Improving the Sensitivity of in Vitro Skin Penetration Experiments

Gerald B. Kasting,[1,2] Thomas G. Filloon,[1] William R. Francis,[1] and Michael P. Meredith[1]

Received January 8, 1994; accepted July 1, 1994

The institution of a readily-implemented sample screening and data handling procedure for in vitro skin penetration studies yields substantial improvements in sensitivity for distinguishing between formulations, treatments, penetrants, etc. The procedure involves four steps: 1) prescreen the tissue samples to determine their intrinsic permeability; 2) apply treatments using a randomized complete block (RCB) design, with blocking by tissue permeability; 3) apply a variance-stabilizing transformation to the penetration data, followed by outlier testing; and 4) analyze the transformed data according to an RCB analysis of variance, using tissue permeability as the blocking variable. For penetration studies in which high sample variability is a concern, the above procedure commonly yields a sensitivity advantage of several-fold versus alternative methods of comparison.

KEY WORDS: in vitro skin penetration; human cadaver skin; statistics; data transformations; data analysis; sensitivity.

INTRODUCTION

Common objectives of in vitro skin penetration studies include the selection of topical or transdermal drug formulations, dermal exposure assessment for environmental toxins, and estimation of the harshness of skin care products via their effect on skin barrier function. The importance of choosing the right model system for such studies (e.g., skin species and thickness, dose, degree of occlusion, and receptor phase composition) in order to obtain relevant results has been well documented (1–7). This paper focuses on a subsequent question: Having chosen the best available model system for a penetration study, how can one maximize the sensitivity of the test for distinguishing between treatments? This question is particularly important when working with a highly variable substrate such as human cadaver skin which, despite its problems, is often the model of choice (5–7). Implementation of the procedure described herein gives up to three-fold higher sensitivity for distinguishing between treatments in a single study. Furthermore, the tissue screening aspect can lead also to greater reproducibility of skin penetration values between studies.

To show why the proposed experimental design and analysis can significantly impact results, we first consider the nature of cadaver skin penetration data. The complex structure of skin with its multiple diffusion pathways and (in vitro) its varying thickness and degree of damage leads to a distribution of experimental penetration rates that is a function of exposure time and of the physicochemical properties of the test compound.

Ionic compounds yield broad and highly right-skewed permeability distributions in skin (8–10). Freshly excised tissue yields less variable penetration data than does frozen tissue (10). Our own experience extends these observations to neutral but poorly lipid-soluble compounds that might be expected to penetrate skin via polar pathways (GBK, unpublished data). These facts alone lead us to associate the right-skewed nature of the permeability distributions for these materials with microscopic damage incurred during tissue collection and storage. The transport of ions and water-soluble compounds that cannot easily penetrate the stratum corneum lipid bilayers would be expected to be highly sensitive to the number of defects present in the lipid barrier.

Figures 1 and 2 present evidence that the permeability distributions for human cadaver skin to other agents, both hydrophilic and lipophilic, are also highly skewed. In many cases they can reasonably be taken to be lognormal, as shown in the insets. Figure 1a shows a typical distribution of water penetration values through samples of excised human cadaver skin obtained from a single donor. The results were obtained using the ^3H_2O penetration test developed by Franz and Lehman (11) and described in the Experimental section. The tissue had been harvested at autopsy by standard methods and stored frozen (9). Other skin samples identically prepared have had either narrower, more sharply peaked distributions or broader, relatively flat distributions. In our experience with this test, the primary difference from study to study lies not in the mode of the distribution (usually about 0.3–0.8 $\mu L/cm^2$), but rather in the percentage of high permeability samples. In other words, the permeability distributions differ in the weight of the right-skewed tail. We have observed that freshly excised skin (either human surgical waste or animal skin) tends to have a narrower, more symmetrical permeability distribution than does skin bank skin. However, the highly skewed distribution of cadaver skin water permeability values obtained over a large number of studies is clearly shown in Fig. 1b.

The data in Fig. 1b are skewed even if highly permeable samples are excluded from the analysis. In fact, a good fit to the lognormal (Shapiro-Wilk $W = 0.981$, $p = 0.157$) is obtained even after excluding samples whose penetration values exceed Franz and Lehman's acceptance criterion of 1.2 $\mu L/cm^2$.

Figure 2 shows that a lognormal distribution may also be obtained with lipophilic compounds. This figure shows the distribution of pooled penetration data from seven cadaver skin penetration studies involving 36 different permeants (12). The data for each compound have been normalized by dividing by the median in order to remove differences related to the center of the distributions and leave only information related to shape. The linearity of the lognormal probability plot and coincidence of the 50th percentile with the median demonstrate how well the lognormal describes the pooled data. Shapiro-Wilk testing supports this conclusion ($W = 0.983$, $p = 0.465$).

The test permeants in Fig. 2 were predominately small, lipophilic compounds, although they included several weak bases dosed as salts. Most of the latter compounds probably diffused through the skin in their neutral form due to pH-partition equilibria; in any case, their permeability distribu-

[1] The Procter & Gamble Company, Miami Valley Laboratories, P.O. Box 538707, Cincinnati, Ohio 45253-8707.

[2] To whom correspondence should be addressed.

0724-8741/94/1200-1747$07.00/0 © 1994 Plenum Publishing Corporation

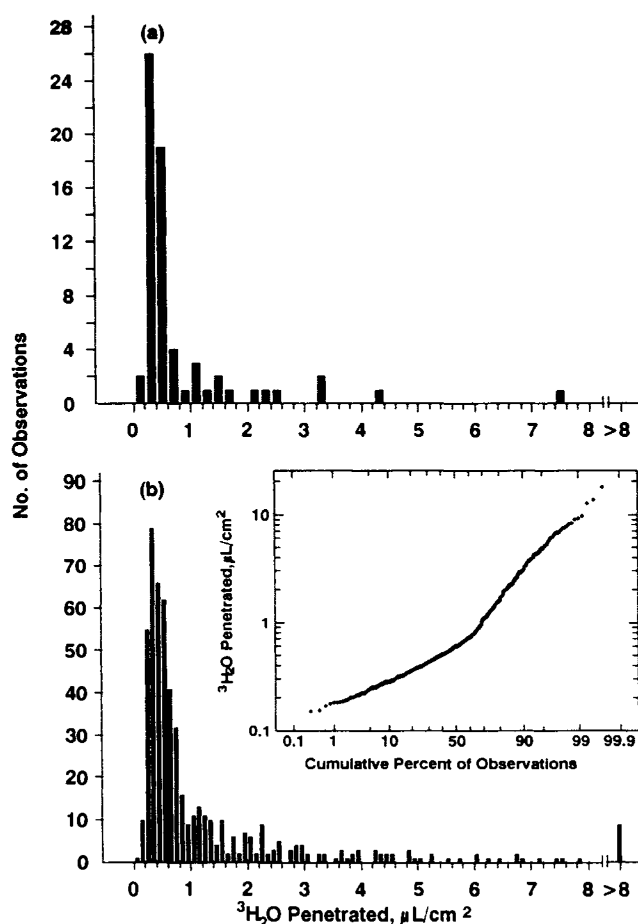

Fig. 1. Frequency distribution of water penetration values through split-thickness human cadaver skin using the ^3H_2O penetration test described in the text. (a) Results from 66 samples derived from a single donor. (b) Results from 539 samples derived from 7 donors. The inset shows a lognormal probability plot of the distribution.

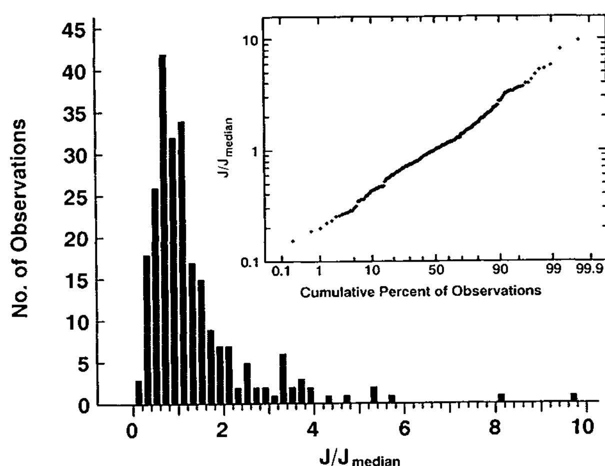

Fig. 2. Frequency distribution of steady state penetration rates of assorted drugs and other lipophilic compounds through split-thickness human cadaver skin about the median value for each compound. The data are taken from Ref. 12. The plot shows results from 3–4 donors and 302 individual samples (242 samples after excluding the medians). The inset shows the lognormal probability plot. The compounds studied were acetaminophen, benzoic acid, benzyl alcohol, caffeine, clonidine hydrochloride, dextromethorphan, dextromethorphan hydrobromide, diazepam, estradiol, ethacrynic acid, 5-fluorouracil, furosemide, griseofulvin, hydralazine hydrochloride, hydrocortisone, ibuprofen, indolyl-3-acetic acid, indomethacin, isosorbide dinitrate, ketoprofen, methyl salicylate, minoxidil, morphine sulfate, naproxen, nicotinic acid, nifedipine, pentazocine, pentazocine hydrochloride, piroxicam, propranolol hydrochloride, salicylamide, salicylic acid, sulindac, terbutaline sulfate, testosterone, and triamcinolone acetonide.

tions were not inordinately broad compared to other compounds in the dataset. Terbutaline sulfate was a notable exception, yielding highly variable penetration data consistent with the findings of Liu et al. (8).

Our subsequent experience with lipophilic compounds has continued to support the distributional trend shown in Fig. 2. Our findings thus differ from those of Liu et al., who considered lipophilic compounds to have fairly symmetrical distributions. However, their data are not inconsistent with our results, as there is a suggestion of skewness to the lipophilic compound distributions despite the small number of observations.

These data suggest three procedures that can profitably be employed to improve both the sensitivity and reproducibility of cadaver skin penetration studies. The first is to screen the tissue samples for permeability prior to each test. This allows one to eliminate unacceptable samples.

The second procedure uses the "accepted" tissue samples and their associated water permeability values to optimally allocate treatments to the diffusion cells. Tissue samples are sequentially placed into the diffusion cells according to the rank order of their water permeability values. Hence, tissue samples are grouped into more inherently homogeneous "blocks." Random assignment of treatment replicates into each of these blocks is called a Randomized Complete Block (RCB) design. The subsequent analysis of variance of this design (RCB ANOVA) is potentially much more precise than a standard ANOVA. The RCB analysis compares treatments within blocks and then pools these comparisons over blocks. This takes the inherent block-to-block tissue variability out of the analysis and, in essence, compares treatments on more homogeneous tissues. Thus one can actually use knowledge of these inherent differences in tissue sample permeability to one's advantage, by designing and analyzing such data appropriately (i.e., RCB ANOVA) so that the most precise treatment comparisons can be made. Otherwise, if a simple ANOVA is performed, the block-to-block tissue variability is not partitioned out and inflates the standard error of the treatment comparisons, yielding less precise comparisons.

The third procedure involves the mathematical transformation of penetration values prior to analysis, followed by outlier testing, so as to obtain more nearly normally distributed, equal variance responses. The transformation step can be accomplished by either taking the logarithm or (more generally, as described in Appendix 1) by applying a quasi-logarithmic transformation. Outlier testing and elimination can then be conducted using standard methods (13,14). This procedure allows common parametric statistical tests to be appropriately applied to the data; hence, valid conclusions can be drawn regarding treatment comparisons. The combi-

nation of these procedures leads to an extremely powerful analysis.

EXPERIMENTAL SECTION

Human Skin and Diffusion Cells. Dermatomed (250 μm) human cadaver skin was obtained from the Ohio Valley Skin and Tissue Center, Cincinnati, OH. The skin was stored frozen in a 10% glycerol solution prior to mounting in modified Franz diffusion cells (0.79 cm^2) as described elsewhere (9,15). The skin specimen provided sufficient tissue for 36 cells. The receptor compartments were stirred and maintained at 37°C by thermostatted heating-stirring modules (15), yielding a skin surface temperature of 30–32°C. The receptor solution was Dulbecco's phosphate-buffered saline, pH 7.4, containing 0.02% sodium azide to retard microbial growth.

Skin Permeability Test. The procedure was similar to that described by Franz and Lehman (11). Following an equilibration period of 1–2 h, 150 μL of ^3H_2O, specific activity 0.4 μCi/mL, was applied to the epidermal surface of each skin sample. At 5 minutes post-dose the tritiated water was removed and the skin carefully blotted dry with a cotton-tipped swab. At 60 minutes post-dose the receptor solution was removed for scintillation counting and replaced with fresh buffer. The penetration samples were mixed with 10 mL of Ultima Gold scintillation cocktail (Packard) and counted for 1 minute each on a Packard Model 1900 TR automatic counter at an efficiency of 21%. All the tissue samples were judged suitable for use based on ^3H_2O penetration values less than 1.6 μL/cm^2.[3] The receptor solution was exchanged a second time at 120 min post-dose and a third time after an overnight equilibration period prior to beginning the formulation study. Tests showed that this procedure reduced the residual radioactivity from ^3H_2O to below the detection level.

We selected the ^3H_2O test because of the extensive database on fresh tissue collected by Franz and Lehman; however, any rapid, nondestructive test could be substituted. Electrical resistivity would seem to be a good alternative test (9,10), especially for studies involving ionic permeants. Permeability to a volatile organic compound might provide a better screen for lipophilic compounds. However, until appropriate validation studies had been conducted, the advantage of comparing penetration values with a known value in undamaged tissue would be lost.

Skin Penetration Study. A proprietary skin care active (MW 215, log $K_{octanol/water}$ = 1.26), radiolabeled with ^{14}C at a specific activity of 30 μCi/mg, was incorporated into a simple solution (Formulation A) at a concentration of 0.1% w/w. This formulation had been shown to be effective in preclinical efficacy testing and thus provided a target delivery rate. Five prototype skin creams (Formulations B–F) consisting of different emulsion bases and varying amounts of dissolved and suspended ^{14}C-labeled active were then prepared. The objective of the study was to identify the cream formulation yielding the greatest delivery of active in 24 h and to compare its delivery rate with that of Formulation A.

The 36 skin samples were arranged in order of permeability (lowest to highest) based on the results of the ^3H_2O skin permeability test. Each of the six formulations was then applied to six skin samples in a randomized complete block (RCB) design. In other words, each formulation was applied at random to one of the six least permeable skin samples, then to one of the six next-least-permeable samples, etc., until all of the 36 samples had been treated. The dose was 3 μL of formulation per cell, corresponding to approximately 0.1 μCi of radioactivity per cell and 4 mg of formulation per square centimeter of skin.

Penetration samples consisting of a complete exchange of the receptor solution in each cell were collected at 2, 4, 6.3, and 24 hours post-dose. The samples were analyzed as in the ^3H_2O test, except that these samples were counted for either 5 minutes or until 2% accuracy (2 SD) had been achieved, whichever was shorter. The counting efficiency of ^{14}C was 90%. Three blank samples were collected at each time point to accurately determine the background radiation level. Results were calculated at each time point as the cumulative amount of active that had appeared in the receptor solution. In no case did the amount penetrated exceed 20% of the applied dose.

Data Analysis. In order to more fully satisfy implicit ANOVA assumptions and, thus, ensure the validity of the subsequent conclusions, a variance-stabilizing transformation was applied to the data. The cumulative penetration values, Y, at each time point were transformed according to Eq. (8) of Appendix 1 using the values $a = 0.01$ $(ng/cm^2)^2$ and $b = 0.004$ ng/cm^2. These values correspond to an analyte detection limit, $\sqrt{a}$, of 0.1 ng/cm^2 or approximately 5 dpm. The parameter c was assigned a value that decreased with time, based on a preliminary evaluation of the data as described in Appendix 2. This reflects the fact that the variance in the penetration data decreased as the study progressed. The values used were $c = 0.49$, 0.16, 0.09, and 0.04 for times 2, 4, 6.3, and 24 h, respectively. For times greater than 4 h, this transformation was equivalent to a simple logarithm [Eq. 2)] to within the experimental error. The pooled residuals of the transformed data, calculated as described in Appendix 2, were approximately normally distributed (e.g., at 24 h, Shapiro-Wilk $W = 0.962$, $p = 0.307$).

We examined the transformed data for outlying values using Dixon's method and the Maximum Normed Residual Test (13,14). No outliers were identified by either method. The transformed data were then analyzed using an RCB ANOVA, with tissue permeability as the blocking variable. Tukey's Studentized Range (HSD) Test with an α value of 0.05 was used to create a multiple comparison table. Means and standard errors were calculated for the transformed data, then converted back to the original units for interpretation using Eqs. (9) and (10). Since the transformation was nearly logarithmic, this procedure estimated the geometric mean, rather than the arithmetic mean, of the original data. No correction for bias was necessary since we considered the geometric mean (which corresponds to the median for a

[3] Franz and Lehman found that freshly excised skin averaged 0.3–0.4 μL/cm^2 and recommended an acceptance criterion of ≤1.2 μL/cm^2. We use the latter value as a guideline, but may alter it depending on whether the interest is in estimating human in vivo penetration rates (adjust downward) or in obtaining relative information regarding formulations (adjust upward). In any case we would suggest excluding samples with water penetration values above about 3 μL/cm^2.

lognormal distribution) to be the more useful estimate of the "center" of the distribution.

For comparative purposes we conducted several alternative analyses at the 24-h time point: RCB ANOVA on the untransformed penetration data, one-way ANOVA (ignoring blocking on tissue permeability) on both transformed and untransformed data, and a nonparametric rank sum (Kruskal-Wallis) test. We also conducted pairwise nonparametric comparisons by the latter method (i.e., Wilcoxon rank sum test).

For data reduction, transformation, and outlier testing we used a BASIC program written by one of the authors. An EXCEL spreadsheet implementing these calculations is also available from the authors. For statistical testing we used procedures GLM, UNIVARIATE, and NPAR1WAY from SAS Release 6.06 (SAS Institute Inc., Gary, NC), selecting the Tukey option in PROC GLM for parametric pairwise comparisons.

RESULTS

The results of the example skin penetration study are shown in Fig. 3. This figure was prepared using the data transformation and back-transformation steps in Eqs. (8–10); hence, the advantages of the data transformation procedure are already realized in this representation of the data. According to this representation Formulation A, the simple solution, yielded the greatest penetration of active compound for all time points. It was of interest to determine which of the cream formulations B, . . . ,F differed significantly from Formulation A or from each other.

The results of statistical tests performed on the cumulative 24-h penetration values are summarized in Table I. Formulations E and F can be seen to deliver significantly less active across the skin than does Formulation A, according to the recommended analysis (Column 1). The alternative methods of analysis are not as sensitive. In particular, direct averaging of the penetration data followed by a one-way ANOVA (Column 4) fails to reveal any significant differences between treatments. The nonparametric analysis (Column 5) and the RCB analysis of the untransformed data (Column 2)

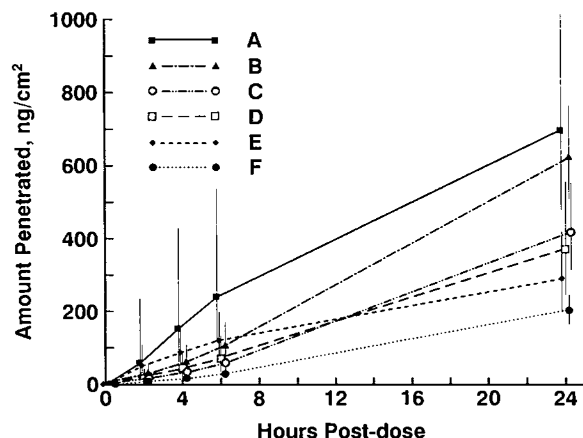

Fig. 3. Cumulative amount of active compound penetrated versus time for the example penetration study. Each point represents the (nearly) geometric mean ± SE (n = 6), calculated by transforming the data according to Eq. (8), averaging, and then back-transforming according to Eqs. (9) and (10).

do distinguish Formulation F from Formulation A, but their lower discriminatory power is evident.

DISCUSSION

The sensitivity advantages of the experimental design and analysis described herein derive from two sources—1) recognition of the potential for a highly skewed distribution of skin permeability values [supported for the frozen, split-thickness cadaver skin substrate by the data presented here and in Refs. 8–10] and 2) the ability to remove some of the tissue-to-tissue variability via an RCB design and analysis. In order to estimate the sensitivity increase afforded by each factor, we discuss below the sample size requirements for a prospective skin penetration study using the example data as pilot information. In this analysis we assume that the distribution of sample permeabilities is approximately lognormal as found in the example penetration study (i.e., $\log_{10}$ permeability is normally distributed). Therefore, we use normal distributional theory for sample size determinations. We are determining what sample sizes are needed to detect a given treatment difference, with 95% confidence and 80% power.

Specifically, we consider the following question: What are the necessary sample sizes for detecting $\log_{10}$ treatment differences of 0.3, 0.6, and 0.9 units using each of the analysis methods in Table I? These differences reflect 50%, 75% and 87.5% reductions in mean level. For the experimental data presented in Fig. 3, where the control (i.e., Formulation A) mean at 24 h was approximately 700 ng/cm², this would correspond to changes of 350, 525, and 610 ng/cm², respectively.

In order to answer this question, we need an estimate of the sample variability in a typical penetration study. We can use the individual sample data 24 h post-dose from the study shown in Fig. 3 for this purpose. Expressed in terms of the Mean Squared Error (MSE), the sample variability is as follows for each ANOVA method:

Untransformed		Transformed	
One-Way	RCB	One-Way	RCB
0.439×10^6	0.218×10^6	0.116	0.045

Using the above variability estimates and a standard power analysis method for normally distributed data (16) and its nonparametric counterpart (17), the sample sizes required to detect various magnitudes of treatment differences (with 95% confidence and 80% power) can be determined. The nonparametric approach implicitly uses the lognormal distributional assumption. For the ANOVA methods, the necessary sample size has been approximated by 16 × (MSE)/(Difference × Difference) as per Lehr's approach (18). This alone allows one to see that sample sizes will be 2–3 fold higher for the one-way versus RCB ANOVA methods. Also, intuitively, the transformed analysis should show an advantage over the untransformed analysis as its variability estimate is not inflated due to its long-tailed distribution (i.e., skewness).

The results of this calculation are shown in Table II. The sensitivity advantage provided by the data-transformed RCB

Table I. Ranking of 24-h Cumulative Penetration Values According to Several Methods of Analysis. Brackets Denote Groups Which are not Statistically Different at $p = 0.05$

Analysis:	RCB ANOVA		One-Way ANOVA		Nonparametric
Data transformed?	Yes	No	Yes	No	
Rank					
1	A	A	A	A	B
2	B	B	B	B	A
3	C	D	C	D	E
4	D	C	D	C	C
5	E	E	E	E	D
6	F	F	F	F	F
p value for overall group differences (F or χ^2 test)	0.0016	0.076	0.088	0.361	0.055

ANOVA approach is clearly evident from this analysis. In particular, if one considers a two-fold difference in mean penetration (i.e., $\log_{10}$ change of 0.3) to be of practical significance in a formulation study, the attractiveness of detecting this difference with a group size of 8 rather than 20–30 is hard to overlook. In such cases the time invested in careful attention to experimental design and analysis can be quickly recovered by reduction of the time spent in execution of the studies.

The data transformation approach described here is derived from the work of Bartlett (19). The motivating factor is variance stabilization in order to yield a more efficient analysis. Such transformations usually normalize the data also so that the transformed data satisfy the implicit ANOVA assumptions of normally distributed, same variance treatment groups. Other transformation approaches to achieve normally distributed data, such as the Box-Cox method (20), $Z = (Y^\lambda - 1)/\lambda$, may also stabilize variances by coincidence, but represent a different tact from what we have attempted here. The strengths of the present approach are that it is data-based and provides for variance stabilization and smooth data handling in both signal-limited and non-signal-limited cases. In the former case the recommended transformation [either Eq. (5) or (8)] smoothly handles the zero and negative data values that frequently arise, using transformation coefficients estimated from the experiment. In the latter case the transformation reduces to the simple, but powerful logarithm [Eq. (2)], which is coefficient-free. The data transformation step can thus be readily automated.

The RCB experimental design and analysis is enabled by the tissue permeability screen at the outset of the study. This procedure also allows for rejection of poor quality samples. Franz and Lehman's work with ^3H_2O (11) has facilitated this step, although alternative screening procedures are possible. The combination of sample rejection, variance stabilization, and RCB ANOVA procedures leads to an extremely powerful analysis. The sensitivity improvement afforded by these procedures can reduce the execution time and improve the quality of the data generated in cadaver skin penetration studies.

APPENDIX 1

Data Transformations for Skin Penetration Studies

In this section we develop data transformations suitable for analyzing experiments in which a variance-stabilizing transformation is desired. In the case where most of the experimental variation arises from a single source (as it may from the tissue samples in a typical penetration study), a simple, coefficient-free transformation such as the logarithm is often appropriate. However, when a second source of error comes importantly into play, the simple transformation may be inappropriate or even impossible to carry out. We

Table II. Group Size (n) Required to Detect Various Magnitudes of Treatment Difference, with 95% Confidence and 80% Power, Using the Example Penetration Data

Analysis:	RCB ANOVA		One-Way ANOVA		Nonparametric
Data transformed? Treatment difference ($\log_{10}$ change)	Yes	No	Yes	No	
0.3	8	14	21	28	25
0.6	2	7	5	13	9
0.9	1[a]	5	2	9	6

[a] A sample size of 1 is displayed here for descriptive purposes only, as at least 2 samples are needed per group for statistical tests to be performed.

give below two quasi-logarithmic data transformations to deal with this issue. Unlike a simple logarithm, these transformations can be used to analyze data in which extremely low signal levels lead occasionally to zero or negative data values.

Derivation of a variance-stabilizing transformation proceeds as follows. Consider a positive-valued random variable Y whose variance is related to the square of its mean value according to Eq. (1):

$$\underline{\text{Model 1}}: \sigma^2 = c\mu^2; c > 0 \quad (1)$$

In this model, the standard deviation, σ, is directly proportional to the mean, μ. The coefficient of variation, $CV = \sigma/\mu$, is stable. It is well known that the appropriate variance-stabilizing transformation for Model 1 is given by:

$$Z = \log_{10} Y \quad (2)$$

where the base 10 logarithm has been chosen as a matter of preference. Typically, the variable Y is assumed to be log-normally distributed; if so, Z is normally distributed. Assumptions necessary for classical ANOVA calculations and inferences are satisfied for the transformed data. As shown in the text, experimental cadaver skin penetration data for both hydrophilic and lipophilic penetrants support the use of Eq. (2) in analyzing skin penetration studies.

The problem is to find suitable transformations when the model is not so simple. In particular, the CV may be large when the response, Y, is small. The signal is lost in the noise of the measurement technique at the lower extreme, resulting in a large CV for small Y even though the CV becomes constant for higher response. Y may possibly assume zero or negative values. The precision of the measurement technique must be incorporated into the model and, therefore, the transformation. We consider two realistic models.

$$\underline{\text{Model 2}}: \sigma^2 = a + c\mu^2; a > 0, c > 0 \quad (3)$$

In this model there are variance components due to noise in the measurement technique (i.e., a) and due to variability in the experimental material under study (i.e., $c\mu^2$). From a variance components point of view, it would appear reasonable to assume that Model 2 holds in the "limit of detectability" problem when the sample being studied has the intrinsic properties of Model 1. Model 2 could be appropriate for a skin penetration study in which the detection technique has a fixed sensitivity limit (e.g., $s = \sqrt{a}$). We have found it to be useful, for example, for studies in which the analysis is conducted by HPLC.

The limitations of a log transform in this case are readily apparent: it will fail if random signal fluctuations yield zero or negative values of Y. This problem can be partially remedied by adding a constant value to Y prior to log transformation (21), $Z = \log_{10}(Y + d)$. However, this approach is difficult to justify on a theoretical basis, and the problem of choosing an appropriate value of d is non-trivial. Variance stabilization at low signal levels cannot be ensured. We describe below an alternative procedure that utilizes the variance components for Model 2. The cost is that the experimenter must estimate the coefficient ratio a/c.

Following the approach of Bartlett (19), the variance-stabilizing transformation for Model 2 can be found from the following integral:

$$\int \frac{dy}{\sqrt{a + cy^2}} = \frac{1}{\sqrt{c}} \log_e \left(y\sqrt{c} + \sqrt{a + cy^2}\right) \quad (4)$$

An appropriate transformation is given by:

$$Z = \log_{10}\left(\tfrac{1}{2}Y + \tfrac{1}{2}\sqrt{Y^2 + a/c}\right) \quad (5)$$

In Eq. (5) we chose the normalization factor of ½ and the base 10 logarithm in order to yield Eq. (2) when $Y^2 \gg a/c$. Thus Z is equal to the base 10 logarithm of the penetration value Y when the signal level is high.

In a similar vein, variance-stabilizing transformations can be derived for more complicated variance models. In Appendix 2 we show that Model 3 can arise from a radiochemical skin penetration experiment in which statistical counting error contributes to the error in the measurement.

$$\underline{\text{Model 3}}: \sigma^2 = a + b\mu + c\mu^2; a > 0, b > 0, c > 0 \quad (6)$$

As before, the variance-stabilizing transformation can be found from the integral:

$$\int \frac{dy}{\sqrt{a + by + cy^2}} = \frac{1}{\sqrt{c}} \log_e \left(2\sqrt{ac + bcy + c^2y^2} + 2cy + b\right) \quad (7)$$

An appropriate transformation is:

$$Z = \log_{10}\left(\tfrac{1}{2}\sqrt{Y^2 + (b/c)Y + a/c} + \tfrac{1}{2}Y + \tfrac{1}{4}b/c\right) \quad (8)$$

Like Eq. (5), Eq. (8) reduces to Eq. (2) for large values of Y. We used Eq. (8) to transform the example penetration data after estimating the values of a, b, and c as described in Appendix 2.

After averaging the data and performing statistical comparisons in the transformed units, it is often desirable to undo the transformation for presentation of the results. The reverse transforms for Eqs. (2), (5), and (8) are easily obtained by algebraic rearrangement. For example, the reverse transform for Eq. (8) is:

$$Y = (4Y_0^2 - a/c)/(4Y_0 + b/c) \quad (9)$$

where Y_0 is defined as:

$$Y_0 = 10^Z - \tfrac{1}{4}b/c. \quad (10)$$

It is prudent to apply a normality test (e.g., the Shapiro-Wilk test) to the transformed data to ensure that an appropriate transformation has been made. If non-normality is detected, the same test may be applied to the untransformed data and the decision as to whether or not to transform the data based on this result. An example of a case in which logarithmic or quasi-logarithmic transforms should often not be made is that of finite dose penetration experiments in which a substantial fraction (i.e., >30%) of the applied dose of active penetrates each skin sample. In this case the distribution of penetration values will no longer be strongly positively-skewed. Although it is possible to construct variance models

which take this feature into account, the corresponding data transformations are not easily obtained. In such cases the best solution is often to eliminate the data transformation step.

APPENDIX 2

Coefficients for Data Transformations

In this section we show how to estimate the value of the coefficients a, b, and c for the quasi-logarithmic data transformations in Appendix 1. Since these transformations reduce to the logarithm [Eq. (2), which is coefficient-free] for all but the lowest signal levels, the coefficients need not be known to a high degree of accuracy for the transformation approach to be of value. The use of the more complex expressions simply yields transformed values Z which are more stable at low signal levels than those calculated from Eq. (2).

Accounting for Measurement Error (Estimation of a and b)

As discussed in Appendix 1, a measurement technique with a fixed sensitivity s adds a constant term to the experimental variance which, when combined with a right-skewed distribution for skin permeability, leads to variance model 2 [Eq. (3)]. The parameter a in this case is calculated from the relationship $s = \sqrt{a}$, and c is estimated as described below. Data transformation is accomplished using Eq. (5).

For radiochemical experiments statistical counting noise at low signal levels can lead to variance model 3 [Eq. (6)]. The appropriate data transformation in this case is given by Eq. (8). To estimate the coefficients we proceed as follows: Let X = sample radioactivity level in counts per minute (cpm). T = sample counting time in minutes, B = background radioactivity level in counts per minute, σ_B = error (standard deviation) in background level, A = conversion factor from cpm to experimental units, Y = sample active level in experimental units. Then, by definition,

$$Y = A(X - B) \tag{11}$$

The error in Y may be assumed to come from three factors: X, B, and the tissue variability. By propagating the errors in Eq. (11) we have:

$$\sigma_Y^2 = \left(\frac{\partial Y}{\partial X}\right)^2 \sigma_x^2 + \left(\frac{\partial Y}{\partial B}\right)^2 \sigma_B^2 + c\mu^2 \tag{12}$$

where the third term on the right derives from the underlying skin permeability distribution [as in Eq. (6)]. The error in X may be calculated from a Poisson statistical model as $\sigma_X^2 = \bar{X}/T = B/T + \mu/AT$. Here we have made use of Eq. (11) and then let μ represent the mean value of Y to obtain the second relationship. Substitution of the appropriate derivatives and error terms into Eq. (12) yields:

$$\begin{aligned} \sigma_Y^2 &= A^2(B/T + \mu/AT) + A^2\sigma_B^2 + c\mu^2 \\ &= A^2(\sigma_B^2 + B/T) + (A/T)\mu + c\mu^2 \end{aligned} \tag{13}$$

Comparing Eq. (13) with Eq. (6) we find:

$$a = A^2(\sigma_B^2 + B/T);\ b = A/T \tag{14}$$

Thus, the coefficients a and b in Eqs. (6–10) are functions of the unit conversion factor A, the background level B, its variance σ_B^2, and the sample counting time T. These factors are readily available to the experimenter. (The experimental background variance estimator, s_B^2, may be substituted for σ_B^2 to calculate a.)

Variance of Skin Permeability Values (Estimation of c)

The variance coefficient c characterizing the skin permeability distribution can be derived from either the ^3H_2O screening data or from the test permeant penetration values at any point in the experiment. We recommend the latter method, as the test permeant results may have quite a different distribution than the water permeation data and, furthermore, their distribution changes over time. Although the number of replicates for each treatment is usually too small to determine the underlying distribution for that treatment, one can obtain distributional estimates under the assumption that the treatments affect only the location, and not the shape, of the permeability distribution. Under this assumption one can combine the data from different treatments after normalizing to account for the change in location. An example was shown in Fig. 2.

Using the pooled data approach, suitable values for c, as well as a normality check on transformed data, can be obtained as follows: 1) Estimate a and b as described above. Make an initial estimate for c at each sampling time from the test permeant penetration results, based on previous experience [the values range from about $c = 0.01$ for narrowly distributed data to $c = 0.5$ for highly variable data]. 2) Transform the penetration data according to either Eq. (5) or (8), and calculate the one-way ANOVA model for transformed data at each sampling time. 3) Calculate the variance of the residuals from each ANOVA model. These values represent refined estimates for c. The residuals may be examined for normality. 4) Retransform and reanalyze the original data according to the RCB ANOVA model using the refined estimates for c. In practice, the retransformation step can often be omitted and the RCB ANOVA model calculated directly as one gains experience with the method.

It is also apparent that c is related to the slope of the lognormal probability plot, e.g., the insets for Figs. 1 and 2. More precisely, the slope is equal to $\sqrt{c}$. Thus, an alternative procedure for estimating c is to regress log-transformed penetration values on normal scores to obtain the slope. For example, regression analysis of the data in Fig. 2 yields $\sqrt{c} = 0.31$, or $c = 0.096$, with an r^2 value of 0.992. Direct calculation of the variance, as suggested above, yields the same result.

ACKNOWLEDGMENTS

We are grateful to A. M. Marrer who contributed significantly to the early stages of this project, E. C. Drago for expert SAS programming, and Dr. T. J. Franz for a preprint of his work and several helpful discussions.

REFERENCES

1. T. J. Franz. Percutaneous absorption. On the relevance of in vitro data. *J. Invest. Dermatol.* **64**:190–195 (1975).

2. R. L. Bronaugh and R. F. Stewart. Methods for in vitro percutaneous absorption studies III: hydrophobic compounds. *J. Pharm. Sci.* 73:1255–1258 (1984).
3. R. L. Bronaugh and R. F. Stewart. Methods for in vitro percutaneous absorption studies VI: preparation of the barrier layer. *J. Pharm. Sci.* 75:487–491 (1986).
4. R. L. Bronaugh and S. W. Collier. In vitro methods for measuring skin permeation. *Cosmet. Toiletries* **105**:86–89, 92–93 (1990).
5. J. P. Skelly, V. P. Shah, H. I. Maibach, R. H. Guy, R. C. Wester, G. Flynn, and A. Yacobi. FDA and AAPS report of the Workshop on Principles and Practices of In Vitro Percutaneous Penetration Studies: relevance to bioavailability and bioequivalence. *Pharm. Res.* 4:265–267 (1987).
6. D. R. Friend. In vitro skin permeation techniques. *J. Control. Release* **18**:235–248 (1992).
7. K-C. T. Hoang et al., Dermal Exposure Assessment: Principles and Applications, Interim Report EPA/600/8-91/011/B of the U.S. Environmental Protection Agency, January, 1992.
8. P. Liu, J. A. S. Nightingale and T. Kurihara-Bergstrom. Variation of human skin permeation in vitro: Ionic vs neutral compounds. *Int. J. Pharm.* **90**:171–176 (1993).
9. G. B. Kasting and L. A. Bowman. DC electrical properties of frozen, excised human skin. *Pharm. Res.* 7:134–143 (1990).
10. G. B. Kasting and L. A. Bowman. Electrical analysis of fresh, excised human skin: a comparison with frozen skin. *Pharm. Res.* 7:1141–1146 (1990).
11. T. J. Franz and P. A. Lehman. The use of water permeability as a means of validation for skin integrity in in vitro percutaneous absorption studies. *J. Invest. Dermatol.* **94**:525 (1990).
12. G. B. Kasting, R. L. Smith, and E. R. Cooper, Effect of lipid solubility and molecular size on percutaneous absorption, in: B. Shroot and H. Schaeffer (eds.), Pharmacology of the Skin, Vol. 1, Karger, Basel, 1987, pp. 138–153.
13. W. J. Dixon and F. J. Massey. Introduction to Statistical Analysis, 3rd ed., McGraw-Hill, New York, 1969, pp. 328–330.
14. G. W. Snedecor and W. G. Cochran. Statistical Methods, 8th ed., Iowa State University Press, Ames, Iowa, 1989, pp. 278–281.
15. E. W. Merritt and E. R. Cooper. Diffusion apparatus for skin penetration. *J. Control. Release* **1**:161–162 (1984).
16. G. W. Snedecor and W. G. Cochran. Statistical Methods, 8th ed., Iowa State University Press, Ames, Iowa, 1989, pp. 68–70.
17. G. E. Noether. Sample size determination for some common nonparametric tests. *J. Amer. Stat. Assoc.* **82**:645–647 (1987).
18. R. Lehr. Sixteen S-squared over D-squared: a relation for crude sample size estimates. *Stat. Med.* **11**:1099–1102 (1992).
19. M. S. Bartlett. The use of transformations. *Biometrics* 3:39–53 (1947).
20. G. F. P. Box and D. R. Cox. An analysis of transformations. *J. Royal Stat. Soc. B* **26**:211–243 (1964).
21. G. W. Snedecor and W. G. Cochran. Statistical Methods, 8th ed., Iowa State University Press, Ames, Iowa, 1989, p. 290.