

FINAL STATISTICAL ANALYSIS PLAN

Establishing Baseline Sysmex UF-5000 Flow Cytometer Results in Healthy Men (USIEV)

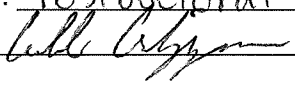
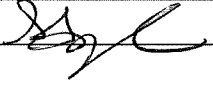
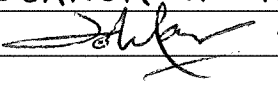
Study Title	Establishing Baseline Sysmex UF-5000 Flow Cytometer Results in Healthy Men
IU Protocol Number	#26774
Principal Investigator	Stephen Jordan, MD, PhD
Co-Investigator	Evelyn Toh, PhD
Sponsor/Support	Sysmex America, Inc.
SAP Version	1.0
SAP Date	May 6, 2026
SAP Author	Caleb Ardizzone, PhD
Statistical Software	R 4.2.1 (lme4 1.1.32, lmerTest 3.1.3, tidyverse 2.0.0, readxl 1.4.3)
Significance Level	$\alpha = 0.05$ (two-sided, unadjusted)

CONFIDENTIAL

This document is intended for regulatory submission purposes only.

Approval Signature Page

By signing below, the individuals listed confirm that they have reviewed and approved this Statistical Analysis Plan.

SAP Author:Name: Caleb ArdizzonePosition/Title: Postdoctoral fellowSignature: Date: 5/7/26**Principal Investigator:**Name: Stephen JordanPosition/Title: Associate ProfessorSignature: Date: May 7, 2026**Co-Investigator:**Name: EVELYN TOHPosition/Title: RESEARCH ASSISTANT PROFESSORSignature: Date: 5/7/26

SAP Roles

SAP Primary Author	Caleb Ardizzone, PhD
Principal Investigator	Stephen Jordan, MD, PhD
Co-Investigator / Senior Reviewer	Evelyn Toh, PhD

SAP Revision History

Version	Date	Author	Description
1.0	May 6, 2026	Caleb Ardizzone, PhD	Original SAP

List of Abbreviations

CI	Confidence Interval
CV%	Coefficient of Variation (%)
df	Degrees of Freedom
FCU	First-Catch Urine
IQR	Interquartile Range
IRB	Institutional Review Board (Indiana University)
LE	Leukocyte Esterase
LME	Linear Mixed-Effects Model
REML	Restricted Maximum Likelihood
SAP	Statistical Analysis Plan
SD	Standard Deviation
SE	Standard Error
STI	Sexually Transmitted Infection
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
WBC	White Blood Cell

1. Introduction and Study Overview

This Statistical Analysis Plan (SAP) documents the statistical methods used in the analysis of data from the USIEV study (IU IRB Protocol #26774), an observational within-subject crossover study evaluating urinary white blood cell (WBC) counts measured by the Sysmex UF-5000 urine flow cytometer using two first-catch urine (FCU) collection methods in healthy male volunteers without sexually transmitted infections (STI). This SAP was prepared retrospectively following completion of the analysis. Deviations from the statistical methods described in the original protocol are documented in Section 13.

The two collection methods under evaluation are:

- Standard sterile urine cup (30–60 mL first-catch urine)
- Colli-Pee device (10 mL first-catch urine)

Each participant provided urine samples under both collection conditions across three visits per week per condition (typically Monday, Wednesday, and Friday), serving as their own control. Collection sequence (which device was used first) was allocated by birth year parity.

This SAP was prepared in accordance with recommendations for observational study SAPs (van der Horst et al., 2019; Watson, 2025) and the SAP template described by Stevens et al. (2023). Results will be reported in accordance with the STROBE guidelines for observational studies. Post-hoc analyses are explicitly labeled as such throughout.

2. Study Objectives and Hypotheses

2.1 Aim 1 (Primary)

Establish the baseline number of urinary white blood cells (WBC/ μ L) by Sysmex UF-5000 flow cytometry in 20 healthy male controls without gonorrhea or chlamydia, comparing first-catch urine collected with a standard sterile urine cup (~30–60 mL) versus the 10 mL Colli-Pee device.

Hypothesis: The 10 mL Colli-Pee will yield a higher WBC/ μ L count and higher LE level compared to the standard-of-care urine cup, reflecting increased sensitivity of the smaller-volume collection.

Note: LE was uniformly negative across all participants and timepoints and was not analyzed inferentially (see Section 4.4).

2.2 Aim 2 (Secondary)

Determine the day-to-day variation in baseline urinary WBC/ μ L levels across three visits per week, comparing the Colli-Pee device and the standard urine cup.

Hypothesis: The 10 mL Colli-Pee will demonstrate greater precision (lower day-to-day variation) in both WBC/ μ L and LE compared to the standard-of-care urine cup.

Note: LE was uniformly negative across all participants and timepoints and was not analyzed inferentially (see Section 4.4).

2.3 Aim 3 (Secondary, Not Completed)

Characterize the change in WBC/ μ L and LE over time (0, 1, 2, 4, and 8 hours) stratified by urinary pH, using prospectively collected urine specimens.

Hypothesis: Urine with lower pH will show more rapid loss of WBC/ μ L and LE over time compared to urine with higher pH.

Note: Aim 3 was not completed in this dataset, which utilized residual urine specimens not suitable for prospective time-course analysis. This aim is planned for a future study using prospectively collected specimens. See Section 4.4 for details.

3. Study Design and Population

3.1 Design

Observational within-subject crossover study. Each participant provided first-catch urine under both collection conditions (Urine Cup and Colli-Pee) across two separate one-week periods, with three visits per week per period (typically Monday, Wednesday, and Friday). Participants were allocated to collection sequence by birth year parity (described in Section 9).

3.2 Eligibility Criteria

Inclusion Criteria

- Male sex
- Age 18 to 50 years
- No urethral symptoms (discharge, dysuria)

Exclusion Criteria

- History of genital surgery or known urethral abnormalities
- Positive test for gonorrhea or chlamydia at enrollment
- Diagnosis of a sexually transmitted infection within the last year
- Unable to collect urine specimens in the morning three times per week for two weeks

3.3 Analysis Population

All 20 enrolled participants completed all six visits and provided urine samples under both collection conditions across all three timepoints per condition, yielding 60 observations per collection method (N = 120 total observations). No per-protocol or intent-to-treat distinction is applicable given the observational crossover design.

3.4 Analytic Refinements

The following analytic decisions were made to appropriately accommodate the repeated-measures crossover design and distributional characteristics of urinary flow cytometry data. Full documentation is provided in Section 13.

3.4.1 Logarithmic Transformation of WBC/ μ L

The original protocol did not specify transformation of WBC/ μ L measurements. A $\log_{10}(\text{WBC}/\mu\text{L} + 1)$ transformation was applied prior to inferential analyses to reduce the influence of right-skewness and improve adherence to model assumptions. Descriptive summaries on the original scale were retained where appropriate for interpretability.

3.4.2 Linear Mixed-Effects Modeling

The original protocol proposed comparison using conventional group-based methods. A linear mixed-effects model with participant-level random intercepts was used instead to appropriately account for within-subject correlation arising from the repeated-measures crossover design. Fixed effects included collection method and visit timepoint.

3.4.3 Sensitivity and Post-Hoc Analyses

Additional post-hoc analyses were conducted, including non-parametric sensitivity analysis, evaluation of collection-sequence effects, and assessment of potential carryover effects. These analyses were considered exploratory and hypothesis-generating.

3.4.4 Missing Data

All planned study visits and urine collections were completed. No imputation procedures were required. Complete-case analysis was conducted using all available observations.

3.5 Protocol Deviations

No protocol deviations were recorded. All 20 participants completed all six visits (three per collection condition) and provided samples under both collection conditions.

3.6 Missing Data

Analyses were conducted using all available data (complete-case analysis). No imputation was performed. The dataset was complete with no missing observations.

4. Outcome Measures and Variable Definitions

4.1 Primary Outcomes

The protocol specified two primary outcomes:

- **Urinary WBC count (cells/ μ L)**, measured by Sysmex UF-5000 urine flow cytometer from first-catch urine samples.
- **Leukocyte esterase (LE)**, measured by dipstick urinalysis. LE was uniformly negative (dipstick score = 0) across all participants and all timepoints under both collection conditions and was not analyzed inferentially (see Section 4.4).

4.2 Secondary Outcome

Day-to-day variation in WBC/ μ L across three timepoints, quantified as the within-subject coefficient of variation (CV%) per collection method (calculated on log-transformed values; see Section 4.3).

4.3 Variable Definitions

Variable	Definition
WBC/ μ L	Raw urinary white blood cell count in cells per microliter, as reported by the Sysmex UF-5000.
WBC.log	$\log_{10}(\text{WBC}/\mu\text{L} + 1)$. The +1 offset is applied to accommodate zero values. Used as the outcome in all parametric analyses.

Variable	Definition
Group	Collection device: Urine Cup (30–60 mL) or Colli-Pee (10 mL). Reference level: Urine Cup.
Time	Visit number within each collection period: Visit 1, Visit 2, or Visit 3. Reference level: Visit 1.
Sequence	Collection order allocation: Colli-Pee-first (Colli-Pee in Period 1, Urine Cup in Period 2) or Cup-first (Urine Cup in Period 1, Colli-Pee in Period 2). Allocated by birth year parity (odd = Colli-Pee-first, even = Cup-first).
Subject.ID	Anonymous participant identifier (SH01–SH20).
CV%	Within-subject coefficient of variation across three timepoints: $(SD / \text{Mean}) \times 100$, calculated on WBC.log values per collection method per subject.
Period 1	The first collection period (three visits within the first week) for each participant. Device used depends on sequence allocation.
Period 2	The second collection period (three visits within the second week) for each participant. Device used is the opposite of Period 1.

4.4 Outcomes Not Analyzed

Leukocyte esterase (LE): LE was uniformly negative (dipstick score = 0) across all participants and all timepoints under both collection conditions. No inferential analysis was performed.

Aim 3 — Time-course evaluation: Evaluation of WBC/ μL and LE over 0–8 hours stratified by urinary pH was not performed due to primary studies showing lower than expected WBC counts and uniformly negative LE measurements. The time-series collection protocol required for Aim 3 (specimens at 0, 1, 2, 4, and 8 hours) was not implemented in this dataset. Aim 3 remains a planned future analysis.

5. General Statistical Principles

All statistical tests are two-sided with a significance level of $\alpha = 0.05$. No adjustment for multiplicity was applied (see Section 8). Continuous variables are summarized with N, mean, SD, median, IQR, minimum, and maximum. All analyses were conducted in R version 4.2.1 using the packages lme4 (v1.1.32), lmerTest (v3.1.3), tidyverse (v2.0.0), and readxl (v1.4.3). Annotated statistical code is documented in an accompanying R Markdown file and available upon request.

6. Primary Analysis — Aim 1

6.1 Distributional Assessment and Transformation

Normality Testing

Normality of raw WBC/ μ L values was assessed using the Shapiro-Wilk test separately for each collection group. Q-Q plots were examined visually to characterize the nature and severity of departures from normality.

Transformation

A $\log_{10}(X+1)$ transformation was applied to WBC/ μ L values prior to parametric analysis. WBC/ μ L values were right-skewed, consistent with flow cytometry count data, necessitating transformation prior to parametric analysis. The +1 offset accommodates one zero value present in the dataset. Shapiro-Wilk testing was repeated on transformed values. Both parametric (LME) and non-parametric (Wilcoxon) analyses were conducted given residual non-normality after transformation.

Note: the decision to apply a $\log_{10}(X+1)$ transformation and to use LME rather than the univariate tests originally described in the protocol are documented as deviations in Section 13.

Group	W (raw)	p (raw)	W (log10)	p (log10)
Urine Cup	0.573	< 0.0001	0.910	0.0003
Colli-Pee	0.479	< 0.0001	0.926	0.0013

6.2 Model Specification

Rationale for Linear Mixed-Effects Model

A standard two-way ANOVA is not appropriate for this study because each participant contributes data under both collection conditions across three timepoints within each condition. This within-subject crossover structure violates the independence assumption required by standard ANOVA. A linear mixed-effects model (LME) with a random intercept for subject appropriately accounts for within-subject correlation.

No additional fixed-effect covariates were included beyond Group and Time in the primary model. Subject was modeled as a random intercept to account for within-subject correlation. Participants were healthy males screened for STI and asymptomatic for urethritis at enrollment. No clinical covariates were expected to vary systematically between collection conditions in this healthy volunteer sample, and no confounders were identified a priori requiring adjustment. Sequence of collection was subsequently evaluated as a post-hoc fixed effect in a sensitivity model (Section 7.2) and found to be non-significant ($F(1,18) = 1.83$, $p = 0.193$).

Primary Model

The following LME was fitted using restricted maximum likelihood (REML):

$$\text{WBC.log} \sim \text{Group} + \text{Time} + (1 \mid \text{Subject.ID})$$

Where:

- Group: collection device (reference = Urine Cup)
- Time: visit number within collection period (Visits 1, 2, 3; reference = Visit 1)
- (1 | Subject.ID): random intercept for subject, capturing within-subject correlation
- WBC.log: $\log_{10}(\text{WBC}/\mu\text{L} + 1)$

P-values for fixed effects were obtained using Satterthwaite’s degrees of freedom approximation (lmerTest package). Type III F-tests are reported.

Model Selection

A full model including the Group × Time interaction was fitted first:

```
WBC.log ~ Group * Time + (1 | Subject.ID)
```

If the Group × Time interaction was non-significant ($p > 0.05$), it was dropped in favor of the parsimonious main-effects model. A non-significant interaction indicates that the device effect is consistent across timepoints, and the simpler model is preferred on grounds of parsimony.

6.3 Primary Results

Effect	F	df	p
Group × Time interaction (full model)	0.30	2, 95	0.740
Group (Colli-Pee vs. Urine Cup) — final model	22.18	1, 97	< 0.0001
Time (Visit 1 vs. 2 vs. 3) — final model	0.30	2, 97	0.745

Group effect: Colli-Pee yielded significantly higher WBC/μL than the urine cup ($\beta = 0.234$ log units, $SE = 0.050$), corresponding to approximately 1.71× higher WBC/μL on the untransformed scale. This supports Aim 1.

Time effect: WBC/μL was stable across the three visits in both groups. No significant day-to-day trend was observed.

7. Sensitivity Analyses

7.1 Non-Parametric Sensitivity — Wilcoxon Signed-Rank Test (Post-Hoc)

As a post-hoc robustness check, given residual non-normality after log transformation, a Wilcoxon signed-rank test was conducted on per-subject mean $\log_{10}(\text{WBC}/\mu\text{L} + 1)$ values (collapsed across three timepoints) to verify the primary Group effect.

Test	Statistic	p
LME — Group effect (primary)	$F(1,97) = 22.18$	< 0.0001
Wilcoxon signed-rank (subject means)	$V = 191$	0.0006

Agreement between the LME and Wilcoxon results confirms the Group effect is robust to distributional assumptions.

7.2 Sequence Effect (Post-Hoc)

As a post-hoc sensitivity analysis, collection sequence (Colli-Pee-first vs. Cup-first) was tested as an additional fixed effect to assess whether collection order confounded the primary Group effect:

WBC.log ~ Group + Time + Sequence + (1 | Subject.ID)

Sequence was non-significant ($F(1,18) = 1.83, p = 0.193$), and the Group effect was unchanged ($F(1,97) = 22.18, p < 0.0001$), confirming that collection order did not confound the primary result. Note that the denominator df for Sequence is 18 (subject-level), correctly reflecting its between-subjects nature in the crossover design.

7.3 Carryover Assessment (Post-Hoc)

As a post-hoc sensitivity analysis, carryover from Period 1 to Period 2 was assessed by comparing Period 2 WBC.log means between sequence groups using a Wilcoxon rank sum test. Colli-Pee-first subjects contributed their Urine Cup data as Period 2; Cup-first subjects contributed their Colli-Pee data as Period 2. No significant difference was detected ($W = 36, p = 0.331$), consistent with no meaningful carryover given the one-week washout interval between collection periods.

8. Secondary Analysis — Aim 2

8.1 Coefficient of Variation

Within-subject day-to-day variation was quantified as CV% calculated across the three timepoints per subject per collection method, using WBC.log values:

$$CV\% = (SD / Mean) \times 100$$

8.2 Descriptive Statistics

Statistic	Urine Cup	Colli-Pee
Median CV%	28.4%	38.4%
IQR	24.2% – 51.0%	23.4% – 53.8%
Mean CV%	42.0%	42.1%

Median and IQR are the primary descriptive statistics given the skewed distribution of CV% values. The wide IQR reflects genuine between-subject variability in day-to-day WBC/μL fluctuation rather than an analytical artifact.

8.3 Inferential Comparison

A paired Wilcoxon signed-rank test was used to compare per-subject CV% between collection methods. A one-sample t-test on paired CV% differences (Colli-Pee minus Urine Cup) was used to derive a 95% confidence interval on the mean difference.

Test	Statistic	p
Paired Wilcoxon signed-rank on CV%	V = 97	0.784
Paired difference t-test (95% CI)	t(19) = 0.01; -17.0%, 17.3%	0.989

There was no significant difference in day-to-day precision between collection methods. The Aim 2 hypothesis that Colli-Pee would improve precision was not supported.

9. Collection Sequence Allocation

Collection sequence was allocated by birth year parity: participants with an odd birth year were assigned to the Colli-Pee-first sequence (n = 9); participants with an even birth year were assigned to the Cup-first sequence (n = 11). Birth year parity is a deterministic rule and does not constitute true randomization. This is acknowledged as a study limitation.

The unequal group sizes (9 vs. 11) reflect the natural distribution of birth year parity in this sample. As reported in Section 7.2, sequence did not significantly affect WBC/ μ L ($F(1,18) = 1.83$, $p = 0.193$), and no carryover was detected ($p = 0.331$). The one-week washout interval between collection periods was considered adequate for this short-term collection-method comparison, as urinary WBC/ μ L is not subject to pharmacokinetic persistence.

10. Sample Size and Power

This study is an exploratory pilot study. Formal a priori power calculations were not performed. The sample size of N = 20 was determined on a practical basis to characterize baseline urinary WBC/ μ L distributions under each collection condition and to estimate within-subject day-to-day variability. Results are intended to inform the design and power calculations of future confirmatory studies.

11. Interim Analyses

No interim analyses were planned or conducted. All analyses were performed on the complete dataset following closure of data collection.

12. Multiplicity

This is an exploratory pilot study. No formal adjustment for multiplicity was applied. Aim 1 (device effect on WBC/ μ L) is the primary inferential focus. Aim 2 (day-to-day precision) is secondary and exploratory. All p-values and confidence intervals are reported unadjusted, and findings should be interpreted as exploratory and hypothesis-generating.

13. Deviations from Protocol Statistical Methods

This SAP was prepared retrospectively following completion of the analysis. The statistical methods described herein deviate from the protocol in the following respects. All deviations were necessitated by the nature of the data or made at the direction of the Principal Investigator. Each is documented with its rationale in accordance with best practice (Stevens et al., 2023).

1. Statistical software: The protocol specified that analyses would be performed using Prism 10 in collaboration with the Indiana University Department of Biostatistics. Analyses were instead performed in R 4.2.1 by the SAP author. This change was made at the direction of the Principal Investigator.

2. Log₁₀(X+1) transformation: The protocol specified descriptive statistics and univariate tests for Aim 1 without specifying a data transformation. To accommodate right-skewness observed in urinary flow cytometry count data, a log₁₀(X+1) transformation was applied prior to parametric analysis to satisfy distributional assumptions to the extent possible.

3. Linear mixed-effects model: The protocol specified univariate tests for the primary comparison. Following recognition that each participant contributed data under both collection conditions across multiple timepoints (crossover design), standard univariate tests were deemed inappropriate due to violation of the independence assumption. A linear mixed-effects model with subject as a random intercept was used instead.

4. Sequence and carryover analyses (post-hoc): Sequence effect and carryover assessment were added as post-hoc sensitivity analyses following clarification of the crossover design structure and the birth year parity allocation rule. These were not described in the original protocol. Results should be interpreted as exploratory.

Any future deviations from this SAP will be documented in the final study report with the rationale for the change.

14. Conventions

Means and medians are presented to one more decimal place than the raw data. Standard deviations and standard errors are presented to two more decimal places than the raw data. Percentages are based on available data with denominators excluding missing values. P-values are reported to four decimal places, except where $p < 0.0001$, which is reported as such. All confidence intervals are 95% two-sided unless otherwise specified. Back-transformed estimates from log-scale models are reported as approximate fold-differences on the original WBC/ μL scale.

15. Summary of Findings

Aim	Hypothesis	Finding	Key Statistic
Aim 1	Colli-Pee yields higher WBC/ μ L	Supported	$F(1,97) = 22.18, p < 0.0001$
Aim 2 (stability)	No day-to-day drift	Consistent with hypothesis	$F(2,97) = 0.30, p = 0.745$
Aim 2 (precision)	Colli-Pee reduces CV%	Not supported	$V = 97, p = 0.784$
Aim 3	Lower pH $\rightarrow$ faster WBC/LE loss	Not completed	See Section 4.4

16. References

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.

Satterthwaite FE (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6), 110–114.

Stevens G, Dolley S, Mogg R, Connor JT (2023). A template for the authoring of statistical analysis plans. *Contemporary Clinical Trials Communications*, 34, 101100.

van der Horst ICC, Hiemstra B, Keus F (2019). DEBATE – statistical analysis plans for observational studies. *BMC Medical Research Methodology*, 19, 233.

Watson HJ (2025). A statistical analysis plan template for observational studies: promoting quality and rigor in research. *Journal of Statistical Theory and Practice*, 19, 91.

von Elm E, Altman DG, Egger M, et al. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Annals of Internal Medicine*, 147(8), 573–577.