Meta-analysis comparing National Children's Study analytical concentrations to comparable literature values

John Rogers[1]
[1]Westat, 1600 Research Blvd., Rockville, MD 20850

Abstract

Biospecimen handling methods proposed for large multi-site epidemiological studies like the National Children's Study (NCS) should be validated to ensure they provide measurements similar to other published results. We developed regression models to compare blood concentrations of HbA1c, fasting glucose, and fasting insulin from pregnant women in the NCS Vanguard Study with concentrations in similar subpopulations in peer-reviewed publications. The data were compared based on means and standard deviations. The models developed account for measurement error, differences between subpopulations reported in the same publication, and differences between publications. The measurements are roughly log-normally distributed; thus the distribution of reported means and standard deviations are skewed. This paper uses simulation to compare various statistical models considered for the analysis and presents the reasons for selecting the final model. We show that the blood concentrations of HbA1c, fasting glucose, and fasting insulin from pregnant women in the NCS Vanguard Study are similar to reported values from other studies.

Key Words: National Children's Study, mixed model

1. Introduction

The National Children's Study (NCS) was a planned large-scale epidemiological cohort study of U.S. children and their parents. NCS biospecimen handling procedures were developed to minimize variability by using standardized protocols across multiple study sites, and centralized processing, aliquoting, and long-term storage of biospecimens. The Vanguard Pilot was a prospective birth cohort study designed to pilot strategies and procedures under consideration for the NCS[1,2]. As one evaluation of the operational quality of biospecimen procedures in the NCS Vanguard Pilot, this analysis compares hemoglobin A1c (HbA1c), fasting glucose, and fasting insulin measurements from NCS Vanguard blood samples with comparable values from peer-reviewed publications based on means (location) and standard deviations (scale).

2. The Data

The NCS data are measured blood concentrations of hemoglobin A1c (HbA1c), fasting glucose, and fasting insulin from pregnant women in the NCS Vanguard Study. Pregnant women were recruited during 2009-2010 in seven locations throughout the United States using a multistage area probability sampling design. The data analysis used the sample mean and standard deviation of the measured concentrations.

The data from peer-reviewed publications are HbA1c, fasting glucose, and fasting insulin concentrations for pregnant women. The publications primarily reported on US and European women. Generally the data reported in the publications are sample size, mean, and standard deviation (or standard error). We did not use publication data from: (1) case, treatment, or intervention groups; (2) plots or graphs; (3) summary statistics based on within-subject averages; or (4) quantiles or ranges. Some data was excluded if the analytical procedures differed from the NCS procedures.

2.1 Data Characteristics

The individual NCS measurements was found to be reasonably described by a lognormal distribution. The analysis assumes that of the underlying data used to calculate the summary statistics in the peer-reviewed publications can also be described by a lognormal distribution.

Sources of variation included:

- Differences between publication;
- Differences between multiple subpopulations reported on in the same publication (for example, one publication reported on one group of women early in pregnancy and one late in pregnancy); and
- Differences between people within a subpopulation, including laboratory analytical error.

3. Basic model for estimating location and scale differences

The basic model for estimating location and scale differences is described below. Appendix A provides additional background on the model and its derivation.

Let T be a measure of location or scale, assumed roughly normally distributed, possibly after being log-transformed.

$$T = \omega + I_{NCS}\alpha + \beta_i + \gamma_{j(i)} + \varepsilon_{k(ij)}$$

ω = intercept, mean across publications.
I_{NCS} = Indicator of NCS data (1=NCS, 0=publication).
α = mean difference between the NCS and other publications.

Sources of variation incorporated as random effects are:

- $\beta_i \sim N(0, \sigma_\beta^2)$ for i^{th} publication;
- $\gamma_{j(i)} \sim N(0, \sigma_\gamma^2)$ for j^{th} subpopulation within publication i; and
- $\varepsilon_{k(ij)} \sim N(0, \hat{V})$ for variation within subpopulation j and publication i associated with analytical error and differences between women (k). $\hat{V}$ is the estimated within-subpopulation error variance of T, i.e., the uncertainty in the reported summary statistics for each subpopulation. $\hat{V}$ is estimated using the delta method[3].

3.1. What statistics (T) should we use for comparing location and scale?

The following summary statistics were considered for comparing location and scale between the NCS data and data from other published studies.

Location:

- Log-transformed mean ($\ln(\text{Mean})$)
- Estimated mean of the log-transformed concentrations ($\hat{\mu} = e^{\text{Mean}} - \frac{\hat{\sigma}^2}{2}$)

Scale:

- Log-transformed standard deviation ($\ln(\text{Std})$)
- Log-transformed coefficient of variation ($\ln(\text{CV}) = \ln\left(\frac{\text{Std}}{\text{Mean}}\right)$)
- Log-transformed standard deviation of the log transformed concentrations ($\ln(\hat{\sigma}) = \ln\left(\sqrt{\ln(\text{CV}^2 + 1)}\right)$)

For interpreting results as indicating location or scale differences, independent estimates of location and scale might be preferable. If the underlying concentrations are log-normally distributed, $\hat{\mu}$ and $\ln(\hat{\sigma})$are independent, if estimated from the log-transformed concentrations. However the available estimates of location and scale calculated from the reported mean and standard deviation are approximate and somewhat correlated. We felt that $\hat{\mu}$ and $\ln(\hat{\sigma})$ as measures of location and scale are difficult to understand. After discussion (see Acknowledgements), we decided to use $\ln(Mean)$ and $\ln(Std)$ as these were considered most understandable.

Two models were fit to the data, one to compare differences in location and one to compare differences in scale. Thus, there are two chances to decide the NCS measurements differ from the measurements from peer-reviewed publications. Based on simulated data, Figure 1 compares the power for detecting differences when using the following summary statistics for assessing location and scale:

1. $\hat{\mu}$ and $\ln(\hat{\sigma})$ estimated from the log-transformed concentrations (an ideal situation since the underlying data are not available) (in red);
2. $\hat{\mu}$ and $\ln(\hat{\sigma})$ estimated from the reported mean and standard deviation (in green); and
3. $\ln(\text{Mean})$ and $\ln(\text{Std})$ (in blue).

The power calculations assume two subpopulations are being compared, one generated with $\mu = 0$ and $\sigma = .5$ (shown on the plot as a dot and corresponding to a somewhat skewed distribution), and the other with various values of μ and σ. Each subpopulation has N = 16. The subpopulations are compared using a t-test with error variance calculated using the delta method. If either the location or scale differences are significant at the 5% level using a two-sided test, a difference is declared. Figure 1 shows a curve surrounding values of μ and σ for the second subpopulation for which the probability of detecting a difference is less than 80%.

This and other calculations suggest that using $\ln(Mean)$ and $\ln(Std)$ provide a reasonable approximation to what might be obtained using the preferred statistics ($\hat{\mu}$ and $\ln(\hat{\sigma})$) estimated from the log-transformed measurements if they were available). When

the other variance components are greater than zero, the choice of statistics to assess location and scale differences becomes relatively less important.

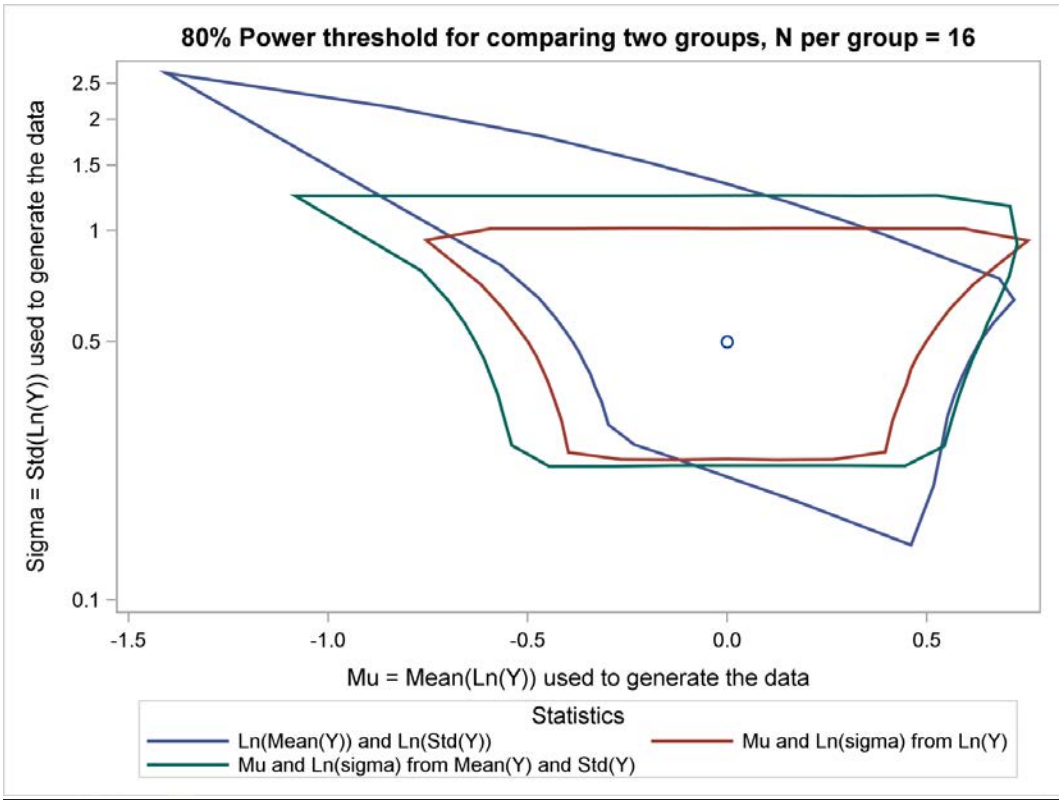

Figure 1: 80% Power threshold for comparing two groups each with 16 subjects.

3.2. Two options for estimating $\widehat{\mathbf{V}}$

We considered two options for estimating the error variance of the reported mean and log-transformed standard deviation ($\widehat{\mathbf{V}}$):

- Delta method
  - The variance formula uses an estimate of σ and assumes large sample sizes. The variance estimate may be biased for small sample sizes and because it is estimated from the reported mean and standard deviation of the untransformed concentrations.

- Empirical model based on simulated data
  - This estimate involves simulating lognormally distributed data using a range of assumed values of μ, σ, and sample size; calculating the variances of the mean and log-transformed standard deviation of the simulated data; deriving a model to predict the variances from the reported mean, standard deviation and sample size; and using the model to predict the error variance of the reported data. This option is somewhat arbitrary because it depends on the model fit to the data and on the range of parameter values that are used in the simulations.

Initial analysis suggested that empirical variance estimates may be less biased than using the delta method. However, the final analysis used the delta method because it was easier to document and because the two variance estimates were highly correlated.

4. Results

Figure 2 shows the data for each subpopulation. The first column shows the means; the second column shows the standard deviations (Sds). The rows correspond to measurements for HbA1c, glucose, and insulin. Within each panel, publications associated with each subpopulation are represented by numbers on the horizontal axis. The NCS value is the first, left most, dot and is shown by the horizontal grey line. The vertical lines represent 95% confidence intervals based on the estimated error variance. As is evident from the figure, there is considerable variation between publications and subpopulations in addition to the uncertainty in the summary statistics for each subpopulation.

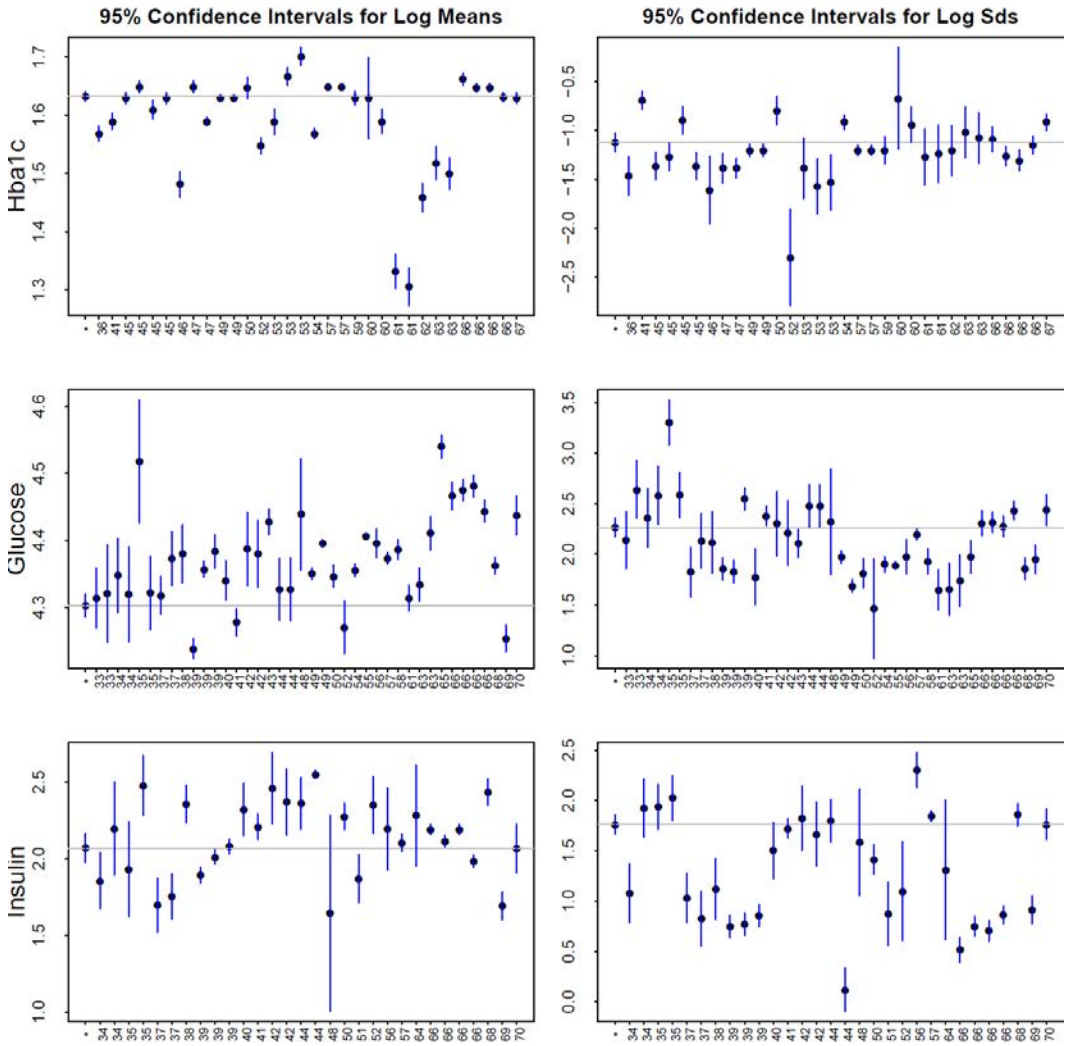

Figure 2: Reported means and standard deviations (with confidence intervals) for each sub-population.

For HbA1c, glucose, and insulin, Table 1 shows the sample size, mean, and standard deviation of the NCS Vanguard Study data, the pooled sample size, mean, and standard deviation across the publications, and p-values for assessing differences between the NCS mean and standard deviation and the reported means and standard deviations from the peer-reviewed publications. Based on the p-values, the mean and standard deviation differences between the NCS and other peer-reviewed publications for HbA1c, glucose, and insulin are not statistically significant.

Table 1: Results comparing the NCS data to data from other publications

Analyte	Vanguard Study		Publications				P-values	
	N	Mean (SD)	N	Sub-popula-tions	Women	Pooled mean (SD)	Mean diff	SD diff
HbA1c (%)	205	5.1 (0.3)	18	32	5,178	4.9 (0.3)	0.52	0.81
Glucose (mg/dL)	195	73.9 (9.6)	27	40	6,293	79.3 (9.1)	0.31	0.62
Insulin µIU/mL	193	7.9 (5.8)	20	27	3,050	8.6 (4.2)	0.73	0.43

Appendix A

The following outlines the derivation of the model used for analysis. If the measured concentrations for individual women were available, the NCS and publication data could be compared using a standard mixed model (Model 1).

Model 1: Ideal model assuming measured concentrations for individuals are available:
Y_{ijk} = concentration for person k in subpopulation j in publication i

$$\ln(Y_{ijk}) = \omega + I_{NCS}\alpha + \beta_i + \gamma_{j(i)} + \delta_{k(ij)}$$
$$\beta_i \sim N(0, \sigma_\beta^2)$$
$$\gamma_{j(i)} \sim N(0, \sigma_\gamma^2)$$
$$\delta_{k(ij)} \sim N(0, \sigma_\delta^2)$$

$\delta_{k(ij)}$ = random error for kth person within subpopulation j and publication i

This model can be modified to incorporate variance differences between publications and subpopulations within publications and thus assess location and scale differences as measured by the mean and variance of the log-transformed measurements.

However, individual data are not available from publications. If the publications provided the mean and standard deviation of the log-transformed concentrations (or equivalently the geometric mean and geometric standard deviation), the following variation on Model 1 can be fit (Model 2).

Model 2: Model predicting mean of log-transformed concentrations, if available:

$\text{Mean}_k\left(\ln\left(Y_{ijk}\right)\right) = \overline{\ln\left(Y_{ij}\right)}$ = Mean log-transformed in subpopulation j in publication i

$$\text{Mean}_k\left(\ln\left(Y_{ijk}\right)\right) = \overline{\ln\left(Y_{ij}\right)} = \omega + I_{NCS}\alpha + \beta_i + \gamma_{j(i)} + \varepsilon_{k(ij)}$$

$$\beta_i \sim N(0, \sigma_\beta^2)$$

$$\gamma_{j(i)} \sim N\left(0, \sigma_\gamma^2\right)$$

$$\varepsilon_{k(ij)} \sim N\left(0, \frac{\text{Var}_k\left(\ln\left(Y_{ijk}\right)\right)}{N_{ij}}\right)$$

This model is similar to Model 1 except the within-subpopulation variance of reported mean (measurement error) is approximated from the reported standard deviation and is assumed to be known rather than estimated.

Fitting model 2 has a possible complication: because the variance of the error variance is assumed to be known, the model cannot be fit if the between publication and between subpopulation variance component estimates are zero. In simulations, this was handled by using weighted regression with weights (assumed known) equal to the inverse of the estimated measurement error variance. This problem did not occur with the study data.

Since summary statistics for log-transformed concentrations are not available, Model 2 cannot be fit. However, the relationships in Model 2 can be approximated by predicting the log-transformed sample mean and estimating the error variance using the delta method (Model 3).

Model 3: Model predicting log-transformed mean concentrations

$\text{Mean}_k\left(Y_{ijk}\right) = \overline{Y_{ij}}$ = Mean concentration in subpopulation j in publication i

$$\ln\left(\overline{Y_{ij}}\right) = \omega + I_{NCS}\alpha + \beta_i + \gamma_{j(i)} + \varepsilon_{k(ij)}$$

$$\beta_i \sim N(0, \sigma_\beta^2)$$

$$\gamma_{j(i)} \sim N\left(0, \sigma_\gamma^2\right)$$

$$\varepsilon_{k(ij)} \sim N\left(0, \frac{\widehat{\sigma^2}}{N}\left(1 + 0.5\widehat{\sigma^2}\right)\right), \qquad \widehat{\sigma^2} = \ln(CV^2 + 1)$$

In this model we assume the log-transformed mean has a normal distribution and the error variance is known and can be calculated using the delta method from the reported mean and standard deviation.

Similar modifications can be used to compare other measures of location and scale using the following basic model.

T = measure of location and scale, assumed roughly normally distributed

$$T = \omega + I_{NCS}\alpha + \beta_i + \gamma_{j(i)} + \varepsilon_{k(ij)}$$

$$\beta_i \sim N(0, \sigma_\beta^2)$$

$$\gamma_{j(i)} \sim N\left(0, \sigma_\gamma^2\right)$$

$$\varepsilon_{k(ij)} \sim N\left(0, \widehat{V}\right)$$

$\hat{V}$ is the estimated within-subpopulation error variance of T which is assumed to be known when fitting the model.

Acknowledgements

The analysis was completed under a contract to NIH and reflects contributions from John Moye (NIH), William Meyer (Quest Diagnostics), and Berry Ryan (Emory University), Pat Chulada, Lori Merrill, and Luke Smith (Westat).

References

Landrigan PJ, Trasande L, Thorpe LE, et al. The National Children's Study: 1 21-year prospective study of 100,000 American children. Pediatrics. 2006;118(5):2173-2186.

Ozkaynak H, Whyatt RM, Needham LL, et al. Exposure assessment implications for the design and implementation of the National Children's Study. Environ Health Perspect. 2005;113(8):1108-1115.

Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2). Pacific Grove, CA: Duxbury.