# Sequential Design of Reliability Studies to Evaluate Measurement Error of Biomarkers

**Aiyi Liu**
**(with Enrique F. Schisterman & Chengqing Wu)**

*Biometry and Mathematical Statistics Branch*
*NICHD/NIH/DHHS*

Liua@mail.nih.gov

# OUTLINE

- A Motivating Example
- ANOVA Model For Measurement Errors
- Sample Size and Power Calculation in A Fixed-Size Reliability Study
- Two-Stage Design of A Reliability Study
- Numerical Computation of the Critical Values
- Discussion

## 1. A Motivating Example

Studies have suggested that oxidative stress might be implicated in the risk of human infertility. For this reason, the BioCycle study, a longitudinal study to assess the effects of endogenous hormones (i.e. estrogen and progesterone) on biomarkers of oxidative stress and antioxidant status during the menstrual cycle, was initiated at NICHD/NIH.

One part of the first phase of the study is to enroll a number of 10 women to assess the variation in measure of F2 Isoprostanes, an important oxidative stress biomarker, during various phases of the menstrual cycle. It was planned that at each specific time point within a menstrual cycle, the F2 Isoprostanes values will be measured simultaneously three times, which is used to assess the consistency of the measurements. Since each assay costs about $130, cost reduction was considered when planning for the study.

This is a typical study of reliability. Let $\rho$ denote the intraclass correlation coefficient (to be defined below) of F2 Isoprostanes measures at a selected time point. Then the null hypothesis to be tested in this study is $H_0 : \rho \leq \rho_0 (= 0.5)$ with level of significance to be $\alpha (= 0.05)$.

To address the cost concern, we proposed a two-stage testing procedure to test $H_0$. At the first stage, we will evaluate F2 Isoprostanes measurements from five women and compare the $\rho$-estimate with a critical value. If data indicate small measurement error, then the testing will be stopped. Otherwise, we will continue to collect data from the other five women. By doing this way, the cost could be potentially reduced.

## 2. ANOVA Model For Measurement Errors 

Let $X_{ij}$ be the $j$th $(j = 1, \ldots, p)$ F2-measurement from the $i$th $(i = 1, \ldots n)$ women. Then the one-way ANOVA model is

$$X_{ij} = \mu + u_i + \epsilon_{ij},$$

where $\mu =$ the grand mean (fixed effect), $u_i \sim N(0, \sigma_u^2)$, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. $\sigma_\epsilon^2$ reflects the variation among the measurements within a subject. Further we assume that $u_i$ are independent of $\epsilon_{ij}$. Thus $Cov(X_{ij}, X_{i'j'}) = 0$ if $i \neq i'$, $Cov(X_{ij}, X_{ij'}) = \sigma_u^2$ if $j \neq j'$, and $Var(X_{ij}) = \sigma_u^2 + \sigma_\epsilon^2$.

The intraclass correlation coefficient is defined as

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}.$$

Larger values of $\rho$ indicates higher coherence among measurements from the same subject since then the within-subject error is relatively smaller as compared to the between subject error; perfect consistency occurs when $\rho = 1$ (then $\sigma_\epsilon = 0$).

Define $S_W^2 = \sum_{i=1}^n \sum_{j=1}^p (X_{ij} - \bar{X}_{i.})^2$, the within sum of squares, and $S_B^2 = \sum_{i=1}^n p(\bar{X}_{i.} - \bar{X}_{..})^2$, the between sum of squares, where $\bar{X}_{i.} = \sum_{j=1}^p X_{ij}/p$, the within-subject average, and $\bar{X}_{..} = \sum_{i=1}^n \sum_{j=1}^p X_{ij}/(np)$, the overall average. Then an estimate of $\rho$ is given by:

$$\hat{\rho} = \frac{n(p-1)F - (n-1)}{n(p-1)F + (n-1)(p-1)},$$

where $F = S_B^2/S_W^2$.

## 3. Sample Size and Power Calculation in A Fixed-Size Reliability Study

Note that $\hat{\rho} > c$ if and only if

$$F > c_F = (n-1)(1+(p-1)c)/(n(p-1)(1-c)).$$

With level of significance $\alpha$ we reject $H_0: \rho \leq \rho_0$ if $\hat{\rho} > c$ with such $0 < c < 1$ that

$$\alpha = P_{\rho=\rho_0}(\hat{\rho} > c) = P_{\rho=\rho_0}(F > c_F).$$

But $F \sim \frac{(n-1)(1+(p-1)\rho)}{n(p-1)(1-\rho)} F_{n-1,n(p-1)}$. Hence

$$c = \frac{(1+(p-1)\rho_0)F^{-1}_{n-1,n(p-1)}(1-\alpha) - (1-\rho_0)}{(p-1)(1-\rho_0) + (1+(p-1)\rho_0)F^{-1}_{n-1,n(p-1)}(1-\alpha)}.$$

The power of such test at $\rho > \rho_0$ is thus given by

$$P_\rho(\hat{\rho} > c) = 1 - F_{-1,n(p-1)}\left(\frac{(1-\rho)(1+(p-1)\rho_0)}{(1-\rho_0)(1+(p-1)\rho)} F^{-1}_{n-1,n(p-1)}(1-\alpha)\right).$$

Use this equation for two study questions:
1) Given $n$, what is the power at some $\rho$?
2)Given power $1-\beta$ at some $\rho$, what is $n$?

# 4. Two-Stage Design of A Reliability Study

First, $p$ repeated measurements are taken from each of the $n_1(< n_1)$ subjects, and the $\rho$-estimate $\hat{\rho}_1$ is computed. If $\hat{\rho}_1 > c_1$, then the test stops and $H_0$ is rejected. Otherwise, measurements from $n - n_1$ additional subjects are taken, and the $\rho$-estimate $\hat{\rho}_2$ is computed based on data from all $n$ subjects. If $\hat{\rho}_2 > c_2$, then $H_0$ is rejected. Otherwise, $H_0$ is not rejected and no more measurements are taken.

The critical values $c_1$ and $c_2$ are chosen to satisfy the error requirements. Suppose the type I error to be spent at stage 1 is $\alpha_1$, and that the overall type I error is $\alpha$ and power is $1-\beta$ at $\rho_1$. Then

$$P_{\rho_0}(\hat{\rho}_1 > c_1) = \alpha_1\,, \quad P_{\rho_0}(\hat{\rho}_1 \le c_1, \hat{\rho}_2 > c_2) = \alpha - \alpha_1\,,$$

$$P_{\rho_1}(\hat{\rho}_1 > c_1) + P_{\rho_1}(\hat{\rho}_1 \le c_1, \hat{\rho}_2 > c_2) = 1 - \beta\,.$$

These three equations together determine the critical values and the sample size assuming the allocation ratio, $n_1/n$, of subjects to the first group is specified.

## 5. Numerical Computation of the Critical Values

$$c_1 = \frac{(1+(p-1)\rho_0)F^{-1}_{n_1-1,n_1(p-1)}(1-\alpha_1)-(1-\rho_0)}{(p-1)(1-\rho_0)+(1+(p-1)\rho_0)F^{-1}_{n_1-1,n_1(p-1)}(1-\alpha_1)}.$$

To compute $c_2$, let $U_1$, $U_2$, $U_3$ and $U_4$ be independent $\sim \chi^2$ with d.o.f $n_1-1$, $n_1(p-1)$, $g_2 = n_2 - n_1$, and $g_2(p-1)$, respectively and define

$$b(n, c; \rho) = \frac{(1-\rho)(n-1)(1+c(p-1))}{n(p-1)(1-c)(1+(p-1)\rho)}.$$

Then

$$P\left(\hat{\rho}_1 \le c_1,\ \hat{\rho}_2 > c_2\right) = P\Big(U_1 \le b_1 U_2,\ U_1 + U_3 > b_2(U_2 + U_4)\Big)$$

$$= \int_0^{b_1} \left( \int_\Delta d\chi_{n_1(p-1)}(u_1) d\chi_{g_2}(u_3) d\chi_{g_2(p-1)}(u_4) \right) dF_{n_1-1, n_1(p-1)} \left( \frac{n_1(p-1)}{n_1 - 1}\, y \right),$$

where $b_1 = b(n_1, c_1; \rho)$, $b_2 = b(n_2, c_2; \rho)$, and $\Delta = \{(u_2, u_3, u_4) : (y - b_2)u_2 + u_3 > b_2 u_4\}$.

Simulation from multivariate normal vs $\chi^2$ vs exact computation using multiple integrals!

*Back to the F2 example:* In this example, $n_1 = 5 = n/2$. The interim analysis was planned when half of the subjects supply their blood samples and the values of F2 Isoprostanes are measured. The interim error $\alpha_1 = \alpha/2 = 0.025$. With $\rho_0 = 0.5$, $n_1 = n_2/2 = 5$, we found via $N = 100,000$ simulations that $c_1 = 0.8493$, and $c_2 = 0.7593$. The power of the test at $\rho_1 = 0.80, 0.85, 0.90$ is 0.65, 0.81 and 0.94, respectively. The average sample size at these alternative values is 9, 8, and 7, respectively, reflecting the cost-effective nature of the design.

## 6. Discussion

*Advantage of Sequential Design:* 1) Cost-effective due to possible early stopping; 2) Ethical when procedures such as taking bone marrow samples may be very painful to the study subjects and thus sampling process should be stopped if early evidence shows the measurement error is within the level of tolerance.

*Extension to Multistage Testing:* Ideas are straightforward but numerical computations for critical values become extremely extensive.
*Other Approaches to Choosing Critical Values:* Various error spending functions, optimal designs that minimize expected sample size.

*More Future Research:* Inference on $\rho$ (point estimate, confidence intervals, etc.) following sequential testing. Extension to multi-level and multivariate measurement errors.

| alpha | beta | r | rho_0 | rho_1 | n_1 | c_1 | c_2 | ASN | fixed size |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | p=2 | | | | |
| 0.05 | 0.1 | 1 | 0.5 | 0.7 | 46 | 0.6874 | 0.6284 | 66.05 | 83 |
| 0.05 | 0.1 | 2 | 0.5 | 0.7 | 44 | 0.7133 | 0.6250 | 69.06 | 83 |
| 0.025 | 0.1 | 1 | 0.5 | 0.7 | 55 | 0.6943 | 0.6354 | 80.74 | 102 |
| 0.025 | 0.1 | 2 | 0.5 | 0.7 | 53 | 0.7156 | 0.6319 | 84.25 | 102 |
| 0.05 | 0.2 | 1 | 0.5 | 0.7 | 33 | 0.7155 | 0.6493 | 51.82 | 61 |
| 0.05 | 0.2 | 2 | 0.5 | 0.7 | 32 | 0.7425 | 0.6423 | 54.08 | 61 |
| 0.025 | 0.2 | 1 | 0.5 | 0.7 | 43 | 0.7153 | 0.6528 | 67.91 | 78 |
| 0.025 | 0.2 | 2 | 0.5 | 0.7 | 40 | 0.7414 | 0.6510 | 68.23 | 78 |
| 0.05 | 0.1 | 1 | 0.5 | 0.8 | 16 | 0.7864 | 0.7049 | 23.13 | 29 |
| 0.05 | 0.1 | 2 | 0.5 | 0.8 | 15 | 0.8211 | 0.7014 | 23.90 | 29 |
| 0.025 | 0.1 | 1 | 0.5 | 0.8 | 20 | 0.7901 | 0.7119 | 29.09 | 36 |
| 0.025 | 0.1 | 2 | 0.5 | 0.8 | 19 | 0.8180 | 0.7084 | 30.19 | 36 |
| 0.05 | 0.2 | 1 | 0.5 | 0.8 | 12 | 0.8176 | 0.7310 | 18.85 | 22 |
| 0.05 | 0.2 | 2 | 0.5 | 0.8 | 11 | 0.8554 | 0.7293 | 18.88 | 22 |
| 0.025 | 0.2 | 1 | 0.5 | 0.8 | 15 | 0.8211 | 0.7397 | 23.90 | 27 |
| 0.025 | 0.2 | 2 | 0.5 | 0.8 | 14 | 0.8515 | 0.7362 | 24.13 | 27 |
| 0.05 | 0.1 | 1 | 0.6 | 0.8 | 29 | 0.7882 | 0.7319 | 41.57 | 52 |
| 0.05 | 0.1 | 2 | 0.6 | 0.8 | 27 | 0.8138 | 0.7293 | 42.68 | 52 |
| 0.025 | 0.1 | 1 | 0.6 | 0.8 | 34 | 0.7959 | 0.7397 | 50.13 | 64 |
| 0.025 | 0.1 | 2 | 0.6 | 0.8 | 33 | 0.8146 | 0.7362 | 52.61 | 64 |
| 0.05 | 0.2 | 1 | 0.6 | 0.8 | 21 | 0.8131 | 0.7536 | 32.92 | 38 |
| 0.05 | 0.2 | 2 | 0.6 | 0.8 | 20 | 0.8383 | 0.7484 | 33.92 | 38 |
| 0.025 | 0.2 | 1 | 0.6 | 0.8 | 27 | 0.8138 | 0.7571 | 42.68 | 48 |
| 0.025 | 0.2 | 2 | 0.6 | 0.8 | 25 | 0.8372 | 0.7553 | 42.77 | 48 |
| 0.05 | 0.1 | 1 | 0.6 | 0.9 | 9 | 0.8851 | 0.8127 | 12.81 | 15 |
| 0.05 | 0.1 | 2 | 0.6 | 0.9 | 8 | 0.9166 | 0.8136 | 12.84 | 15 |
| 0.025 | 0.1 | 1 | 0.6 | 0.9 | 10 | 0.8978 | 0.8301 | 14.91 | 19 |
| 0.025 | 0.1 | 2 | 0.6 | 0.9 | 10 | 0.9149 | 0.8231 | 16.04 | 19 |
| 0.05 | 0.2 | 1 | 0.6 | 0.9 | 7 | 0.9069 | 0.8336 | 10.81 | 12 |
| 0.05 | 0.2 | 2 | 0.6 | 0.9 | 6 | 0.9394 | 0.8379 | 10.37 | 12 |
| 0.025 | 0.2 | 1 | 0.6 | 0.9 | 8 | 0.9166 | 0.8509 | 12.84 | 14 |
| 0.025 | 0.2 | 2 | 0.6 | 0.9 | 8 | 0.9327 | 0.8440 | 13.71 | 14 |
| | | | | | p=3 | | | | |
| 0.05 | 0.1 | 1 | 0.5 | 0.7 | 28 | 0.6826 | 0.6250 | 40.08 | 53 |
| 0.05 | 0.1 | 2 | 0.5 | 0.7 | 27 | 0.7074 | 0.6215 | 41.91 | 53 |
| 0.025 | 0.1 | 1 | 0.5 | 0.7 | 35 | 0.6863 | 0.6319 | 50.35 | 64 |
| 0.025 | 0.1 | 2 | 0.5 | 0.7 | 33 | 0.7088 | 0.6284 | 51.55 | 64 |
| 0.05 | 0.2 | 1 | 0.5 | 0.7 | 21 | 0.7059 | 0.6423 | 32.41 | 38 |
| 0.05 | 0.2 | 2 | 0.5 | 0.7 | 20 | 0.7339 | 0.6380 | 33.27 | 38 |
| 0.025 | 0.2 | 1 | 0.5 | 0.7 | 26 | 0.7106 | 0.6493 | 40.76 | 48 |
| 0.025 | 0.2 | 2 | 0.5 | 0.7 | 25 | 0.7333 | 0.6458 | 41.88 | 48 |
| 0.05 | 0.1 | 1 | 0.5 | 0.8 | 11 | 0.7656 | 0.6884 | 15.35 | 19 |
| 0.05 | 0.1 | 2 | 0.5 | 0.8 | 10 | 0.8027 | 0.6875 | 15.40 | 19 |
| 0.025 | 0.1 | 1 | 0.5 | 0.8 | 13 | 0.7756 | 0.7014 | 18.45 | 23 |
| 0.025 | 0.1 | 2 | 0.5 | 0.8 | 12 | 0.8060 | 0.6997 | 18.63 | 23 |
| 0.05 | 0.2 | 1 | 0.5 | 0.8 | 8 | 0.7984 | 0.7154 | 12.22 | 14 |
| 0.05 | 0.2 | 2 | 0.5 | 0.8 | 7 | 0.8409 | 0.7154 | 11.75 | 14 |
| 0.025 | 0.2 | 1 | 0.5 | 0.8 | 10 | 0.8027 | 0.7223 | 15.40 | 18 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.025 | 0.2 | 2 | 0.5 | 0.8 | 9 | 0.8365 | 0.7258 | 15.09 | 18 |
| 0.05 | 0.1 | 1 | 0.6 | 0.8 | 19 | 0.7810 | 0.7275 | 26.93 | 35 |
| 0.05 | 0.1 | 2 | 0.6 | 0.8 | 18 | 0.8045 | 0.7240 | 27.81 | 35 |
| 0.025 | 0.1 | 1 | 0.6 | 0.8 | 22 | 0.7897 | 0.7362 | 32.05 | 42 |
| 0.025 | 0.1 | 2 | 0.6 | 0.8 | 22 | 0.8057 | 0.7310 | 34.16 | 42 |
| 0.05 | 0.2 | 1 | 0.6 | 0.8 | 14 | 0.8033 | 0.7449 | 21.55 | 25 |
| 0.05 | 0.2 | 2 | 0.6 | 0.8 | 13 | 0.8299 | 0.7414 | 21.68 | 25 |
| 0.025 | 0.2 | 1 | 0.6 | 0.8 | 17 | 0.8089 | 0.7536 | 26.66 | 32 |
| 0.025 | 0.2 | 2 | 0.6 | 0.8 | 16 | 0.8304 | 0.7501 | 26.94 | 32 |
| 0.05 | 0.1 | 1 | 0.6 | 0.9 | 6 | 0.8718 | 0.8023 | 8.44 | 11 |
| 0.05 | 0.1 | 2 | 0.6 | 0.9 | 6 | 0.8936 | 0.7936 | 9.05 | 11 |
| 0.025 | 0.1 | 1 | 0.6 | 0.9 | 7 | 0.8808 | 0.8162 | 10.03 | 13 |
| 0.025 | 0.1 | 2 | 0.6 | 0.9 | 7 | 0.8981 | 0.8092 | 10.68 | 13 |
| 0.05 | 0.2 | 1 | 0.6 | 0.9 | 5 | 0.8871 | 0.8170 | 7.45 | 8 |
| 0.05 | 0.2 | 2 | 0.6 | 0.9 | 4 | 0.9261 | 0.8231 | 6.73 | 8 |
| 0.025 | 0.2 | 1 | 0.6 | 0.9 | 6 | 0.8936 | 0.8266 | 9.05 | 10 |
| 0.025 | 0.2 | 2 | 0.6 | 0.9 | 5 | 0.9248 | 0.8353 | 8.42 | 10 |
| | | | | p=4 | | | | | |
| 0.05 | 0.1 | 1 | 0.5 | 0.7 | 23 | 0.6773 | 0.6215 | 32.58 | 43 |
| 0.05 | 0.1 | 2 | 0.5 | 0.7 | 22 | 0.7022 | 0.6180 | 33.77 | 43 |
| 0.025 | 0.1 | 1 | 0.5 | 0.7 | 28 | 0.6831 | 0.6284 | 40.09 | 52 |
| 0.025 | 0.1 | 2 | 0.5 | 0.7 | 27 | 0.7034 | 0.6250 | 41.58 | 52 |
| 0.05 | 0.2 | 1 | 0.5 | 0.7 | 17 | 0.7010 | 0.6380 | 26.05 | 31 |
| 0.05 | 0.2 | 2 | 0.5 | 0.7 | 16 | 0.7297 | 0.6350 | 26.43 | 31 |
| 0.025 | 0.2 | 1 | 0.5 | 0.7 | 21 | 0.7061 | 0.6458 | 32.63 | 39 |
| 0.025 | 0.2 | 2 | 0.5 | 0.7 | 20 | 0.7292 | 0.6432 | 33.23 | 39 |
| 0.05 | 0.1 | 1 | 0.5 | 0.8 | 9 | 0.7581 | 0.6832 | 12.50 | 16 |
| 0.05 | 0.1 | 2 | 0.5 | 0.8 | 8 | 0.7970 | 0.6823 | 12.27 | 16 |
| 0.025 | 0.1 | 1 | 0.5 | 0.8 | 10 | 0.7743 | 0.6980 | 14.35 | 19 |
| 0.025 | 0.1 | 2 | 0.5 | 0.8 | 10 | 0.7962 | 0.6910 | 15.23 | 19 |
| 0.05 | 0.2 | 1 | 0.5 | 0.8 | 6 | 0.7991 | 0.7154 | 9.31 | 12 |
| 0.05 | 0.2 | 2 | 0.5 | 0.8 | 6 | 0.8272 | 0.7049 | 9.87 | 12 |
| 0.025 | 0.2 | 1 | 0.5 | 0.8 | 8 | 0.7970 | 0.7188 | 12.27 | 14 |
| 0.025 | 0.2 | 2 | 0.5 | 0.8 | 8 | 0.8194 | 0.7101 | 12.96 | 14 |
| 0.05 | 0.1 | 1 | 0.6 | 0.8 | 16 | 0.7759 | 0.7223 | 22.46 | 29 |
| 0.05 | 0.1 | 2 | 0.6 | 0.8 | 15 | 0.7997 | 0.7206 | 22.93 | 29 |
| 0.025 | 0.1 | 1 | 0.6 | 0.8 | 19 | 0.7828 | 0.7327 | 27.11 | 35 |
| 0.025 | 0.1 | 2 | 0.6 | 0.8 | 18 | 0.8024 | 0.7301 | 27.75 | 35 |
| 0.05 | 0.2 | 1 | 0.6 | 0.8 | 12 | 0.7962 | 0.7397 | 18.20 | 21 |
| 0.05 | 0.2 | 2 | 0.6 | 0.8 | 11 | 0.8232 | 0.7362 | 18.10 | 21 |
| 0.025 | 0.2 | 1 | 0.6 | 0.8 | 14 | 0.8048 | 0.7501 | 21.80 | 26 |
| 0.025 | 0.2 | 2 | 0.6 | 0.8 | 13 | 0.8271 | 0.7466 | 21.74 | 26 |
| 0.05 | 0.1 | 1 | 0.6 | 0.9 | 5 | 0.8650 | 0.7971 | 7.04 | 10 |
| 0.05 | 0.1 | 2 | 0.6 | 0.9 | 5 | 0.8867 | 0.7884 | 7.49 | 10 |
| 0.025 | 0.1 | 1 | 0.6 | 0.9 | 6 | 0.8718 | 0.8092 | 8.47 | 11 |
| 0.025 | 0.1 | 2 | 0.6 | 0.9 | 6 | 0.8891 | 0.8005 | 8.96 | 11 |
| 0.05 | 0.2 | 1 | 0.6 | 0.9 | 4 | 0.8834 | 0.8127 | 6.02 | 7 |
| 0.05 | 0.2 | 2 | 0.6 | 0.9 | 4 | 0.9048 | 0.8049 | 6.38 | 7 |
| 0.025 | 0.2 | 1 | 0.6 | 0.9 | 5 | 0.8867 | 0.8231 | 7.49 | 8 |
| 0.025 | 0.2 | 2 | 0.6 | 0.9 | 5 | 0.9036 | 0.8144 | 7.91 | 8 |