TABLE 4: Phase 3 trial primary success by Phase 2 trial signal and target/mechanism validation.

Predictor category	Trial pairs	Phase 3 trial primary successes	Success %
Clear positive Phase 2 trial signal	34	27	79.4%
All other Phase 2 trial signal categories	27	8	29.6%
Strong/validated target or class	24	21	87.5%
Other target/mechanism categories	37	14	37.8%

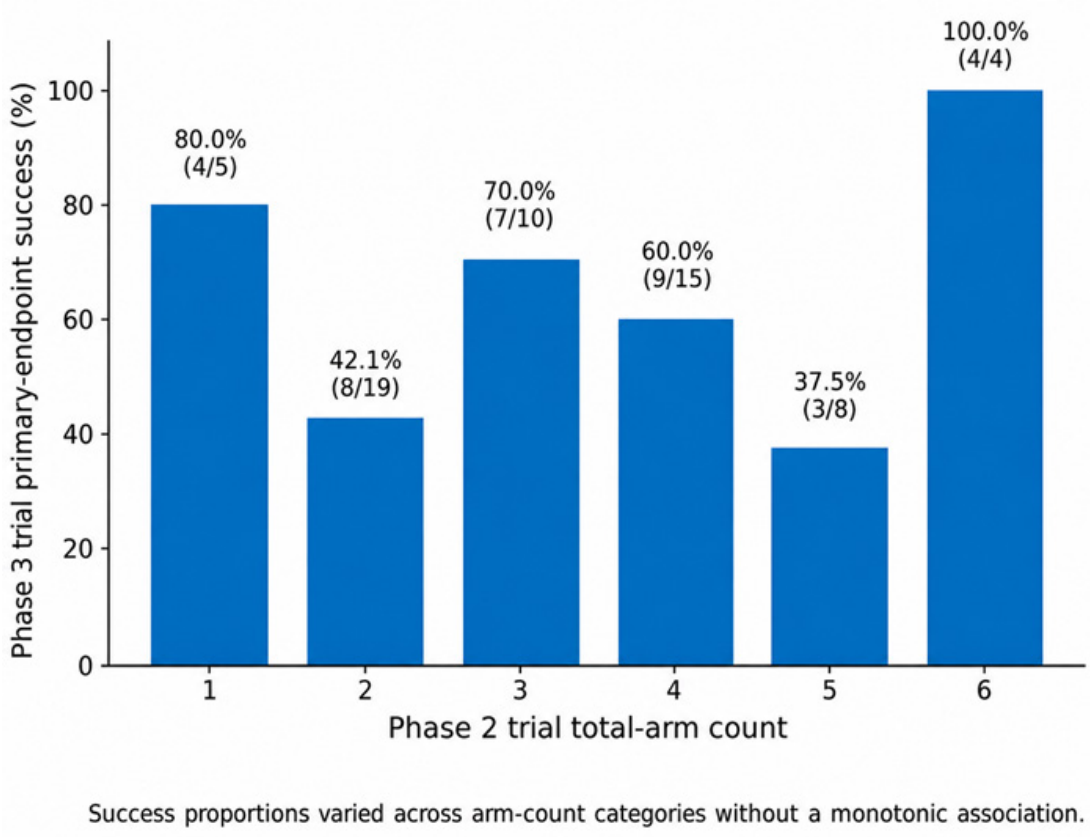


FIGURE 1: Phase 3 trial success by Phase 2 trial total-arm count.

Search strategy

Searches combined terms for phase, disease area, and trial outcome: ("phase II" OR "phase 2" OR "phase 2b") AND (Alzheimer OR multiple sclerosis OR Parkinson OR ALS OR migraine OR epilepsy OR Huntington OR "myasthenia gravis" OR "spinal muscular atrophy" OR stroke OR "traumatic brain injury") AND ("phase III" OR "phase 3" OR pivotal OR confirmatory OR approval OR failed). Sources included PubMed-indexed trial publications, PubMed Central, ClinicalTrials.gov, journal pages, and sponsor or regulator primary readout announcements when peer-reviewed Phase 3 trial publications were not available. Searches were current through June 3, 2026. Row-level source URLs are included in the accompanying extraction workbook (see appendix).

Data extraction and coding

For each linked pair, extraction captured disease area, indication, intervention, therapy class, Phase 2 trial design, Phase 2 trial total arms, Phase 2 trial active arms, approximate Phase 2 trial sample size, randomized/control status, Phase 2 trial signal, endpoint type, biomarker/genetic enrichment, target/mechanism validation, Phase 3 trial/readout, Phase 3 trial design, approximate Phase 3 trial sample size, Phase 3 trial primary-endpoint success, regulatory/practice success, linkage confidence, and source Uniform Resource Locator (URLs).

Target/mechanism validation was coded as strong when the program involved a genetic disease target, a clinically validated immune pathway, or a mature therapeutic class; moderate when the mechanism was plausible but not yet clinically validated; and weak when the target was largely unvalidated neuroprotection, mitochondrial/metabolic, or broad symptomatic biology.

Hypothesis tests and complexity proxies

The primary information-value exposure was Phase 2 trial total-arm count. Supporting exposures were active-arm count and multi-arm/randomized Phase 2 trial design. Because trial complexity is multidimensional and could not be fully reconstructed from all public sources, we used two transparent arm-count proxies: high total-arm complexity, defined as >=4 total arms, and high active-arm complexity, defined as >=3 active arms. These thresholds were chosen before analysis of the reframed question to capture dose-ranging or multi-dose designs beyond a simple active-versus-control comparison. The primary outcome was Phase 3 trial primary-endpoint success. Analyses used descriptive success proportions, Spearman rank correlation for ordinal arm count, and Fisher exact tests with descriptive odds ratios for binary comparisons. Results were interpreted as hypothesis-generating because of nonindependence, disease-area confounding, and selective advancement from Phase 2 trial to Phase 3 trial.

Results

Included trial pairs

The review included 61 linked trial pairs across 11 neurological disease areas. The overall Phase 3 trial primary-endpoint success proportion was 35/61 (57.4%), as shown in table 1.[12–23]

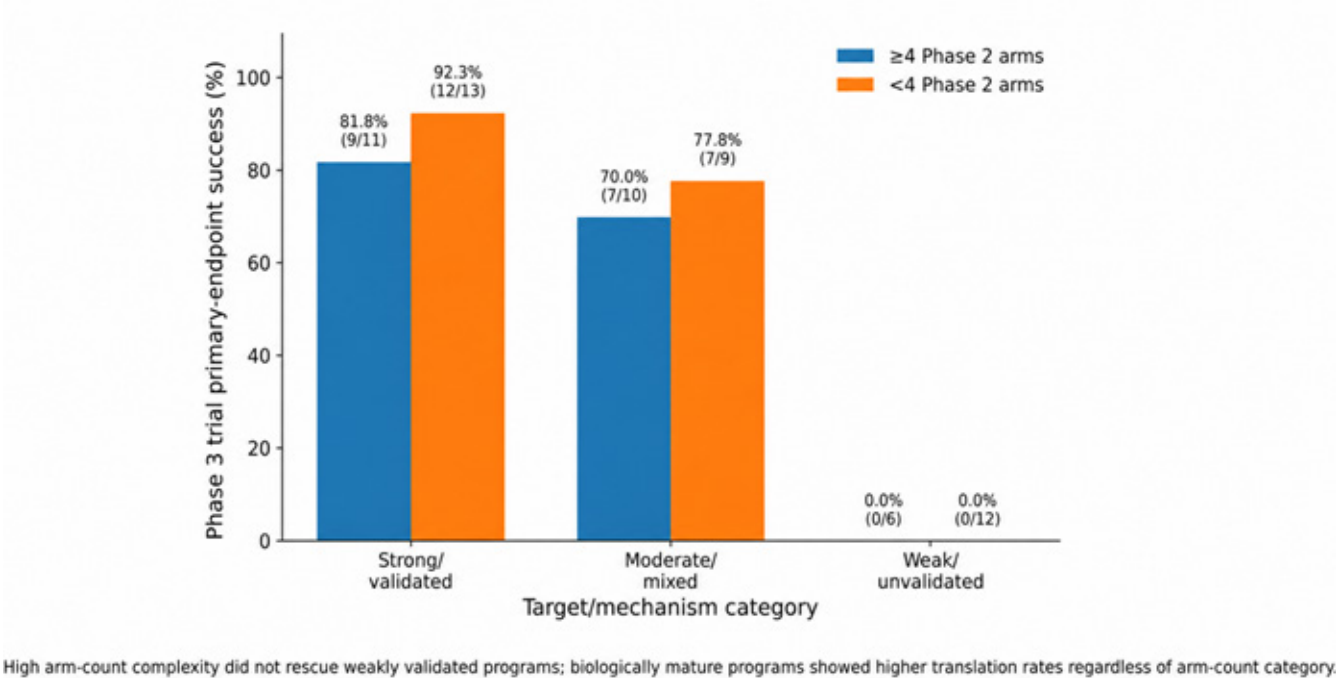


FIGURE 2: Phase 3 trial success by target validation and complexity.

Hypothesis 1: information value of additional arms

Phase 2 trial total-arm count did not show a monotonic association with later Phase 3 trial success (Spearman rho=0.06, p=0.66) as shown in table 2 and figure 1. The apparent high success among single-arm pairs reflected rare-disease or highly selected settings with strong target validity rather than an advantage of single-arm design. Multi-arm/randomized Phase 2 trial evidence had 31/56 (55.4%) Phase 3 trial primary successes, compared with 4/5 (80.0%) for single-arm evidence (Odds ratio (OR) 0.31, 95% CI 0.03-2.95; Fisher p=0.38).

Hypothesis 2: complexity as a possible barrier to translation

Using >=4 Phase 2 trial total arms as a high-complexity proxy, as shown in Table 3 Phase 3 trial success was 16/27 (59.3%) in high-complexity pairs and 19/34 (55.9%) in lower-complexity pairs (OR 1.15, 95% CI 0.41-3.20; Fisher p=1.00). Using >=3 active treatment arms gave the same split in this dataset: 16/27 (59.3%) versus 19/34 (55.9%) (OR 1.15, 95% CI 0.41-3.20; Fisher p=1.00). Thus, we did not observe aggregate evidence that arm-count complexity prevented Phase 3 trial success.

Biological and clinical information outweighed arm count

Clear Phase 2 trial positivity was much more informative than arm count. As shown in table 4 clear positive Phase 2 trial evidence had 27/34 (79.4%) Phase 3 trial primary successes, compared with 8/27 (29.6%) for all other Phase 2 trial signal categories (OR 9.16, 95% CI 2.84-29.57; Fisher p=0.0002). Strong target/mechanism validation was similarly associated with translation: 21/24 (87.5%) successes for strong/validated targets versus 14/37 (37.8%) for all other target categories (OR 11.50, 95% CI 2.89-45.72; Fisher p=0.0002).

Complexity interacted descriptively with biological maturity

When stratified by target maturity , complexity did not rescue weakly validated programs and did not appear necessary for strong programs. As shown in Table 5 and Figure 2, Strong/validated programs had high Phase 3 trial success in both lower-arm and higher-arm settings; weak or unvalidated programs had no Phase 3 trial primary-endpoint successes regardless of arm-count category. This pattern is consistent with the interpretation that multiple arms are useful when they refine a plausible program, but do not overcome poor target validity.

Representative individual trial-pair examples

Table 6 highlights examples showing why arm count requires clinical interpretation. Multi-arm designs supported successful translation when paired with validated targets and credible efficacy signals, while similar or greater arm complexity did not prevent failure when target validity or clinical proof-of-concept was weak. Single-arm successes were concentrated in rare or genetically defined settings and should not be interpreted as evidence that single-arm designs are generally preferable.

Discussion

This individual trial-pair review reframes Phase 2 trial arm count as a question of information value rather than as a simple structural predictor. The theoretical case for multiple treatment arms is strong: more arms can support dose selection, regimen selection, common-control comparisons, enrichment, and early futility. However, the empirical hypothesis that more Phase 2 trial arms would translate into greater Phase 3 trial success was not supported in this dataset. Arm count showed no monotonic association with Phase 3 trial primary-endpoint success, and high arm-count complexity was not harmful in aggregate. The findings also

TABLE 5: Phase 3 trial primary success for high versus lower arm-count complexity, stratified by target maturity.

Target/mechanism category	≥4 Phase 2 trial arms	<4 Phase 2 trial arms
Strong/validated	9/11 (81.8%)	12/13 (92.3%)
Moderate/mixed	7/10 (70.0%)	7/9 (77.8%)
Weak/unvalidated or weakly validated	0/6 (0.0%)	0/12 (0.0%)

TABLE 6: Representative trial-pair examples by Phase 2 trial arm count, Phase 2 trial signal, and Phase 3 trial outcome.

ID	Disease	Intervention	P2 arms	P2 signal	P3 primary	Interpretive note
1	Alzheimer/dementia	Bapineuzumab	4	Mixed/positive subgroup	No	Failure: early signal or multi-arm exploration did not overcome weak/ uncertain translation.
3	Alzheimer/dementia	Tarenflurbil	3	Positive/mixed	No	Failure: early signal or multi-arm exploration did not overcome weak/ uncertain translation.
5	Alzheimer/dementia	Lecanemab	6	Mixed/positive	Yes	Success: strong biology and/or informative dose-population selection.
6	Alzheimer/dementia	Donanemab	2	Positive	Yes	Success: strong biology and/or informative dose-population selection.
18	Multiple sclerosis	Ublituximab	1	Positive	Yes	Success: strong biology and/or informative dose-population selection.
24	Parkinson disease	Coenzyme Q10	4	Positive/mixed	No	Failure: early signal or multi-arm exploration did not overcome weak/ uncertain translation.
25	Parkinson disease	Isradipine	4	Mixed/tolerability-selected	No	Failure: early signal or multi-arm exploration did not overcome weak/ uncertain translation.
35	Migraine	Erenumab	2	Positive	Yes	Success: strong biology and/or informative dose-population selection.
53	Spinal muscular atrophy	Nusinersen	1	Positive	Yes	Success: strong biology and/or informative dose-population selection.
59	Traumatic brain injury	Progesterone	2	Positive	No	Failure: early signal or multi-arm exploration did not overcome weak/ uncertain translation.

did not support a strong aggregate version of the complexity hypothesis. Phase 2 trial designs with >=4 arms did not have lower Phase 3 trial success than less complex designs. Nevertheless, the theoretical risks of complexity remain clinically important. Multiplicity, adaptive selection, dose picking, subgroup exploration, endpoint heterogeneity, and operational burden can all make a Phase 2 trial program appear more informative than it truly is. These risks are most concerning when the underlying mechanism is weak, the Phase 2 trial signal is mixed, or the selected dose/subgroup is chosen post hoc. The most important descriptive predictors were clear Phase 2 trial clinical positivity and strong target/mechanism validation. This suggests that complexity is best understood as an effect modifier and decision tool. In biologically mature programs, multiple arms may improve Phase 3 trial readiness by optimizing dose, regimen, or population. In biologically uncertain programs, more arms may simply multiply opportunities for noisy signals without improving the probability of reproducible clinical benefit.

Several disease-area patterns support this interpretation. calcitonin gene-related peptide (CGRP)-pathway migraine therapies, sphingosine-1-phosphate (S1P) and cluster of differentiation 20 (CD20) therapies in multiple sclerosis, survival motor neuron (SMN)-targeting therapies in spinal muscular atrophy, and neonatal Fc receptor (FcRn)/complement-targeted therapies in myasthenia gravis often translated successfully.[15,21–26] Many used randomized or dose-ranging Phase 2 trial designs, but their success plausibly reflected biological validation and coherent clinical endpoints more than arm count itself. By contrast, weakly validated neuroprotective, mitochondrial, antioxidant, or broad disease-modification programs in Parkinson disease, amyotrophic lateral sclerosis (ALS), Huntington disease, stroke, and traumatic brain injury failed consistently in this extraction despite supportive early signals in some cases.[27–31]

The practical implication is that Phase 2 trial complexity should be justified prospectively by a specific decision. Examples include choosing between doses with a plausible exposure-response relationship, testing whether a biomarker-positive population is required, comparing a regimen with an appropriate control, or dropping futile arms early. Complexity is unlikely to help when the scientific question is vague, per-arm sample sizes are too small, or the design creates multiple opportunities for optimistic post hoc selection. Adaptive comparative-effectiveness methodology provides a complementary rationale for multi-arm Phase 2 trial designs.[32] Gajewski et al used a planned amyotrophic lateral sclerosis sialorrhea trial to show that selecting only two arms from several plausible options before enrollment can create a type III selection error by excluding the best treatment.[33] Their Bayesian adaptive simulations suggested that keeping more candidate arms and using response-adaptive randomization, adaptive equal randomization, or arm dropping can reduce this risk with little loss of power relative to an ideal two-arm comparison and appreciable gains when the two arms would otherwise be prescreened at random. This supports our conceptual framing: additional arms are not merely burdensome complexity when they are tied to prespecified adaptive selection, adequate per-arm information, and statistical control. However, our empirical results suggest that raw arm count alone is insufficient; what likely matters is whether the extra arms are embedded in a decision-informative design.

Limitations

This analysis is limited because it used linked Phase 2 trial-to-Phase 3 trial pairs rather than a full Phase 2 inception cohort. As a result, it may miss Phase 2 trials that led to no-go decisions, were abandoned for futility or toxicity, or remain pending before Phase 3 trial, creating selective-advancement bias. Therefore, the results estimate Phase 3 trial success conditional on advancement, not the overall probability that a Phase 2 trial program ultimately succeeds. A more definitive study would include all eligible Phase 2 trials, classify programs as no-go, pending, advanced, Phase 3 trial successful, or Phase 3 trial unsuccessful, and use time-to-event methods to account for delayed or ongoing development. Consequently, the findings should be considered hypothesis-generating rather than causal. The included dataset focused on prominent and readily linkable neurological trial pairs across major disease areas, which improve interpretability but limits completeness. Arm count was used as a practical proxy for trial complexity, although true protocol complexity would require detailed coding of endpoints, visit schedules, procedures, eligibility criteria, countries, sites, amendments, adaptive rules, and decision algorithms. The observed associations are descriptive rather than causal, because disease area, target class, sponsorship, endpoint type, regulatory context, and selective advancement from Phase 2 trial to Phase 3 trial are strongly confounded. Publication and reporting biases may influence which Phase 2 trial signals and failed programs are publicly available for linkage.[34] In addition, some programs with multiple Phase 3 trials were summarized at the program level, and the analysis was based on study-level rows rather than patient-level data, without incorporating effect-size precision, endpoint harmonization, or time from Phase 2 trial to Phase 3 trial. Arm count is also an imperfect proxy for trial complexity: multi-arm trials may improve decision-making when well designed, but poor per-arm power, inadequate multiplicity control, post hoc selection, or endpoint mismatch may limit their translational value.

Conclusions

Multiple Phase 2 trial treatment arms may make neurological trials more informative, and more arms did not reduce Phase 3 trial success in this individual study-level extraction. Nor did arm-count complexity show evidence of preventing successful translation in aggregate. The tested hypothesis is therefore best stated as follows: multiple arms are valuable when they improve the reliability of Phase 3 trial decisions without adversely affecting success factor. Strong biological rationale, clear Phase 2 trial efficacy, appropriate dose and population selection, and clinically valid endpoints were more informative than the number of treatment arms.

Data availability

The companion extraction workbook contains the row-level dataset, codebook, summary tables, descriptive predictor analyses, search strategy, and source URLs for each included trial pair. This revised manuscript uses the same row-level dataset and adds arm-count complexity proxies and hypothesis framing.

REFERENCES

1. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics 2019;20(2):273–86. doi: 10.1093/biostatistics/kxx069.
2. Harrison RK. Phase II and phase III failures: 2013-2015. Nat Rev Drug Discov 2016;15(12):817–8. doi: 10.1038/nrd.2016.184.
3. Moyer H, Mellett R, Vigneault K, et al. Prevalence and Impact of Bypassing or Overriding Phase 2 Trials in Neurologic Drug Development. Neurology 2024;103(1):e209533. doi: 10.1212/WNL.0000000000209533.
4. Ghosh P, Liu L, Mehta C. Adaptive multiarm multistage clinical trials. Stat Med 2020;39(8):1084–102. doi: 10.1002/sim.8464.
5. Wason JMS, Jaki T. Optimal design of multi-arm multi-stage trials. Stat Med 2012;31(30):4269–79. doi: 10.1002/sim.5513.
6. Research C for DE and. Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry. April 21, 2020. Accessed June 15, 2026. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry
7. Multi-Arm Multi-Stage (MAMS) platform trials. UCL Innovative Clinical Trials Unit. Accessed June 15, 2026. https://www.innovative-ctu.ucl.ac.uk/our-research/methodology/design/multi-arm-multi-stage-mams-platform-trials/
8. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013;14(5):365–76. doi: 10.1038/nrn3475.
9. Robertson DS, Choodari-Oskooei B, Dimairo M, et al. Point estimation for adaptive trial designs I: A methodological review. Stat Med 2023;42(2):122–45. doi: 10.1002/sim.9605.
10. Robertson DS, Choodari-Oskooei B, Dimairo M, et al. Point estimation for adaptive trial designs I: A methodological review. Stat Med 2023;42(2):122–45. doi: 10.1002/sim.9605.
11. Markey N, Howitt B, El-Mansouri I, et al. Clinical trials are becoming more complex: a machine learning analysis of data from over 16,000 trials. Sci Rep 2024;14(1):3514. doi: 10.1038/s41598-024-53211-z.
12. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in Early Alzheimer's Disease. N Engl J Med 2023;388(1):9–21. doi: 10.1056/NEJMoa2212948.
13. Mintun MA, Lo AC, Duggan Evans C, et al. Donanemab in Early Alzheimer's Disease. N Engl J Med 2021;384(18):1691–704. doi: 10.1056/NEJMoa2100708.
14. Kappos L, Li D, Calabresi PA, et al. Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. Lancet 2011;378(9805):1779–87. doi: 10.1016/S0140-6736(11)61649-8.
15. Hauser SL, Bar-Or A, Comi G, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. N Engl J Med 2017;376(3):221–34. doi: 10.1056/NEJMoa1601277.
16. Shults CW, Oakes D, Kieburtz K, et al. Effects of coenzyme Q10 in early Parkinson disease: evidence of slowing of the functional decline. Arch Neurol 2002;59(10):1541–50. doi: 10.1001/archneur.59.10.1541.
17. Parkinson Study Group QE3 Investigators, Beal MF, Oakes D, et al. A randomized clinical trial of high-dosage coenzyme Q10 in early Parkinson disease: no evidence of benefit. JAMA Neurol 2014;71(5):543–52. doi: 10.1001/jamaneurol.2014.131.
18. Wright DW, Kellermann AL, Hertzberg VS, et al. ProTECT: a randomized clinical trial of progesterone for acute traumatic brain injury. Ann Emerg Med 2007;49(4):391–402, 402.e1-2. doi: 10.1016/j.annemergmed.2006.07.932.
19. Skolnick BE, Maas AI, Narayan RK, et al. A clinical trial of progesterone for severe traumatic brain injury. N Engl J Med 2014;371(26):2467–76. doi: 10.1056/NEJMoa1411090.
20. Frerichs LM, Friedman DI. Galcanezumab for the prevention of migraine. Pain Manag 2021;11(2):101–12. doi: 10.2217/pmt-2020-0030.
21. Goadsby PJ, Reuter U, Hallström Y, et al. A Controlled Trial of Erenumab for Episodic Migraine. N Engl J Med 2017;377(22):2123–32. doi: 10.1056/NEJMoa1705848.
22. Finkel RS, Mercuri E, Darras BT, et al. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. N Engl J Med 2017;377(18):1723–32. doi: 10.1056/NEJMoa1702752.
23. Howard JF, Bril V, Vu T, et al. Safety, efficacy, and tolerability of efgartigimod in patients with generalised myasthenia gravis (ADAPT): a multicentre, randomised, placebo-controlled, phase 3 trial. Lancet Neurol 2021;20(7):526–36. doi: 10.1016/S1474-4422(21)00159-9.
24. Kappos L, Radue E-W, O'Connor P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. N Engl J Med 2010;362(5):387–401. doi: 10.1056/NEJMoa0909494.
25. Mercuri E, Deconinck N, Mazzone ES, et al. Safety and efficacy of once-daily risdiplam in type 2 and non-ambulant type 3 spinal muscular atrophy (SUNFISH part 2): a phase 3, double-blind, randomised, placebo-controlled trial. Lancet Neurol 2022;21(1):42–52. doi: 10.1016/S1474-4422(21)00367-7.
26. Howard JF, Utsugisawa K, Benatar M, et al. Safety and efficacy of eculizumab in anti-acetylcholine receptor antibody-positive refractory generalised myasthenia gravis (REGAIN): a phase 3, randomised, double-blind, placebo-controlled, multicentre study. Lancet Neurol 2017;16(12):976–86. doi: 10.1016/S1474-4422(17)30369-1.
27. Parkinson Study Group STEADY-PD III Investigators. Isradipine Versus Placebo in Early Parkinson Disease: A Randomized Trial. Ann Intern Med 2020;172(9):591–8. doi: 10.7326/M19-2534.
28. Cudkowicz ME, van den Berg LH, Shefner JM, et al. Dexpramipexole 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.
29. McColgan P, Thobhani A, Boak L, et al. Tominersen in Adults with Manifest Huntington's Disease. N Engl J Med 2023;389(23):2203–5. doi: 10.1056/NEJMc2300400.
30. Ginsberg MD, Palesch YY, Hill MD, et al. High-dose albumin treatment for acute ischaemic stroke (ALIAS) Part 2: a randomised, double-blind, phase 3, placebo-controlled trial. Lancet Neurol 2013;12(11):1049–58. doi: 10.1016/S1474-4422(13)70223-0.
31. Dávalos A, Alvarez-Sabín J, Castillo J, et al. Citicoline in the treatment of acute ischaemic stroke: an international, randomised, multicentre, placebo-controlled study (ICTUS trial). Lancet 2012;380(9839):349–57. doi: 10.1016/S0140-6736(12)60813-7.
32. Gajewski BJ, Statland J, Barohn R. Using Adaptive Designs to Avoid Selecting the Wrong Arms in Multiarm Comparative Effectiveness Trials. Stat Biopharm Res 2019;11(4):375–86. doi: 10.1080/19466315.2019.
33. Gajewski BJ, Statland J, Barohn R. Using Adaptive Designs to Avoid Selecting the Wrong Arms in Multiarm Comparative Effectiveness Trials. Stat Biopharm Res 2019;11(4):375–86. doi: 10.1080/19466315.2019.1610044.
34. Mitra-Majumdar M, Kesselheim AS. Reporting bias in clinical trials: Progress toward transparency and next steps. PLoS Med 2022;19(1):e1003894. doi: 10.1371/journal.pmed.1003894.