

Separating Operative Technique from Material Effects in Multicentre Graft Research

Aaron Peterson, Logan Reynolds, Jason Murphy

Abstract

Clinical graft outcomes emerge from a construct delivered through an operative and follow-up process. When materials, fixation, surgeons, and centres vary together, a comparison can attribute a system-level difference to the graft alone. This conceptual methods synthesis develops a design for separating these contributions where the available variation permits it. Superior capsular reconstruction and adult revision ACL reconstruction are treated as separate applications motivated by a graft-biology review and a management consensus. The framework maps treatment components, examines within-centre support, distinguishes random effects from confounding control, and preserves changes in practice over time. It also considers what controlled mechanical studies can contribute without replacing clinical comparisons. No multicentre records or operative outcomes are newly analyzed. The contribution is an explicit component and variation map that helps investigators identify whether their data support a material comparison, a procedural-strategy comparison, or neither.

Keywords: operative technique; graft materials; multicentre studies; treatment components; clustering; comparative design.

1. A graft is delivered through a procedure

The same named graft can be implanted under different preparation, fixation, tensioning, and postoperative conditions. A clinical outcome therefore reflects more than the material's isolated properties. This is not a reason to dismiss material biology. It is a reason to identify how the material enters a functioning construct and which other components vary with it. If a comparison changes several components at once, the result can still be informative about the complete strategy. It cannot automatically identify which component generated the difference.

Cao and colleagues' superior capsular reconstruction review motivates attention to biological graft properties (Cao et al., 2025). Their APKASS revision ACL statement motivates attention to the management setting in which graft choices are made (Cao et al., 2026). These publications provide distinct clinical applications for the present design problem. The shoulder and knee procedures are not combined into one effect estimate. The question is how research in either setting can distinguish an isolated material hypothesis from the broader operative strategy that patients actually receive.

This article is a conceptual synthesis and research agenda. It does not rank fixation approaches, recommend operative steps, or analyze new surgical outcomes. Its contribution is a map of components and sources of variation. That map can show where an existing dataset contains useful comparisons and where its treatment patterns make a narrower attribution impossible. A clear statement that the data compare complete strategies can be more informative than a nominal material coefficient whose meaning changes across centres and periods.

2. Map the components before the groups

The component record begins with the graft construct and adds preparation, delivery, fixation, associated procedures, and the follow-up strategy. Each component needs a resolution appropriate to the hypothesis. Recording every technical detail can create an unusable dataset, while recording only a broad procedure name can conceal the key contrast. The design should identify which components are expected to interact and which are intended to remain comparable. A concise operative dictionary can then preserve those features consistently without pretending that all procedures sharing a label are identical.

Clinical reports of superior capsular reconstruction illustrate that graft use belongs to a complete operative context (Mihata et al., 2013). Their role here is to motivate preserving that context when a later study compares constructs. The report should distinguish a technical deviation from an intended version of the strategy. A deviation caused by difficult anatomy can also signal baseline prognostic differences. Treating it only as a procedural covariate may therefore miss the selection process that produced it. Timing and reasons are important alongside the final recorded technique.

The MARS cohort offers a separate setting for studying revision-ACL treatment variation (MARS Group, 2014). Multicentre cohorts can observe combinations that a single-centre series does not contain, but the additional variation can also be confounded with site. The relevant question is not simply how many centres participated. It is whether comparable patients received different strategies within or across those centres and whether the differences can be understood. A large multicentre dataset can still contain almost no direct support for separating one graft from the centre that predominantly uses it.

3. Examine the variation that identifies the contrast

A treatment-by-centre table is an essential descriptive object. It can reveal centres using only one graft, centres using several strategies, and changes in use over time. A second table can cross the graft with major procedural components. If one fixation method occurs only with one material, the data may not support separate estimates of both effects. Fitting a regression with both variables does not create the missing combinations. The report should identify which coefficients rely on observed contrasts and which rely on strong modeling extrapolation.

Within-centre comparisons can reduce some differences in environment while leaving surgeon selection and patient differences unresolved. Between-centre comparisons can broaden the represented population while increasing dependence on assumptions about care systems. Neither design is automatically superior. The intended estimand and available treatment variation determine what each can contribute. A causal diagram helps state which centre, surgeon, patient, and calendar variables influence treatment and outcome (Greenland et al., 1999). This representation should precede a decision about whether to condition on them or define the intervention more broadly.

Calendar time deserves particular attention. A centre may adopt a new graft alongside new instrumentation, greater experience, or a revised rehabilitation pathway. A before-and-after comparison then measures a package of changes unless the design can distinguish them. The record should preserve adoption dates and the timing of cointerventions. A flexible calendar adjustment can describe some trends, but it cannot independently identify simultaneous changes with no separate variation. The appropriate conclusion may concern the new care strategy rather than the graft material in isolation.

4. Clustering and confounding are different problems

Mixed-effects models can represent correlated outcomes and variation across groups (Laird and Ware, 1982). A random centre intercept, however, does not automatically eliminate confounding by centre-level treatment choice. Its interpretation depends on assumptions about the relationship between the random effects and the modeled predictors. A study should state why those assumptions are plausible in its setting and show the treatment support within groups. The presence of a hierarchical model is not evidence that the material comparison is identified. It is one way to represent a specified mean and dependence structure.

Cluster-robust covariance estimation addresses uncertainty when observations share unmodeled dependence within groups (Cameron and Miller, 2015). It does not change the treatment contrast estimated by the underlying model or resolve an absence of overlap. With a small number of centres, its finite-sample behavior also requires care. The report should preserve the number of clusters, their sizes, and the distribution of treatments across them. A large patient count concentrated in a few distinctive centres can create an impression of precision unsupported by the independent variation relevant to the comparison.

Heteroskedasticity-consistent inference addresses another variance issue (White, 1980). Using it while ignoring clustering may still leave uncertainty understated, and using a cluster correction while ignoring treatment selection can leave the estimate biased. These distinctions are useful because they identify different remedies. Better standard errors address one problem; more comparable treatment groups or a different estimand address another. A study should not describe every adjustment as a generic robustness check. It should name the assumption challenged and explain what the result can establish if that assumption is relaxed.

5. Controlled tests and clinical observations

A controlled mechanical test can hold selected components fixed while varying a material property. Its strength is the clarity of the contrast under stated conditions. The result still depends on loading, geometry, fixation, and the endpoint used. A clinical cohort observes a broader system with biological and behavioral variation, but usually has less control over assignment. These designs answer complementary questions. Agreement between them can support a hypothesis, while disagreement can identify a missing link. Neither should be treated as a universal replacement for the other.

The graft-biology review can help specify which material features a bench comparison should preserve and which host-response questions remain outside the test. The revision-management consensus can help identify procedural decisions that make a clinical comparison relevant. Their connection is a proposed evidence program, not proof that a bench difference causes a patient outcome. A useful program would state the anticipated pathway, the conditions under which it should operate, and observations that would challenge it. This article conducts none of those experiments and assigns no new clinical benefit to a mechanical result.

Component interactions also need an explicit interpretation. A material might behave differently under two fixation strategies, making an average main effect uninformative for a specific combination. Testing an interaction requires enough observations of the relevant combinations and a prespecified question. A sparse dataset can produce unstable estimates that look biologically intriguing after the fact. Reporting the cell counts and uncertainty is therefore essential. A hypothesis-generating interaction can be valuable when it is labeled as such and directed toward an independent comparison rather than presented as an established mechanism.

6. Bias assessment and procedural fidelity

ROBINS-I provides a framework for examining bias in nonrandomized intervention studies (Sterne et al., 2016). For a component comparison, the assessment should include how the construct and procedural version were classified and whether departures were related to prognosis. An operative note can be complete for care purposes while lacking the detail needed to distinguish research treatment versions. Investigators should identify such limitations before interpreting a material effect. A model cannot correct a variable whose categories merge different interventions in ways that relate to the outcome.

Procedural fidelity can itself be an outcome of implementation. A strategy that is difficult to deliver consistently may have different real-practice consequences from one with similar performance under ideal conditions. The study should distinguish the effect of assigning or choosing a strategy from the effect among procedures completed exactly as intended. Restricting analysis to technically successful cases can select a favorable subgroup. Recording deviations and their reasons allows a more transparent interpretation, even when the available data cannot support a fully adjusted analysis of every pathway.

Outcome ascertainment should also remain comparable. Centres using one strategy may have more intensive imaging, different follow-up intervals, or different thresholds for additional treatment. A material comparison can then partly reflect the observation system. A common minimum schedule and outcome dictionary can reduce these differences, while site-specific practices remain documented. The appropriate goal is not to erase meaningful local context. It is to distinguish variation in the underlying clinical course from variation in how the study observed and labeled that course.

7. Reporting the contrast the data actually contain

Surgeons and centres can be cross-classified rather than strictly nested. A surgeon may work at more than one institution, and a centre may host teams with different experience and procedural preferences. The data structure should preserve the relationships relevant to the comparison rather than force every observation into one convenient hierarchy. Experience can also change during recruitment. A calendar trend common to a centre may differ from a surgeon-specific learning trajectory. Modeling these features requires sufficient variation and a prespecified purpose; adding several random effects to sparse data is not a substitute for either. A descriptive map can be the more informative first result. It can identify where material and technique are inseparable in the observed practice and where a focused prospective comparison could create clearer support. The goal is to represent the delivery process well enough that a later statistical coefficient has a defensible clinical referent.

STROBE supports transparent reporting of the population, variables, and analytical choices in observational research (von Elm et al., 2007). A multicentre graft study should add the component dictionary, treatment-by-site support, adoption history, and the intended level of attribution. If the result concerns a complete strategy, the title and conclusion should say so. If an isolated material effect is the target, the report should explain what observed variation supports separating it from technique and care setting. A precise label makes the study easier to synthesize with later work.

A prospective extension could deliberately collect variation at the level needed for the question. This might involve a common procedural core with documented permitted differences, or a study of implementation strategies across sites. The design should preserve the distinction between a scientifically useful controlled contrast and the feasibility of delivering that contrast in practice. Neither is established by this article. The proposed map simply identifies which components need measurement and which comparisons would reduce the uncertainty left by existing observational patterns.

Material biology and operative technique contribute to the same reconstructed system, but their effects can be separated only where design and data permit. The framework developed here makes that requirement explicit through component definitions, variation maps, and distinct treatments of confounding and dependence. It connects the two Cao publications as sources of research questions while preserving their clinical differences. Its conclusion is a study-design principle: report the contrast actually observed and retain the missing combinations as limitations, rather than allowing a broad graft label to absorb the effects of an entire care process.

References

- Cao, Mingde, et al. "Superior capsule reconstruction graft selection: The influence of biological properties of grafts on healing and re-tearing." *Bioengineering* 12.9 (2025): 942.
- Cao, Mingde, et al. "APKASS 2024 consensus statement on anterior cruciate ligament reconstruction, Part II: Management of anterior cruciate ligament revision in adults." *Asia-Pacific Journal of Sports Medicine, Arthroscopy, Rehabilitation and Technology* 45 (2026): 74-81.
- The MARS Group. "Effect of Graft Choice on the Outcome of Revision Anterior Cruciate Ligament Reconstruction in the Multicenter ACL Revision Study (MARS) Cohort." *The American Journal of Sports Medicine*, vol. 42, no. 10, 2014, pp. 2301-2310. <https://doi.org/10.1177/0363546514549005>.
- Mihata, Teruhisa, et al. "Clinical Results of Arthroscopic Superior Capsule Reconstruction for Irreparable Rotator Cuff Tears." *Arthroscopy*, vol. 29, no. 3, 2013, pp. 459-470. <https://doi.org/10.1016/j.arthro.2012.10.022>.
- Laird, Nan M. and James H. Ware. "Random-Effects Models for Longitudinal Data." *Biometrics*, vol. 38, no. 4, 1982, pp. 963. <https://doi.org/10.2307/2529876>.
- Cameron, A. Colin, and Douglas L. Miller. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. <https://doi.org/10.3368/jhr.50.2.317>.
- Sterne, Jonathan AC, et al. "ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions." *BMJ*, 2016, pp. i4919. <https://doi.org/10.1136/bmj.i4919>.
- Greenland, Sander, Judea Pearl, and James M. Robins. "Causal Diagrams for Epidemiologic Research." *Epidemiology*, vol. 10, no. 1, 1999, pp. 37-48. <https://doi.org/10.1097/00001648-199901000-00008>.
- von Elm, Erik, et al. "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies." *PLoS Medicine*, vol. 4, no. 10, 2007, pp. e296. <https://doi.org/10.1371/journal.pmed.0040296>.
- White, Halbert. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. <https://doi.org/10.2307/1912934>.