

Divergent Analytical Demands Across Clinical Trial Phases: A Comparative Practice Framework for Statistical Programming Resource Allocation

Radhika Aitha

Independent Researcher, USA.

Abstract: Statistical programming teams are often staffed as though the work performed at every clinical trial phase were interchangeable. Headcount, timeline, and study volume tend to drive resourcing decisions far more than the underlying nature of the analytical task itself does. We examined published regulatory guidance, technical practice literature, and documented industry reporting describing statistical programming activity across Phase I through Phase IV of drug development, comparing the analytical, validation, and regulatory demands characteristic of each phase against four resourcing-relevant dimensions. This study aims to determine whether a structured, phase-specific practice framework can better inform how programming resources, particularly seniority and specialization, should be allocated across a development program. Dataset complexity, validation burden, and regulatory exposure increase sharply and unevenly from Phase I through Phase III, where dual-programming validation practice alone can increase programming effort by 60 to 100 percent, then shift in kind rather than degree at Phase IV, where real-world evidence competencies begin to dominate. Generic, headcount-based resourcing approaches appear to systematically understate the seniority and specialization requirements of later-phase and post-marketing programming work. We outline the practical implications for how statistical programming managers plan staff assignments across a drug's lifecycle.

Keywords: Statistical programming, clinical trial phases, CDISC standards, resource allocation, drug development lifecycle, regulatory submission.

INTRODUCTION

A new drug moves from first human dose to routine clinical use through four regulatory-defined phases, each generating data that must become analysis-ready datasets, statistical outputs, and submission packages before it can inform the next stage of development (IntuitionLabs, 2024). The standardized formats underlying that transformation are themselves a relatively recent development: the Clinical Data Interchange Standards Consortium was founded in 1997, and the tabulation and analysis-dataset standards that now govern nearly every submission were not mandated until the FDA required their use for studies starting after December 2016 (CDISC, 2024; U.S. Food and Drug Administration, 2024). The statistical programmer sits at the point in this process where a single dataset derivation choice can determine whether a primary endpoint result withstands regulatory scrutiny, yet project planning still commonly treats the profession as one undifferentiated function.

Programming resource plans are typically built around headcount and calendar time: a given number of programmers assigned to a given number of studies over a given number of months. What such plans rarely capture explicitly is that the character of the analytical work, not merely its volume, changes from phase to phase. A Phase I pharmacokinetic dataset and a Phase III integrated summary of safety draw on the same underlying

skill set in name only. The former rewards speed and independent judgment under compressed timelines. The latter depends on a dual-programmer validation architecture and pooled-dataset reconciliation logic that a programmer trained exclusively in early-phase work may never have practiced (WJARR, 2025). Treating these as equivalent units of programming effort risks two failures at once: under-resourcing the phases carrying the greatest regulatory exposure, and over-specifying generalist staff for phases that would be better served by speed and flexibility.

The consequences of getting this wrong are not merely administrative. A Phase III submission delayed because dual-programming discrepancies were resolved too slowly, or a Phase IV real-world evidence study that fails FDA's data-quality expectations because it was staffed by trial-native programmers unfamiliar with claims-data conventions, both trace back to the same root cause: a resourcing model that did not distinguish between phases that look superficially similar on a project timeline but differ substantially in what they actually demand of a programmer.

This study aims to: characterize the analytical, validation, and regulatory demands documented in the literature and industry practice for each of the four clinical trial phases; compare these demands across a consistent set of dimensions to identify where they diverge most sharply; and translate the

resulting comparison into practical implications for how statistical programming managers allocate staff seniority and specialization across a development program.

MATERIALS AND METHODS

We conducted a structured synthesis of published regulatory guidance, technical practice literature, and documented industry reporting describing statistical programming activity across Phase I through Phase IV of clinical drug development. Sources were drawn from official FDA and ICH guidance documents, CDISC standards documentation, PHUSE technical conference papers describing current validation and automation practice, and peer-reviewed literature addressing estimands implementation, CDISC case studies, and real-world evidence methodology. Twenty sources were initially screened against these criteria; seventeen were ultimately retained in the reference list following verification that each was directly drawn upon in the analysis presented here.

Four comparison dimensions were selected before the literature review began, chosen specifically because they map onto the practical variables a programming manager must weigh when assigning staff, rather than onto dimensions of purely academic interest. Dataset complexity captures the structural and derivation complexity of the datasets a programmer must construct at a given phase. Validation burden captures the quality-control architecture required, including whether independent dual programming applies as a matter of practice rather than formal regulatory requirement. Regulatory submission exposure captures the degree to which phase-specific outputs are directly scrutinized in a regulatory filing, as distinct from outputs used only for internal decision-making. Required analytical specialization captures the degree to which phase-specific work depends on skills that are not transferable from other phases, which is the dimension most directly relevant to training and hiring decisions.

Each of the four phases was assessed against all four dimensions using the sources described above, and phase-level findings were then compared to identify points of convergence and divergence. Divergence was defined, for purposes of this comparison, as a dimension where the documented practice at one phase differs materially in either magnitude or kind from the practice documented at an adjacent phase. This

comparative structure, rather than a statistical test, forms the basis of the framework presented in the Results section.

Because this synthesis draws on documented practice and regulatory guidance rather than on primary clinical or operational data collected by the authors, no primary statistical analysis was performed. The contribution is analytical and comparative in nature, in the same sense that a structured literature synthesis in health services research is analytical: it organizes existing, verified evidence into a framework that did not previously exist in consolidated form, without generating new primary observations.

RESULTS

Phase I: Speed and Independent Judgment Under Compressed Timelines

Phase I first-in-human studies enroll comparatively small cohorts, typically twenty to one hundred subjects (IntuitionLabs, 2024), and operate under a compressed review cycle in which dose-escalation decisions are frequently made from near-real-time data review rather than a fully specified, locked analysis plan. The centerpiece of Phase I programming is pharmacokinetic dataset construction: concentration-time datasets that must correctly align dose administration timing with sample collection timing to support non-compartmental analysis parameters, including area under the concentration-time curve, maximum observed concentration, time to maximum concentration, and terminal half-life. Constructing these datasets correctly requires the programmer to reconcile two independently collected timing variables, dosing time and sampling time, in a way that later pharmacokinetic modeling depends on entirely; an error here does not surface as an obvious data anomaly but as a subtly distorted exposure parameter that may not be caught until modeling produces an implausible result.

Validation burden in Phase I is comparatively lighter than in later phases. The emphasis on speed for dose-committee decisions means programmers frequently work with less extensive protocol specification for any individual analysis than would be tolerated in a locked, pre-specified Phase III analysis plan. Regulatory submission exposure in Phase I is nonetheless foundational rather than minimal. The SDTM-formatted datasets and variable-naming conventions established at this stage typically persist through the remainder of the program, meaning early derivation decisions carry consequences well beyond the phase in which they

are made. A programmer establishing treatment-emergent flag logic or laboratory normal-range handling in Phase I is, in effect, setting a precedent that Phase II and Phase III programmers will inherit, often without revisiting the original rationale. A published academic use-case analysis of SDTM implementation outside a traditional sponsor-CRO setting reached a similar conclusion from a different angle: even when SDTM-compliant datasets are built without dedicated commercial infrastructure, as in an investigator-initiated academic trial, the same upstream data-standardization decisions determine downstream analysis quality, underscoring that this is a structural feature of SDTM-based programming rather than an artifact of any one organizational setting (JMIR Medical Informatics, 2023).

Phase II: Confounded Data and the Introduction of Efficacy Structures

Phase II shifts the study population from healthy volunteers to patients with the target disease, introducing confounding from disease state, concomitant medication, and prior treatment history that Phase I datasets did not need to accommodate. Efficacy analysis datasets appear for the first time in the development program, structured according to therapeutic area: tumor response datasets in oncology, disease activity score datasets in immunology, and comparable structures in other therapeutic areas. Validation burden increases correspondingly, since programmers must now build datasets that withstand scrutiny not only for accuracy but for fidelity to a more complex, protocol-specific statistical analysis plan than Phase I typically requires.

A defining feature of Phase II programming is the data cutoff and snapshot process supporting interim analyses. Many Phase II trials include planned interim looks at accumulating data, typically reviewed by an independent data monitoring committee, to decide whether the program should continue, stop for futility, or adapt its design and patient population ahead of Phase III (Cook *et al.*, 2025). Producing a frozen, reproducible dataset version at a specified cutoff, independent of ongoing data cleaning and database updates, is a technical requirement whose integrity functions as a direct safeguard against bias in the interim decision; guidance on interim analysis practice emphasizes that the analysis must be conducted on fully pre-specified, acceptably clean data to avoid introducing operational bias into a decision that is, by design, made before the study

has finished (PMC, 2024). This snapshot discipline has no meaningful Phase I analogue, since Phase I dose-escalation decisions typically operate on rolling, unlocked data rather than a formally frozen cutoff.

Phase II is also where the ICH E9(R1) estimands framework typically first requires substantive programming translation. An estimand specifies the population, endpoint, handling of intercurrent events such as treatment discontinuation or rescue medication use, and summary measure that a trial is designed to estimate (International Council for Harmonisation, 2019). Programmers must translate a chosen intercurrent-event strategy, whether treatment policy, hypothetical, composite, while-on-treatment, or principal stratum, into concrete dataset derivation rules determining which observations contribute to an analysis and how affected observations are handled (PHUSE/CDISC, 2023). Because each strategy produces a different analytical result from the same underlying data, this is a decision with no direct Phase I equivalent, and getting it right in Phase II establishes the analytical precedent the confirmatory Phase III analysis will follow.

Phase III: Maximum Validation Burden and Regulatory Exposure

Phase III represents the point of maximum convergence across all four comparison dimensions. A single Phase III program may comprise twenty or more parallel studies submitted together in one regulatory filing, with an ADaM library spanning thirty or more datasets, thousands of table, listing, and figure outputs, and a define.xml metadata file documenting tens of thousands of variable-level entries. Two features distinguish Phase III programming from the two phases preceding it.

First, integrated summaries of safety and efficacy require pooling data collected under different protocols, at different sites, and under potentially different adverse-event coding conventions into a single coherent analytical structure. Reconciling variable definitions across studies, aligning laboratory normal-range standards across sites and countries, and constructing a treatment-emergent flag that applies uniformly across studies with different dosing durations and comparator arms are reconciliation tasks with no direct Phase I or Phase II equivalent, since neither earlier phase requires harmonizing data across multiple independently conducted studies into one pooled structure. Regulators now expect this reconciliation to be

documented against the same CDISC-standardized backbone established at Phase I, a requirement that industry commentary has described as one of the more consequential shifts in the FDA's submission expectations over the past decade (Applied Clinical Trials Online, 2024).

Second, independent dual-programming validation, in which every primary efficacy output and key safety output is reproduced by a second programmer working from a separately written program and compared cell by cell against the original, becomes close to universal practice at this phase even though it is not a formal regulatory requirement. Recent practice literature estimates that dual programming increases programming resource requirements by roughly 60 to 100 percent relative to single programming, consuming 1.6 to 2.0 times the primary programming effort, and that a typical Phase III trial's 200 to 500 table, listing, and figure outputs extend development timelines by an additional 20 to 40 percent once iterative discrepancy-resolution cycles are included (WJARR, 2025). Regulatory exposure at this phase is correspondingly at its highest: these outputs are the direct evidentiary basis for an approval decision, and any unresolved discrepancy must be investigated and documented before the output can be accepted.

Phase IV: A Shift in Kind Rather Than in Degree

Phase IV programming does not simply continue the escalating pattern of Phase I through Phase III. It introduces a distinct competency requirement that the earlier phases do not build toward in any cumulative sense. Post-marketing commitment and requirement studies must meet the same CDISC and submission-quality standards as pre-approval studies, but the fastest-growing component of Phase IV work involves real-world evidence: analyses of electronic health records, medical claims databases, and patient registries that were designed for billing and care coordination rather than clinical trial analysis.

FDA's August 2023 guidance on real-world data and real-world evidence distinguishes interventional studies, which remain subject to Investigational New Drug regulations, from non-interventional studies that analyze data reflecting a marketed drug's use in routine practice (U.S. Food and Drug Administration, 2023). For non-interventional studies, the guidance identifies data provenance, variable completeness, and

generalizability as central quality dimensions, and requires that sponsors document audit trails from data extraction through final analysis dataset, describe the patient characteristics of the source population against the study population, and ensure that RWD and associated programming code are documented well enough for FDA to replicate the analysis independently. None of these requirements has a direct analogue in trial-native SDTM and ADaM programming, where the underlying dataset structure is standardized from the outset rather than reverse-engineered from a source never intended for research use. A 2023 tutorial on generating analysis-ready data from electronic health records for real-world evidence purposes describes this reverse-engineering step explicitly: producing a usable analysis dataset from EHR sources typically requires applying informatics techniques, including natural language processing over unstructured clinical notes, that have no counterpart anywhere in a trial-native SDTM-to-ADaM pipeline (Tan *et al.*, 2023).

Extracting a valid analytic cohort from a claims database, defining exposure and outcome variables from diagnosis and procedure codes rather than protocol-specified case report form fields, and applying propensity-score matching or other causal-inference methods to compensate for the absence of randomization together constitute a specialization that a programmer trained exclusively on trial-native data does not automatically possess, regardless of how many years of SDTM and ADaM experience that programmer has accumulated. Propensity-score methods estimate each patient's probability of receiving a given treatment based on observed baseline characteristics, then use that estimate to approximate the balance between treatment groups that randomization would otherwise provide; a cardiology-focused comparison of propensity-score and covariate-adjustment methods found this approach reduces observed confounding but cannot substitute for a randomized trial where unmeasured confounding remains uncontrolled (Elze *et al.*, 2017), and more recent methodological guidance aimed specifically at regulatory settings confirms the same limitation applies when propensity-score methods are used to support drug-approval decisions rather than purely academic causal-inference questions (Li *et al.*, 2023). This distinction matters directly for how confidently a Phase IV programmer's output can be characterized in a regulatory submission.

Table 1: Comparative analytical demand across clinical trial phases, by dimension (author's synthesis of cited sources).

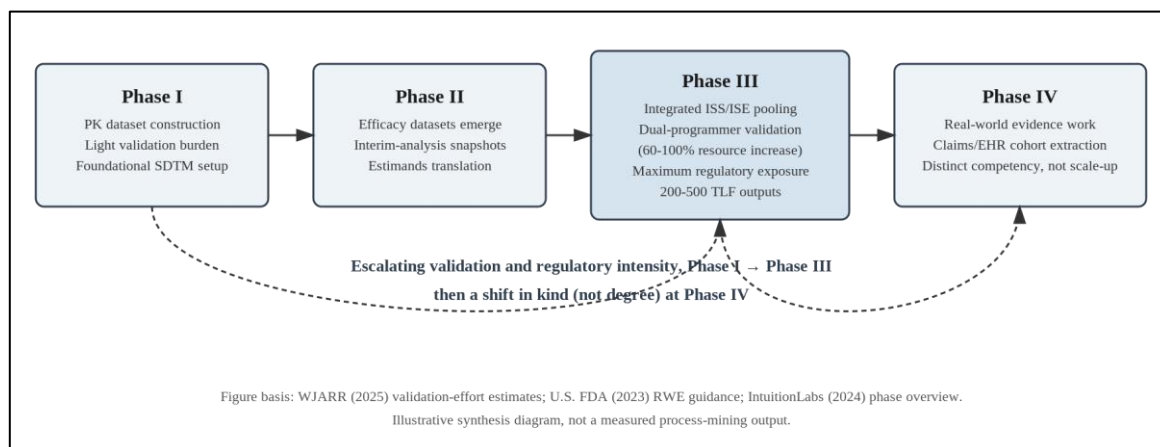
Dimension	Phase I	Phase II	Phase III	Phase IV
Dataset Complexity	Low-moderate (PK-focused)	Moderate (efficacy structures introduced)	High (integrated ISS/ISE pooling across studies)	Distinct (claims/EHR-derived, non-native structure)
Validation Burden	Lighter; speed-prioritized	Increasing; interim-analysis snapshot discipline	Highest; dual-programming near-universal (60-100% resource increase)	High but distinct; RWD audit-trail and provenance requirements
Regulatory Submission Exposure	Foundational (sets SDTM precedent)	Moderate (dose-selection evidence)	Maximum (direct approval evidentiary basis)	Variable (PMC/PMR and label-expansion dependent)
Specialization Required	PK/NCA modeling literacy	Estimands/ICH E9(R1) translation	Pooled-dataset reconciliation, validation leadership	RWE/causal-inference methods (propensity scoring, cohort extraction)

DISCUSSION

The comparison above indicates that analytical demand does not increase uniformly across the four phases. It increases sharply from Phase I through Phase III and then changes character altogether at Phase IV. This pattern carries a direct, and we would argue underappreciated, implication for how statistical programming managers plan staff assignments across a development program.

A resourcing model built primarily around headcount and study count implicitly treats a programmer's time as fungible across phases. That assumption understates the seniority requirement of Phase III work, where dual-programmer validation alone can consume 60 to 100 percent more effort than the primary programming task and extend timelines by a further 20 to 40 percent

once discrepancy resolution is included (WJARR, 2025), and it simultaneously risks assigning trial-native programmers to Phase IV real-world evidence work for which their SDTM and ADaM training provides only partial preparation. Both errors are costly in different ways. Under-resourcing Phase III risks the kind of validation shortfall that produces exactly the discrepancies the dual-programming framework exists to catch, arriving later in the timeline when they are more expensive to resolve. Over-assuming transferability into Phase IV risks a real-world evidence study that fails FDA's stated data-provenance and generalizability expectations not because the underlying data was inadequate, but because the programming team approached it with trial-native assumptions a claims database does not support.

**Figure 1:** Escalation of programming and validation intensity across clinical trial phases, culminating in a competency shift (not a scale-up) at Phase IV.

The workforce literature offers a complementary perspective on why this distinction matters more

acutely now than in earlier decades of drug development. Recent industry reporting on the

statistical programming labor market describes a widening divide between experienced programmers whose expertise is concentrated in SAS and an emerging workforce entering the field with stronger open-source language backgrounds, a divide that functions as a source of resourcing strain independent of any individual study's complexity (Phares, 2023). Layered onto the phase-specific demand differences described above, this suggests that programming managers are navigating two compounding resourcing variables at once: which phase a study occupies, and which generation of language proficiency a given programmer brings to it. A manager who

resources Phase III validation work with a programmer proficient in a newer open-source language but unfamiliar with the dual-programming reconciliation conventions accumulated in SAS-based validation practice over two decades may be introducing exactly the kind of discrepancy that dual-programming validation exists to catch, rather than avoiding it. This is not an argument against open-source adoption; it is an argument that language proficiency and phase-specific validation fluency are two separate competencies that a resourcing plan should track separately rather than treating as a single skill dimension.

Table 2: Validation framework requirements by clinical trial phase.

Phase	Typical Validation Practice	Basis
Phase I	Single programming; light QC given compressed dose-committee timelines	Practice literature (WJARR, 2025)
Phase II	Single programming with heightened QC around interim-analysis snapshots	Practice literature (WJARR, 2025)
Phase III	Independent dual programming near-universal; 60-100% resource increase; 1.6-2.0x primary effort	WJARR, 2025
Phase IV (RWE)	Documented audit trails, provenance and reproducibility requirements in place of dual programming	U.S. FDA, 2023

Therapeutic area specialization compounds the phase-based differences described here rather than operating independently of them. An oncology programmer's familiarity with the Response Evaluation Criteria in Solid Tumors, the categorical response scheme (complete response, partial response, stable disease, progressive disease) it defines, and the censoring conventions used to derive progression-free survival from that scheme via Kaplan-Meier estimation is most consequential precisely in Phase II and Phase III, where efficacy datasets and integrated summaries are constructed (National Cancer Institute/PMC,

2024). That same specialization contributes comparatively little direct value to a Phase I dose-escalation programmer's pharmacokinetic dataset work, where the analytical demands are pharmacological rather than disease-specific. Programming managers who treat therapeutic area experience as a single, portable credential, valuable everywhere a programmer is assigned, risk misallocating that expertise to phases where it adds comparatively little, while under-resourcing the phases where it is most consequential.

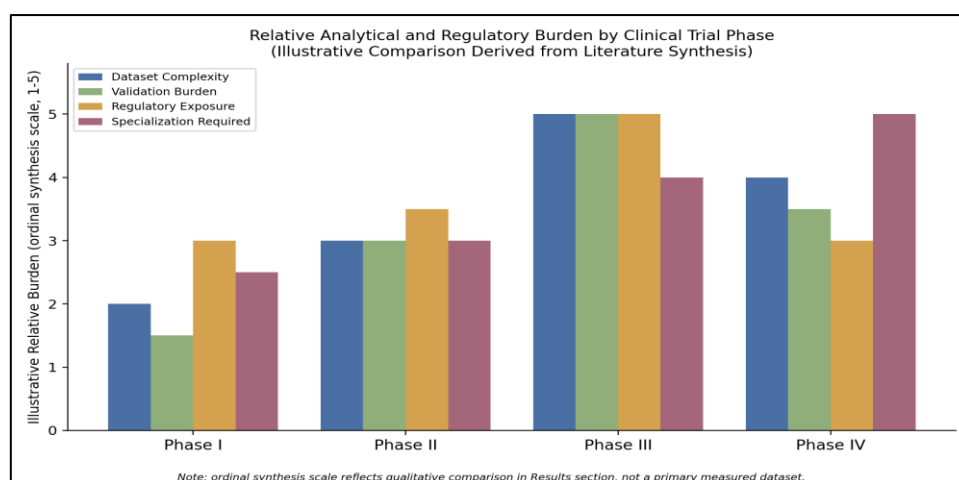


Figure 2: Relative analytical and regulatory burden by clinical trial phase — illustrative ordinal synthesis derived from the literature comparison in Results, not a primary measured dataset.

Two limitations qualify these conclusions and should inform how the framework is used in practice. First, this synthesis draws on documented practice literature and regulatory guidance rather than on primary time-and-motion data collected across programming teams. The comparative dimensions presented here are analytically derived from published sources rather than empirically measured against a controlled resourcing dataset, and the framework is best read as a structured basis for resourcing discussion rather than a validated staffing formula that specifies exact headcount ratios. Second, the real-world evidence competency described for Phase IV occupies a rapidly evolving regulatory area. FDA first articulated a formal framework for evaluating real-world evidence in regulatory decision-making in 2018 (U.S. Food and Drug Administration, 2018), and the specific technical requirements documented here reflect guidance current as of 2023 that has already superseded parts of that original framework and is itself subject to further revision as the agency's experience with real-world data submissions accumulates (U.S. Food and Drug Administration, 2023).

CONCLUSION

Statistical programming demand across the clinical trial lifecycle is not a matter of degree alone. It changes in kind as a development program moves from Phase I through Phase IV. Resourcing plans that allocate programming staff primarily by headcount and calendar time understate the seniority and validation infrastructure that Phase III work demands, and they understate the distinct, non-transferable competency that Phase IV real-world evidence work requires. A phase-aware resourcing framework, of the kind outlined here, offers programming managers a more defensible basis for staff assignment than volume-based planning alone, and provides a starting structure for the kind of resourcing conversation that headcount-based planning tends to foreclose before it begins.

Recommendations

Programming managers should treat phase transition points, not only study start dates, as resourcing decision points, and should evaluate real-world evidence competency as a distinct specialization requiring its own staffing and training track rather than an extension of trial-native programming skill.

CRedit Author Statement

Radhika Aitha: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.

ACKNOWLEDGMENTS

The authors used Claude (Anthropic, claude-sonnet model) to assist with manuscript drafting and language refinement. The authors take full responsibility for the accuracy, originality, and integrity of all content presented in this manuscript, and confirm that all data, analysis, and conclusions reflect their own independent work and verified sources.

REFERENCES

1. Nelson, B. "FDA binding guidance: a pivotal milestone for CDISC standards." *Applied Clinical Trials* 25.12 (2016): 42.
2. CDISC. "History of CDISC." Clinical Data Interchange Standards Consortium (2024).
3. Mossop, H., Walmsley, Z., Wilson, N., Agbeleye, O., Bardgett, M., Bevin-Nicholls, A., & Wason, J. M. "Practical guidance for conducting high-quality and rapid interim analyses in adaptive clinical trials." *BMC medicine* 23.1 (2025): 528.
4. Elze, M. C., Gregson, J., Baber, U., Williamson, E., Sartori, S., Mehran, R., & Pocock, S. J. "Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies." *Journal of the American College of Cardiology* 69.3 (2017): 345-357.
5. Ich, C. H. M. P. "E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials." *Proceedings of the international conference on harmonisation of technical requirements for registration of pharmaceuticals for human use*. (2019).
6. IntuitionLabs. "Clinical Trial Phases: A Guide to Phase I, II, III & IV." IntuitionLabs (2024).
7. Matsuzaki, K., Kitayama, M., Yamamoto, K., Aida, R., Imai, T., Ishida, M., & Suzuki, Y. "A pragmatic method to integrate data from preexisting cohort studies using the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model: case study." *JMIR medical informatics* 11 (2023): e46725.
8. Li, H., & Yue, L. Q. "Propensity score-based methods for causal inference and external data leveraging in regulatory settings: From basic

- ideas to implementation." *Pharmaceutical Statistics* 22.4 (2023): 721-738.
9. Ruchalski, K., Anaokar, J. M., Benz, M. R., Dewan, R., Douek, M. L., & Goldin, J. G. "A call for objectivity: Radiologists' proposed wishlist for response evaluation in solid tumors (RECIST 1.1)." *Cancer Imaging* 24.1 (2024): 154.
 10. PHUSE / CDISC. "Implementation of the ICH E9(R1) Estimands Framework Using Data Standards." PHUSE US Connect (2023).
 11. PMC. "Guidance on Interim Analysis Methods in Clinical Trials." PMC (2024).
 12. Phares, C. "Exploring the Skills Gap in Statistical Programming Organizations." *Pharmaceutical Technology* 18.8 (2023).
 13. Hou, J., Zhao, R., Gronsbell, J., Lin, Y., Bonzel, C. L., Zeng, Q., & Cai, T. "Generate analysis-ready data for real-world evidence: tutorial for harnessing electronic health records with advanced informatic technologies." *Journal of medical Internet research* 25 (2023): e45662.
 14. Unnamed author(s). "Validation Strategies for Clinical Trial Programming: A Comprehensive Review of QC, Dual Programming, and Automated Validation Checks." *World Journal of Advanced Research and Reviews* (2025).
 15. Mahendraratnam, N., Mercon, K., Gill, M., Benzing, L., & McClellan, M. B. "Understanding use of real-world data and real-world evidence to support regulatory decisions on medical product effectiveness." *Clinical Pharmacology & Therapeutics* 111.1 (2022): 150-154.
 16. FDA, U. "Framework for FDA's real-world evidence program." *Silver Spring, MD: US Department of Health and Human Services Food and Drug Administration* (2018).
 17. U.S. Food and Drug Administration. "Study Data Standards Resources." FDA CDER/CBER (2024).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:

Aitha, R. "Divergent Analytical Demands Across Clinical Trial Phases: A Comparative Practice Framework for Statistical Programming Resource Allocation." *Sarcouncil journal of Medical sciences* 5.10 (2026): pp 1-8.