Pharmaceutical Research, Vol. 25, No. 7, July 2008 (© 2008)
DOI: 10.1007/s11095-008-9577-4

Research Paper

Optimizing Metrics for the Assessment of Bioequivalence Between Topical Drug Products

Berthe N'Dri-Stempfer,[1] William C. Navidi,[2] Richard H. Guy,[3] and Annette L. Bunge[1,4]

Received October 30, 2007; accepted March 25, 2008; published online May 1, 2008

Purpose. Stratum corneum tape stripping post-application of a drug product followed by analysis of the active agent in this tissue layer is an approach being seriously considered for the comparative assessment of topical bioavailability. Key issues revolve around how best to perform this experiment and interpret the data.
Methods. Using previously published results from a comparative study of three 0.025% tretinoin gel products, alternative data analysis approaches are presented that may render the technique more accessible to the evaluation of new and generic topical dosage forms.
Results. For the tretinoin gel study, the conclusions for bioequivalence from measurements of drug levels at only one uptake and one clearance time were the same as those from the original study, which required measurements at eight different treatment times. Furthermore, comparisons of drug levels at one uptake and one clearance time discriminated differences in bioequivalence for clearance and uptake, which had previously been missed. Half-life estimates, derived from time course data of drug clearance, can be related to lag time for drug penetration through the SC.
Conclusions. This new data analysis demonstrates that comparative bioequivalence might be assessed more easily.

KEY WORDS: dermatopharmacokinetics; skin; stratum corneum; tape stripping; topical drug bioequivalence.

INTRODUCTION

The US Food and Drug Administration (FDA) is mandated by law to identify procedures to facilitate the commercialization of bioequivalent, generic drug products. With respect to oral delivery, the accepted approach is relatively straightforward and is principally based on matching blood level profiles (rate and extent of absorption). For topical drug products, apart from those containing corticosteroids, a clinical trial is the only route for approval of a generic or for replacement of an already approved dermatological product that has appreciable compositional changes. Comparative clinical trials are relatively insensitive, time-consuming and costly. To gain adequate statistical power to clearly evaluate bioequivalence can require several hundred subjects (1).

As a result, the FDA has been seeking an assessment protocol that might replace clinical efficacy studies and, in 1998, a dermatopharmacokinetic (DPK) approach was proposed as an alternative for assessing bioequivalence (BE) (2). In this minimally invasive method, drug levels in the stratum corneum (SC) are measured as a function of time post-application and post-removal of the formulation using tape-strip sampling *in vivo* in humans.

The DPK protocol (2) specifies that each formulation is applied to at least eight sites divided equally for studying the kinetics of drug uptake and clearance. To assess drug uptake, the formulation is cleaned from four treatment sites at four different times post-application and then immediately tape stripped. Clearance is measured at the four remaining sites by removing the drug from the skin surface at the longest of the post-application times in the uptake measurements, and then tape stripping at four different times post-removal. Each of the 8 treatment sites is tape stripped 12 times and tape strips 3–12 are combined and quantified for drug. The amount of drug in the first two tape strips is not included in the assessment due to the possibility of incomplete removal of the topically applied product from the skin surface (2). The time points for uptake and clearance are not specified, except for the longest uptake time, which is supposed to be long enough that drug uptake is at steady state (2). Although other study sites are allowed (2), most DPK studies are conducted on the volar forearm for reasons of convenience.

The FDA guidance recommended that the performance of a topical formulation should be quantified in a manner which mirrors BE tests for oral drug bioavailability in which blood levels are measured (2). That is, the amount of drug in the tape-strip samples is characterized in terms of the three pharmacokinetic parameters illustrated in Fig. 1: (1) the time

[1] Chemical Engineering Department, Colorado School of Mines, 1500 Illinois Street, Golden, Colorado 80401, USA.
[2] Mathematical and Computer Science Department, Colorado School of Mines, Golden, Colorado 80401, USA.
[3] University of Bath, Bath, BA2 7AY, UK.
[4] To whom correspondence should be addressed. (e-mail: abunge@mines.edu)

integration of the amount of drug (AUC); (2) the maximum amount of drug (A_{max}), which is somewhat confusingly referred to as C_{max} in the FDA guidance (2) and by Pershing *et al.* (3,4); and (3) the time (T_{max}) at which A_{max} is first observed. Bioequivalence between the test and reference formulations is then assessed by constructing the 90% confidence interval for the ratio of the geometric means of the test and reference products (i.e., the geometric means based on the log-transformed data) using the two one-sided hypotheses at the α=0.05 level of significance (2). For the test and referenced products to be deemed BE, the confidence interval for the ratio of the means should fall within 80–125% for the AUC and 70–143% for A_{max} (2). No BE criteria are specified for T_{max}, probably because it usually coincides with the time of drug removal in the clearance studies [e.g., see (3)] and is therefore an insensitive indicator of differences in products. While the FDA guidelines allow for the use of metrics other than AUC and A_{max}, no other options are mentioned.

In 2002, the FDA withdrew the DPK guidance (5) citing two principal concerns: (a) the adequacy of the method to assess the bioequivalence of topical dermatological drug products, and (b) the reproducibility of the method between laboratories. The first concern was based on the fact that the DPK method measures penetration through healthy SC. For drugs treating target sites other than the SC, it is possible that the DPK method may not accurately reflect therapeutic effectiveness if penetration through another pathway, such as hair follicles, is important. Similarly, because skin barrier function is perturbed in many dermatological diseases, therapeutic effectiveness could be unrelated to penetration through healthy SC. In light of this concern, the FDA has chosen to restrict studies of the DPK method which began in 2003 (6,7) to drugs for which the SC is the site of activity (e.g., antifungals). The second concern was based on the contradictory results of a comparison study, performed by two expert laboratories (only one of these studies has been published), of a reference product with a bio-inequivalent product. The disagreement between these two studies has been explained convincingly as differences in the control of lateral spread from the application site (8–10), a problem that can be managed by an improved protocol.

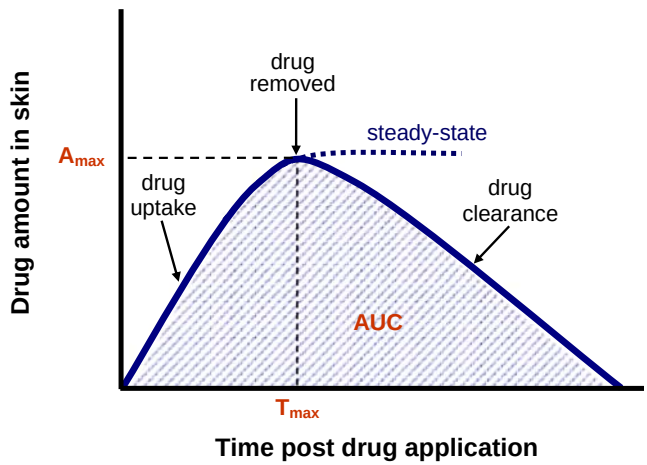

Fig. 1. Schematic diagram illustrating the DPK analysis of drug in the SC sampled by the tape-stripping methodology recommended in the 1998 FDA guidance (2).

There are other practical problems associated with the use of DPK when used as recommended in the 1998 FDA protocol. These include: (a) the variability in the amount of SC collected on each tape strip, (b) the associated variability in the amount of drug that is discarded with the first two tape strips, and (c) the variability in cleaning effectiveness. As a result, in the only published study of topical dermatological products that followed the FDA protocol, variability was large, necessitating a large number of subjects to achieve adequate statistical power (3). For three 0.025% tretinoin gel products, drug levels were determined in nearly 1,200 sites (three formulations at eight sites per formulation in 49 subjects), an amount of effort that clearly reduces the potential advantage of DPK over a classic clinical trial.

The number of tests required for BE assessment using DPK can be reduced by decreasing the number of determinations per formulation as well as the measurement variability so that fewer subjects would be required. The strategy of reducing variability is the focus of an on-going study, which is examining improved cleaning procedures, inclusion of drug from the first two tape strips, and an increase in the number of tape strips combined with a methodology to insure that nearly all of the SC is collected. Here, we examine the requirement of 8 treatment sites per formulation and the AUC analysis for BE assessment.

There is no experimental evidence that reliable BE assessment of topical products requires drug level determinations in eight treatment sites measured at four uptake and four clearance times. In fact, in several studies using tape-strip sampling of the SC, meaningful comparisons between formulations have been made using measurements collected at only one time post-application, chosen so that steady-state was not yet achieved (11–14) or at only one time post-removal (15). While these studies differed from the FDA protocol in that drug concentration was determined as a function of position within the SC, they provide encouraging evidence that a BE assessment might be possible with measurements at fewer time points. Although never justified, the recommendation for eight treatment sites was presumably chosen to provide a sufficient number of time points for the AUC analysis. This number of points is not needed to determine A_{max}, which is often found in the treatment site that is tape stripped immediately after the longest uptake time.

While the parallel with oral drug assessment is appealing, there are important differences between topical and oral drug delivery that are pertinent to the appropriateness of the strategy of using AUC or A_{max} for evaluating topical BE. In oral delivery, blood levels typically increase until the delivered dose is exhausted (with the exception of some prolonged-release formulations), after which the drug's concentration decreases as a result of metabolism and/or elimination. The AUC and A_{max} are measures of total exposure and maximal acute exposure, respectively, and both are affected by the kinetics of drug uptake, distribution, metabolism and elimination. This is not the case for topical drug delivery to the skin. Compared with oral delivery, the mechanisms of uptake and elimination to the SC are simplified and nearly always controlled by penetration across the skin, often through the SC itself (16,17). Furthermore, only a small fraction of the applied dose is delivered (often less than a few percent) (18). Perhaps more important, in DPK, drug levels usually do not

decrease until the applied dose is removed. As a result, the time course is influenced artificially by the time chosen to remove excess drug from the skin.

In addition, the factors that control drug amount in the SC during uptake and clearance phases are different. During uptake, partitioning into and diffusion through the SC both control whereas, during clearance, the mechanisms include diffusion through the SC as well as partitioning from the SC to the deeper skin tissues and, in some cases, binding within the SC. A BE assessment should therefore examine both uptake and clearance. However, A_{max} is a poor choice because it primarily measures performance during uptake. Equally, although AUC includes both uptake and clearance, the two are usually weighted unequally and the relative weighting depends on the chosen maximum application time. For example, in the tretinoin study the clearance time was eight times longer than the uptake time (3). Depending on the study protocol, AUC could be dominated by either the uptake or the clearance, making it poorly sensitive to differences in the other.

Also, because some drugs, designed to target the SC, are relatively large (e.g., molecular weights between 350 to 450 for many antifungals), drug levels in the SC are low and the uptake and clearance rates in skin are both slow. For example, the slow dermatopharmacokinetics of econazole mean that if sufficient time is allowed during the clearance phase for drug levels in the SC to have changed significantly, the absolute levels present approach the limits of detection of the analytical method (6). If the treatment area is increased to achieve adequate drug detection, then only six treatment sites per arm (6,7) can be studied and it becomes impossible, therefore, to develop an 8-point concentration–time curve to compare two formulations unless the experiment is conducted on two occasions separated by the time needed for the skin to recover.

We therefore initiated an investigation into alternatives to AUC or A_{max} in an attempt to identify more suitable metrics for BE assessment of the rate and extent of drug delivery to the skin, in particular for drugs that might absorb or clear slowly. It was a further objective to find an approach requiring fewer than eight treatment sites per formulation. Three different strategies were considered: (a) comparisons of drug levels at one uptake time and one clearance time; (b) comparisons of drug levels at one uptake time and one clearance time with each replicated at least twice; (c) comparisons of drug levels summed from all uptake times and also from all clearance times. The last strategy is similar to AUC except that the drug levels from each measurement are weighted equally rather than with respect to time. Also, the second and third strategies would be the same if all treatment sites are studied at the same uptake or clearance times. Here we compare the results from the first and last strategy to the AUC and A_{max} approach by reanalyzing the published (3,4) DPK study of the tretinoin gel products.

MATERIALS AND METHODS

In the earlier investigation (3,4), applications of 4.4 μL cm^{-2} for each of three 0.025% tretinoin gel products (an applied dose of 750 ng tretinoin cm^{-2} on an area of 1.13 cm^2) were evaluated by tape stripping the SC with D-Squame disks (1.33 cm^2 area from Cuderm Corp, Dallas, TX, USA) immediately after cleaning the residual drug from the skin surface at four *uptake* times (0.25, 0.5, 1 and 1.5 h), and at four *clearance* times (3, 6, 9 and 12 h) following the longest application time of 1.5 h. This study deviated from the FDA protocol in that it was not evident that steady state was achieved at or before the longest uptake time. Tape strips were extracted and tretinoin and its isotretinoin metabolite were individually quantified by HPLC with UV detection (3,4). Determinations below the limit of reliable quantification (4 ng ml^{-1} of extract or 3 ng cm^{-2} of skin) were reported as zero (4). Consistent with the FDA guidelines, retinoid on the first two tape strips was not included in the total; this material was assumed to represent additional, unabsorbed drug left after excess formulation was cleaned off the skin at the end of the application period. The AUC reported (3) was calculated for the time interval from zero to the longest clearance time (13.5 h after drug was applied and 12 h after it was removed) using the trapezoidal rule (19); A_{max} was the maximum observed drug amount during the uptake phase (3). For two of the products, the maximum amount of drug was observed during the clearance period in a few subjects (e.g., total retinoid was largest for post-application times of 4.5 h or larger for five and four subjects for products identified later as A and C, respectively). The AUC and A_{max} were determined for tretinoin, isotretinoin and the combination of tretinoin and isotretinoin in each subject (3). Statistical analyses were performed (3,4) for both AUC and A_{max} using the criterion that drug products are BE if the 90% confidence interval for the ratio of the population geometric averages is contained completely within the 0.8 to 1.25 interval using the 2, 1-sided *t*-test (20).

Using the numbers from the final report of the tretinoin study to the FDA (4) for each subject at each time, we were able to duplicate the original AUC and A_{max} results for tretinoin, isotretinoin and the combination of tretinoin and isotretinoin, except for a few minor differences. We then calculated values for four alternative metrics: (a) the sum of the drug amount from all four uptake times, (b) the sum of the drug amount from all four clearance times, (c) the sum of drug amount from all uptake and clearance times combined, and (d) the amount of drug at each time point assessed separately. The results are reported as the geometric mean ratio (R) of the selected metric (Z) for the test and reference drug formulations and the projected upper and lower 90% confidence intervals for the population mean ratio ($R_{90\%,upper}$ and $R_{90\%,lower}$, respectively) calculated as follows (20):

$$R = 10^{\bar{y}} \quad R_{90\%,upper} = 10^{(\bar{y}+\delta)} \quad R_{90\%,lower} = 10^{(\bar{y}-\delta)} \tag{1}$$

where $\bar{y}$ and δ are specified in Eqs. 2, 3 and 4:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \log\left(Z_{test}/Z_{ref}\right)_i \tag{2}$$

$$\delta = \frac{s \cdot t_{0.05,n-1}}{\sqrt{n}} \tag{3}$$

$$s = \sqrt{\sum_{i=1}^{n} \left[\log\left(Z_{test}/Z_{ref}\right)_i - \bar{y}\right]^2 \Big/ (n-1)} \tag{4}$$

In these equations, n is the total number of subjects and $t_{0.05,n-1}$ is the t-value of the two-tailed student distribution for a probability of 0.1 and $n-1$ degrees of freedom. Because ratios involving zeroes are problematic, the total number of subjects considered for a given metric did not include subjects in which the selected metric was zero for either Z_{test} or Z_{ref}.

A similar procedure was followed in calculating, for each product, the population averages of the ratio of the amount of drug in the SC after a clearance time $t-t_o$ (A_{t-t_o}), where t_o is the longest uptake time, to the amount of drug in the SC after no clearance time (A_{t_o}), denoted as W, and the upper and lower 90% confidence intervals for the population mean ratio ($W_{90\%,upper}$ and $W_{90\%,lower}$, respectively). That is,

$$W = 10^{\overline{w}} \quad W_{90\%,upper} = 10^{(\overline{w}+\delta_w)} \quad W_{90\%,lower} = 10^{(\overline{w}-\delta_w)} \tag{5}$$

where $\overline{w}$ and δ_w are specified in Eqs. 6, 7 and 8

$$\overline{w} = \frac{1}{n}\sum_{i=1}^{n} \log(A_{t-t_o}/A_{t_o})_i \tag{6}$$

$$\delta_w = \frac{s_w \cdot t_{0.05,n-1}}{\sqrt{n}} \tag{7}$$

$$s_w = \sqrt{\sum_{i=1}^{n} \left[\log(A_{t-t_o}/A_{t_o})_i - \overline{w}\right]^2 \Big/ (n-1)} \tag{8}$$

Thus, as specified by Eq. 6, the ratio $((A_{t-t_o}/A_{t_o})_j$ for a given product is calculated within a subject and then averaged across all subjects. Calculated values for the drug products can then be compared with theoretical predictions of W (see the Appendix) to derive estimates of the lag time for penetration through the SC.

RESULTS AND DISCUSSION

Table I lists the mean values of AUC and A_{max} calculated for tretinoin, isotretinoin and the combination of tretinoin and isotretinoin for each of the three topical products (4). As the ratio of the metabolite to the parent compound logically increased with time after the formulation was removed from the skin surface (3), it was decided that the sum of tretinoin and isotretinoin would be the most sensible quantity with which to represent the total amount of retinoid in the SC. Figure 2 shows this total retinoid level in the collected tapes as a function of time.

Table I. Mean Untransformed AUC and A_{max} Results Calculated for Tretinoin, Isotretinoin and Total Retinoid for the Three Topical Products (n=49)

Parameter	Product	Tretinoin	Isotretinoin	Total Retinoid
A_{max} (ng/cm^2)	A	115.4	55.1	168.2
	B	74.9	41.1	114.5
	C	118.2	58.2	173.4
AUC[a] (ng-h/cm^2)	A	437.3	306.6	743.9
	B	216.8	164.6	381.4
	C	449.4	327.2	776.6

[a] AUC was re-calculated using the trapezoidal rule.

From comparative clinical studies with the reference listed drug (product A), it is known that product C is equivalent (3,21,22), while product B is less efficacious (3,22). The AUC and A_{max} results for the total retinoid were consistent with these clinical observations. However, T_{max} in the tretinoin gel study was equated with the longest uptake time (see Fig. 1), and was therefore insensitive to the difference between products. Similar observations have been reported for T_{max} in other dermatopharmacokinetic studies (9,23).

The results of the comparisons of drug levels summed from all uptake times and from all clearance times are also consistent with the clinical results (Fig. 3). Interestingly, the difference between products B and A is greater during the clearance phase than in the uptake phase. It is possible that the clearance rate for product B was more rapid than from products A or C, suggesting that a difference in diffusion, rather than partitioning, is responsible for the observed difference between formulations B and A. Another possible explanation is that cleaning prior to the clearance phase was more effective for product B compared to products A and C. Whatever the mechanism, the results for product B illustrate that the relative performance of a drug formulation during uptake and clearance could be different. There is the potential for a drug product to be assessed as BE based on a metric that combines uptake and clearance, while assessment based on metrics that evaluate uptake and clearance separately would conclude that the products are different.

Figure 4 shows the ratios of the amounts of total retinoid in the SC from products B and C relative to product A for each time point. At the longer clearance times, there are number of determinations below the limit of reliable quantification (3 ng cm^{-2}). At each clearance time, we included only those subjects with measurable amounts of total retinoid [i.e., those with an amount of either tretinoin or isotretinoin above the limit of reliable quantification (LOQ)]. The actual number of subjects included at each time point is indicated in Fig. 4.

It is possible that dropping subjects with measurements below the LOQ may introduce bias in the ratios shown in Fig. 4. To assess the magnitude of this bias, we redid our calculations, using imputed values for the measurements below LOQ. The results of this simulation did not differ in any important way from those obtained by dropping the subjects with measurements below the LOQ. Details of this calculation are given in Appendix B.

When each time point was assessed separately (Fig. 4), the results were again generally consistent with the clinical findings. For measurements at a few times, however, the 90% confidence interval of the ratio did not fall entirely within or outside the 0.8 to 1.25 window. If the 0.8 to 1.25 window was the accepted criterion for BE, then these points would be classified as inconclusive (20). The largest variability was observed in data from some of the uptake times. In general, the variability of each time point was larger than the variability in the combined time results shown in Fig. 3. Thus, when several measurements from a single subject were averaged, the variability was noticeably reduced. This suggests that intrasubject variability, which is likely to be dominated by the experimental variability, is a substantial component of the overall variability.

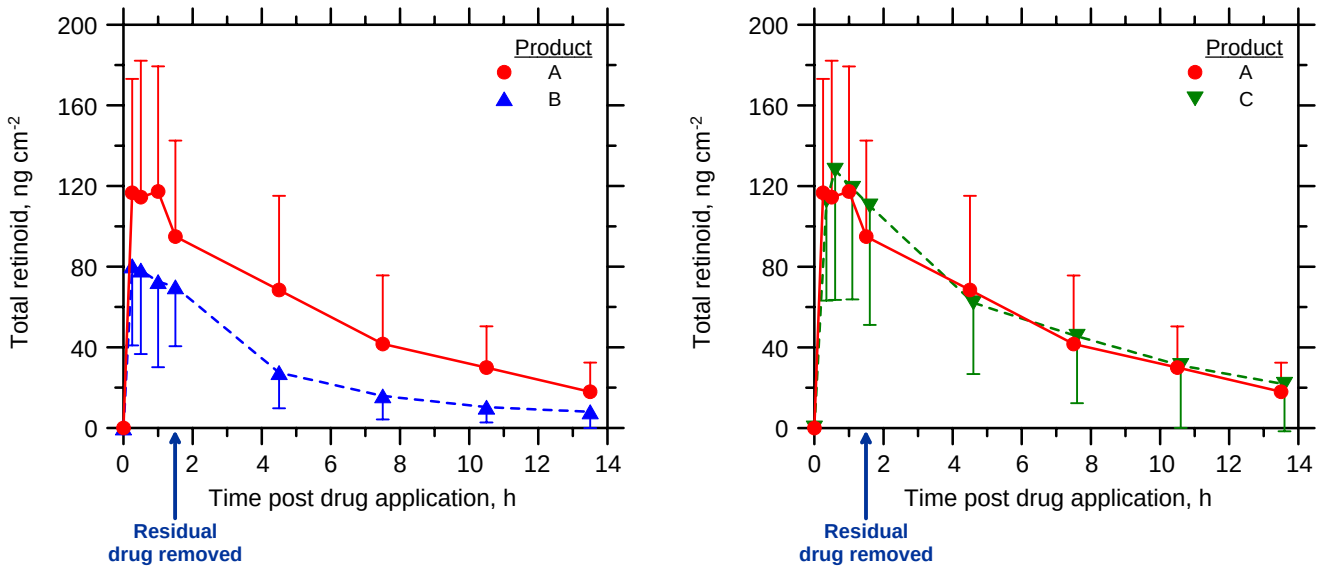

Fig. 2. The mean amounts of total retinoid (i.e., tretinoin plus isotretinoin) in human SC following application of three topical 0.025% tretinoin gel products are compared. Product A is the reference listed drug; Product B is known from clinical studies to be bio-inequivalent, while Product C has been shown to be bioequivalent. Mean of $n=49$ subjects; error bars designate 1 standard deviation (*upper half* for product A and *lower half* for products B and C). Data from Pershing *et al.* (3) have been re-drawn using results in the final report to the FDA (4) with amounts for product C shifted 0.1 h to facilitate viewing of the error bars for product A compared with product C.

Sources of variability that were probably causes for the inconclusive points in Fig. 4 include large inconsistency in the total amount of skin stripped, and exclusion of the drug from the first two tape strips in the drug total. With respect to skin stripping, there is evidence from an earlier experiment that the variation in the amount of skin collected is due to the tape-stripping procedure itself, rather than to differences between the subjects (20). In that study, the mass of SC was determined on nine 1.3-cm diameter D-Squame disks applied to four sites on each arm of three subjects (23). The average amount of SC collected per cm^2 was 199.2 μg with a

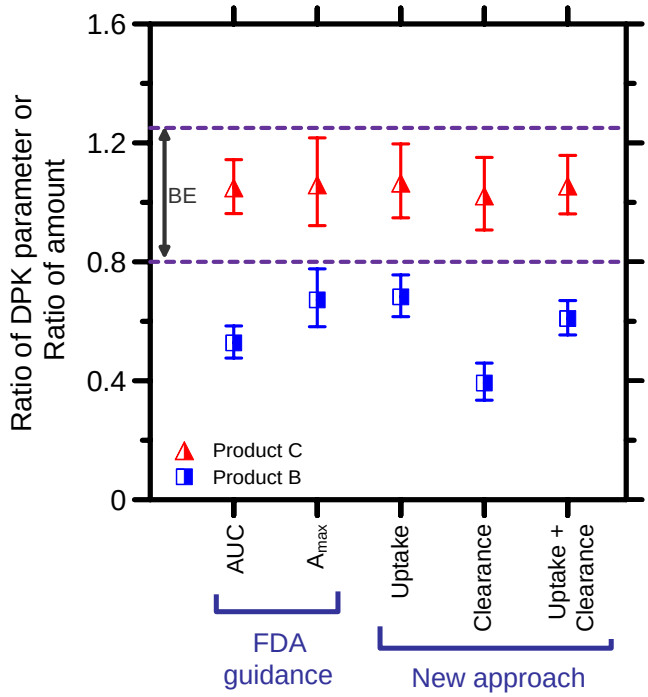

Fig. 3. Bioequivalence assessment of topical 0.025% tretinoin gel products B or C compared with product A using determinations of the total amount of retinoid in human SC to calculate the log-transformed ratios (mean and 90% confidence intervals) for: (a) the dermatopharmacokinetic parameters AUC and A_{max} according to the 1998 Food and Drug Administration guidance (2), or (b) the sum of the amounts in SC during uptake, clearance, or uptake plus clearance; 90% confidence interval of 0.8 to 1.25 for bioequivalence is shown for comparison.

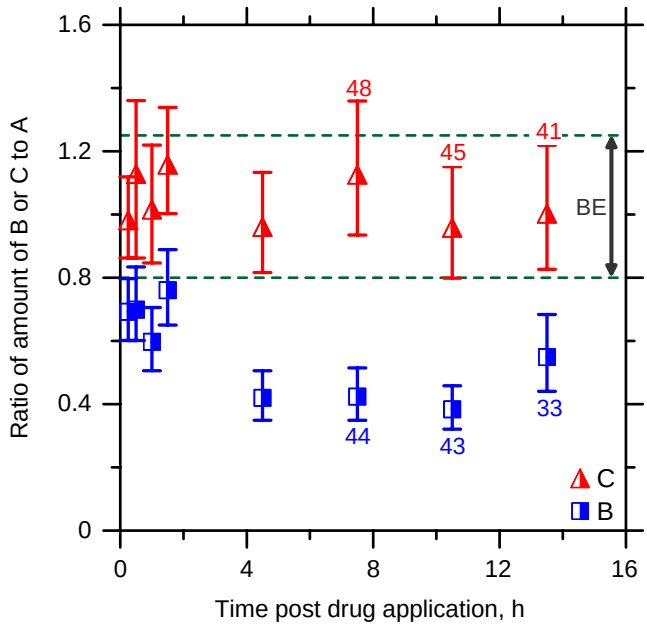

Fig. 4. Bioequivalence assessment of topical 0.025% tretinoin gel products B or C compared with product A using the total amount of retinoid in human SC to calculate the log-transformed ratio (mean and 90% confidence interval) at each time point. The integers indicate the number of subjects included in ratios calculated with fewer than 49 subjects because analytical insensitivity prevented calculation of a meaningful ratio for some subjects.

coefficient of variation of all sites in all subjects of 27.7%. This coefficient of variation is not much larger than those calculated separately for each subject (27.3%, 17.9%, 21.8%), suggesting that the total variation (intersubject plus intrasubject) is not much larger than the intrasubject variation alone. Also, these same measurements indicate that a significant fraction of the SC was not collected. If the thickness of the SC is estimated as 10 μm and the density of SC is 1 g cm^{-3}, then the mass of the entire SC is 1,000 μg cm^{-2}, which is five times larger than the average amount of SC harvested on the nine D-Squame disks. Although 12 D-Squame disks were applied in the tretinoin study instead of nine, it is still unlikely that the SC was completely removed.

Potentially, exclusion of drug from the first two tape strips was the largest contributor to experimental variability in the tretinoin gel study. This is because the concentration of drug is largest in the outermost tape strips (even if all drug present has been absorbed), and the amount of SC harvested on the first two disks is variable. After drug has had time to clear from the skin, a smaller fraction of all the drug in the SC will be present in the first two tape strips. Thus, variability arising from this factor should be greater for the uptake determinations, which is consistent with the observations shown in Fig. 4.

Although the amounts of drug in the SC changed over time (see Fig. 2), the ratios of the test to reference formulations were nearly constant, except between uptake and clearance for formulation B (see Fig. 4). For both products B and C, there is little difference in the ratios between the four individual uptake times or between the four individual clearance times. Also, the results from the individual times shown in Fig. 4 are consistent with the combined uptake results and combined clearance results shown in Fig. 3. Thus, BE during the clearance phase could have been conclusively assessed from any one of the four clearance determinations for product B, and from three of the four clearance determinations for product C. Consistent with the combined time results shown in Fig. 3, product B is nearly BE during uptake, but clearly not BE when assessed during clearance. Quite possibly, the BE assessment from any of the time points would have been conclusive if the experimental variability was reduced by protocol changes that ensured most of the drug in the SC is collected. These would include more complete harvesting of the SC, reliable cleaning of drug from the skin surface (so that unabsorbed drug is not left behind), and inclusion of the first two tape strips in the drug total. Notably, because drug concentration in the SC varies with position, knowing the amount of SC collected from each site (e.g., by weighing the tapes or quantifying the amount of protein on the tapes) is not sufficient to adjust the measured drug amounts for differences in the amount of SC collected when a significant amount of the drug has been left in the SC after tape stripping is completed.

For the tretinoin gel study, it is difficult to justify the eight different treatment times required for the AUC analysis when the same conclusions for BE could be obtained from measurements of drug levels at only one uptake and one clearance time. Furthermore, comparisons of drug levels at one uptake and one clearance time can discriminate differences in BE for clearance and uptake, which the AUC, as a time-weighted average, might miss.

If BE assessment could be reliably evaluated with as few as two treatment times, then the measurements at these times can be replicated at least twice in each subject (7). Although the number of treatment sites is the same whether measurements are made at two different time points or two measurements are made at the same time point, the latter strategy has two practical advantages. First, volunteers are required to make fewer return visits for sample collection. Second, the intrasubject variability can be considerably reduced because the amount of drug in the SC at each time point is represented by the average of two measurements. The latter advantage may be substantial, because, as discussed above, the intrasubject variability is likely to be a substantial component of the overall variability.

Time course determinations, at least during clearance, could be used to derive additional information, such as lag time for drug penetration through the SC, which is not available from the AUC and A_{max} metrics. This idea has been explored by mathematically modeling (as described in Appendix A) how drug levels in the SC should vary with time during the clearance phase for drug removed from the skin at different times. The results of the modeling are shown in Fig. 5. According to these calculations, the ratio (W) of the amount of drug in the SC during clearance to the amount in the SC at the time the drug was removed from the skin surface (t_o) is insensitive to (t_o) as long as t_o is greater than about 1.2 times the lag time (t_{lag}) for diffusion across the SC. Furthermore, the logarithm of W is linear and represented reasonably well by the following expression:

$$\log_{10} W \cong -(t - t_o)/(6t_{lag}) \qquad (9)$$

Significantly, 1.2 t_{lag} is about half the time required to reach steady state, which occurs at approximately 2.4 t_{lag} (24). Based

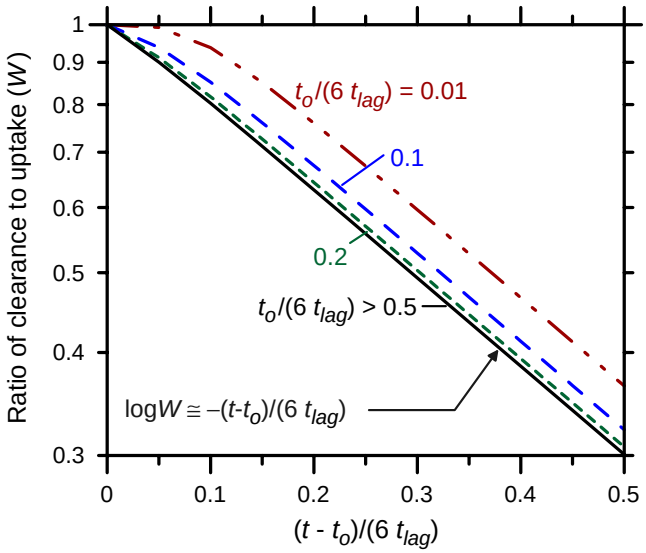

Fig. 5. Log of the calculated mass of drug in the SC at time t after cleaning the skin surface at time t_o post application, normalized by the mass of drug in the SC at t_o, plotted as a function of normalized time after drug removal, i.e., $(t - t_o)/(6\ t_{lag})$. Curves are shown for different values of the time at cleaning (t_o) normalized by t_{lag}. The logarithm of the drug mass ratio is linear with respect to $(t - t_o)/(6\ t_{lag})$ and essentially insensitive to the cleaning time [i.e., $t_o/(6\ t_{lag})$] as long as $t_o > \sim 1.2\ t_{lag}$.

on Eq. 9, the half-life for clearance occurs at $(t_{1/2}-t_o)\approx 1.8\ t_{lag}$. For $t_o \le 1.2\ t_{lag}$, the half-life occurs at somewhat longer times and W will be somewhat larger than estimates calculated from Eq. 9.

The results shown in Fig. 5 were derived assuming that skin layers beneath the SC provide little resistance to clearance and that the drug's diffusivity and partition coefficient in the SC do not change significantly over the time-course of the experiment. For chemicals with octanol–water partition coefficients greater than about 10^5, the skin layers beneath the SC can limit mass transfer through the skin (25), and the calculations shown in Fig. 5 need to be adjusted to include this effect. Such calculations are not currently available. However, based on prior work, we anticipate that the main effect will be that the lag time estimated from clearance using Eq. 9 could be as much as twice the lag time due to the SC alone (25). Since uptake into the SC will be largely unaffected by the mass transfer limits of the deeper skin layers for sample times less than about 2.4 times the lag time due only to the SC, it is possible that the uptake rates of highly lipophilic compounds might seem to be larger than expected from the lag time estimate determined from the clearance results. Detailed calculations are needed to confirm that this is the case.

The mean values and 90% confidence intervals of the logarithm of the clearance-to-uptake ratios (log$W$) are shown in Fig. 6 for the three tretinoin gel products plotted as a function of time since drug was removed. When examined this way, it is evident that the slopes (i.e., −0.054, −0.047, and −0.054 h^{-1} for products A, B and C, respectively) and hence, the clearance rates, are essentially the same. These values correspond to a lag time of 3–3.5 h, calculated by dividing the negative reciprocal of the slope [i.e., $-\Delta t/\Delta(\log W)$] by 6 following Eq. 9. Based upon this lag time, we would estimate that steady state might not be reached for almost 8 h (i.e., 2.4 × t_{lag}). This prediction disagrees with the results in Fig. 2, which on average shows a decline in the amount of drug even before the uptake period ended at 1.5 h. However, because tretinoin is highly lipophilic (the logarithm of its octanol-water partition coefficient is 6.30), it is possible that one-half of the estimated lag time of 3–3.5 h (i.e., 1.5–1.75 h or about 3 h to reach steady state) is more representative of the uptake period. Other explanations are that uptake from the gels slowed as the vehicle changed due to dehydration or depletion of drug in the film directly contacting the skin. A lag time of 3 to 3.5 h would correspond to a half-life for clearance of 5.4 to 6.3 h, which is similar to the values (5.6, 9.4 and 5.7 h for products A, B and C, respectively) calculated based on clearance rates by Pershing *et al.* (3) using a different procedure than that described here.

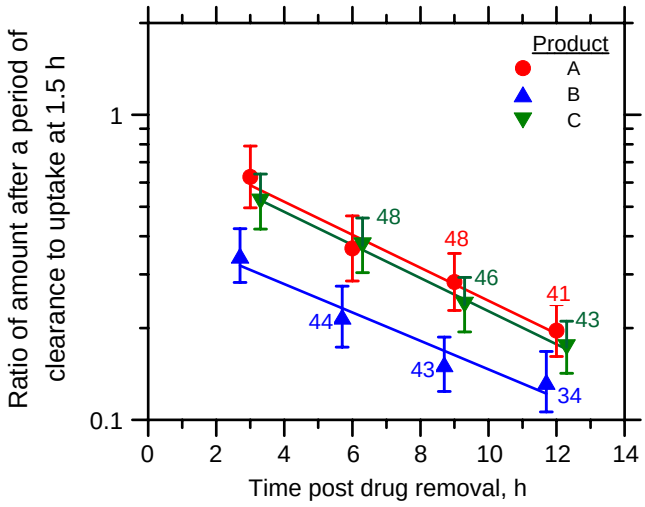

Fig. 6. Mean ratio (and 90% confidence intervals) of the amount of total retinoid in the SC at a specified time after drug is removed to the amount in the SC when drug is first removed plotted as a function of time post drug removal. To clarify viewing of the error bars, the data for products B and C have been plotted at the time of the measurement plus or minus 0.3 h, respectively. Best fit lines through the data are also shown; due to the time shift, the apparent W-intercepts for products B and C differ slightly from the actual values. While the number of subjects evaluated at each time point was 49, for some times the ratio was calculated using less than 49 due to insufficient analytical sensitivity preventing calculation of a meaningful ratio. When less than 49, the actual number of measurements is indicated by the integer adjacent to the data point.

It is interesting that the W-intercept at $(t-t_o)=0$ for all three products are less than the predicted value of one (i.e., 0.85, 0.44, and 0.76 for A, B and C, respectively). In part, this may be due to bias introduced by dropping subjects with measurements below the LOQ. Using imputed values for the measurements below LOQ, as described in Appendix B, the estimated amount of bias could increase the intercept values by as much as 10–15%, which still would be insufficient to adjust the W-intercept for product B to deviate insignificantly from the predicted intercept of 1. An additional factor is that the fraction of drug in the first two tape strips (which were discarded) is larger at the end of the uptake period than during clearance, and is highly variable.

If drug levels are determined at as few as one uptake time and one clearance time, guidelines for judicious selection of these times will be needed. Further mathematical simulations of drug uptake and clearance similar to those described in Appendix A would be useful for assessing the sensitivity of the results to sample time choices, from which criteria for suitable (or unsuitable) choices can be developed. We expect that these criteria would be based on estimates of a lag time derived from a pilot study, combined with the need for detectable drug levels. Also, the chosen clearance time needs to provide a reduced drug level compared to uptake. It would be sensible to choose uptake and clearance times based on the time between repeated applications specified for normal prescribed use. Related to this, bioavailability and BE of dermatological topical products might be more appropriately assessed by measuring the total amount of drug in the SC after repeated applications as prescribed for normal use and then again after an appropriate clearance time.

CONCLUSIONS

The re-analysis of the tretinoin DPK data supports the ability of a protocol using one uptake time and one clearance time to assess BE reliably with far fewer analyses. This streamlined scheme offers the significant advantage that drug is measured from two rather than eight treatment sites. As a result, there would normally be space on the treated forearms to at least duplicate measurements, which has significant

statistical advantages permitting a conclusive BE determination with fewer volunteers. The proposed scheme of assessing BE using measurements at one uptake and one clearance time would be most appropriate for dermatological formulations, such as anti-fungal preparations, which absorb into and penetrate through the SC slowly.

NOTATION

A_k Amount of drug in the SC at the start or after a period of clearance, i.e., $k=t_o$ or $k=(t-t_o)$, respectively
A_{max} Maximum observed drug amount per area of application during the uptake phase
AUC Area under the mass of drug collected on the tape strips per area of drug application *versus* time curve for the time interval from drug application to the longest clearance time
BE Bioequivalence
$\tilde{C}$ Drug concentration in the SC (C, given in units of mass of drug per volume of SC) normalized by $K \cdot C_v$, where K is the SC-vehicle partition coefficient and C_v is the drug concentration in the vehicle
$\langle \tilde{C} \rangle$ Normalized average concentration within the SC
DPK Dermatopharmacokinetic
L Thickness of the SC
m Counter for summation
n Number of volunteers included in calculation or counter for summation
R Geometric mean ratio of the selected metric for the test and reference drug formulations; $R_{90\%,upper}$ and $R_{90\%,lower}$ are the upper and lower 90% confidence intervals, respectively for the population mean ratio
s Population standard deviation for the log-transformed ratio of the selected metric
s_w Population standard deviation for the log-transformed ratio of the drug level in tape strips collected during clearance and at the end of drug uptake period
SC Stratum corneum
t Time since drug was applied
t_o Time when drug is cleaned from the skin surface
t_{lag} Lag time for diffusion through the SC
$\overline{w}$ Arithmetic mean of the log-transformed ratio of the drug level in tape strips collected during clearance and at the end of the uptake period
W Geometric mean ratio of the drug level in tape strips collected during clearance and at the end of drug uptake period; $W_{90\%,lower}$ and $W_{90\%,lower}$ are the upper and lower 90% confidence intervals, respectively for the population mean ratio
x Position within the SC, $x=0$ at the surface of the SC
$\overline{y}$ Arithmetic mean of the log-transformed ratio of the selected metric for the test and reference drug formulations
Z Selected metric for assessing BE

Greek

δ Projected 90% confidence interval for the log-transformed ratio of the selected metric
δ_w Projected 90% confidence interval for the log-transformed ratio of the drug level in tape strips collected during clearance and at the end of drug uptake period
ξ Position within the SC normalized with respect to the thickness of the SC L
τ Time normalized by $(6t_{lag})$; τ_o is t_o normalized by $(6t_{lag})$

ACKNOWLEDGEMENTS

We acknowledge support from the FDA. No endorsement from the FDA or US government should be inferred.

APPENDIX

Appendix A

The concentration of drug at a given location within the SC will vary with time relative to the lag time (t_{lag}) during the uptake and clearance periods as described by Eqs. 10 and 11, respectively:

$$\tilde{C} = 1 - \xi - 2\sum_{n=1}^{\infty} \frac{\sin(n\pi\xi)}{n\pi} \exp\left[-n^2\pi^2\tau\right] \qquad \text{for } t \le t_o \tag{10}$$

$$\tilde{C} = 2\sum_{n=0}^{\infty}\left(\frac{1}{\lambda_n^2} - 2\sum_{m=1}^{\infty}\frac{\exp\left[-m^2\pi^2\tau_o\right]}{m^2\pi^2-\lambda_n^2}\right)\exp\left[-\lambda_n^2(\tau-\tau_o)\right]\cos(\lambda_n\xi), \tag{11}$$

$$\lambda_n = \frac{2n+1}{2}\pi \text{ for } t > t_o$$

where t_o is the time at which drug is removed from the skin surface. In these equations $\tilde{C}$ is the concentration in the SC (C, with units of drug mass per volume of SC) normalized by $K \cdot C_v$, where K is the SC-vehicle partition coefficient and C_v is the drug concentration in the vehicle. The relative position (ξ) is the location x within the SC measured from the skin surface normalized by the SC thickness L. The dimensionless time (τ) is defined as the time since the drug was applied to the skin normalized by $6 \cdot t_{lag}$ [i.e., $\tau=(t/(6 \cdot t_{lag})]$; τ_o is the dimensionless time at which the drug is removed from the skin surface $[t_o/(6 \cdot t_{lag})]$, initiating the clearance phase.

Equations 10 and 11 were developed assuming that: (1) drug permeation in the SC can be treated as Fickian diffusion with a constant diffusion coefficient in a pseudo-homogeneous membrane of thickness L, (2) constant drug concentration in the vehicle applied to the skin surface, and (3) sink conditions (i.e., drug concentration is zero) at the inside boundary of the SC. According to Eq. 10, at a given position within the SC, the concentration will increase in time until steady state is established, which will occur when τ is approximately 0.4 (i.e., at $t \approx 2.4 \times t_{lag}$). After this time, the concentration profile will be linear (i.e., $\tilde{C} = 1 - x/L$) and will not vary in time unless the drug concentration in the vehicle depletes or the drug is cleaned from the skin surface.

The predicted average concentration of drug in the SC at a given time can be derived by integrating the concentration expressions represented by Eqs. 10 and 11 over the SC

thickness. The resulting expressions for the average normalized concentration $\langle \tilde{C} \rangle$ during uptake and clearance are given by Eqs. 12 and 13, respectively.

$$\langle \tilde{C} \rangle = \frac{1}{2} - \frac{4}{\pi^2} \sum_{n=0}^{\infty} \frac{\exp\left[-(2n+1)^2 \pi^2 \tau\right]}{(2n+1)^2} \text{ for } t \le t_o \tag{12}$$

$$\langle \tilde{C} \rangle = 2 \sum_{n=0}^{\infty} \frac{(-1)^n \exp\left[-\lambda_n^2 (\tau - \tau_o)\right]}{\lambda_n} \left(\frac{1}{\lambda_n^2} - 2 \sum_{m=1}^{\infty} \frac{\exp\left[-m^2 \pi^2 \tau\right]}{m^2 \pi^2 - \lambda_n^2} \right) \quad \lambda_n = \frac{2n+1}{2} \pi \text{ for } t > t_o \tag{13}$$

The ratio of the amount of drug in the SC normalized to the amount in the SC when the drug was cleaned from the skin surface is denoted by the symbol *W*. The ratio of the drug amounts is also the ratio of the average concentrations. Thus, the curves in Fig. 5 showing the variation in *W* with clearance time ($t - t_o$) were derived by dividing the average drug concentration in the SC (i.e., $\langle \tilde{C} \rangle$ calculated from Eq. 13 at a given time (t) that is after drug was removed from the SC surface), by the average drug concentration in the SC at the time drug was removed (t_o) [(i.e., $\langle \tilde{C} \rangle$ calculated from Eq. 12 for $\tau = t_o/(6\ t_{lag})$].

Appendix B

For times at which some measurements were below the LOQ (equal to 3 ng cm^{-2}), the remaining measurements, on the log scale, can be considered to come from a truncated Gaussian distribution, with a truncation point of log (3 ng cm^{-2}). In order to impute values for the measurements below the LOQ, we needed to estimate the mean and standard deviation of the full distribution. We chose estimates so that the mean of the truncated distribution matched the sample mean of the measurements above the LOQ, and so that the area under the tail of the curve to the left of log 3 matched the sample proportion of subjects whose measurements were below the LOQ. For each subject whose measurement was below the LOQ, we then generated a random value from the tail of the curve to the left of log 3.

After values were imputed for all subjects whose original measurements were below the LOQ, we computed means and confidence intervals as described above. Because the imputed values are random, the means and confidence intervals will differ when the process is repeated. To determine how large this variation would be, we repeated the process several times. The results were similar across replications. For the ratios of the total amount of retinoid of product B to product A and product C to product A (Fig. 4), the results at t=13.5 h, for which the proportion of measurements below LOQ was greatest, differed by 10–20% from the values obtained when the values below LOQ were dropped. Smaller differences were observed for other times. For the ratios of the amount of total retinoid in the SC at a specified time after the drug is removed compared with the amount in the SC when the drug is first removed (Fig. 6), the geometric mean ratios decreased by about 30% at 12 h after removal, for which the proportion of measurements below LOQ was greatest. Smaller decreases were observed for earlier times.

REFERENCES

1. V. P. Shah, G. L. Flynn, A. Yacobi, H. I. Maibach, C. Bon, N. M. Fleischer, T. J. Franz, S. A. Kaplan, J. Kawamoto, L. J. Lesko, J.-P. Marty, L. K. Pershing, H. Schaefer, J. A. Sequeira, S. P. Shrivastave, J. Wilkins, and R. L. Williams. Bioequivalence of topical dermatological dosage forms—methods of evaluation of bioequivalence. *Pharm. Res* **15**:167–171 (1998).
2. US FDA. Guidance for Industry: topical dermatological drug product NDAs and ANDAs-*in vivo* bioavilability, bioequivalence, *in vitro* release, and associated studies. Draft Guidance, June 1998, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), 1998.
3. L. K. Pershing, J. L. Nelson, J. L. Corlett, S. P. Shrivastave, D. B. Hare, and V. P. Shah. Assessment of dermatopharmacokinetic approach in the bioequivalence determination of topical tretinoin gel products. *J Am Acad Dermatol* **48**:740–751 (2003).
4. L. K. Pershing. Final Report to Food and Drug Administration (FDA): Dermatopharmacokinetic Bioequivalence Study on Three Tretinoin Gel, 0.025% Products, University of Utah, Salt Lake City, 2000.
5. US FDA. Guidance for industry on special protocol assessment; availability. *Fed. Reg*. **67**:35122 (2002).
6. A. L. Bunge, B. N'Dri-Stempfer, W. C. Navidi, and R. H. Guy. Dermatopharmacokinetics: Improvement of Methodology for Assessing Bioequivalence of Topical Dermatological Drug Products, Revised Final Report, Award No. D3921303, Submitted to Department of Health and Human Services, Food and Drug Administration, Colorado School of Mines, Golden, CO, September 2, 2006.
7. A. L. Bunge, B. N'Dri-Stempfer, W. C. Navidi, and R. H. Guy. Therapeutic Equivalence of Topical Products, Final Report, Award No. 223-04-3004, Submitted to Department of Health and Human Services, Food and Drug Administration, Colorado School of Mines, Golden, CO, January 30, 2007.
8. L. K. Pershing. Bioequivalence Assessment of three 0.025% tretinoin gel products: Dermatopharmacokinetic *vs*. Clinical Trial Methods. *Advisory Committee for Pharmaceutical Sciences Meeting, Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA)*, Rockville, MD, 2001.
9. T. J. Franz. Study #1, Avita Gel 0.025% *vs* Retin-A Gel 0.025%, *Advisory Committee for Pharmaceutical Sciences Meeting, Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA)*, Rockville, MD, 2001.
10. D. P. Conner. Differences in DPK Methods. *Advisory Committee for Pharmaceutical Sciences Meeting, Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA)*, Rockville, MD, 2001. http://www.fda.gov/ohrms/dockets/ac/01/slides/3804s2_05_conner/index.htm, pp 71–75 in transcript.
11. I. Alberti, Y. Kalia, A. Naik, J.-D. Bonny, and R. H. Guy. *In vivo* assessment of enhanced topical delivery of terbinafine to human stratum corneum. *J Control. Release*. **71**:319–327 (2001).
12. I. Alberti, Y. N. Kalia, A. Naik, J. D. Bonny, and R. H. Guy. Assessment and prediction of the cutaneous bioavailability of topical terbinafine *in vivo*. *Pharm. Res*. **18**:1472–1475 (2001).
13. I. Alberti, Y. N. Kalia, A. Naik, J.-D. Bonny, and R. H. Guy. Effect of ethanol and isopropyl myristate on the availability of topical terbinafine in human stratum corneum, *in vivo*. *Int. J. Pharm* **219**:11–19 (2001).
14. F. Pirot, Y. N. Kalia, A. L. Stinchcomb, G. Keating, A. Bunge, and R. H. Guy. Characterization of the permeability barrier of human skin *in vivo*. *Proc. Nat. Acad. Sci. USA*. **94**:1562–1567 (1997).
15. M. B. Reddy, A. L. Stinchcomb, R. H. Guy, and A. L. Bunge. Determining dermal absorption parameters *in vivo* from tape stripping data. *Pharm. Res* **19**:292–297 (2002).
16. R. J. Scheuplein and I. H. Blank. Permeability of the skin. *Physiol. Rev* **51**:702–747 (1971).
17. M. S. Roberts, and K. A. Walters. The relationship between structure and barrier function of skin. In M. S. Roberts, and K. A. Walters (eds.), *Dermal Absorption and Toxicity Assessment*, Vol 91., Drugs and the Pharmaceutical Sciences, Marcel Dekker, New York, 1998, pp. 1–42.

18. J. Hadgraft. Skin, the final frontier. *Int. J. Pharm* **224**:1–18 (2001).
19. E. Kreyszig. *Advanced Engineering Mathematics*, 8th Edition, Wiley, New York, NY, 1999.
20. D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm* **15**:657–680 (1987).
21. US FDA. Avita Gel (Tretinoin Gel), NDA Application No.: 20-400. http://www.fda.gov/cder/foi/nda/98/20400.htm, 1998.
22. US FDA. Approved drug products with therapeutic equivalence evaluations (Electronic Orange Book), US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Office of Pharmaceutical Science, Office of Generic Drugs. http://www.fda.gov/cder/orange/obannual.pdf, 2007.
23. L. K. Pershing, J. L. Corlett, and J. L. Nelson. Comparison of dermatopharmacokinetic *vs*. clinical efficacy methods for bioequivalence assessment of miconazole nitrate vaginal cream, 2% in humans. *Pharm. Res* **19**:270–277 (2002).
24. A. L. Bunge, R. L. Cleek, and B. E. Vecchia. A new method for estimating dermal absorption from chemical exposure. 3. Compared with steady-state methods for prediction and data analysis. *Pharm. Res* **12**:972–982 (1995).
25. R. L. Cleek and A. L. Bunge. A new method for estimating dermal absorption from chemical exposure. 1. General approach. *Pharm. Res* **10**:497–506 (1993).