Microb Health Dis 2026; 8: e1592

Microbiota in Health and Disease
A Journal of the European Helicobacter & Microbiota Study Group

European Helicobacter & Microbiota Study Group

BIOLOGICAL STRATIFICATION IN SMALL MICROBIOME COHORTS: LIMITED VARIANCE REDUCTION AND LOSS OF STATISTICAL POWER

G. Karapetyan, A. Pepoyan

Department of Food Safety and Biotechnology, Armenian National Agrarian University, Yerevan, Armenia

Corresponding Author: Astghik Pepoyan, MD; e-mail: apepoyan@gmail.com

Abstract – *Objective:* Biological stratification is widely used in microbiome research to improve interpretability by reducing heterogeneity, particularly in the small cohorts typical of precision medicine and rare disease studies. Because microbiome composition is strongly influenced by host-specific factors, including genetics, diet, and lifestyle, substantial inter-individual variability is common. Although stratification is frequently employed to address this variability, its consequences for statistical inference under small-sample conditions remain insufficiently quantified.

Materials and Methods: We evaluated the effects of biological stratification on variance, effect-size distributions, and statistical power using a real disease-control microbiome dataset comprising patients with familial Mediterranean fever (FMF) and healthy controls (n = 28). Stratification was performed sequentially by disease status and sex. Variance reduction, effect sizes, minimum detectable effects, and statistical power were assessed across stratification levels.

Results: Stratification produced modest reductions in within-group variability (~3-7%) but substantially reduced effective sample size (n = 4-10 per subgroup). Consequently, minimum detectable effect sizes increased markedly ($d \approx 1.25$-1.98 for 80% power), whereas observed effect sizes remained small (median $|d| \approx 0.2$). Although stratification can reveal subgroup-specific patterns that may be obscured in pooled analyses, particularly when biologically distinct groups exhibit opposing trends, the gain in biological resolution was generally insufficient to offset the accompanying loss of statistical power.

Conclusions: These findings demonstrate that modest variance reduction does not compensate for power loss when underlying effect sizes are small. Under such conditions, stratified analyses are more appropriately interpreted within an exploratory framework emphasizing effect-size estimation and variability assessment rather than confirmatory hypothesis testing. Importantly, improved biological resolution does not necessarily translate into improved statistical detectability. More broadly, the results suggest that the adequacy of a given sample size depends on the biological characteristics of the system under investigation. Although derived from an FMF cohort, the findings have broader implications for microbiome study design, subgroup analyses, and the interpretation of microbiome-based interventions.

Keywords: Microbiome, Biological stratification, Statistical power, Small sample size, Effect size, Inter-individual variability, Variance-power trade-off, Precision medicine.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

INTRODUCTION

Statistical inference in small biomedical cohorts presents a persistent challenge, particularly in settings characterized by high biological variability and limited sample size. Such conditions are increasingly common in precision medicine, early-phase clinical research, and rare disease studies, where cohort size is often constrained by practical, ethical, and biological factors rather than experimental design[1,2]. In these contexts, traditional assumptions of statistical robustness based on large sample sizes are frequently not satisfied, necessitating alternative approaches to study design, analysis, and interpretation[3].

In response to high inter-individual variability, biological stratification is widely employed as a strategy to improve biological interpretability[4]. By grouping individuals according to disease status, sex, genotype, or other clinically relevant variables, stratification is expected to reduce within-group heterogeneity and facilitate the detection of biologically meaningful differences[5,6]. This rationale is particularly prominent in microbiome research, where substantial compositional variability across individuals can obscure subtle biological signals and complicate statistical inference[7-9].

In applied contexts, such as probiotic development, microbiome profiling is frequently used to support targeted interventions and personalized therapeutic strategies[10,11]. Moreover, accumulating evidence suggests that microbiome responses may differ across biological subgroups, including sex-dependent patterns observed in probiotic modulation studies[12]. These observations further strengthen the rationale for stratified analyses and individualized approaches to microbiome-based interventions.

However, the statistical consequences of stratification in small cohorts remain insufficiently quantified. While reducing variability may intuitively improve statistical inference, stratification simultaneously fragments the cohort into smaller subgroups, thereby reducing effective sample size and constraining statistical power[13]. Consequently, biological resolution and statistical detectability do not necessarily improve in parallel. This introduces a fundamental trade-off between variance reduction and statistical power, particularly in high-dimensional datasets where effect sizes are often modest and biological heterogeneity remains substantial.

This limitation is especially relevant in microbiome and probiotic research, where subgroup-specific responses – such as sex-dependent host-microbiome interactions or disease-associated microbial shifts – are frequently hypothesized[12,14]. Although stratification may clarify such patterns biologically, it can also reduce statistical reliability when applied to small cohorts, thereby complicating interpretation and limiting reproducibility across studies[7,8].

Familial Mediterranean fever (FMF) provides a particularly informative context for examining these constraints. Although the disease is relatively prevalent in certain regions, clinical management requires continuous colchicine therapy, limiting the feasibility of controlled interventional studies and large-scale cohort expansion[14,15]. As a result, FMF research is frequently conducted under small-sample conditions, making it a representative model for investigating inferential limitations inherent to precision medicine and rare disease settings.

Here, we argue that biological stratification in small microbiome cohorts is a conditional, rather than universally beneficial, analytical strategy. Although stratification may improve biological resolution by reducing within-group variability, it simultaneously reduces statistical power through sample-size fragmentation. Using an illustrative microbiome dataset from an FMF cohort, we demonstrate that modest reductions in variance are insufficient to offset the corresponding loss of statistical detectability when underlying effect sizes are small.

More broadly, this perspective highlights a general principle relevant to precision medicine, microbiome research, and other high-dimensional biological disciplines: variance reduction alone does not guarantee improved statistical inference. The inferential value of stratification depends on the balance between variance reduction, effect-size magnitude, and sample-size fragmentation.

More generally, our findings suggest that the adequacy of a given sample size cannot be evaluated independently of the biological system under investigation. Sample sizes that may be sufficient for conventional biomedical measurements may remain inadequate for microbiome inference because of the combination of high inter-individual variability and modest effect sizes. Under small-cohort conditions, stratification may therefore be most appropriately viewed as a tool for enhancing biological interpretability within an exploratory framework rather than as a mechanism for achieving confirmatory statistical significance.

MATERIALS AND METHODS

Study Design and Cohort

This study analyzed short-term longitudinal microbiome data from a fully annotated cohort with metadata on disease status and biological sex. Only subjects with complete disease status and biological sex annotations were included (n = 28), comprising 15 FMF patients and 13 healthy controls.

Paired fecal samples collected in previously published studies were analyzed in the present secondary data analysis. Written informed consent had been obtained from all participants in the original studies. All FMF patients carried the homozygous MEFV M694V/M694V mutation and received stable colchicine therapy. The original studies involving human participants were approved by the Ethics Committee at the Higher Education and Science Committee of Armenia (approval No. 10-15-21AG; 21 October 2021)[14,15].

Microbiome Data Processing

Fecal DNA was extracted using commercial kits (ZR Fecal DNA MiniPrep, Zymo Research, USA). Microbial community composition was profiled using the Affymetrix PhyloChip™ G3 microarray, covering 18,724 operational taxonomic units (OTUs).

Signal intensities were normalized using the robust multi-array average (RMA) algorithm[16], which performs background correction, quantile normalization, and log transformation.

OTUs were aggregated to the genus level, and taxa were retained based on predefined prevalence (≥50% of subjects) and abundance (median $\geq 10^{-3}$) thresholds, resulting in 538 genera for downstream analysis.

Stratification Strategy

Three levels of biological stratification were evaluated:
- Level 1: unstratified (n = 28)
- Level 2: disease-stratified (FMF vs healthy)
- Level 3: disease × sex stratification (four subgroups; n = 4-10).

This hierarchical design enabled direct quantification of the trade-off between variance reduction and sample-size fragmentation under complete metadata annotation.

Statistical Analysis

Paired Δ-Profiles

For each genus and subject, paired differences were calculated as

$$\Delta = \ln(1 + \text{abundance_b}) - \ln(1 + \text{abundance_a}),$$

where $\ln(1 + x)$ denotes a natural logarithmic transformation applied to stabilize variance across abundance scales and accommodate zero values.

Rationale for Δ-Profiles

Paired within-subject differences serve as an intra-subject control that reduces baseline inter-individual variability inherent to cross-sectional microbiome comparisons. By measuring temporal change (Δ) rather than absolute abundances, this approach strengthens inference in longitudinal designs by accounting for subject-specific baselines. Importantly, the absence of stratification benefits even under this variance-reducing framework – which already controls for a major source of heterogeneity – reinforces our central conclusion that stratification provides minimal additional gain in small cohorts when effect sizes are modest.

Variance Metrics

For each stratification level, variance was characterized using the interquartile range (IQR; Q75–Q25), standard deviation (SD), and median Δ as a measure of central tendency of temporal change.

Variance reduction was calculated as variance reduction (%) = 100 × (IQR_Level1 – IQR_LevelN) / IQR_Level1, where positive values indicate a decrease in variance and negative values indicate an increase.

Effect Size Quantification

Between-group differences were quantified using both parametric and non-parametric effect size measures. Cohen's *d* was calculated as a standardized mean difference (d = (mean$_1$ – mean$_2$) / SD pooled) to facilitate comparison with existing literature, while acknowledging potential deviations from normality in Δ-distributions. Cliff's delta (δ) was computed as a non-parametric effect size [δ = (n_greater – n_less) / (n_1 × n_2)][17], providing a distribution-independent complement to Cohen's d. Consistent conclusions were obtained across both metrics.

The N-reduction factor was calculated based on variance-based sample size theory as (IQR_Level1 / IQR_LevelN)². This metric provides an estimate of the theoretical reduction in required sample size associated with variance reduction following stratification. Values greater than 1 indicate increased statistical efficiency, whereas values less than 1 indicate reduced efficiency. This estimate represents an upper bound, as realized gains depend on the magnitude of effect sizes and the extent of variance reduction.

Power analysis

Minimum detectable effect sizes (Cohen's *d*) were calculated for each stratification level at 80% statistical power and a significance level of α = 0.05 using the two-sample *t*-test formulation. The minimum detectable effect size was defined as

$$d_min = (z_\{\alpha/2\} + z_\beta) / \sqrt{(n/2)}, \text{ where } z_\{\alpha/2\} = 1.96 \text{ and } z_\beta = 0.84 \text{ for 80\% power.}$$

Assumptions and Robustness

Power calculations were performed under the assumptions of approximate normality of Δ-distributions and balanced group sizes. Unequal group sizes (n = 4-15) and deviations from normality may further reduce effective statistical power relative to these theoretical estimates. However, the qualitative conclusion – that observed effect sizes (median |d| ≈ 0.2) are substantially smaller than the minimum detectable effects (d_min ≥ 1.08-1.98) – remains robust to these deviations.

Power analysis was conducted under parametric (*t*-test) assumptions, providing a conservative approximation for non-parametric tests (e.g., Mann-Whitney U) under similar effect sizes and approximately symmetric distributions.

Statistical Testing

Between-group differences were assessed using the Mann-Whitney U test for disease comparisons (FMF *vs.* healthy controls), selected for robustness to non-normality. Within-subject temporal changes were evaluated using the Wilcoxon signed-rank test, while overall differences across four groups (disease × sex) were analyzed using the Kruskal-Wallis test. *p*-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) procedure at α = 0.05[18].

All analyses were performed in Python 3.10 using the NumPy[19], pandas[20], and SciPy[21] libraries. Analysis code and Jupyter notebooks are publicly available at https://doi.org/10.5281/zenodo.18449541.

The dataset was accessed for research purposes in January 2026. All data were fully de-identified prior to analysis, and the authors did not have access to any information that could identify individual participants at any stage of the study.

RESULTS

Cohort Characteristics

After applying complete-annotation criteria, 28 subjects were included in the analysis (Table 1), comprising 15 FMF patients and 13 healthy controls. Group sizes under progressive stratification ranged from n = 4 to n = 10 at the most granular level. Sex distribution was comparable between disease groups.

TABLE 1. SAMPLE SIZES UNDER PROGRESSIVE STRATIFICATION.

Stratification Level	Group(s)	n
Level 1: All pooled	All	28
Level 2: Disease	Healthy	13
	FMF	15
Level 3: Disease+Sex	Healthy Female	4
	Healthy Male	9
	FMF Female	5
	FMF Male	10

Note: Subject numbering is independent between FMF (P prefix) and Healthy (H prefix) groups.

Effect of Stratification on Variance

Variance reduction under progressive stratification was quantified for selected genera (Figure 1; Table 2). Disease stratification resulted in modest decreases in interquartile range (IQR), including *Faecalibacterium* (~7%), *Bacteroides* (~6%), *Prevotella* (~4%), and *Akkermansia* (~1%).

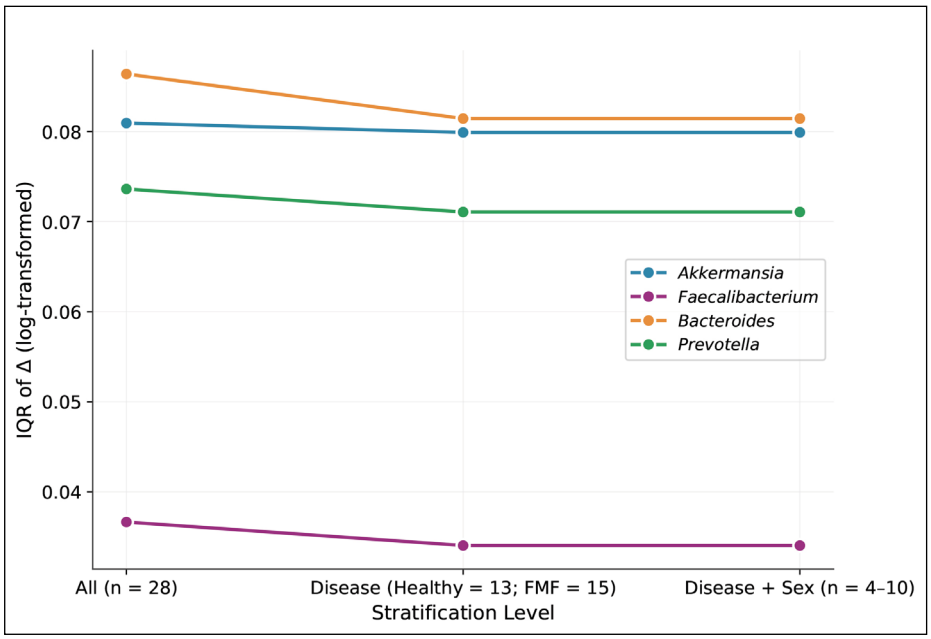

Figure 1. Variance reduction under progressive stratification. Interquartile range (IQR) of log-transformed abundance differences (Δ) for selected genera (A*kkermansia, Faecalibacterium, Bacteroides, Prevotella*) across increasing levels of stratification: all samples (n = 28), disease-based stratification (healthy = 13, FMF = 15), and combined disease and sex stratification (n = 4-10 per subgroup). Variance reduction across stratification levels is modest (~3-7%), indicating that stratification yields limited decreases in within-group variability despite increasing biological resolution.

TABLE 2. VARIANCE REDUCTION FOR KEY GENERA*.

Genus	IQR (All)	IQR (Disease + Sex)	Variance Reduction (%)	N-reduction factor
Faecalibacterium	0.037	0.034	7.1	1.16
Bacteroides	0.086	0.081	5.7	1.13
Prevotella	0.074	0.071	3.5	1.07
Akkermansia	0.081	0.080	1.3	1.03

**Positive values indicate variance decrease.*

No additional variance reduction was observed under further stratification by sex, with IQR values remaining stable between Level 2 and Level 3. Corresponding N-reduction factors ranged from 1.03 to 1.16, indicating limited gains in statistical efficiency.

Statistical Power Constraints

Despite modest variance reductions, effective sample sizes decreased substantially following stratification (n = 4-10). Power analysis demonstrated a corresponding increase in minimum detectable effect sizes (Table 3).

TABLE 3. MINIMUM DETECTABLE EFFECT SIZE (COHEN'S *d*) AT 80% POWER, α=0.05.

Stratification Level	Group Size	Min *d* (80% power)
Disease (Healthy)	n = 13	$d \geq 1.10$
Disease (FMF)	n = 15	$d \geq 1.02$
Healthy Female	n = 4	$d \geq 1.98$
Healthy Male	n = 9	$d \geq 1.32$
FMF* Female	n = 5	$d \geq 1.77$
FMF* Male	n = 10	$d \geq 1.25$

Note: Power calculations assume balanced two-group comparisons. For the complete cohort (N = 28 total subjects), a hypothetical balanced split would yield groups of n ≈ 14 per group, requiring d ≥ 1.08 for 80% power.
**Subjects with familial Mediterranean fever.*

For disease-stratified groups (n = 13-15), detectable effects were d ≥ 1.02-1.10, whereas for fully stratified groups (n = 4-10), thresholds increased to d ≥ 1.25-1.98. These values indicate severe limitations in statistical detectability under small-sample conditions (Figure 2).

Observed Effect Sizes

Observed effect sizes for disease comparisons were uniformly small across all examined genera (Table 4). Median absolute values were $|d| < 0.2$ and $|\delta| < 0.15$, with no statistically significant differences detected after multiple-testing correction. These observed effect sizes were substantially below the minimum detectable thresholds estimated for all stratification levels.

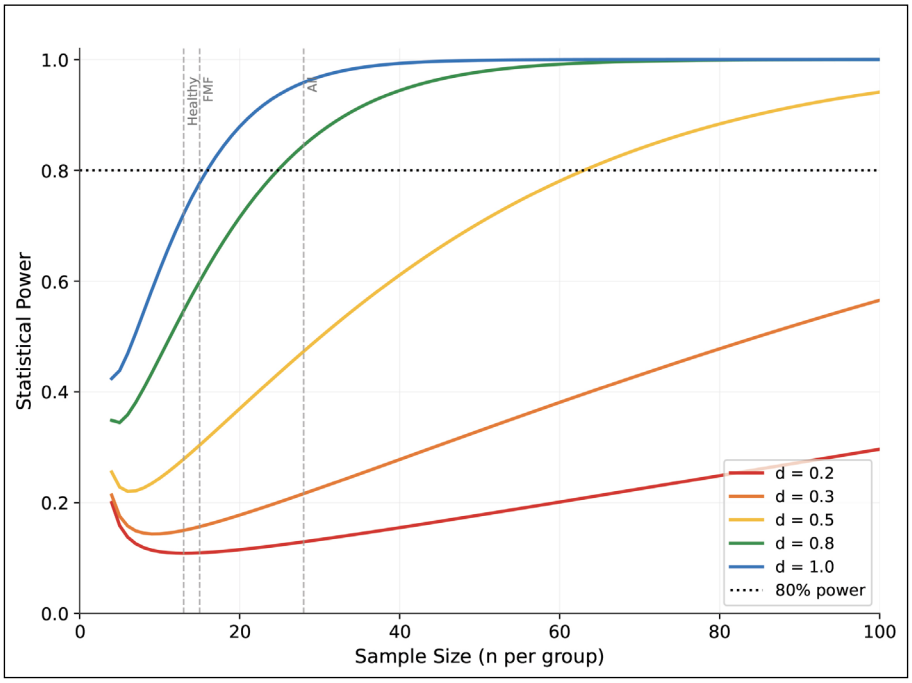

Figure 2. Statistical power as a function of sample size and effect size. Statistical power (1 – β) as a function of sample size per group (n) for a two-sided test at α = 0.05, shown for multiple effect sizes (Cohen's d = 0.2, 0.3, 0.5, 0.8, and 1.0). The dashed horizontal line indicates the conventional 80% power threshold. Vertical reference lines denote sample sizes corresponding to the stratification levels evaluated in the present study. The curves illustrate that effect sizes typical of microbiome datasets (d ≈ 0.2-0.3) require substantially larger cohorts to achieve adequate statistical power. In contrast, the sample sizes obtained after stratification (n = 4-15 per group) remain underpowered for detecting small effects, despite modest reductions in within-group variability.

TABLE 4. OBSERVED DISEASE EFFECT SIZES IN THIS FMF* COHORT.

Genus	Median Δ (Healthy)	Median Δ (FMF)	Cohen's *d*	Cliff's *δ*	Mann-Whitney *p*
Akkermansia	0.001	0.033	0.09	0.08	0.91
Faecalibacterium	−0.003	0.010	0.16	0.14	0.50
Bacteroides	0.004	0.007	0.09	0.07	0.64
Prevotella	−0.004	0.019	0.14	0.12	0.33

**Subjects with familial Mediterranean fever.*

Genome-Wide Effect Size Distribution

Across all 538 genera, the median absolute effect size for disease comparisons was approximately |d| ≈ 0.2, with a 95th percentile of |d| ≈ 0.7. A total of 52 genera (9.7%) showed nominal significance ($p < 0.05$), but none remained significant after false discovery rate (FDR) correction (Figure 3).

Sex-based comparisons showed similar distributions, with median |d| ≈ 0.2 and no significant findings after correction.

Quantifying the Variance-Power Trade-off

Variance reduction and corresponding changes in statistical power were evaluated across stratification levels.

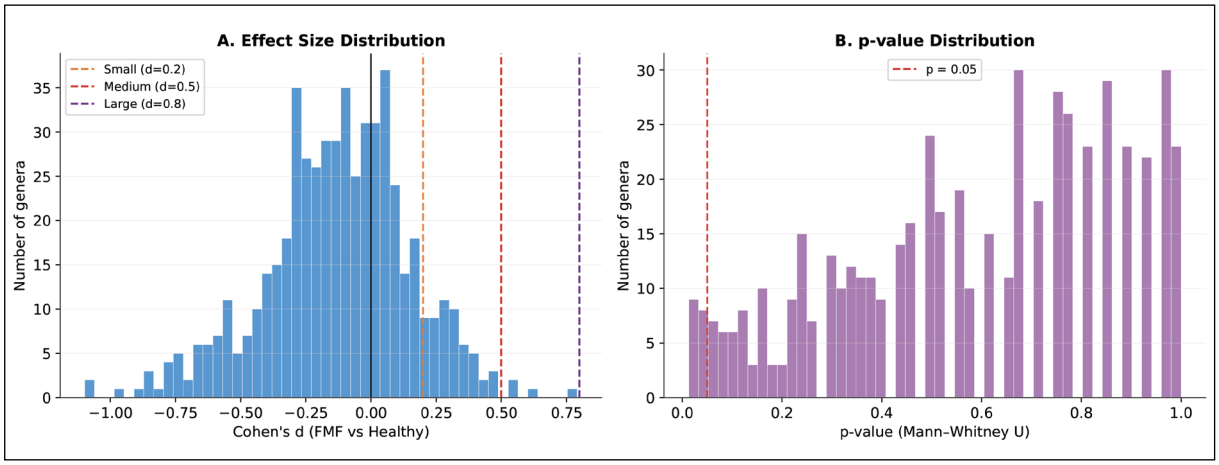

Figure 3. Genome-wide effect size and p-value distributions. **A**, Distribution of effect sizes (Cohen's d) for genus-level comparisons between FMF patients and healthy controls across 538 genera. Vertical dashed lines indicate conventional thresholds for small (d = 0.2), medium (d = 0.5), and large (d = 0.8) effects. Most observed effect sizes fall within the small-effect range, with a median absolute effect size of approximately |d| ≈ 0.2. **B**, Distribution of *p*-values obtained from Mann-Whitney U tests across genera. The dashed vertical line indicates the nominal significance threshold ($p = 0.05$). A total of 52 of 538 genera (9.7%) reached nominal significance, but none remained significant after false discovery rate (FDR) correction, consistent with the limited statistical power associated with small cohort size.

For *Faecalibacterium*, variance reduction reached 7.1% (N-reduction factor = 1.16). However, resulting group sizes (n = 4-10) required minimum detectable effect sizes of d ≈ 1.25-1.98, whereas the observed effect size was |d| ≈ 0.16.

For *Akkermansia*, variance reduction was limited to 1.3% (N-reduction factor = 1.03), indicating negligible gains in statistical efficiency.

Across taxa, modest variance reductions were insufficient to offset the increase in minimum detectable effect sizes associated with sample-size fragmentation.

DISCUSSION

The results presented above establish a consistent pattern across stratification levels, linking modest variance reduction with substantial loss of statistical power. This relationship demonstrates that variance reduction and statistical power are intrinsically coupled rather than independent: gains in within-group homogeneity must be evaluated relative to the accompanying reduction in effective sample size.

This study addresses a central challenge in small-sample biomedical research – the trade-off between variance reduction achieved through stratification and the loss of statistical power caused by sample-size fragmentation [22,23]. Using a fully annotated cohort and a stepwise stratification framework, we show that stratification functions primarily as a variance-refinement mechanism rather than a power-enhancing strategy. Although reductions in within-group dispersion were observed, these gains were insufficient to offset the substantial increase in minimum detectable effect sizes as subgroup sizes decreased. Consequently, many biologically plausible effects remained statistically undetectable despite improved within-group coherence.

This distinction is important because biological interpretability and statistical detectability are often implicitly assumed to improve together. Our results indicate that, in small microbiome cohorts, these objectives may diverge – stratification can enhance biological resolution while simultaneously reducing the ability to detect statistically reliable effects.

In practical terms, these findings demonstrate that modest variance reduction cannot compensate for power loss when underlying effect sizes are small[13]. This observation is particularly relevant for microbiome studies, where biological heterogeneity is high, and effect sizes are frequently modest. Under such conditions, improvements in biological resolution do not necessarily translate into improvements in statistical detectability.

The implications extend beyond the specific dataset examined here. Rather than being interpreted as a means of increasing statistical significance, stratification should be viewed as a tool for improving biological interpretability and structuring exploratory inference, including effect size estimation, characterization of variability, and identification of consistent directional trends[23,24]. When expected effect sizes are modest, and cohort size is limited, analyses are more appropriately framed as exploratory rather than confirmatory[23]. In this context, emphasis on effect size distributions and robustness across stratification levels provides a more informative inferential framework than reliance on significance testing alone.

Although derived from microbiome data, the variance-power trade-off identified here is not domain-specific and is likely relevant to a broad range of high-dimensional biological datasets characterized by substantial inter-individual variability and small sample sizes[22,24].

Evidence from studies in familial Mediterranean fever cohorts further supports the existence of biologically meaningful microbiome heterogeneity, including sex-dependent probiotic responses[25] and differential microbial responses associated with therapeutic exposures[26]. These findings reinforce the biological rationale for stratification while simultaneously illustrating the challenges associated with detecting such effects under small-sample conditions.

Despite the limitations inherent in a single-cohort analysis, the central conclusion remains robust: the inferential value of stratification depends on whether variance reduction is sufficient to offset sample-size fragmentation relative to the underlying effect-size distribution. In the present dataset, modest variance reductions (~3-7%) were insufficient to compensate for power loss given the predominance of small effect sizes.

Taken together, these results define a practical boundary for the utility of stratification in small-sample settings. Rather than serving as a strategy for improving statistical detectability, stratification should be understood as a tool for enhancing biological interpretability under explicitly constrained inferential conditions. This perspective provides a quantitative framework for study design and data interpretation in settings where increasing sample size is not feasible.

Practical Implications for Study Design

These findings have important implications for study design in small-cohort biomedical research. Deep stratification should be applied cautiously when total sample size is limited ($n < 50$), as modest variance reductions are unlikely to offset the resulting loss of statistical power. Under such conditions, emphasis should shift toward effect size estimation, consistency of directional trends, and structured exploratory inference rather than statistical significance alone[13,23].

These observations are consistent with prior microbiome studies demonstrating high inter-individual variability and generally modest effect sizes[7-9], as well as broader analyses of statistical power in small-sample research[13]. In this context, variance reductions on the order of tens of percent would likely be required to meaningfully improve detectability, whereas the reductions observed here (~3-7%) are insufficient to compensate for the corresponding loss of statistical power.

In contrast to assumptions commonly embedded in precision medicine frameworks, which often implicitly rely on moderate-to-large biological effect sizes[27,28], the present study indicates that genus-level microbiome effects are typically small (median $d \approx 0.2$). Consequently, substantially larger cohorts may be required to achieve adequate statistical power when stratification is incorporated into study design.

An important implication of these findings is that statistical adequacy cannot be evaluated solely on the basis of sample size. Cohort sizes considered sufficient in some areas of biomedical research do not necessarily confer comparable inferential power in microbiome studies. Unlike many conventional biological measurements, microbiome datasets are characterized by substantial inter-individual heterogeneity, compositional structure, and generally modest effect sizes. Under these conditions, sample sizes that might appear adequate from a traditional biomedical perspective can remain insufficient for reliable microbiome inference. The present results therefore highlight that the informational value of a given sample size depends not only on the number of subjects studied but also on the variability structure and effect-size distribution of the biological system under investigation.

Limitations

The present study has several limitations. It is based on a single cohort, a specific microbiome profiling platform, and a limited set of stratification variables. Additional biological factors, including genotype, diet, treatment history, or environmental exposures, may yield different variance structures and modify the observed variance-power trade-off. Furthermore, microbiome-specific data properties such as compositionality and sparsity may influence variance behavior and effect size estimation[24].

An additional source of complexity arises from heterogeneity within the FMF cohort itself. Although all patients shared the same MEFV mutation (M694V/M694V), variability in disease severity, treatment response, and associated microbiome characteristics may still be present. Prior studies have shown that microbiome composition in FMF can be influenced by host-microbe interactions, treatment effects, and shifts in specific microbial taxa, including opportunistic microorganisms[11,12,15]. Such intra-group variability may not be fully captured by the stratification variables considered in this study (disease status and sex). Consequently, residual heterogeneity may persist even after stratification, further limiting statistical detectability.

Future Directions

These limitations highlight several directions for future methodological and empirical research. Future studies should evaluate whether alternative study designs, higher-resolution microbiome profiling approaches, or diseases characterized by larger effect sizes alter the observed variance-power trade-off. In addition, methodological strategies such as hierarchical modeling, simulation-based study design, and Bayesian frameworks may provide more efficient approaches for managing stratification under small-sample constraints[23]. Validation in larger and more diverse cohorts will be essential to assess the generalizability of the proposed framework.

CONCLUSIONS

This study demonstrates that the inferential value of biological stratification in small microbiome cohorts is constrained by a fundamental trade-off between variance reduction and statistical power. Although stratification reduced within-group variability, the observed reductions (~3-7%) were insufficient to compensate for the loss of statistical power associated with sample-size fragmentation. Consequently, statistical detectability remained limited despite improvements in biological resolution.

These findings highlight an important distinction between biological interpretability and statistical inference. While stratification may reveal biologically meaningful subgroup structure and support mechanistic interpretation, it does not necessarily improve the ability to detect effects when cohort size is limited, and underlying effect sizes are small. In such settings, increased biological resolution may be accompanied by reduced inferential capacity.

More broadly, the present results suggest that the utility of stratification should not be assumed *a priori*. Rather, its value depends on the quantitative balance between variance reduction, effect-size magnitude, and the degree of sample-size fragmentation. This principle is likely relevant not only to microbiome research but also to other areas of precision medicine and high-dimensional biology characterized by substantial inter-individual variability.

Taken together, our findings support a framework in which biological stratification in small cohorts is viewed primarily as a tool for enhancing biological interpretability and structuring exploratory inference rather than as a strategy for improving statistical detectability. Future studies should explicitly evaluate the variance-power trade-off when designing stratified analyses and prioritize effect-size estimation alongside conventional significance testing.

Conflict of Interest

The authors declare no competing interests.

Ethics Approval and Informed Consent

The present study involved secondary analysis of previously collected and fully de-identified microbiome data; no new participants were recruited, and no new biological samples were collected. The original studies involving human participants were approved by the Ethics Committee at the Higher Education and Science Committee of Armenia (approval No. 10-15-21AG; 21 October 2021)[14,15], and written informed consent was obtained from all participants prior to sample collection. All procedures were conducted in accordance with the Declaration of Helsinki (2013 revision). The current analytical study was approved by the Ethics Committee of the Armenian National Agrarian University (Yerevan, Armenia) under protocol number 1, dated 11/01/2026.

Availability of Data and Materials

Microbiome profiling data and code are publicly available in the Zenodo (https://doi.org/10.5281/zenodo.18448855, https://doi.org/10.5281/zenodo.18449541).

Funding

This work was supported by the State Committee of Science, Armenia (projects 25DD-4D055 and 25AA-4A009). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

The authors are grateful to Prof. Sh. Kamiya and Prof. T. Torok for their insightful scientific advice and collaborative support. The authors also acknowledge Drs. V. Tsaturyan, M. Balayan, and A. Manvelyan for their dedicated technical assistance and contributions to the experimental workflow.

Authors' Contributions

A.P.: Conceptualization, supervision, interpretation of results, writing – review and editing. G.K.: Investigation, methodology, formal analysis, data curation, visualization, writing – original draft preparation. Both authors read and approved the final manuscript.

Use of Artificial Intelligence

The authors utilized Claude (Anthropic) to check the grammar and vocabulary of this manuscript. All authors reviewed and edited the generated suggestions and accept full responsibility for the final content.

ORCID ID

Gurgen Karapetyan: 0009-0005-5509-2576
Astghik Pepoyan: 0000-0002-1935-5341

REFERENCES

1. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. Best practices for analysing microbiomes. Nat Rev Microbiol 2018; 16: 410-422.
2. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nat Rev Immunol 2016; 16: 341-352.
3. Jovel J, Dieleman LA, Kao D, Mason AL, Wine E. The human gut microbiome in health and disease. Metagenomics 2018: 197-213.
4. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. The human oral microbiome. J Bacteriol 2010; 192: 5002-5017.
5. Wade WG. The oral microbiome in health and disease. Pharmacol Res 2013; 69: 137-143.
6. Hajishengallis G. Periodontitis: from microbial immune subversion to systemic inflammation. Nat Rev Immunol 2015; 15: 30-44.
7. Curtis H, Blaser MJ, Dirk G, Kota KC, Rob K, Liu B, Wang L, Sahar A, White JR, Badger JH, Chinwalla AT. Structure, function and diversity of the healthy human microbiome. Nature (London) 2012; 486: 207-214.
8. Naspolini NF, Natividade AP, Asmus CI, Moreira JC, Dominguez-Bello MG, Meyer A. Early-life gut microbiome is associated with behavioral disorders in the Rio birth cohort. Sci Rep 2025; 15: 8674.

9. Renwick S, Allen-Vercoe E. Manipulating the gut microbiota. The Developing Microbiome 2020: 195-215.
10. Walker LS, Caplan-Dover A, Rasquin-Weber A. Validation of a gastrointestinal symptom questionnaire for pediatric populations. J Pediatr Gastroenterol Nutr 2000; 31: 36-42.
11. Brook I. Fusobacterial infections in children. Curr Infect Dis Rep 2013; 15: 288-294.
12. Komiya Y, Shimomura Y, Higurashi T, Sugi Y, Arimoto J, Umezawa S, Uchiyama S, Matsumoto M, Nakajima A. Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity. Gut 2019; 68: 1335-1337.
13. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, Neuberg D, Huang K, Guevara F, Nelson T, Chipashvili O, Hagan T, Walker M, Ramachandran A, Diosdado B, Serna G, Mulet N, Landolfi S, Ramon Y Cajal S, Fasani R, Aguirre AJ, Ng K, Élez E, Ogino S, Tabernero J, Fuchs CS, Hahn WC, Nuciforo P, Meyerson M. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science 2017; 358: 1443-1448.
14. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell Host Microbe 2013; 14: 195-206.
15. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 2012; 9: 811-814.
16. Francavilla R, Ercolini D, Piccolo M, Vannini L, Siragusa S, De Filippis F, De Pasquale I, Di Cagno R, Di Toma M, Gozzi G, Serrazanetti DI, De Angelis M, Gobbetti M. Salivary microbiota and metabolome associated with celiac disease. Appl Environ Microbiol 2014; 80: 3416-3425.