VDSP Statistics Manual

A Guide to the Standardization of 25(OH)D Data Measured in the Past

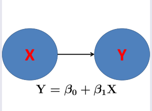
Ramón Durazo, PhD
Lu Tian, PhD

November 15, 2013

Overview

Introduction: Task

Task



Given 25(OH)D measurements obtained using two distinct methods, a "test method"(X) and a "reference method"(Y), estimate their linear association

Precisely

For a given 25(OH)D value X obtained with the "test method" estimate the corresponding "reference method" value, Y

Introduction

Notation

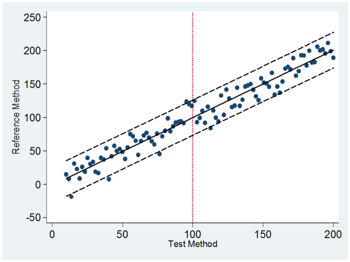
- For the **VDSP**, **Reference Method** stands for Ghent University and NIST reference method procedures.
- **Test Method** relates to the measurement procedure used to obtain 25(OH)D values in "past" studies.
- The units of measurement through this presentation are **nmol/L**

The Measurement Model

Linear Regression

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$



General Error in Regression

$$\begin{aligned} Y &= \tilde{Y} + \epsilon_y \\ X &= \tilde{X} + \epsilon_x \\ \tilde{Y} &= \beta_0 + \beta_1 \tilde{X} + \epsilon \end{aligned}$$

$$\begin{aligned} \epsilon_y &\sim N(0, \sigma_y^2) \\ \epsilon_x &\sim N(0, \sigma_x^2) \\ \epsilon &\sim N(0, \sigma^2) \end{aligned}$$

An estimate of:

$$\beta_0 + \beta_1 X$$

is given by

$$a + bX,$$

where **a,** and **b** estimate the intercept and slope, respectively.

Estimation

Naıve Case: 95% CI for $\beta_0 + \beta_1 X$

$$(a+bX)\pm 1.96\times\sigma_{y\cdot x}\sqrt{\frac{1}{N}+\frac{(X-\overline{X})^2}{(N-1)S_x^2}}$$

- $\overline{X}$ mean of test values
- S_x^2 variance of the test values.
- The error of the reference method for a given X value is:

$$\sigma_{y\cdot x}$$

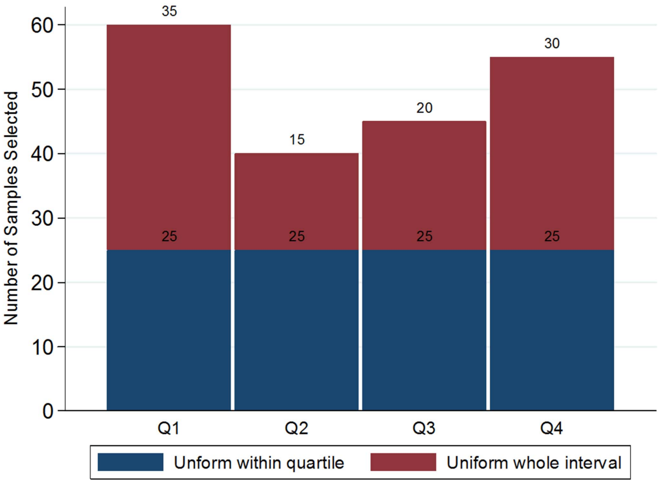
Precision (width) depends on:

- The distribution of the test method values
  - Value of X and its distance from the mean $(\overline{X})$
  - Standard deviation of X values (S_x)
- The error of the reference method
- Sample size, **N**

Sample Selection Algorithm

- Random sampling
- Uniform random sampling
- Uniform random sampling in each of the 4 quartiles of the distribution

Sample Selection
Uniform vs Uniform quartile selection



Sample Selection Algorithm

Example: Irish National Adult Nutrition Survey

- Estimate the required sample size, say **N=100**
- Select **k** samples in each of the 4 quartiles, **k = N/4**
- Estimate quartiles of distribution: **42, 57, 74** (nmol/L)
- For each interval, **[A,B],** compute $D = (B - A)$ and divide by **k** (the number of samples to be selected from each of the intervals).
  - $[20\ nmol/L, 42\ nmol/L)$, D $= 22$, $\frac{D}{25} = 0.88$
  - $[42\ nmol/L, 57\ nmol/L)$, D $= 15$, $\frac{D}{25} = 0.60$
  - $[57\ nmol/L, 74\ nmol/L)$, D $= 17$, $\frac{D}{25} = 0.68$
  - $[74\ nmol/L, 200\ nmol/L)$,D $= 126$, $\frac{D}{25} = 5.04$

Sample Selection Algorithm
Example: Irish National Adult Nutrition Survey

- For each interval, **[A,B],** sort the samples from smallest to largest
- In each of the intervals, select the **k** values by sequentially adding **D/k** to **A.** For example for the the interval from **20 nmol/L** to **42 nmol/L**, **D/k = 0.88**. The **k=25** values are:

$$20, 20.88, 21.76, \ldots, 40.24, 41.12, \; nmol/L$$

- When the target 25(OH)D value does not exist in your sample, select the nearest value
- Record the lot number for the selected samples. Manual adjustment may be necessary if the selected lots appear non-representative

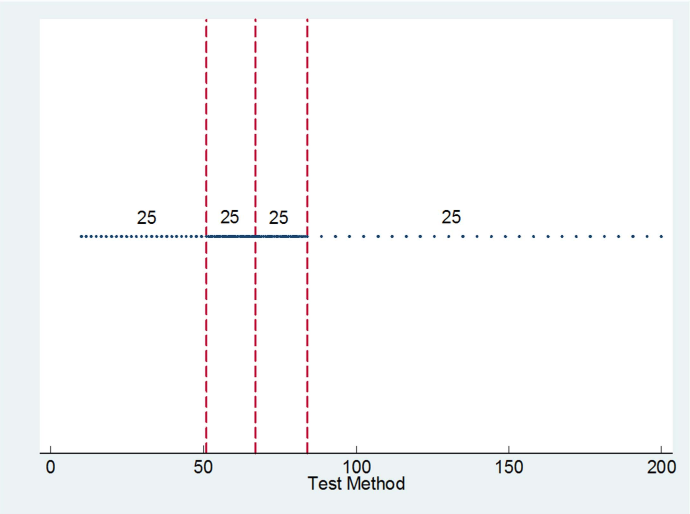
Sample Selection

Statistics Coordinating Center

- Prepare the de-identified data including (and only including) the following information
  - Individual ID
  - 25-Hydroxyvitamin D measurement value
  - Lot and date of the measurement
- The study members will keep the link between the individual **ID** sent out and the study sample **ID,** which can be used to identify the actual samples. That information will not be sent to the VDSP statistics coordinating center
- Send the data to the VDSP coordinating center
- Receive individual ID representing selected samples and link the individual ID with actual sample

Sample Selection

Example



Example: Assumptions

- X, the test method value is measured without error
- The error of the reference method, Y, is independent of X
- Sample will be selected uniformly from $[20\ nmol/L, 200\ nmol/L]$
- The CV% for the reference method at the mean value $(\overline{X} = 110,\ nmol/L)$, is known (5%)

Example: Assumptions

- Mean X value $= \overline{X} = 110\ nmol/L$
- Standard deviation of X values, $S_x = \frac{180}{\sqrt{12}} = 51.96$
- The CV% for the reference method at the mean value ($\overline{X} = 110\ nmol/L$), is known ($5\%$)
- Error of reference method: $\sigma_{y \cdot x} = 0.05 \times \overline{X} = 0.05 \times 110 = 5.5\ nmol/L$

Sample Size (N) Estimation

Test method error-free

Sample Size (N) Estimation: 95% Prediction CI

Assumptions:

- Uniform sampling
- Width = 3 nmol/L
- Min = 20 nmol/L, Max = 200 nmol/L
- $\sigma_{y \cdot x} = 5.5\ nmol/L$
- $CV_y = 5\%$ at 110 nmol/L

25(OH)D Value (nmol/L)	Sample Size (N)
30	175
50	121
75	76
100	54

Statistical Analysis

Test method non-error-free

General Error in Regression

$$
\begin{aligned}
Y &= \tilde{Y} + \epsilon_y \\
X &= \tilde{X} + \epsilon_x \\
\tilde{Y} &= \beta_0 + \beta_1 \tilde{X} + \epsilon
\end{aligned}
$$

$$
\begin{aligned}
\sigma_y &= CV_y \times \tilde{Y} \\
\sigma_x &= CV_x \times \tilde{X}
\end{aligned}
$$

Estimation Methods

- Deming Regression(Homogeneous Measurement Error)
- Weighted Deming Regression (Heterogeneous Measurement Error)
- Other Methods (Heterogeneous Measurement Error) Lu T, Durazo-Arvizu RA, Myers G, Brooks S, Sarafin K, Sempos TS. *The Estimation of Calibration Equations for Variables with Heteroscedastic Measurement Error.* Under review (Statistics in Medicine)

Sample Size (N) Estimation: 95% Prediction CI

Assumptions:

- Width = 3 nmol/L
- Min = 20 nmol/L, Max = 200 nmol/L
- Quartiles: 41.6, 57.2, 74.1 nmol/L
- $CV_x = 6.6\%$
- $CV_y = 5\%$

25(OH)D Value (nmol/L)	Uniform (N)	Quartile (N)
30	231	**97***
50	131	39
75	86	104
100	130	323

* Irish National Adult Nutrition Survey

Sample Size (N) Estimation

Test method non-error-free

Sample Size (N) Estimation: 95% Prediction CI

Assumptions:

- Maximum value = 125 nmol/L
- Width = 5 nmol/L
- Uniform Sampling
- $CV_y = 5\%$
- 25(OH)D = 30 nmol/L

Minimum Value	**Coefficient of Variation Test Method**			
	5%	10%	12%	15%
10	13	29	**37**	51
15	17	40	**52**	72
20	24	56	**73**	102

Sample Size (N) Estimation

Test method non-error-free

Sample Size (N) Estimation: 95% Prediction CI

Assumptions:

- Maximum value = 150 nmol/L
- Width = 5 nmol/L
- Uniform Sampling
- $CV_y = 5\%$
- 25(OH)D = 30 nmol/L

Minimum Value	Coefficient of Variation Test Method			
	5%	10%	12%	15%
10	20	46	**60**	81
15	26	61	**79**	108
20	34	81	**105**	145

Sample Size (N) Estimation: 95% Prediction CI

Assumptions:

- Maximum value = 175 nmol/L
- Width = 5 nmol/L
- Uniform Sampling
- $CV_y = 5\%$
- 25(OH)D = 30 nmol/L

Minimum Value	Coefficient of Variation Test Method			
	5%	10%	12%	15%
10	30	68	**88**	119
15	37	87	**112**	152
20	47	110	**142**	195

Sample Size Estimation: 95% Prediction CI

Assumptions:

- Maximum value = 200 nmol/L
- Width = 5 nmol/L
- Uniform Sampling
- $CV_y = 5\%$
- 25(OH)D = 30 nmol/L

Minimum Value	**Coefficient of Variation Test Method**			
	5%	10%	12%	15%
10	41	95	**122**	164
15	50	117	**150**	204
20	62	144	**185**	253

Acknowledgments

- Dr. Chris Sempos, NIH–ODS (VDSP Director)
- Dr. Gary Myers G, AACC
- Dr. Stephen Brooks, Health Canada
- Dr. Kurtis Sarafin, Health Canada
- Dr. Liping Tong, Loyola University Chicago

Thank You!

Questions?