White paper

Olink™ Sample Index: A tool to optimize data quality by addressing pre-analytical variation

Olink Sample Index empowers large-scale proteomic studies

Capturing dynamic changes in the proteome, provides deeper insights into disease mechanisms, enables the discovery of reliable biomarkers, and supports the development of targeted therapies. It is this dynamic nature of the proteome that makes it so informative, yet also susceptible to influence from sample handling.

After collection, biological samples remain biochemically active, with processes such as protein aggregation, denaturation, and hemolysis potentially altering protein levels. The temperature at which samples are handled and the time between collection and centrifugation influence both the rate and nature of these proteomic changes. Protein measurements made using accurate, sensitive technologies such as Olink therefore reflect both the original biological status and any additional effects that may occur post-processing.

Pre-analytical variation refers to the variability introduced during sample collection, handling, processing, and storage. Olink provides an additional white paper titled "*Pre-analytical variation in protein biomarker research*" (1), which offers practical guidelines to reduce pre-analytical variation.

The Olink Sample Index (OSI) is available for all Olink™ Explore HT users, and is a sample integrity & variation index that provides a standardized way to assess whether sample handling differences could influence interpretation of protein data. OSI provides a structured evaluation of how delay of centrifugation and handling temperature impact proteomic measurements and the structure of the data. OSI helps users isolate and reduce the impact of pre-analytical variation in proteomics results, remove technical artifacts, and reveal true biological signals without discarding valuable data.

RELATED WHITE PAPER

The design and validation of the OSI models, and technical transparency on how the OSI framework was developed and evaluated are described in the related technical white paper *Design and validation of Olink™ Sample Index* (2), available on the Olink website.

This white paper introduces OSI from a practical perspective: what the OSIs mean, how they are derived, and how to use them in downstream analysis.

IMPORTANT

OSI is not a pass–fail measure of sample quality, but it provides users with an actionable way to explore and account for specific sources of pre-analytical variation in their dataset. Evaluating the impact of pre-analytical variation, enables an informed decision to correct the model for these effects if needed, increasing confidence in the results either way.

Why is sample handling important?

How samples are handled before analysis can make a difference to the quality of the results. Ideally, a well-designed study ensures that all samples are collected and handled in the same way to prevent technical artifacts caused by differences in the sample collection process. Consistent handling is key to reliable data. However, in real-world scenarios, this is not always feasible. This is especially true for large-scale studies or those involving multiple collection sites, where variations in procedures are more likely to occur.

Sample handling inconsistencies can introduce variability in protein measurements, making it harder to trust the results. These inconsistencies can compromise study design and create apparent trends that lead to misleading conclusions. For a well-planned study, pre-analytical variation is not usually the most critical covariate, but it can subtly influence downstream results and is therefore an important factor to consider alongside standard variables like age or sex.

OSI provides an actionable solution to these potential issues, helping users ensure best-practice data analysis and paving the way for reliable findings that deliver high value from their studies.

What is the OSI feature, and why use it?

The OSI feature provides a sample-specific metric that reflects its handling in terms of sample preparation temperature and time

from sample draw to centrifugation. The OSI model is currently validated for EDTA plasma and is available in the data output files from NPX™ Map (version 2.0 and higher) for all Olink Explore HT users, giving them the flexibility to assess its utility in their analysis. This empowers users to make informed decisions about potential variation caused by deviations in sample collection, providing deeper trust in the acquired data.

Lab-recorded sample-handling data is always preferred when available, but in the absence of such data, OSI can be used as a substitute to estimate handling conditions.

OSI variables

- **Time to Centrifugation**
  The elapsed time between sample collection and sample centrifugation used to separate plasma from cells.
- **Preparation Temperature**
  The temperature at which the sample is handled prior to centrifugation.
- **OSI Summary (continuous)**
  A combination of OSI for Time to Centrifugation and Preparation Temperature. A function of the two individual OSIs that aims to provide an overview of pre-analytical consistency in the data.
- **OSI Category**
  A categorical representation of OSI Summary that aims at simplifying interpretation of OSI for non-expert users. OSI Category is encoded as integer numbers ranging from 0–4, see Table 1.

Table 1. OSI Category: categorical representation of OSI Summary.

OSI Category	OSI Summary
0	Missing value
1	0–0.25
2	0.25–0.5
3	0.5–0.75
4	0.75–1.0

Machine learning model

OSI applies machine learning models to detect patterns indicative of pre-analytical effects related to Time to Centrifugation and Preparation Temperature, and to predict sample handling conditions during collection. The resulting predictions are scaled and standardized to a 0–1 range for interpretability. For Time to Centrifugation an OSI close to 0 indicates centrifugation very soon after sample collection, while an OSI closer to 1 indicates a long delay. For Preparation Temperature, an OSI close to 0 indicates handling on ice (4°C), while an OSI closer to 1 indicates handling at room temperature.

Further details about the model and the validation is available in the technical white paper (2).

How can OSI elevate or increase confidence in your results?

OSI can help explain why certain samples deviate from the overall sample cohort.

Visualization

As a first step during the data quality assessment process, there are visualizations that can help detect deviations in pre-analytical sample handling:

- **Distribution of OSI**
  Illustrates how OSIs are distributed along the range 0–1. Lack of a clear "peak" in the distribution warrants further investigation of pre-analytical variation effects.
- **Principal Component Analysis (PCA) colored by OSI**
  Overlaying OSI (numerical or categorical values) in PCA visualizations allows users to determine effects from pre-analytical variation on their data. Alignment of OSI with the top PCA components indicates an impact of pre-analytical variation, whereas the absence of such patterns suggests consistent pre-analytical sample handling.

OSI increases accuracy in your data analysis

When there is no or only minimal impact from pre-analytical variation on the data, the statistical analysis may proceed without accounting for OSI. However, when OSI indicates impact from pre-analytical variation in the data, it is strongly recommended to incorporate the corresponding OSI variable into the statistical model (e.g. linear regression). This matters because unaccounted pre-analytical variation can distort relationships between variables and lead to biased results or incorrect conclusions by acting as an uncontrolled source of variability. Instead, adjusting for these factors reduces noise, improves accuracy and strengthens confidence in the results.

Proteins that remain significant after accounting for pre-analytical factors typically exhibit clearer disease associations once sample handling differences have been considered.

WORKFLOW OVERVIEW

Visualizations: Use distribution plots for OSI, and/or overlay OSI to PCA visualization to determine the value of OSI in downstream statistical analysis.

Statistical models: Include one or both of the OSIs for Time to Centrifugation or Preparation Temperature as covariates in the statistical analysis to account for the effects when warranted.

The following studies illustrate when and how OSI can be effectively applied. These examples include one study with minimal pre-analytical variation and another with higher variation.

Example 1: A study with consistent sample handling

This study identified proteins that can distinguish benign ovarian tumors from ovarian cancer using plasma samples from 233 women (3). Of these, 80 had benign tumors and 153 had ovarian cancer. Participants ranged from 22 to 88 years old (median age 63).

Impact of pre-analytical variation

The distribution of sample handling differences (measured by OSI) and their impact on overall data patterns were assessed using distribution plots (Fig. 1) and PCA plots (Fig. 2) to see whether these differences are reflected in the structure of the data.

- The Time to Centrifugation OSI showed some variation in sample handling as indicated by the distribution plot (Fig. 1), but did not represent noticeable patterns in the PCA projection (Fig. 2).
- The Preparation Temperature OSI was generally low, with only a few samples showing higher values, as indicated by the distribution plot (Fig. 1). A small subset of samples with elevated Preparation Temperature OSI showed a mild tendency to sub-cluster in the PCA projection colored by Preparation Temperature OSI. However, in the PCA projection based on combined Time to Centrifugation and Preparation Temperature OSI, these samples no longer formed sub-clusters but were dispersed across the distribution.

This suggests that Time to Centrifugation and Preparation Temperature have little to no impact on this dataset, and downstream analysis can proceed without accounting for these parameters. For demonstration purposes, we developed two models: one that does not account for these factors and one that does.

Results from statistical models

Protein levels were compared between benign and ovarian cancer groups using two different models:

- Model A: Corrected for subject age (standard practice in statistical analysis).
- Model B: Corrected for subject age, OSI Time to Centrifugation and OSI Preparation Temperature.

Comparison between the significant findings in the models is illustrated in the Venn diagram in Fig. 3b.

In summary, 81.1% of the significant proteins in model A (1,364 out of the 1,681) remained significant in model B, indicating great replication between models. Meanwhile, only 317 proteins were significant only in model A, and 32 were only significant in model B.

The effect sizes from both models were highly similar as shown by the correlation plot in Fig. 3a. This indicates that the observed differences are driven by cancer status rather than being

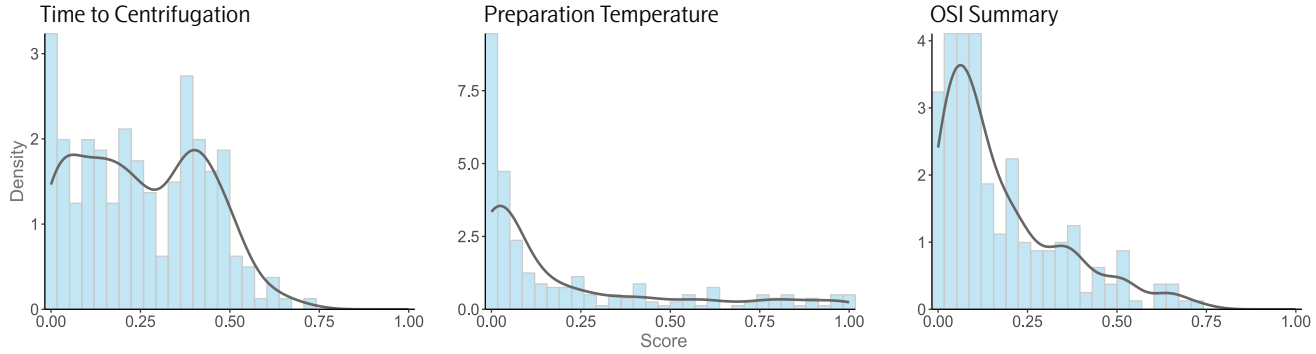

Figure 1. Distribution plots of OSI values across all samples for Time to Centrifugation, Preparation Temperature and OSI Summary.

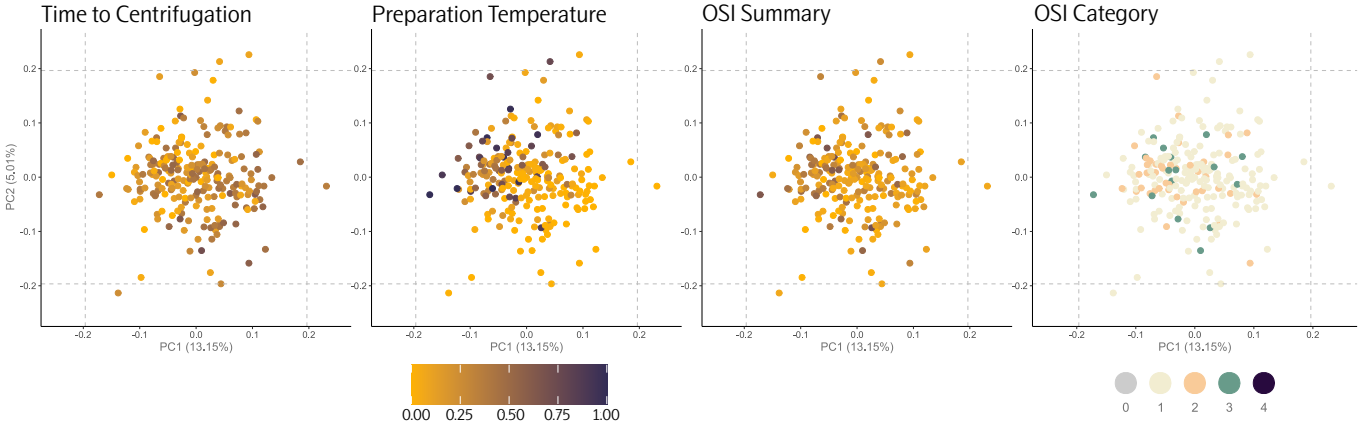

Figure 2. A projection along the first two principal components from a principal component analysis. Each dot represents a sample, with its position based on the measured protein values and its color based on the sample's OSI. Dashed lines indicate ±3 standard deviations from the mean value for each component.

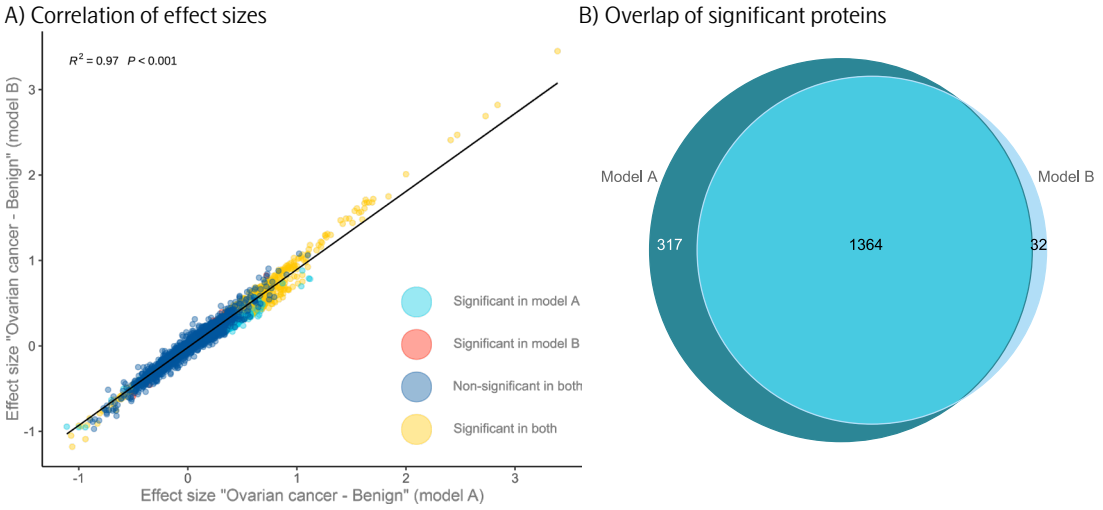

Figure 3. Correlation of protein effect sizes estimated from two diffrent linear models (Model A: corrected for age, and Model B: corrected for age, Time to Centrifugation and Preparation Temperature) and a Venn diagram showing the overlap of significant proteins between models. A) In the correlation plot each point represents a protein and is colored according to its significance group. The regression line summarizes the relationship between effect sizes. B) Circles represent the number of proteins significant in each model and the intersection indicates those significant in both.

influenced by pre-analytical variation.

All top 1000 markers from Model A were retained as significant in Model B, confirming that the sample-handling differences captured by OSI had minor impact on the results of this dataset. The following example demonstrates a case where these factors had a significant effect on results.

Example 2: A study with pre-analytical variation where OSI makes a difference

This study analyzed plasma samples from subjects with different cancer types and healthy controls. The dataset included colorectal (n=53), lung (n=55), and prostate cancer (n=76) samples, as well as healthy controls (n=226).

Impact of pre-analytical variation

The distribution of sample handling differences (measured by OSI) and their impact on overall data patterns were assessed using distribution plots (Fig. 4), while PCA plots (Fig. 5) were used to see whether these differences are reflected in the structure of the data.

OSI Summary and Category showed a clear pattern of samples with higher OSIs located on the bottom-right side of the PCA in Fig. 5, confirming that sample handling varied across the cohort.

- The distribution of Time to Centrifugation OSI showed clear indications of variation in sample handling (Fig. 4), with most samples handled soon after collection, while others were delayed. This was also seen clearly in the corresponding PCA plot (Fig. 5), with the lower-right part of the plot containing primarily samples with delayed centrifugation.
- The distribution of the Preparation Temperature OSI showed that most samples were handled under colder conditions, but a few were handled under warmer conditions. In the PCA, samples handled at room temperature were more abundant in the lower part of the plot.

Visual assessment of Fig. 4 and Fig. 5 suggests that inconsistent sample handling has affected the protein profile. Therefore, it is recommended to account for OSI in downstream analysis. For illustrational purposes, we prepared two analysis of variance (ANOVA) models: one without adjustment for OSI variables, and one with these factors included. ANOVA is a statistical method that simultaneously compares means across several groups to determine if observed differences are due to chance or reflect genuine distinctions.

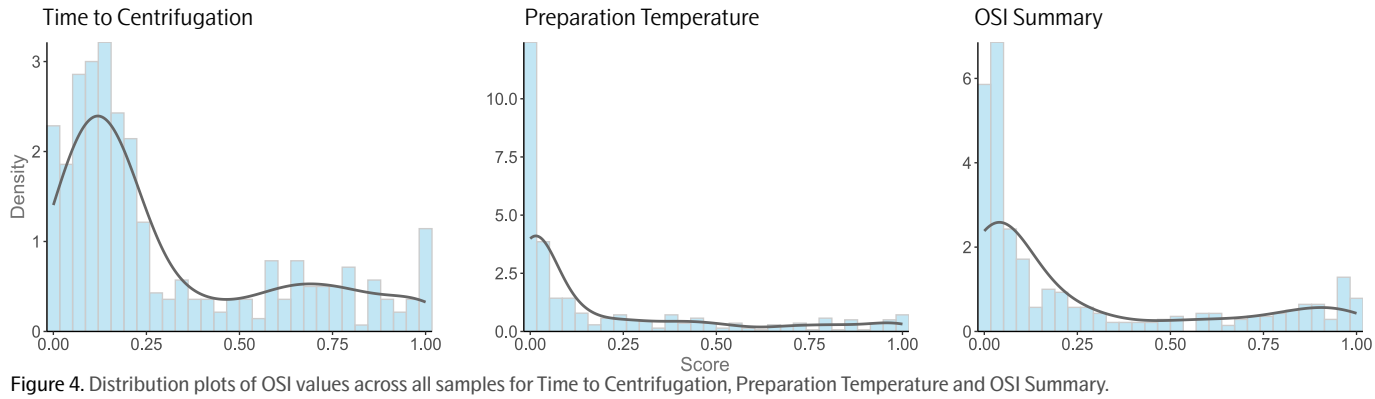

Figure 4. Distribution plots of OSI values across all samples for Time to Centrifugation, Preparation Temperature and OSI Summary.

Results from statistical models

Protein levels were compared between control subjects and subjects with various cancer types in two different models:

- Model A: ANOVA model corrected for subject age (standard practice in statistical analysis).
- Model B: ANOVA model corrected for subject age, OSI Time to Centrifugation and OSI Preparation Temperature.

As seen in Fig. 6A, there is a moderate correlation between the estimated effect sizes from the two respective models, suggesting that differences in sample handling influence the statistical results. This is further reflected in Fig. 6B, which shows proteins identified as significant in Model A but not in Model B, consistent with their association with variation in pre-analytical sample handling accounted for in the latter. To improve reliability, OSI should be included in the statistical analysis. Similar results were also observed when comparing prostate and lung cancer samples with healthy controls.

Even so, out of the top 1000 significant markers in Model A, a substantial fraction were retained as significant in Model B:

- 727 for colorectal cancer
- 570 for prostate cancer
- 505 for lung cancer

The proteins remaining significant after correcting for OSI variables are more likely to reflect true biological variance, while those losing significance are likely driven by sample handling variability.

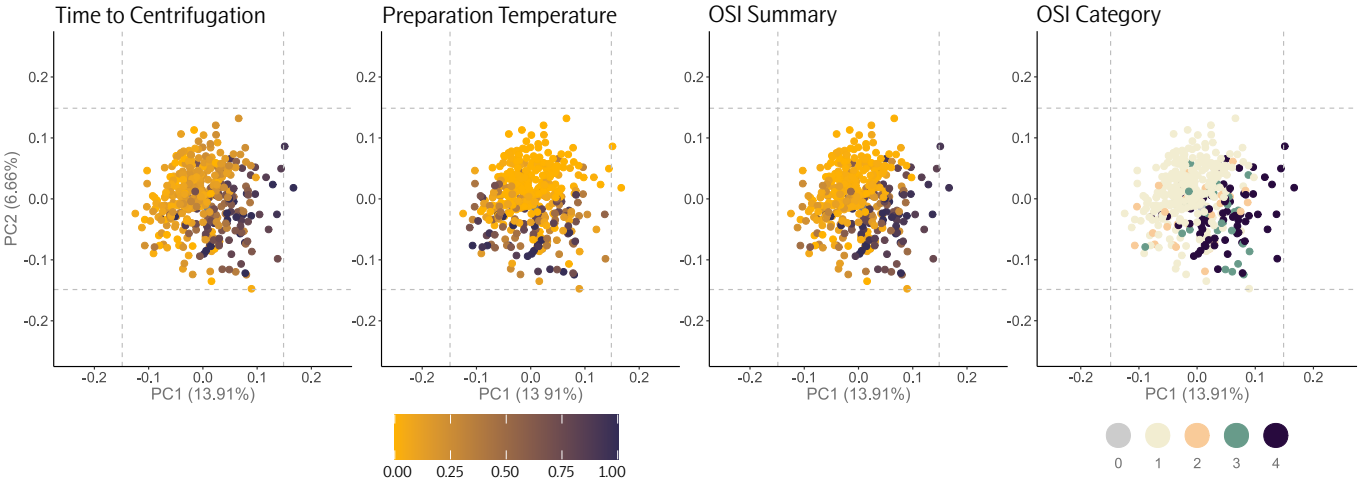

Figure 5. Projection along the first two principal components from a principal component analysis. Each dot represents a sample, with its position based on the measured protein values and its color based on the sample's OSI. Dashed lines indicate ±3 standard deviations from the mean value for each component.

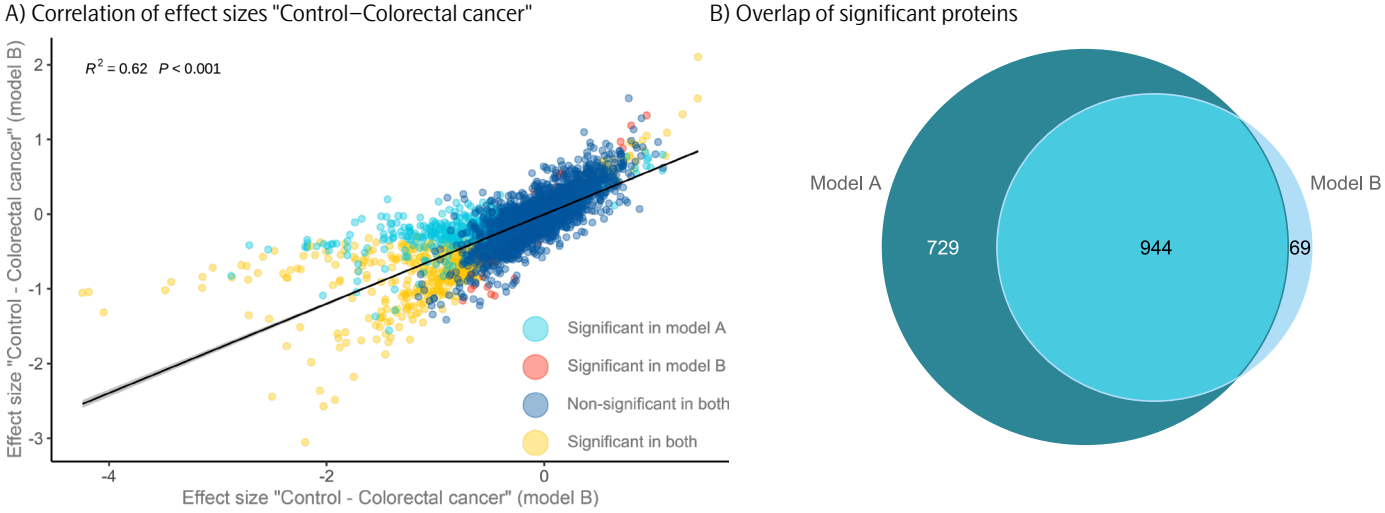

Figure 6. Correlation of protein effect sizes for colorectal cancer estimated from the two linear models (Model A: corrected for age, and Model B: corrected for age, Time to Centrifugation and Preparation Temperature), and a Venn diagram showing the overlap of significant proteins between models. Effect sizes represent the difference between cancer and healthy samples. A) Each point represents a protein and is colored according to its significance group. The black regression line summarizes the relationship between effect sizes. B) Circles represent the number of proteins significant in each model and the intersection indicates those significant in both.

How do I know if I need to take action on OSI?

IMPORTANT

OSI should not be the reason to exclude samples from the analysis. Instead, treat OSI as a signal representing an additional layer of information about sample variance that you need to explore for effects in the data before proceeding with downstream analysis.

OSI provides an additional layer of information about the measured samples. It is recommended to evaluate OSI using the methods described above, such as density plots and PCA plots, as an initial quality assessment step. While this white paper showcases PCA, the same approach applies to other visualization techniques (e.g. other dimensionality reduction techniques).

If OSI indicates substantial differences in sample handling, these should be accounted for in downstream statistical analyses, for example by including them as model covariates to improve the accuracy of findings. However, OSIs should be considered supportive indicators rather than definitive factors, as they may partially overlap with biological or disease-related signals. For example, poor study designs in which pre-analytical sample handling is linked to the biological hypothesis (e.g. cases handled consistently different from controls) raise the risk of misleading findings.

Conclusions

Pre-analytical variation is a natural part of sample handling, especially in large studies, and can be a source of bias in protein biomarker research. Inconsistencies in sample handling, such as delays before centrifugation or variations in preparation temperature, can affect the results of the statistical analysis. Reporting Olink Sample Index together with the measured protein data provides researchers with an additional layer of insight into their data, helping to identify patterns that may relate to sample handling rather than biology.

While pre-analytical variation should never overshadow the primary biological question, incorporating OSI as a covariate or using it as a supportive assessment tool will ensure more robust and reproducible results. This feature provides researchers with confidence in the results from the statistical analysis and allows them to answer the scientific questions that matter the most.

OLINK SAMPLE INDEX SUMMARY

What OSI does: Flags potential sample handling effects that may influence protein data.

When to use OSI: During QC, troubleshooting unexpected results, and as covariates in statistical models.

What OSI does not do: It does not replace biological interpretation or serve as an absolute measure of sample quality.

References

1. Olink white paper Pre-analytical variation in protein biomarker research (2025).
2. Olink technical white paper Design and validation of Olink™ Sample Index (2026).
3. Moskov *et al.* Deep plasma proteomics identifies and validates an eight-protein biomarker panel that separate benign from malignant tumors in ovarian cancer. *Nature Communications Medicine* (2025).

www.olink.com

1681, v1.0, 2026-03-05