Hyperspectral retinal imaging for micro- and nanoplastics detection: a conceptual and methodological framework

Tan Aik Kah

Eye Clinic, Normah Medical Specialist Centre, Kuching, Sarawak, Malaysia

Appendix B. Statistical Power Calculations for *Ex Vivo* and *In Vivo* Pilot Experiments

1. Overview

To guide experimental design for the proposed phantom, ex vivo, and in vivo validation of hyperspectral retinal imaging (HSRI) for detecting microplastics and nanoplastics (MNPs), we conducted prospective statistical power calculations. Calculations address two scenarios:

1. **Binary detection outcomes** (e.g., presence/absence of spectral anomaly or polymer classification success), analyzed using two-proportion tests.[1]
2. **Continuous spectral metrics** (e.g., mean reflectance shift, classifier probability scores), analyzed using two-sample t-tests with standardized effect sizes (Cohen's *d*).[2]

All calculations assume two-sided tests, significance threshold $\alpha = 0.05$, and statistical power $(1-\beta) = 0.80$ unless otherwise specified.[3]

2. Methods for Power Analysis

2.1 Two-proportion tests

Sample sizes were calculated for varying control false-positive rates (p_2) and treated/spiked detection rates (p_1). Required sample size per group was derived as [1]:

$$n = \frac{(z_{1-\frac{a}{2}} \cdot \sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta} \cdot \sqrt{p_1(1-p_1) + p_2(1-p_2)})^2}{(p_1 - p_2)^2}$$

where $\bar{p} = (p_1+p_2)/2$

2.2 Continuous outcomes (Cohen's *d*)

For continuous effect sizes, the per-group sample size was calculated as [2]:

$$n = \frac{2(z_{1-\frac{a}{2}} + z_{1-\beta})^2}{d^2}$$

where d is Cohen's standardized mean difference.

2.3 Design adjustment

For clustered data (e.g., multiple region of interest (ROI)s per retina), sample sizes must be inflated by the **design effect**:

$$DE = 1 + (m-1) \times ICC$$

where m = number of repeated measurements per subject and ICC = intraclass correlation coefficient. Final enrolled n should also account for anticipated attrition (e.g., +15%).

3. Results

3.1 Two-proportion power table

Sample sizes required per group across a range of detection probabilities.

Control (p_2)	Treated (p_1) = 0.20	0.30	0.40	0.50	0.60	0.80
0.05	76	36	22	15	11	6
0.1	199	62	32	20	14	7
0.2	—-	294	82	39	23	10

Note: "—" indicates no positive effect to test ($p_1 \leq p_2$).

Interpretation: For example, if HSRI detects MNPs in 60% of spiked samples while false positives occur in 5% of controls ($p_1 = 0.60$, $p_2 = 0.05$), only **~11 samples per group** are required. Conversely, smaller effects ($p_1 = 0.30$, $p_2 = 0.05$) require **~36 per group**.

3.2 Continuous outcomes (Cohen's *d*)

Sample sizes required per group for continuous spectral differences.

Cohen's d	n per group
0.30 (small)	175
0.50 (medium)	63
0.80 (large)	25
1.00 (very large)	16

Interpretation: Detecting a large standardized effect (d = 0.8) requires ~25 samples per group, while small effects (d = 0.3) require substantially larger sample sizes (~175 per group).

4. Adjustments for Design Complexity

- **Repeated measures example:** If each retina yields 4 ROIs and ICC = 0.25, then DE = 1 + (3 × 0.25) = 1.75. For a nominal requirement of n = 25 per group, adjusted n = 44 per group.

- **Attrition correction:** With 15% expected data loss, enrolment should be inflated to n = 52 per group [3].

5. Recommended Pilot Sample Sizes

- **Phantom experiments:** 5–10 replicates per condition for classifier training and variance estimation.

- **Ex vivo validation:** ~20–30 tissues per condition, corresponding to medium-large effect sizes and providing sufficient classifier training sets.

- **In vivo animal pilot:** 8–15 animals per group if large binary detection effects are expected ($p_1 \geq 0.5$). Adjust upwards with DE and attrition correction as above [1-3].

References

1. Chow S-C, Shao J, Wang H. Sample Size Calculations in Clinical Research. 2nd ed. Chapman & Hall/CRC Biostatistics Series; 2008.
2. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York, NY: Routledge; 1988. Available from: https://doi.org/10.4324/9780203771587
3. Lenth RV. Some practical guidelines for effective sample size determination. Am Stat 2001;55:187-93.