## Supplementary material to

# Estimation and Testing Based on Data Subject to Measurement Errors: From Parametric to Non-Parametric Likelihood Methods

Albert Vexler[a*†], Wan-Min Tsai[a] and Yaakov Malinovsky[b]

## Appendix

### A1. Parametric likelihood based on the hybrid design

Under the normal assumption, the log likelihood function based on the pooled-unpooled data is in the form of

$$l_H(Z|\mu_x,\sigma_x^2,\sigma_m^2) = -\sum_{i=1}^{n_p}\frac{\left(Z_i^p-\mu_x\right)^2}{2(\sigma_m^2+\sigma_x^2/p)} - \sum_{j=1}^{n_{up}}\frac{\left(Z_j-\mu_x\right)^2}{2(\sigma_m^2+\sigma_x^2)}$$

$$-\frac{n_p}{2}log\left(2\pi\left(\sigma_m^2+\frac{\sigma_x^2}{p}\right)\right) - \frac{n_{up}}{2}log\left(2\pi(\sigma_m^2+\sigma_x^2)\right).$$

Taking the corresponding first derivatives of the log likelihood function $l_H(Z|\mu_x,\sigma_x^2,\sigma_m^2)$ yields

$$\frac{\partial l_H}{\partial \mu_x} = \sum_{i=1}^{n_p}\frac{Z_i^p-\mu_x}{\sigma_m^2+\sigma_x^2/p} + \sum_{j=1}^{n_{up}}\frac{Z_j-\mu_x}{\sigma_m^2+\sigma_x^2},$$

$$\frac{\partial l_H}{\partial \sigma_x^2} = \sum_{i=1}^{n_p}\frac{\left(Z_i^p-\mu_x\right)^2}{2p(\sigma_m^2+\sigma_x^2/p)^2} + \sum_{j=1}^{n_{up}}\frac{\left(Z_j-\mu_x\right)^2}{2(\sigma_m^2+\sigma_x^2)^2}$$

[a] *Department of Biostatisics, The State University of New York, Buffalo, NY 14214, U.S.A.*
[b] *Divison of Epidemiology, Statistics and Prevention Research, National Institute of Child Health and Human Development, NIH/DHHS, 6100 Executive Blvd., Bethesda, MD 20892, USA*
[*] *Correspondence to: Albert Vexler, Department of Biostatisics, The State University of New York, Buffalo, NY 14214, U.S.A.*
[†] *Email: avexler@buffalo.edu*

$$-\frac{n_p}{2p(\sigma_m^2+\sigma_x^2/p)}-\frac{n_{up}}{2(\sigma_m^2+\sigma_x^2)},$$

$$\frac{\partial l_H}{\partial\sigma_m^2}=\sum_{i=1}^{n_p}\frac{\left(Z_i^p-\mu_x\right)^2}{2(\sigma_m^2+\sigma_x^2/p)^2}+\sum_{j=1}^{n_{up}}\frac{\left(Z_j-\mu_x\right)^2}{2(\sigma_m^2+\sigma_x^2)^2}-\frac{n_p}{2(\sigma_m^2+\sigma_x^2/p)}-\frac{n_{up}}{2(\sigma_m^2+\sigma_x^2)}.$$

Let these first derivatives be equal to zero. Then we can obtain the following system of equations whose solutions are the maximum likelihood estimators.

$$\hat{\mu}_x=\frac{(\hat{\sigma}_m^2+\hat{\sigma}_x^2)\sum_{i=1}^{n_p}Z_i^p+(\hat{\sigma}_m^2+\hat{\sigma}_x^2/p)\sum_{j=1}^{n_{up}}Z_i}{n_{up}(\hat{\sigma}_m^2+\hat{\sigma}_x^2/p)+n_p(\hat{\sigma}_m^2+\hat{\sigma}_x^2)},$$

$$\hat{\sigma}_x^2=\frac{p}{p-1}\left[\frac{\sum_{j=1}^{n_{up}}\left(Z_j-\hat{\mu}_x\right)^2}{n_{up}}-\frac{\sum_{i=1}^{n_p}\left(Z_i^p-\hat{\mu}_x\right)^2}{n_p}\right],$$

$$\hat{\sigma}_m^2=\frac{\sum_{i=1}^{n_p}\left(Z_i^p-\hat{\mu}_x\right)^2}{n_p}-\frac{\hat{\sigma}_x^2}{p}.$$

To present the matrix **I**, we obtain the second derivatives of the log likelihood function as follows:

$$\frac{\partial^2 l_H}{\partial(\mu_x)^2}=-\frac{n_p}{\sigma_m^2+\sigma_x^2/p}-\frac{n_{up}}{\sigma_m^2+\sigma_x^2},$$

$$\frac{\partial^2 l_H}{\partial(\sigma_x^2)^2}=-\frac{\sum_{i=1}^{n_p}\left(Z_i^p-\mu_x\right)^2}{p^2(\sigma_m^2+\sigma_x^2/p)^3}-\frac{\sum_{j=1}^{n_{up}}\left(Z_j-\mu_x\right)^2}{(\sigma_m^2+\sigma_x^2)^3}+\frac{n_p}{2p^2(\sigma_m^2+\sigma_x^2/p)^2}+\frac{n_{up}}{2(\sigma_m^2+\sigma_x^2)^2},$$

$$\frac{\partial^2 l_H}{\partial(\sigma_m^2)^2}=-\frac{\sum_{i=1}^{n_p}\left(Z_i^p-\mu_x\right)^2}{(\sigma_m^2+\sigma_x^2/p)^3}-\frac{\sum_{j=1}^{n_{up}}\left(Z_j-\mu_x\right)^2}{(\sigma_m^2+\sigma_x^2)^3}+\frac{n_p}{2(\sigma_m^2+\sigma_x^2/p)^2}+\frac{n_{up}}{2(\sigma_m^2+\sigma_x^2)^2},$$

$$\frac{\partial^2 l_H}{\partial\mu_x\partial\sigma_x^2}=-\sum_{i=1}^{n_p}\frac{Z_i^p-\mu_x}{p(\sigma_m^2+\sigma_x^2/p)^2}-\sum_{j=1}^{n_{up}}\frac{Z_j-\mu_x}{(\sigma_m^2+\sigma_x^2)^2},$$

$$\frac{\partial^2 l_H}{\partial\mu_x\partial\sigma_m^2}=-\sum_{i=1}^{n_p}\frac{Z_i^p-\mu_x}{(\sigma_m^2+\sigma_x^2/p)^2}-\sum_{j=1}^{n_{up}}\frac{Z_j-\mu_x}{(\sigma_m^2+\sigma_x^2)^2},$$

$$\frac{\partial^2 l_H}{\partial\sigma_x^2\partial\sigma_m^2} = -\frac{\sum_{i=1}^{n_p}\left(Z_i^p-\mu_x\right)^2}{p(\sigma_m^2+\sigma_x^2/p)^3} - \frac{\sum_{j=1}^{n_{up}}\left(Z_j-\mu_x\right)^2}{(\sigma_m^2+\sigma_x^2)^3} + \frac{n_p}{2p(\sigma_m^2+\sigma_x^2/p)^2} + \frac{n_{up}}{2(\sigma_m^2+\sigma_x^2)^2}.$$

The Fisher Information matrix $\mathbf{I}$ can be calculated as

$$\mathbf{I} = -\lim_{N\to\infty}\frac{1}{N}E\begin{bmatrix} \frac{\partial^2 l_H}{\partial(\mu_x)^2} & \frac{\partial^2 l_H}{\partial\mu_x\partial\sigma_x^2} & \frac{\partial^2 l_H}{\partial\mu_x\partial\sigma_m^2} \\ \frac{\partial^2 l_H}{\partial\sigma_x^2\partial\mu_x} & \frac{\partial^2 l_H}{\partial(\sigma_x^2)^2} & \frac{\partial^2 l_H}{\partial\sigma_x^2\partial\sigma_m^2} \\ \frac{\partial^2 l_H}{\partial\sigma_m^2\partial\mu_x} & \frac{\partial^2 l_H}{\partial\sigma_m^2\partial\sigma_x^2} & \frac{\partial^2 l_H}{\partial(\sigma_m^2)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\alpha}{\sigma_m^2+\sigma_x^2/p}+\frac{1-\alpha}{\sigma_m^2+\sigma_x^2} & 0 & 0 \\ 0 & \frac{\alpha}{2p^2(\sigma_m^2+\sigma_x^2/p)^2}+\frac{1-\alpha}{2(\sigma_m^2+\sigma_x^2)^2} & \frac{\alpha}{2p(\sigma_m^2+\sigma_x^2/p)^2}+\frac{1-\alpha}{2(\sigma_m^2+\sigma_x^2)^2} \\ 0 & \frac{\alpha}{2p(\sigma_m^2+\sigma_x^2/p)^2}+\frac{1-\alpha}{2(\sigma_m^2+\sigma_x^2)^2} & \frac{\alpha}{2(\sigma_m^2+\sigma_x^2/p)^2}+\frac{1-\alpha}{2(\sigma_m^2+\sigma_x^2)^2} \end{bmatrix}.$$

The asymptotic distribution of the maximum likelihood estimators (1) can be derived

$$\sqrt{N}\begin{pmatrix} \hat{\mu}_x-\mu_x \\ \hat{\sigma}_x^2-\sigma_x^2 \\ \hat{\sigma}_m^2-\sigma_m^2 \end{pmatrix}\sim MVN(\mathbf{0},\mathbf{I}^{-1}).$$

**A2. The asymptotic distribution of the log empirical likelihood ratio test**

*A2.1. The repeated measurements*

Under the null hypothesis, the empirical likelihood function based on repeated measures data is

$$L_R(\mu_x) = \prod_{i=1}^{t}\frac{1}{t}\frac{1}{1+\lambda(\bar{Z}_i-\mu_x)}.$$

Thus, the EL ratio based on repeated measures data is given by

$$EL_R = \frac{\max_{p:H_1}\prod_{i=1}^{t}p_i}{L_R(\mu_x)} = \prod_{i=1}^{t}\frac{\frac{1}{t}}{\frac{1}{t}\frac{1}{1+\lambda(\bar{Z}_i-\mu_x)}} = \prod_{i=1}^{t}1+\lambda(\bar{Z}_i-\mu_x),$$

where $\lambda$ is a root of

$$\sum_{i=1}^{t} \frac{\bar{Z}_i - \mu_x}{1 + \lambda(\bar{Z}_i - \mu_x)} = 0.$$

By applying the Taylor expansion, we have

$$\sum_{i=1}^{t} \frac{\bar{Z}_i - \mu_x}{1 + \lambda(\bar{Z}_i - \mu_x)} = \sum_{i=1}^{t} (\bar{Z}_i - \mu_x) - \lambda \sum_{i=1}^{t} (\bar{Z}_i - \mu_x)^2 + \mathrm{R}_1 = 0,$$

where $R_1$ is a remainder term (for details, see Owen [25], Vexler et al. [26]).

Therefore,

$$\lambda \approx \frac{\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)}{\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)^2}, \quad \text{as } t \to \infty.$$

Now consider the log empirical likelihood ratio statistic in the form of

$$l_R(\mu_x) = 2\log EL_R = 2 \sum_{i=1}^{t} \log\left[1 + \lambda(\bar{Z}_i - \mu_x)\right].$$

Again, by applying the Taylor expansion with $\lambda(\bar{Z}_i - \mu_x)$ around zero, we obtain

$$l_R(\mu_x) = 2 \sum_{i=1}^{t} \left[\lambda(\bar{Z}_i - \mu_x) - \frac{\lambda^2}{2} (\bar{Z}_i - \mu_x)^2 + R_2\right],$$

where $R_2$ is a remainder term (for details, see Owen [25], Vexler et al. [26]).

Substituting the approximate solution of λ that we obtained yields

$$l_R(\mu_x) \approx 2 \frac{\left[\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)\right]^2}{\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)^2} - \frac{\left[\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)\right]^2}{\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)^2} = \left[\frac{\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)}{\sqrt{\sum_{i=1}^{t} (\bar{Z}_i - \mu_x)^2}}\right]^2,$$

where

$$\frac{\sum_{i=1}^{t}(\bar{Z}_i - \mu_x)}{\sqrt{\sum_{i=1}^{t}(\bar{Z}_i - \mu_x)^2}} \xrightarrow{p} N(0,1), \text{ as } t \to \infty.$$

Hence,

$$l_R(\mu_x) \approx [\frac{\sum_{i=1}^{t}(\bar{Z}_i - \mu_x)}{\sqrt{\sum_{i=1}^{t}(\bar{Z}_i - \mu_x)^2}}]^2 \xrightarrow{p} \chi_1^2, \text{ as t} \to \infty.$$

Therefore, we complete to outline the proof of Proposition 3.2.1.

*A2.2 The hybrid design*

In a similar manner to Section A2.1., we present the empirical likelihood ratio based on the pooled-unpooled data in the form of

$$EL_H = \frac{\max_{p:H_1} \prod_{i=1}^{n_p} p_i \prod_{j=1}^{n_{up}} q_j}{L_H(\mu_x)} = [\prod_{i=1}^{n_p} \frac{\frac{1}{n_p}}{\frac{1}{n_p}\frac{1}{1+\lambda_1(Z_i^p - \mu_x)}}][\prod_{j=1}^{n_{up}} \frac{\frac{1}{n_{up}}}{\frac{1}{n_{up}}\frac{1}{1+\lambda_2(Z_j - \mu_x)}}]$$

$$= [\prod_{i=1}^{n_p} 1 + \lambda_1(Z_i^p - \mu_x)]\,[\prod_{j=1}^{n_{up}} 1 + \lambda_2(Z_j - \mu_x)],$$

where $\lambda_1$ and $\lambda_2$ are roots of

$$\sum_{i=1}^{n_p} \frac{(Z_i^p - \mu_x)}{1+\lambda_1(Z_i^p - \mu_x)} = 0\,, \quad \sum_{j=1}^{n_{up}} \frac{(Z_j - \mu_x)}{1+\lambda_2(Z_j - \mu_x)} = 0,$$

respectively.

By applying the Taylor expansion, we have

$$\sum_{i=1}^{n_p} \frac{(Z_i^p - \mu_x)}{1+\lambda_1(Z_i^p - \mu_x)} = \sum_{i=1}^{n_p}(Z_i^p - \mu_x) - \lambda_1 \sum_{i=1}^{n_p}(Z_i^p - \mu_x)^2 + \mathrm{R}_1 = 0,$$

where $R_1$ is a remainder term (for details, see Owen [25], Vexler et al. [26]).

Consequently,

$$\lambda_1 \approx \frac{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)}{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)^2}, \text{as } n_p \to \infty.$$

Similarly, one can obtain

$$\lambda_2 \approx \frac{\sum_{j=1}^{n_{up}}(Z_j - \mu_x)}{\sum_{j=1}^{n_{up}}(Z_j - \mu_x)^2}, \text{as } n_{up} \to \infty.$$

Now consider the log EL ratio statistic, $l_H(\mu_x) = 2\log EL_H$, presented in the equation (3).

By applying the Taylor expansion with $\lambda_1(Z_i^p - \mu_x)$ and $\lambda_2(Z_j - \mu_x)$ around zeros, we obtain

$$l_H(\mu_x) = 2\sum_{i=1}^{n_p}[\lambda_1(Z_i^p - \mu_x) - \frac{{\lambda_1}^2}{2}(Z_i^p - \mu_x)^2 + R_2] +$$

$$2\sum_{j=1}^{n_{up}}[\lambda_2(Z_j - \mu_x) - \frac{{\lambda_2}^2}{2}(Z_j - \mu_x)^2 + R_3],$$

where $R_2$ and $R_3$ are remainder terms (for details, see Owen [25], Vexler et al. [26]).

Substituting the approximate solutions of $\lambda_1$ and $\lambda_2$ yields

$$l_H(\mu_x) \approx 2\frac{\left[\sum_{i=1}^{n_p}(Z_i^p - \mu_x)\right]^2}{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)^2} - \frac{\left[\sum_{i=1}^{n_p}(Z_i^p - \mu_x)\right]^2}{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)^2} + 2\frac{\left[\sum_{j=1}^{n_{up}}(Z_j - \mu_x)\right]^2}{\sum_{j=1}^{n_{up}}(Z_j - \mu_x)^2}$$

$$-\frac{\left[\sum_{i=1}^{n_{up}} Z_j - \mu_x\right]^2}{\sum_{i=1}^{n_{up}}(Z_j - \mu_x)^2} = [\frac{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)}{\sqrt{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)^2}}]^2 + [\frac{\sum_{j=1}^{n_{up}}(Z_j - \mu_x)}{\sqrt{\sum_{i=1}^{n_{up}}(Z_j - \mu_x)^2}}]^2,$$

where

$$\frac{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)}{\sqrt{\sum_{i=1}^{n_p}(Z_i^p - \mu_x)^2}} \xrightarrow{p} N(0,1) \text{ as } n_p \to \infty, \frac{\sum_{j=1}^{n_{up}}(Z_j - \mu_x)}{\sqrt{\sum_{i=1}^{n_{up}}(Z_j - \mu_x)^2}} \xrightarrow{p} N(0,1) \text{ as } n_{up} \to \infty.$$

Hence,

$$2\sum_{i=1}^{n_p}\log\left[1+\lambda_1\left(Z_i^p-\mu_x\right)\right]+2\sum_{j=1}^{n_{up}}\log\left[1+\lambda_2\left(Z_j-\mu_x\right)\right]$$

$$\approx [\frac{\sum_{i=1}^{n_p}\left(Z_i^p-\mu_x\right)}{\sqrt{\sum_{i=1}^{n_p}\left(Z_i^p-\mu_x\right)^2}}]^2+[\frac{\sum_{j=1}^{n_{up}}\left(Z_j-\mu_x\right)}{\sqrt{\sum_{i=1}^{n_{up}}\left(Z_j-\mu_x\right)^2}}]^2\xrightarrow{p}\chi_2^2 \text{ as } n_p, n_{up}\to\infty.$$

Thus, we complete the sketchy proof of Proposition 3.3.1.