Standard Operating Procedure

Biospecimen Collection & Lab Partnership Management

Guidance for study PIs and coordinators working with research labs

Version 1.0 · April 2026

PURPOSE & SCOPE

As biomarker collection becomes increasingly central to population health and aging research, the success of studies depends on strong, well-coordinated collaborations between study teams and laboratories.

This SOP provides standardized guidance for population study teams planning and executing biospecimen collection and managing lab partnerships. It covers how to engage labs effectively, prepare for collection, support the lab during processing, and conduct your own quality control on returned data. Developed from collective experience shared at the CBPH Population Study-Lab Partnerships Meeting, April 2026.

Who this is for: Study PIs, data managers, field coordinators, and program staff working on population-based longitudinal studies, cohort surveys, and multi-site research that involves biospecimen collection and laboratory analysis.

SECTION 1 · STUDY DESIGN

1.1 Engage your lab partner at the proposal stage

This is the single most impactful thing a study team can do. Labs consulted before budgets are finalized can flag what's feasible, what sample volumes each assay requires, what storage conditions are necessary, and what decisions made at design stage will affect data quality years later.

- Include lab collaborator as a named subcontract in the grant where possible
- Share your scientific aims, not just your biomarker list
- Ask what specimen type, volume, and storage each proposed assay requires
- Ask what bridging samples you'll need to hold for longitudinal comparisons
- Ask what metadata the lab will need for plate randomization (age, sex, key vars)

⚠ The most common mistake

Copying a biomarker list from another study without understanding what each assay requires is the single most frequent source of downstream problems. Labs should be involved before any list is finalized.

1.2 Get multiple bids — and evaluate them carefully

Cheapest is rarely best. The price difference between academic and commercial labs often reflects what you actually get: commercial labs may be cheaper but typically return end products without transparency into QC, methodology, or raw data. Importantly, you cannot expect a fee-for-service arrangement to have the same qualities as a full scientific-lab partnership.

- Ask for references from academic longitudinal studies of comparable scale
- Ask each lab to itemize cost by component: assay, QC, analytical support
- Ask what documentation and raw data you will receive
- Ask how the lab handles platform changes and longitudinal continuity
- Ask what happens to your samples when the grant ends

1.3 Set collaboration terms in writing before work begins

Misaligned expectations are the biggest source of study-lab friction. A clear written agreement at the start prevents most of it.

- Scope of work and what is explicitly excluded
- What analytical support is included vs. charged separately
- Authorship expectations for papers using the assay data
- Data access: what can the lab access, and through what mechanism?
- Cost structure broken down by component
- Communication protocol: who is the point of contact, expected response time
- What happens if the lab changes platforms or key personnel
- Sample storage policy and what happens at grant end

SECTION 2 · BIOSPECIMEN COLLECTION

2.1 Pilot your collection protocol in your target population

Collection protocols developed in controlled settings often fail in the field, especially in older adult populations. Pilot with your actual target population before large-scale deployment.

- Test failure rates in your specific population (e.g., older adults, rural settings, non-clinical environments)
- Go beyond manufacturer instructions — develop supplementary field protocols (e.g., for TASSO devices it was found after testing blood flow mobilization techniques that standard instructions are insufficient for many participants – detailed guidance available from gero.usc.edu/cbph/biomarker-resources/)
- Assess how much sample you reliably collect across the distribution of your participants
- Plan for what happens when collection fails — have backup protocols ready

2.2 Train field staff systematically

Collection quality varies substantially by collector. Systematic training and refresher protocols matter — especially for studies that span years or use multiple field sites.

- Have collectors demonstrate collection technique before going into the field
- If possible, invite the lab to demonstrate what a good sample looks like
- Develop clear visual guides for acceptable vs. unacceptable samples
- Log collector ID on every sample — this is essential for QC later
- Run refresher training at wave transitions or after extended breaks

2.3 Track pre-analytical factors systematically

Pre-analytical factors are the biggest source of lab error (approximately 75%). They need to be tracked prospectively, not reconstructed after the fact.

- Collection date and time
- Collector ID and site
- Device lot number
- Any deviations from protocol — write them down at the time, not later
- Shipping carrier, date shipped, and ambient conditions
- Participant notes relevant to collection (fasting status, medications, hydration, difficulty with collection)

This data is your QC

Pre-analytical tracking variables become covariates in downstream analyses. A collector effect, a lot-number effect, or a shipping delay effect can only be detected and corrected if you tracked it. It cannot be reconstructed later.

SECTION 3 · LABP ROCESSING

3.1 Insist on embedded QC samples — and budget for them

Embedded QC samples protect your data. They are the mechanism by which problems are caught before they propagate across a full data collection. Study teams typically don't know to ask for them and don't want to pay the overhead – but this can be the key to a successful study!

- Request blinded duplicates, replicates, and reference standards in every batch
- Ask to receive QC sample performance data in the return package
- For longitudinal studies: consider creating your own bridging sample set from pooled samples held across waves
- Budget approximately 5–10% overhead for QC samples in your grant

3.1 Check in regularly — especially when there is no news

The lab is managing many projects. A brief check-in keeps your study visible and gives you information you need to manage funders and participants. Silence from either side is where problems can arise.

- Establish a regular cadence for updates at the project outset
- If you haven't heard anything in 2–3 weeks, send a brief check-in
- Share information relevant to the lab: collection delays, participant issues, changes in study design
- Give the lab enough scientific context to flag problems you might not anticipate

3.1 Give the lab scientific investment where possible

Labs that are involved only as sample processors may not be full invested and engaged (e.g., lab is given samples and expected to return results, but excluded from the science). Labs that have something scientifically meaningful to contribute stay invested.

- Where feasible, offer co-authorship on papers using the assay data
- Allow the lab to use your population data for methods validation if useful
- Include the lab in scientific discussions about what the results mean
- Ask the lab what questions they find interesting in your data — they may see things you don't

SECTION 4 · DATA RETURN

4.1 Always request the raw data

Never accept only derived outputs, composite scores, or normalized values without the underlying raw data. Standards for processing and normalization change over time. Studies that only received derived metrics have found themselves unable to reprocess when methods were revised.

- Raw signal data (intensities, Ct values, raw counts) before normalization
- Processed data with the normalization method documented
- Derived outputs (clocks, composite scores) as separate files
- Full QC report with sample-level flags and QC sample performance
- Processing log documenting every decision made during analysis

4.2 Conduct your own QC independent of the lab's

The lab's QC addresses internal processes. Your QC addresses whether the results make sense given everything else you know about your sample. These are different questions and both are necessary.

- Check distributions by age, sex, race/ethnicity, education, health status — do they look right?

- Compare to prior waves, related measures, or published reference ranges
- Cross-check against survey responses where possible (e.g., self-reported health, medication use)
- Look for outliers, implausible values, and missing data patterns
- Examine flagged samples from the lab — what share of your sample do they represent?
- Run analyses with and without flagged samples to assess sensitivity

4.3 Structure multi-group QC for omics data

For complex omics data (methylation, RNA-seq, proteomics), the study-side QC process requires a structured approach. This applies regardless of what QC the lab has already performed.

- Run QC with at least two, ideally three, independent analysis groups simultaneously
- Do not share code between groups until a discrepancy is diagnosed — share only conceptual steps
- Circulate a living document tracking decisions, filters applied, and sample/probe counts at each step
- Use known biological associations (sex-linked expression, established phenotype links) to adjudicate normalization choices
- Document every decision, even when there is no established standard — the goal is documented decisions, not perfect ones

SECTION 5 · DATA ANALYSIS

5.1 Treat platform and batch effects as primary analytic concerns

Platform differences and batch effects are not nuisances—they are often among the largest sources of variation in biomarker data.

- Include platform, batch, and assay version as covariates where appropriate
- Use bridging or replicate sample to estimate systematic bias and variance differences.
- Apply normalization methods (e.g., batch correction) cautiously; confirm you are not removing real biological variation.
- Check whether exposure or outcome variables are confounded with batch

5.2 Use known biological relationships as validation checks

Biomarker data should reproduce well-established biological patterns. When they do not, the default assumption should be a measurement or processing issue.

- Check expected relationships with characteristics that have well-known biomarker distributions (e.g., age, sex, education)

- Compare distribution to prior waves, other population studies, or known references ranges
- Verify internal consistency by examining correlations among related biomarkers

> ⚠ **If you are not seeing expected patterns**
> *Investigate assay performance, examine QC flags and batch effects, and review sample handling and preprocessing. This is also a good time to incorporate information on collection and environmental context into analysis. Pre-analytic conditions are often important sources of confounding – examine variables related to collection and handling of biospecimen data, such as shipping time and temperature, as well as potential systemic variation by site/field team.*

5.3 Be explicit about measurement scale and interpretation

Misinterpretation often arises from unclear assumptions about what the data represent.

- Determine whether measures are in absolute or relative units (do not compare relative values across studies without calibration)
- For longitudinal analyses confirm that scaling is consistent across waves and document any data transformations (e.g., log, z-score) and justify their use

5.4 Use sensitivity analyses to assess robustness

Given the number of potential sources of variation, no single analytic specification should be treated as definitive. Thus, multiple analytic approaches should be used.

- Analyses should be conducted with and without sample that were flagged by the lab (e.g., quality concern, undetectable value), with alternative normalization approaches, and with and without batch adjustment
- For longitudinal data test consistency of findings across waves
- Make sure to document all sensitivity analyses, even when results are unchanged

5.5 Document analytic decisions and assumptions

Transparency in analysis is critical given the number of judgment calls required. Document all aspects of data analysis. The goal is not a “perfect” analysis, but a fully traceable one.

- Document normalization methods, handling of batch/platform differences, inclusion/exclusion criteria applied to samples, and any other QC decisions
- Record rationales for chosen transformations and sensitivity analyses
- Maintain a versioned record of code, datasets, and outputs