SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 874

October 1989

REPORTING THE ACCURACY OF
BIOCHEMICAL MEASUREMENTS FOR
EPIDEMIOLOGIC AND NUTRITION STUDIES

Lisa M. McShane¹
Larry C. Clark²
Gerald F. Combs, Jr.³
Bruce W. Turnbull⁴

¹Present address: National Institutes of Health, Bethesda, Maryland 20892

²Arizona Cancer Center, University of Arizona, Tucson, Arizona 85724

³Division of Nutritional Sciences, Cornell University, Ithaca, New York 14853

⁴School of Operations Research and Industrial Engineering, Cornell University,
Ithaca, New York 14853

Correspondence and reprint requests should be addressed to:
Dr. G. F. Combs, Jr., Department of Poultry and Avian Sciences and Division of
Nutritional Sciences, Cornell University, Ithaca, New York 14853

Supported in part by Grants R01 GM 28364 and R01 CA49764 for the U.S. National
Institutes of Health and a fellowship from the U.S. Army Research Office through the
Mathematical Sciences Institute at Cornell University.

Reporting the Accuracy of Biochemical
Measurements for Epidemiologic and Nutrition Studies

Lisa M. McShane¹

Larry C. Clark²

Gerald F. Combs, Jr.³

Bruce W. Turnbull⁴

¹Present address: National Institutes of Health, Bethesda, Maryland 20892

²Arizona Cancer Center, University of Arizona, Tucson, Arizona 85724

³Division of Nutritional Sciences, Cornell University, Ithaca, New York 14853

⁴School of Operations Research and Industrial Engineering, Cornell University,
Ithaca, New York 14853

Correspondence and reprint requests should be addressed to:

Dr. G. F. Combs, Jr., Department of Poultry and Avian Sciences and Division of Nutritional
Sciences, Cornell University, Ithaca, New York 14853

Supported in part by Grants R01 GM 28364 and R01 CA49764 for the U.S. National Institutes of
Health and a fellowship from the U.S. Army Research Office through the Mathematical Sciences
Institute at Cornell University.

Running head: REPORTING LABORATORY ACCURACY

ABSTRACT

Procedures for reporting and monitoring the accuracy of biochemical measurements are presented. They are proposed as standard procedures for laboratory assays for epidemiologic and clinical nutrition studies. It is also proposed that explanations of any quality control procedures be reported along with accuracy reports in such studies. These should identify all major sources of variability and model the total variability using variance component techniques. It is shown how the total variance model can be used to compute a maximum percent error that provides an easily understandable measure of laboratory accuracy and accounts for all sources of variability. This avoids ambiguities that may arise when reporting a standard deviation that may take into account only a few of the potential sources of variability. This model can also be used to estimate the precision of laboratory methods or for planning replication schemes in which estimates produced for individuals have a specified level of precision. It also allows for the development of effective quality control plans in which analytical results from control samples inserted into batches of experimental samples are compared against “control limits” that are calculated from the variance model and estimated variance components. An example of the use of these procedures is presented with the case of the analysis of α -tocopherol in human plasma using high-performance liquid chromatography.

KEY WORDS: Laboratory accuracy, quality control, variance components,
assay maximum percent error, high-performance liquid chromatography,
vitamin E

Introduction

Biochemical measurements of plasma or serum are often used to compare the nutrient status of different groups, to rank individuals within a group, or to assess an individual's status (1).

Clearly, the accuracy of these measurements will directly influence the validity of the conclusions drawn from the clinical, environmental, and epidemiologic studies that employ them. It may also affect the accuracy of medical diagnoses and, thus, the appropriateness of medical treatments of individual patients.

The intent of this paper is to propose standards for reporting both the accuracy of biochemical measurements and quality control procedures, and to advocate such reporting for all measurements used in biomedical investigations. Traditionally, many laboratories use both internal and external standards to assess and monitor the quality of their analyses. Methods include making periodic "checks" by inserting a series of controls, standards, or blind replicate samples into batches of experimental samples being analyzed, or by reanalyzing a certain fraction of experimental samples. The methods proposed in this paper should help researchers to estimate and report the accuracy of their laboratory assays. These methods provide a test for proper system functioning and yield detailed information regarding the precision with which measurements for individuals are determined.

In any analytical procedure, the researcher must identify the major sources of variability and take them into account when reporting the accuracy of assays. Sources of variability are any factors that may cause replicate readings to differ; each source contributes a component of variability to the total variability in the overall process. Many of the laboratory procedures used to measure nutrients are very complex, i.e., they are composed of multiple stages. A particular stage of such a complex process is an important source of variability if replicate readings at that stage show substantial differences. In addition to such internal sources of variability, other sources of variability may be external to the procedure, for example, day-to-day, batch-to-batch, aliquot-to-aliquot effects. Any of these sources of variability may, in turn, be comprised of more than one minor source of variability.

In this paper, a variance components model is proposed in which the total variability in the measurement process is expressed in terms of the variance components contributed by the various stages in the process (i.e., internal sources) and other significant external factors. Variance components analysis is a standard statistical technique; a thorough expository review was given by Searle (2). The present purpose is to show how these techniques can be used to develop easily interpretable measures of the components of variability of an assay and its accuracy.

From the variance components model, it is possible to compute the number of replicates required on each variance source (including stages of the measurement procedure) to produce readings with a desired accuracy. Conversely, for a given replication scheme, reported laboratory data can be accompanied by estimated maximum percent errors. The present model will focus only on the variability of the readings, and not on experimental bias; it is assumed that calibrations using standards ensure that measurements are reasonably unbiased. The variance components method of assessing measurement accuracy will be explained by use of an example of a high-performance liquid chromatography procedure for measuring plasma vitamin E.

Methods

Laboratory Procedures and Data Collection

Alpha-tocopherol was measured in human plasma using the technique of high-performance liquid-liquid partition chromatography (HPLC) that employed a reverse-phase system with a 3.9 mm × 300 mm C-18 column as the stationary phase and methanol:deionized water (96:4 v/v) as the mobile phase (2 ml/min). Alpha-tocopherol was detected by molecular fluorescence (excitation: 290 nm; emission: 330 nm) using a flow-through spectrofluorometer (Perkin-Elmer 650-10S, Norwalk, CT); quantitation was effected using a recording integrator (HP 3390A, Hewlett-Packard Instruments, Avondale, PA) by comparison to a series of authentic standards of all-rac- α -tocopherol. Samples were applied to the HPLC system automatically using a programmable sample injector (WISP 710B, Waters Associates, New Milford, MA). Each sample was placed in a sealed vial (with

puncturable cap) in a carousel tray holding a total of 36 experimental samples and 4 standard samples. The samples were injected in triplicate (20 μ l each) to the HPLC system. Each carousel tray of samples constituted an analytical “batch”.

Blood was obtained by venipuncture from each of two volunteers into tubes previously treated with ethylenediaminetetraacetic acid (EDTA). Plasma was prepared by centrifugation (1000 $\times$ g) and pooled by subject. For each pool, multiple samples (1 ml each) of plasma were frozen (-70° C) in sealed vials (1.5 ml capacity; Sarstedt, Princeton, NJ) until they were to be analyzed. On the day of analysis, five aliquots of each volunteer’s plasma were thawed in a room temperature water bath. The following procedures were followed under low-intensity red light to extract vitamin E from those samples. Duplicate sub-samples (i.e., 2 $\times$ 200 μ l) of each aliquot were each treated with an equal volume of absolute ethanol; the resulting mixture was then extracted by vortex-mixing (1 min) with an equal volume of hexane. The phases were separated by low-speed centrifugation (500 $\times$ g), and 200 μ l of the organic phase was transferred to a clean glass vial in which it was first evaporated to dryness at room temperature under a stream of nitrogen gas, and then redissolved in absolute ethanol. In this way, two “extractions” were obtained from each aliquot. The ethanolic solution was applied to the HPLC system as described above. On each of five days, α -tocopherol was determined in duplicate extracts of each of five aliquots of each of the two pools; these samples were loaded immediately after the analytical standards in adjacent positions in the carousel for the automatic injector.

This replication scheme was designed in view of our experience which indicated the major sources of variability in the process to be:

1. Batch (B)
2. Aliquot (A)
3. Extraction (X)
4. Random error (including injection) (E)

A batch can be thought of as some natural grouping of analyses. In this example, the measurement

procedure was partially automated, and the analyses were grouped in autosampler carousel trays. Each tray of approximately 20 extractions, plus a control sample and the 4 analytical standards, comprised a batch. For simplicity, only one such tray was run per day in this experiment; therefore, “batch” actually includes many factors such as daily climatic conditions (e.g., temperature, humidity, etc.), equipment effects (e.g., column degradation), and the effects of different sets of standards. For laboratories that run more than one batch per day, a “day” effect can be separated out from the batch effect. The sources of variation “aliquot” and “extraction” correspond to those stages in the measurement process. The last source of variation, random error, includes variability due to injections and any other variability not accounted for by the other identified sources of variability.

The replication scheme (five days with five aliquots per pool per day, two extractions per aliquot, and three injections per extraction) for the preliminary run was designed according to which sources of variability were thought to be the largest contributors, based on previous experience. The preliminary set of quality control samples need not necessarily be run separately from routine laboratory analyses; in fact, it may be preferable for the analyst to remain blinded as to the identity of the quality control samples. The preliminary data could be collected from control samples inserted into several different days’ runs; however, if the laboratory procedure is newly developed, it would be wise to perform analyses on a set of preliminary quality control samples prior to using the procedure on “real” samples. Sources of variability may be identified which would allow further refinement of the methods.

Statistics

Variance Components Model

A variance components model is proposed for the variability in this type of laboratory data. Each source of variability identified in the procedure (i.e., batch, aliquot, extraction, and random sources for the HPLC/vitamin E procedure) is assumed to contribute a random amount of error to

the overall error in an individual reading. These random contributions are assumed to be normally distributed with mean zero and variance depending upon the source of variability.

For this HPLC/vitamin E procedure, $Y_i(j, k, m, n)$ represents the α -tocopherol concentration obtained from the n th injection of the m th extraction of the k th aliquot of pool i present in the j th batch. That is, pools are crossed with batches, aliquots are nested within pool $\times$ batch combinations, extractions are nested within aliquots, and injections are nested within extractions.

We recommend that this analysis be based on the logarithms of the actual data, as our examination of several data sets from different types of laboratory procedures has indicated that, frequently, the variances in analytical results are dependent on the mean values of the samples measured. For these vitamin E data and other types of laboratory measurements that we have studied, evidence suggests that the standard deviations of replicate measurements on pools are proportional to the means for these pools. Therefore, the “logarithm” is the appropriate “variance stabilizing transformation”.

For the example in which from each of I individuals (i.e., pools), K aliquots of plasma are placed in each of J batches, M extractions are taken from each aliquot, and N injections are taken from each extraction. We define $W_i(j, k, m, n) = \log(Y_i(j, k, m, n))$ to be a sum of independent terms:

$$W_i(j, k, m, n) = \theta_i + B(j) + A_i(j, k) + X_i(j, k, m) + E_i(j, k, m, n) \quad (1)$$

$$i = 1, 2, \dots, I; \quad j = 1, 2, \dots, J;$$

$$k = 1, 2, \dots, K; \quad m = 1, 2, \dots, M;$$

$$n = 1, 2, \dots, N$$

where

θ_i = true log-vitamin E value for pool i ; $B(j)$ is a normally distributed random variable with mean zero and variance σ_B^2 , representing the effect of batch j ; $A_i(j, k)$ is a normally distributed random variable with mean zero and variance σ_A^2 , representing the effect of the k th aliquot of pool i in batch j ; $X_i(j, k, m)$ is a normally distributed random variable with mean zero and variance σ_X^2 ,

representing the effect of the m th extraction of the k th aliquot of pool i in batch j ; $E_i(j, k, m, n)$ is a normally distributed random variable with mean zero and variance σ_E^2 , representing the random error (including injection effect) associated with the n th injection of the m th extraction of the k th aliquot of pool i in batch j ;

For several laboratory procedures that we have investigated, variance components models similar to (1) shown above have been reasonable, and we feel confident that similar variance components models are appropriate for many laboratory procedures. However, before applying the following methods to new laboratory procedures, the assumptions of this model require verification. The assumption of constant variance can be checked by plotting actual data against pool means. Normality assumptions can be checked by constructing normal probability plots of the estimated effects, $\hat{B}$, $\hat{A}$, $\hat{X}$, and $\hat{E}$ (see (3), p. 177), or by using the Kolmogorov-Smirnov test or similar quantitative goodness-of-fit tests (see (4), p. 293).

Modelling Precision of Summary Readings

Often, one of the major goals in modelling the variability in the measurement process is to determine how precisely the summary measurements for each pool estimate the true pool values. The summary measurement for an individual, e.g., pool i , is defined as the arithmetic mean of all readings for that individual. That is,

$$\bar{Y}_i = \frac{\sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \sum_{n=1}^N Y_i(j, k, m, n)}{JKMN}.$$

The variance of $\bar{Y}_i$, which we denote as $\text{Var}(\bar{Y}_i)$, supplies a measure of how precisely the summary measurement estimates the pool value. It depends on the variance in the procedure as well as the replication scheme (the values of J , K , M , N).

From model (1) we can derive an expression for $\text{Var}(\bar{W}_i)$ where $\bar{W}_i$ is the summary reading

for the log-readings for pool i . This expression is given by

$$\text{Var}(\bar{W}_i) = \frac{\sigma_B^2}{J} + \frac{\sigma_A^2}{JK} + \frac{\sigma_X^2}{JKM} + \frac{\sigma_E^2}{JKMN}. \quad (2)$$

However, interest is really in $\text{Var}(\bar{Y}_i)$, the variance of the arithmetic mean of all untransformed readings made on pool i . Using calculus techniques (the delta method) an approximate relationship can be derived between the variance of $\bar{Y}_i$ and the variance of $\bar{W}_i$ (see (5), p. 321). The relationship is given by

$$\text{Var}(\bar{Y}_i) = \mu_i^2 \text{Var}(\bar{W}_i) \quad (3)$$

where $\mu_i \approx \exp(\theta_i)$ is the true mean reading for pool i . Then it follows from Equations (2) and (3) that

$$\text{Var}(\bar{Y}_i) = \mu_i^2 \left[\frac{\sigma_B^2}{J} + \frac{\sigma_A^2}{JK} + \frac{\sigma_X^2}{JKM} + \frac{\sigma_E^2}{JKMN} \right]. \quad (4)$$

An advantage in modelling $\text{Var}(\bar{Y}_i)$ in terms of the variability in the log-readings is that the variance components for the log-readings are independent of the pool mean. Therefore, the variance components for the process can be estimated from preliminary data obtained from pools which may not have the same mean as pools we may later be interested in analyzing. Expression (4) states that the ratio of the variance of the summary reading for any pool to the squared mean of the pool is constant, and that constant can be estimated from preliminary data.

If it is assumed that $\bar{Y}_i$ is approximately normally distributed and $\text{Var}(\bar{Y}_i)$ is given by (4), we can make the following statement:

$$\Pr \left[\mu_i \in \bar{Y}_i \pm z_{\alpha/2} \mu_i \left[\frac{\sigma_B^2}{J} + \frac{\sigma_A^2}{JK} + \frac{\sigma_X^2}{JKM} + \frac{\sigma_E^2}{JKMN} \right]^{1/2} \right] \approx 1 - \alpha \quad (5)$$

where $z_{\alpha/2}$ is the value such that if Z is a standard normal random variable, then

$\Pr[Z > z_{\alpha/2}] = \alpha/2$. Typically we choose $\alpha = 5\%$ and thus $z_{\alpha/2} = 1.96 \approx 2$. While $\bar{Y}_i$ might not be expected to have an exact normal distribution if the model (1) were exactly true, the Central Limit Theorem (see Feller, (6)) implies that the normal distribution would be a reasonable approximation because $\bar{Y}_i$ is a mean of several readings.

The approximations (1) and (5) are useful in that they provide a convenient method to make easily interpretable statements about the variability in the measurement process. For the vitamin E measurement process, one can say that if plasma samples from a pool are run in J batches, with K aliquots per pool per batch, M extractions per aliquot, and N injections per extraction, then the summary reading for that pool will be within approximately 100p% of the true pool value with probability $1 - \alpha$, where p is given by

$$p = z_{\alpha/2} \left[\frac{\sigma_B^2}{J} + \frac{\sigma_A^2}{JK} + \frac{\sigma_X^2}{JKM} + \frac{\sigma_E^2}{JKMN} \right]^{\frac{1}{2}}. \quad (6)$$

That is, given values for the variance components σ_B^2 , σ_A^2 , σ_X^2 , and σ_E^2 , Equation (6) allows one to quantify the precision of summary readings obtained under specific replication schemes by supplying a “maximum percent error”. The maximum percent error can be used to describe the accuracy of readings already made or it can be used to plan replication schemes which achieve desired percent errors.

Results

Estimation of Variance Components

Because the values of the variance components are not known, they must be estimated from the preliminary quality control data set. The estimates were obtained using the PROC VARCOMP procedure in the SAS statistical package (7). The method used was the restricted maximum-likelihood method (REML). For this particular data, the replication scheme was balanced, so we also computed the variance components by the method of moments (type I method), and the

answers were identical to those obtained using REML (7)¹. The estimates of the variance components (with associated standard errors given in parentheses) were computed to be

$$\begin{aligned}\hat{\sigma}_B^2 &= 0.014538 && \text{(batch variance component)} \\ (\text{s.e.}(\hat{\sigma}_B^2)) &= 0.001330 \\ \hat{\sigma}_A^2 &= 0.004558 && \text{(aliquot variance component)} \\ (\text{s.e.}(\hat{\sigma}_A^2)) &= 0.000412 \\ \hat{\sigma}_X^2 &= 0.016731 && \text{(extraction variance component)} \\ (\text{s.e.}(\hat{\sigma}_X^2)) &= 0.001530 \\ \hat{\sigma}_E^2 &= 0.000463 && \text{(error variance component)} \\ (\text{s.e.}(\hat{\sigma}_E^2)) &= 0.000000.\end{aligned}$$

Estimates are rounded to the sixth decimal place.

From these estimates, it can be seen that the components contributing the most to the total variability in the process are “batch” (equivalent to “day” in this experiment) and “extraction”. The relative sizes of the variance components may be interesting in themselves in that they may point to “problem areas” in the measurement process.

Another interesting aspect of this preliminary quality control data set was that it revealed another potential problem with our particular laboratory system. The above estimates are actually based on only four days of data because on the second day of the five day run, the HPLC column failed and was replaced. Most of the data for that day was lost, and what remained was considered to be highly unreliable; therefore, it was decided to omit that day from the analysis. The column failure served as an incentive to plan regular checks during future analyses to detect impending column failure.

¹The SAS program is given in the Appendix.

Calculating Maximum Percent Errors

Having estimated the variance components, examples are now presented showing how Equation (6) can be used. Substituting in the estimated values of the variance components, the expression for p in Equation (6) becomes

$$p = z_{\alpha/2} \left[\frac{0.01454}{J} + \frac{0.00456}{JK} + \frac{0.01673}{JKM} + \frac{0.00046}{JKMN} \right]^{\frac{1}{2}}. \quad (7)$$

Suppose that it is desired to find a replication scheme such that with 95% probability, then the summary measurement for a pool will be within 20% of the true value. In this example, it was possible to obtain sample volumes sufficient for only two extractions per aliquot; therefore, attention was restricted to schemes in which $M = 2$.

First, it is noted that since the variance contribution due to injections is very small; further, as it is divided by a large number ($JKMN$), little is gained by performing multiple injections. Therefore, it was decided to make only two injections ($N = 2$) from each extraction. Taking $z_{\alpha/2} = 2$ ($\alpha \approx 0.05$), $M = 2$, and $N = 2$ in (7), several combinations of J and K were tried to find a replication scheme which meets the 20% criterion and requires as few extractions as possible. To obtain more precision, more replication would be required. Table 1 summarizes the findings and is useful for designing the analysis scheme.

[Insert Table 1 about here.]

Because the values used for the variance components were only estimated, the sensitivity of the maximum percent errors ($100p$) to perturbations of the values for the variance components was investigated. It was found that if every variance component estimate was replaced by itself plus or minus twice its standard error, then the change in the calculated percent error was approximately 10 percent of the original percent error. For example, for the replication scheme

$(J,K,M,N) = (1,1,2,2)$, when every variance component estimate was replaced by itself minus twice its standard error, the calculated percent error changed from 33.2% to 30.0%. When each estimate was replaced by itself plus twice its standard error, the calculated percent error was 36.1%. For the replication scheme $(J,K,M,N) = (3,2,2,2)$ the calculated percent errors changed to 15.2% and 18.2% when reducing and increasing the estimates by twice their standard errors.

From Table 1, it can be seen that the replication scheme requiring the fewest number of extractions (JKM) and still meeting the 20% criterion is given by $(J,K,M,N) = (3,1,2,2)$. That is, samples from each pool should be run in 3 batches with 1 aliquot per pool per day, 2 extractions per aliquot, and 2 injections per extraction. That plan requires 6 extractions per pool. If a percent error of 23.5% could be tolerated, the plan $(2,1,2,2)$ could be used, and it only requires 4 extractions per pool. Thus, it can be seen that distributing the total number of extractions for a given pool over several different batches gives a lower maximum percent error than including all extractions for a pool in the same batch.

The total variability in this example is higher than usually reported; however, it should be noted that this includes *all* variability. It would have been much lower had we reported only the injection or extraction variability, as often occurs in published reports. If an estimated standard deviation (SD) is reported based on readings from several extractions, each averaged over 2 injections, from the same aliquot, run on the same day, that estimate would be estimating the quantity $\sqrt{\sigma_X^2 + \frac{1}{2}\sigma_E^2}$. From these data, we would estimate that quantity to be approximately 0.13024. A more appropriate “standard deviation” based on our *total* variability model would be $\sqrt{\sigma_B^2 + \sigma_A^2 + \sigma_X^2 + \frac{1}{2}\sigma_E^2}$, which we would estimate from our data to be approximately 0.18989. The first standard deviation, which measures only extraction and injection variability, is one-third lower than the standard deviation based on the total variability model.

It is clear from this example that the term “standard deviation” has little meaning unless all of the major sources of variability are identified, and it is made clear exactly what component or components of variability are being estimated. This ambiguity can be avoided by reporting all

major sources of variability and using result (6) to provide an easily understandable measure of laboratory accuracy, namely the maximum percent error.

Planning Quality Control Checking Schemes

Another benefit of this variance component analysis is that it can aid in the development of quality control checking schemes. As an example, the following shows how to develop a quality control checking scheme for the HPLC/vitamin E procedure.

The estimates of the variance components indicated that there was particularly large batch-to-batch variability. In an effort to identify “unacceptable” batches, control plasma samples were placed in each batch. These controls came from large pools of human plasma for which the true α -tocopherol concentration was well-estimated. It was planned that, if the control measurement for a particular batch was aberrant (i.e., far from the known pool value), then the samples in that batch would be reanalyzed. Due to “normal” variability it was known that some of the control results will stray reasonably far from the target value, and the laboratory analysts wanted some assurance that not too many batches would be falsely identified as aberrant, requiring many unnecessary reanalyses.

A checking scheme can be developed as follows. Numbers p and α ($0 < p < 1$, $0 < \alpha < 1$) can be specified and the number of aliquots, extractions, and injections needed from the control pool in each batch can be determined so that, if a batch is identified as aberrant when the control mean for that batch differs from the known value by at least $100p\%$, the probability of that batch being falsely identified as aberrant (“false alarm”) is approximately α . The batch rejection criterion is based on “control limits” computed from the estimated variance components for the process. The control limits provide a measure of the acceptable amount of variability for the control readings. These control limits will be widest for procedures with greater “normal” (acceptable) variability (corresponding to large estimated variance components).

The variance of the mean of the controls for a given batch can be computed from Equation (4) where all readings are made from a single batch, and the purpose is to test whether the batch

effect is excessive. If there truly is no batch effect, then $\sigma_B^2 = 0$. Substituting $\sigma_B^2 = 0$ into Equation (4) produces Equation (8) which is given by

$$\text{Var}(\bar{Y}_c) \approx \mu_c^2 \left[\frac{\sigma_A^2}{K} + \frac{\sigma_X^2}{KM} + \frac{\sigma_E^2}{KMN} \right]. \quad (8)$$

If a batch is declared to be aberrant when

$$|\bar{Y}_c - \mu_c| \geq p \mu_c,$$

then the probability of falsely declaring a batch as aberrant (“false alarm”) is

$$\alpha = 2 \Phi(z), \quad \text{where}$$

$$z = -p \left[\frac{\sigma_A^2}{K} + \frac{\sigma_X^2}{KM} + \frac{\sigma_E^2}{KMN} \right]^{-\frac{1}{2}}$$

and $\Phi(z)$ is the standard normal cumulative distribution function. Substituting in the estimates for the variance components, yields

$$z = -p \left[\frac{0.00456}{K} + \frac{0.01673}{KM} + \frac{0.00046}{KMN} \right]^{-\frac{1}{2}}.$$

The use of Equation (8) is rather conservative, as it is probably reasonable to allow for some small batch effect, i.e., let $\sigma_B^2 = \delta$ (δ small) instead of taking $\sigma_B^2 = 0$ in Equation (4). By setting $\sigma_B^2 = 0$, $\text{Var}(\bar{Y}_c)$ will be underestimated and control limits that are somewhat too narrow will be produced, resulting in a relatively high rate of false batch rejections. For purposes of this example, the more conservative approach is taken, i.e., using $\sigma_B^2 = 0$. Table 2 gives false alarm probabilities for a variety of checking plans.

[Insert Table 2 about here.]

This information can be used to select a reasonable quality control checking scheme. If it is desired to have the “false alarm” probability less than 5% or 6%, and if it is planned to use only one aliquot (extracted in duplicate) of the control per batch, then a batch would be identified as aberrant if the mean of the two control extractions were to differ from the known pool value by more than 25%; however, if two aliquots with two extractions per aliquot on each batch were to be used, then the rejection criterion could be tightened to 15% - 20%. If $\sigma_B^2 = \delta$ were to be used instead of $\sigma_B^2 = 0$ when computing the “false alarm” probabilities in Table 2, then all such false alarm probabilities would be greater; thus, to maintain a false alarm rate of 5% - 6%, a larger value of p would be selected, i.e., wider control limits would be used.

Discussion

The primary goal of this paper was to develop an informative and uniform method for presenting the inherent variability of laboratory analytical results such as those used in epidemiologic and clinical nutritional investigations. The issue of quality control is of paramount importance to investigators; however, the implementation, reporting, and interpretation of quality control procedures have not been standardized. We have proposed a relatively straightforward statistical procedure for investigating and reporting the variability in biochemical assays. This approach is based on variance component analysis and a summary measure of the variability of biochemical assays, the maximum percent error.

In addition to producing a summary measure of variability, the variance components procedure facilitates the investigation of the sources of analytical variability for the improvement of the analytic techniques. The procedure also produces information with which the relative efficiencies of different analytical replication schemes can be compared. For example, Table 1 demonstrates that substantial reductions in the maximum percent error can be achieved by placing duplicate samples in different analytical batches rather than within the same analytical batch. The variance components procedure also facilitates the rational determination of the rejection criteria for

analytical batches of samples. The use of a batch rejection criteria is an important aspect of a quality control scheme.

The standardized and informative reporting of the quality control information described here was developed for a double-blind clinical trial presently in progress in which several chemical and biochemical analyses are being employed to investigate the effects of micronutrient status on the risk of new non-melanoma skin cancer. The quality control report for the vitamin E assay would state: "For each individual sample, two extractions with two injections per extraction were analyzed within the same batch. The maximum percent error from the true value for each sample was estimated to be 33.2 % using a variance components model. The variance components model estimated the variance for each major source of variability in the assay procedure which included effects for batch (0.014538, s.e.= 0.001330), aliquot (0.004558, s.e.= 0.000412), extraction (0.016731, s.e.= 0.001530), and random error (including injection) (0.000463, s.e.= 0.000000). The variance component model was also used to develop a quality control procedure for this assay. Two injections from each of two extractions from each of two aliquots from a standard pool of plasma with known vitamin E concentration were obtained in each analytical batch. For this design, a batch rejection criterion of 15 % was used, resulting in an estimated false alarm probability of 6 %."

The statistical power of a study is a function of several factors, including the number of subjects, the variability of the variable of interest in the population, and the analytical precision of the assay methods. Because the cost of collecting samples from human subjects is often extremely high, the efficacies that may be realized by optimizing laboratory precision through a well-defined quality control program may offer a relatively low-cost means of increasing statistical power.

Appendix

SAS Program to Compute Variance Component

Estimates for Vitamin E Data

```
OPTIONS LS=64;

CMS FILEDEF VE DISK VITE DATA A;

DATA VEDATA;

INFILE VE;

INPUT VITE P T A X E;

LVITE = LOG(VITE);

PROC PRINT;

PROC VARCOMP METHOD = TYPE1;

CLASS P T A X;

MODEL LVITE = P T A(P T) X(P T A);

PROC VARCOMP METHOD = REML;

CLASS P T A X;

MODEL LVITE = P T A(P T) X(P T A);
```

The output for this program and examples of other SAS programs for analyzing different laboratory data sets are available from the first author.

Acknowledgements

The authors gratefully acknowledge the technical assistance of Ms. Lynne P. Deuschle in conducting the vitamin E analyses.

References

1. Nelson M, Black AE, Morris JA, Cole TJ. Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *Am J Clin Nutr* 1989;50:155-167.
2. Searle SR. Topics in variance component estimation. *Biometrics* 1971;27:1-76.
3. Draper N, Smith H. *Applied regression analysis*. 2nd edition. New York: John Wiley, 1981.
4. Conover WJ. *Practical nonparametric statistics*. New York: John Wiley, 1971.
5. Rao CR. *Linear statistical inference and its applications*. New York: John Wiley, 1965.
6. Feller W. *An introduction to probability theory and its applications*, Vol II. New York: John Wiley, 1966.
7. SAS Institute Inc. *SAS User's Guide: Statistics, Version 5 Edition*. Cary, North Carolina: SAS Institute Inc., 1985.

Table 1

Replication Schemes and Approximate Maximum Percent Errors

Achieved with 95% Probability for Vitamin E HPLC Procedure

Replication Scheme				total no. of extractions	Maximum % Error (100p)
no. of batches	no. of aliqu./bat.	no. of ext./aliqu.	no. of inj./ext.		
J	K	M	N	JKM	
1	1	2	2	2	33.2
1	2	2	2	4	29.0
1	3	2	2	6	27.5
1	4	2	2	8	26.7
2	1	2	2	4	23.5
2	2	2	2	8	20.5
3	1	2	2	6	19.2
3	2	2	2	12	16.8

Maximum percent errors were computed from Equation (7).

Table 2

Probabilities (α) of Falsely Identifying Batches as Aberrant in the Vitamin E HPLC Procedure

Replications of Control			Batch Rejection Criterion			
no. of aliq./bat.	no. of ext./aliq.	no. of inj./ext.	p = 0.10	p = 0.15	p = 0.20	p = 0.25
K	M	N				
Probability of false alarm (α)*						
1	2	2	0.38	0.19	0.08	0.03
2	2	2	0.22	0.06	0.01	0.00
3	2	2	0.13	0.02	0.00	0.00

* Probabilities are rounded to second decimal place.