

## Research paper

# The determination of stratum corneum thickness An alternative approach

Lisa M. Russell, Sandra Wiedersberg, M. Begoña Delgado-Charro \*

*Department of Pharmacy & Pharmacology, University of Bath, Bath, UK*

Received 30 November 2007; accepted in revised form 6 February 2008

Available online 14 February 2008

---

**Abstract**

The individual thickness of the stratum corneum is required to normalise drug permeation profiles in dermato-pharmacokinetic studies. The thickness is often estimated using tape-stripping combined with transepidermal water loss measurements. A linear transformation of Fick's first law is used to relate the progressively thinner barrier with the corresponding increase in transepidermal water loss and to estimate the thickness by linear regression. However, the data from an important subset of subjects are poorly fitted to this linear model. This is typically due to the removal of loose outer layers of stratum corneum, which do not contribute significantly to barrier function. This work proposes two alternative non-linear models. All three models were used to fit data from 31 *in vivo* tape-stripping experiments and their outcomes and goodness-of-fit compared. The results suggest that the linear model may overestimate the stratum corneum thickness and is open to subjectivity regarding the selection of data points to be fitted. The non-linear models satisfactorily fitted all the data, including all data points. No significant differences were found between the thicknesses derived from the two non-linear models. However, the analysis of the goodness-of-fit of the models to the data suggests a preference for a baseline-corrected approach. © 2008 Elsevier B.V. All rights reserved.

**Keywords:** Tape-stripping; Transepidermal water loss (TEWL); Stratum corneum; Thickness; Dermato-pharmacokinetics

---

**1. Introduction**

The individual thickness of the stratum corneum (SC) is required in order to normalise drug-penetration profiles from different volunteers during dermato-pharmacokinetic studies. Bioavailability describes the rate and extent to which a drug, in an active form, reaches its target site. In the case of topical formulations seeking a local effect, the target site is the skin. By measuring the 'rate' and 'extent' of drug penetration into the skin, the so-called dermato-pharmacokinetic parameters may be derived. These parameters and the drug-penetration profiles provide information with which to assess bioavailability and demonstrate bio-

equivalence between different formulations. It has been suggested that since the stratum corneum, SC, is the principal barrier to drug absorption, the kinetic data of drug passage through this layer can be related to bioavailability in the target tissue [1–3].

The tape-stripping technique has attracted considerable interest from regulatory bodies, such as the US Food and Drug Administration, as a means to determine dermato-pharmacokinetic parameters, and ultimately test bioequivalence between formulations [2,3]. Tape-stripping involves the sequential removal of layers of the stratum corneum using adhesive tapes. It can be performed with minimum discomfort and relative ease *in vivo*. Typically, formulations are applied to the accessible volar forearm, which has a SC thickness shown to be consistently between 10 and 20  $\mu\text{m}$  for most volunteers [4]. Restoration of 80–90% of SC's barrier thickness is complete within 3 days [5] and transepidermal water loss (TEWL) values return to normal after 8 [6] to 11 [7] days.

---

\* Corresponding author. Department of Pharmacy & Pharmacology, University of Bath, Claverton Down, Bath BA2 7AY, UK. Tel.: +44 (0) 1225 383969; fax: +44 (0) 1225 386114.

E-mail address: [B.Delgado-Charro@bath.ac.uk](mailto:B.Delgado-Charro@bath.ac.uk) (M. Begoña Delgado-Charro).

In 1998, the FDA issued guidance on the tape-stripping technique for evaluating drug penetration through the SC [2]. It was later withdrawn when the comparison of three products by two laboratories resulted in contradictory outcomes. Since then, some problems with the original guidance have been identified: (i) discarding two tape strips has lost favour, as drug in these outer layers would probably become available for absorption eventually [1,8], (ii) in an attempt to relate the drug concentration to a particular depth within the SC, and in order to normalise the data derived from different volunteers [1,8–12], the total thickness of the SC is now measured [13,14].

An exact measurement of the SC thickness ( $H$ ) is impossible in a completely non-invasive way. However, a version of Fick's first law allows relating the easily measurable TEWL value from an intact SC to several SC parameters [15,16]:

$$\text{TEWL}_0 = \frac{D \cdot K \cdot \Delta C}{H} \quad (1)$$

where  $\text{TEWL}_0$  is the baseline TEWL,  $D$  is the diffusion coefficient of water in the SC;  $K$  is the SC-viable tissue partition coefficient of water;  $\Delta C$  is the water concentration gradient, and  $H$  is the thickness of the SC. This equation assumes that the SC is the main barrier to water loss, and that it provides a homogeneous barrier to water diffusion, justified by Kalia et al. [16].

Therefore, once a certain cumulative thickness of the SC,  $x$ , is removed by tape-stripping, the TEWL will increase as follows:

$$\text{TEWL}_x = \frac{D \cdot K \cdot \Delta C}{H - x} \quad (2)$$

$x$  is calculated from area of strip; mass of SC on tape; and the density of the SC ( $\sim 1 \text{ g/cm}^3$  [17]). A reciprocal transformation of Eq. (2) provides a linear expression which relates  $1/\text{TEWL}_x$  to  $x$ :

$$\frac{1}{\text{TEWL}_x} = \frac{H}{D \cdot K \cdot \Delta C} - \left( \frac{1}{D \cdot K \cdot \Delta C} \right) x \quad (3)$$

Experimental data expressed according to Eq. (3) should yield a single straight line that can be analyzed by simple linear regression. Extrapolation of this linear regression to  $1/\text{TEWL} = 0$  yields  $x = H$ , allowing  $H$  to be found.

However, data from a significant subset of subjects are not satisfactorily fitted by this linear model (LM). In these cases, representation of the data according to Eq. (3) yields an initial plateau followed by a declining straight line. It could be hypothesised that this initial plateau corresponds to the removal of the stratum corneum disjunctum; the looser outer stratum corneum layer [18–21] which does not provide a significant contribution to the barrier to water loss. It follows that its removal would contribute a significant mass change without a corresponding increase in TEWL. Thus, the linear model (Eq. (3)), which assumes that the SC barrier properties are homogeneous, cannot reflect experimental findings accurately.

Another disadvantage of the LM results from the mathematical procedure: upon inversion of the TEWL values into  $1/\text{TEWL}$ , there is also an inversion of the errors associated with each data point, which can skew the goodness-of-fit of the linear regression [22].

The aim of this work was to develop an improved model which avoids the problems mentioned above. We propose first the baseline-corrected non-linear (BC-NL) model (Eq. (4)) which (a) fits directly TEWL versus cumulative thickness data, thus avoiding the errors associated with the reciprocal transformation [22], and (b) incorporates a baseline parameter,  $B$ , to reflect the initial plateau during which TEWL remains constant despite the removal of some layers of SC.

$$\text{TEWL}_x = B + \frac{D \cdot K \cdot \Delta C}{H - x} \quad (4)$$

The second model proposed is the "Simple Non-Linear" (S-NL) model (Eq. (2)) which also fits the data directly to Fick's first law equation (Eq. (2)) but does not incorporate a baseline.

To decide on the best approach for estimating the stratum corneum thickness, we examined data sets collected from different volunteers on different occasions. We evaluated the performance of the three models and compared their outcomes. Each set of data (TEWL versus cumulative thickness removed) has been analyzed separately using each of the three models: LM, S-NL and BC-NL. The SC thickness ( $H$ ) estimated for each volunteer by each model is compared, along with a discussion of the goodness-of-fit, and the relative merits and weaknesses of each method.

## 2. Materials and methods

### 2.1. Subjects

Eighteen healthy volunteers (3 male, 15 female, age range 22–43 years), with no history of dermatological disease, participated in the study. Ethical approval was granted by Salisbury Local Research Ethics Committee, the Declaration of Helsinki protocols were followed, and written informed consent was obtained from all the volunteers. A total of 31 sites were examined over a period of 2 years. Participants refrained from using any topical products on the test area on the day of the experiments. As TEWL measurements may be affected by sweating or the changes to the relative humidity or temperature of the laboratory, the volunteers rested in the room for at least 15 min prior to taking first TEWL measurement; the experiment was completed in less than 1 h; and repeated initial TEWL measurements were taken until stabilised. The mean ( $\pm$ SD) temperature and relative humidity in the study room were  $21.9 \pm 1.5^\circ\text{C}$  and  $38.2 \pm 9.4\%$ , respectively. Subjects were given code numbers (1–18) and nine subjects participated only once in the study. Six subjects participated twice; subjects 9 and 14 participated thrice,

and subject 8 participated four times. Repeated participation was coded with a letter, for example, subjects 1a and 1b for the first and second participation of subject 1. Repeat participation using the same arm was delayed by at least 1 month, which is sufficient for barrier regeneration [5–7].

## 2.2. Tape-stripping procedure

Two tapes, with pressure applied using a roller, were taken and discarded. These pre-tapes are also taken before all dermato-pharmacokinetic experiments in our laboratories to remove exogenous substances and prepare the skin surface in a systematic way. A plastic template was applied to delimit a constant area to be stripped. An initial TEWL measurement was taken with a closed-chamber evaporimeter (Biox Aquaflux AF102, Biox Systems Ltd., London, UK; measurement range 0–100 g m<sup>2</sup> h<sup>−1</sup>; resolution ±0.05 g m<sup>2</sup> h<sup>−1</sup>; the probe was applied for a minimum 60 s, the TEWL was obtained as mean of 10 successive measurements having a CV < 1%).

A preweighed (Sartorius Microbalance SE-2F, precision 0.1 µg; Sartorius AG, Goettingen, Germany) piece of Scotch Book tape 845 (3M, St. Paul, MN) was placed over the template and adhesion to the skin was assured systematically with a set number of rolls with a roller. The tape was removed swiftly and another TEWL measurement taken. The sequence was repeated until the TEWL value was 3–4 times its initial value (usually 60–80 g m<sup>−2</sup> h<sup>−1</sup>). All tapes were reweighed following the completion of the stripping procedure, with prior removal of any hairs as necessary. Static electricity was discharged from the tapes prior to weighing, using an Eltex R50 discharging bar, with an Eltex ES50 power supply (Eltex Elektrostatik GmbH, Weil am Rhein, Germany). Three to five blank tapes were weighed at the same time as the tapes used for the tape-stripping experiments. Any change in the mass of these blank tapes was used to correct the calculated mass of SC for any variations in weight due to environmental or other conditions. For a detailed review on the tape-stripping technique the reader is referred to Herkenne et al. [23].

## 2.3. Linear model (LM)

Linear regressions were performed on each data set of 1/TEWL versus cumulative SC thickness removed, using GraphPad Prism® (version 4.00 for Windows, GraphPad Software, San Diego, CA). All the data points were fitted into the regression unless stated otherwise. All the slopes were significantly non-zero ( $p < 0.0001$ ) which confirmed a statistically significant relationship between 1/TEWL and cumulative SC thickness removed.

## 2.4. Non-linear (S-NL and BC-NL) models

All data sets were evaluated separately using WinNonLin® software (Version 5.1, Pharsight Corporation, Moun-

tain View, CA) using an ASCII user-defined model written for models S-NL and BC-NL. In all the cases, no data points were excluded; uniform weighting was applied; no bounds were used; and the initial parameter estimates were  $H = \text{cumulative thickness removed} + 1$ ;  $(D \cdot K \cdot \Delta C) = 30$ ;  $B = \text{initial TEWL value before stripping}$  (for BC-NL model only). Iterations continued until the relative change in weighted sum of squares was <0.000100. The model derived parameter estimates were used as the new ‘initial’ values in the model, and the model was rerun until there was no change in the parameter output. All parameter estimates are those from the final iteration of the model, with the smallest resultant residual sum of squares; however, usually, no change was seen after the first run, suggesting that the modelling was stable. All statistical tests for comparisons between parameter outputs for different data sets were done using GraphPad Prism®.

## 2.5. Evaluation of goodness-of-fit

The statistical evaluation of ‘goodness-of-fit’ of a model is not a trivial matter, especially in the case of non-linear models [22,24]. Obviously, a first step involves a visual graphical assessment of how well the model fits the experimental values. Models are usually evaluated for their *accuracy*; but this was not possible as there are no independent methods to obtain SC thickness that can be considered the “gold standard”. Thus, in an effort to evaluate objectively all three models, a series of statistical tools were considered concurrently.

First of all, the *precision* of the parameter estimates is evaluated using the coefficient of variation (CV, %), which relates the parameter’s standard error of the regression (SER) to its estimate ( $p$ ):

$$CV(\%) = \frac{SER}{p} \cdot 100 \quad (5)$$

The SER is an absolute measure whereas CV is a relative error that can be used to compare the three models. Normally, the model resulting in the smaller CV should be preferred.

Next, the two non-linear models were compared; it would be expected that the BC-LN would fit the data better than the S-NL model, simply because it has one extra parameter. Thus, two statistical tools, the Akaike Information Criterion and an  $F$ -test that take into account the difference in the number of parameters, were used to compare objectively the two non-linear models.

The Akaike Information Criterion (AIC) for the S-NL and BC-NL models was calculated for each of the 31 data sets separately as follows [22]:

$$AIC = N_{\text{obs}} \cdot \ln(WRSS) + 2 \cdot N_{\text{par}} \quad (6)$$

where  $N_{\text{obs}}$  is the number of data points of each data set; WRSS is the weighted residual sum of squares (provided by WinNonLin®);  $N_{\text{par}}$  is the number of parameters which were 2 (“DKAC” and “H”) and 3 (“DKAC”, “B” and

“ $H$ ”) for the S-NL and the BC-NL models, respectively. The absolute value of the AIC for a single model on its own is meaningless; hence the AIC is always used to compare several models. Briefly, the model with the lowest AIC is more likely to be correct. Furthermore, the probability that the BC-NL model is correct, rather than the S-NL model, for a given data set can be calculated via the difference in the AIC scores,  $\Delta AIC = AIC_{BC-NL} - AIC_{S-NL}$ , as follows [24]:

$$\text{Probability} = \frac{e^{-0.5 \cdot \Delta AIC}}{1 + e^{-0.5 \cdot \Delta AIC}} \quad (7)$$

The S-NL and BC-NL models can be considered as nested models; that is, the BC-NL model can be considered an extension of the other. Crucially, both models would be identical if a single parameter,  $B$ , is set to zero. Under these conditions, an  $F$ -test [22,25] can examine the effect of the additional parameter on the WRSS. The  $F^*$  value was calculated separately for each data set:

$$F^* = \frac{df_2 \cdot |WRSS_1 - WRSS_2|}{|df_1 - df_2| \cdot WRSS_2} \quad (8)$$

where  $df_1$  and  $df_2$  are the degrees of freedom for the S-NL and BC-NL models, respectively, and  $WRSS_1$  and  $WRSS_2$  the corresponding weighted residual sum of squares (both provided by WinNonlin®). The  $F^*$  was then compared to critical values ( $F_{table}$ ) taken from  $F$  tables [26] for a two-tailed test, with a  $p$  value of 0.05, column value =  $|df_1 - df_2|$ ; and row value =  $df_2$ . If  $F^*$  is greater than

$F_{table}$ , it can be concluded that the full model is better than the reduced model.

The AIC and the  $F$ -test consider which model, S-NL or BC-NL, may be more appropriate for each data set individually. However, an overall preferred model may be suggested based on the preferred model selected for all 31 cases.

### 3. Results and discussion

TEWL was measured before and after each tape strip in 31 experiments. The average baseline (before stripping) and final (after stripping) TEWL values, along with their standard deviations, were  $10.12 \pm 2.65 \text{ g m}^{-2} \text{ h}^{-1}$  and  $61.48 \pm 18.78 \text{ g m}^{-2} \text{ h}^{-1}$ , respectively. This increase in TEWL is due to the barrier disruption and is in agreement with previous work [13,14,16,27].

The first method considered was the linear model, LM, which has been widely used [1,8,10–13,28] since its introduction [14,16]. The linear transformation of Fick's first law (Eq. (3)) predicts a straight line when the  $1/\text{TEWL}$  values are plotted versus the cumulative thickness of the stratum corneum removed ( $x$ ). Fig. 1a shows example plots for two volunteers: a single straight line is observed for both subjects; the linear regression fits the experimental data well ( $R^2 = 0.99$ ) and the thickness of the stratum corneum is easily extrapolated from the regression line.

However, there were several sets of data which were not satisfactorily described by the linear model. Fig. 2a illus-

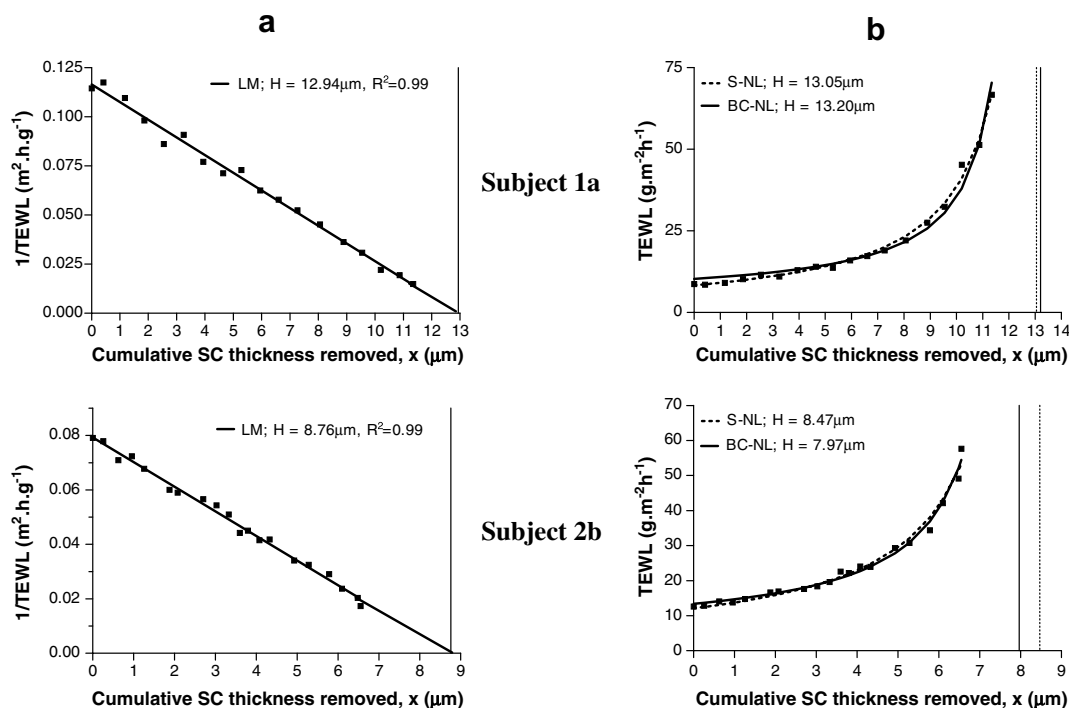


Fig. 1. Estimation of stratum corneum thickness ( $H$ ) by the linear model (panel a) and by the simple (S-NL) and baseline-corrected (BC-NL) non-linear models (panel b) for subjects 1a and 2b. [■] Experimental data points. Panel a: [—] prediction by LM. Panel b: [---] prediction by BC-NL; [...] prediction by S-NL model. The vertical lines mark the value of  $H$  estimated by the LM (panel a, solid line), the S-NL (panel b, dotted line), and the BC-NL (panel b, solid line).

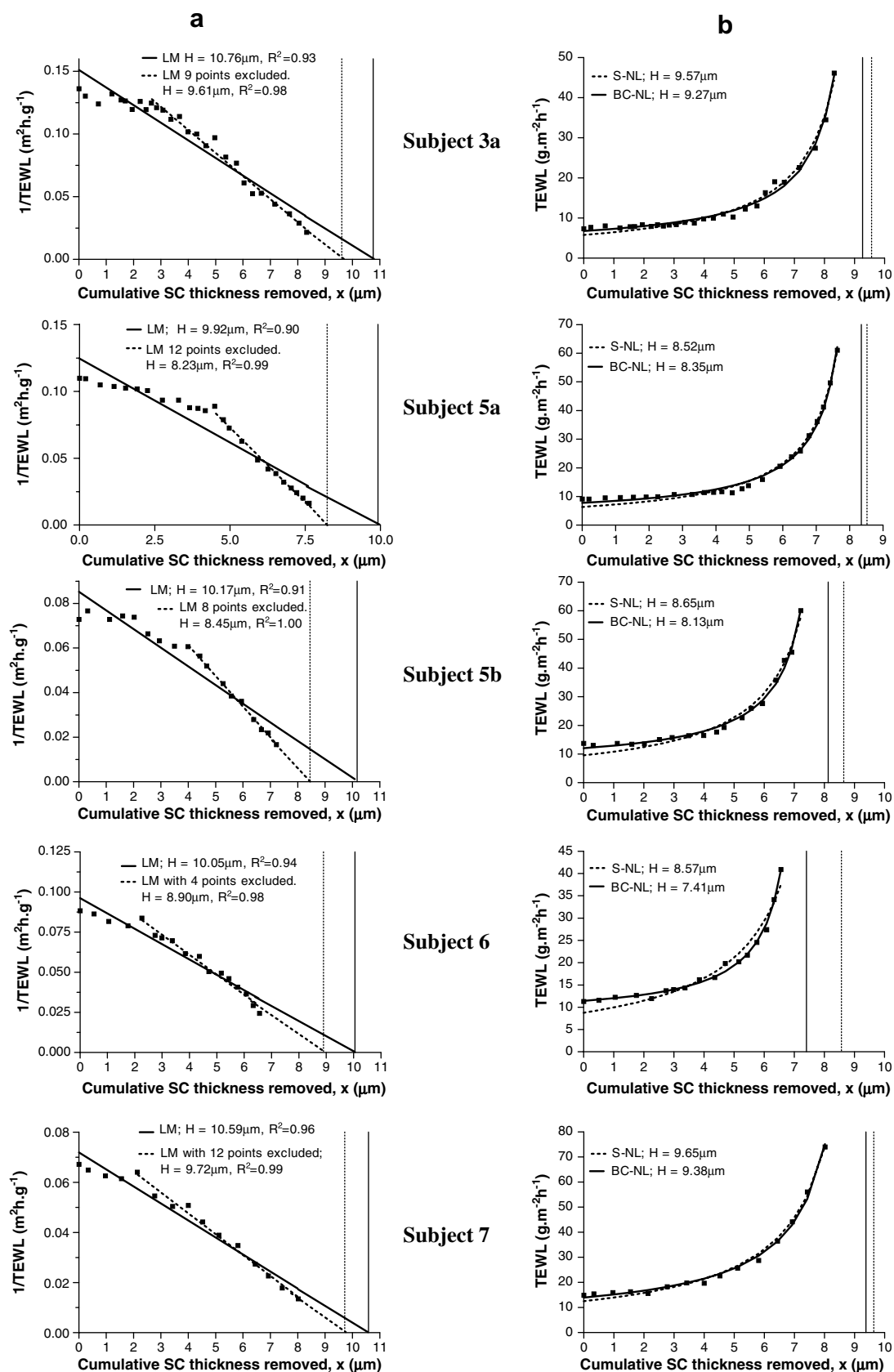


Fig. 2. Estimation of stratum corneum thickness ( $H$ ) by the linear model (panel a) and by the simple (S-NL) and baseline-corrected (BC-NL) non-linear models (panel b) for five subjects. [■] Experimental data points. Panel a: [—] prediction by LM including all experimental data; [...] prediction by LM excluding a number of data points. Panel b: [—] prediction by BC-NL; [...] prediction by S-NL model. The vertical lines mark the value of  $H$  estimated by LM – all data points (panel a, solid line), LM – excluding data points (panel a, dotted line), the S-NL (panel b, dotted line) and the BC-NL (panel b, solid line).



trates five of these examples which show an initial plateau before a clear linear descent is observed. Clearly, the LM fails to fit all the experimental data to the regression line. Worryingly, there is a clear potential for overestimation of  $H$  as the slope of the regression line is shifted upwards to include the initial plateau as Fig. 2 illustrates. The dilemma is how to deal with this type of data which was apparent in approximately half of the 31 experiments performed.

Obviously, the key question is the reason for this initial plateau. We could not find any trend (age, gender, etc.) that would assign this plateau to any “skin type”. In any case, we hypothesised that the initial plateau is due to loose outer portions of SC, which are soon to be lost through natural shedding. These loose portions seem to constitute a minor barrier to water loss, but do represent a considerable mass when removed. This agrees with the heterogeneous nature of the SC structure [18–21]; where the SC compactum evolves into the SC disjunctum as the corneocytes migrate towards the surface, and progressively lose their corneodesmosome links. The cells of the SC disjunctum do not contribute significantly to the water barrier function of the SC. Therefore, when the SC disjunctum layers are removed, a large mass change, and hence cumulative SC thickness removed,  $x$ , is registered, without a corresponding increase in TEWL, resulting in the plateau. Since these layers represent a poor barrier to water loss, it can be argued that they would also constitute a reduced barrier with regard to drug ingress, and hence should be considered separately from the main SC barrier during the estimation of  $H$ .

A potential solution is to subjectively exclude these initial plateau values and fit the rest with the LM. An example of such procedure is illustrated in Fig. 2a. The exclusion of the initial plateau values provides higher  $R^2$  values, and an 8.2–17% change ( $0.87$ – $1.72 \mu\text{m}$ ) in  $H$ . Because the regression line fits only the declining straight line, its gradient shifts downward, resulting in the lower values of  $H$  as shown in Fig. 2a. It should be noted that these fits are produced through the exclusion of a considerable proportion of data points: 9 out of (27) total data points; 12 (24); 8 (18); 4 (17); and 4 (15) for the subjects sequentially shown in Fig. 2a. Unfortunately, it is impossible to define an algorithm to exclude these points objectively, thus, a researcher would be forced to subjectively exclude points by eye in order to improve the linear fit if this approach was adopted. This potential subjectivity in deciding upon the portion to be fitted by the regression constitutes an important disadvantage of the LM. Clearly, any step of a future dermato-pharmacokinetic “modus operandi” aiming to compare topical formulations in an objective way should allow minimal room for inter-laboratories or inter-researcher variability. Therefore, the crucial step of  $H$  determination, in order to normalise data from different volunteers, should clearly be standardised.

To attempt to remedy the problems associated with the LM, namely the poor fit and the potential subjectivity of

removing data points, two alternative non-linear models have been proposed: the simple non-linear (S-NL) (Eq. (2)) and the baseline-corrected non-linear models (BC-NL) (Eq. (4)) which were applied to all data sets with no data points excluded. The first step to evaluate their performance is a visual assessment of the model's predicted fit as compared to the raw data points. Figs. 1 and 2, section b show 7 examples of data fitted with the S-NL and BC-NL models. The predicted line fits the experimental data closely in all cases, independent of the existence of an initial plateau (Fig. 2). A comparison of the two non-linear models, suggests that the BC-NL model fits the earlier TEWL values slightly better, which was the rationale behind the introduction of the baseline parameter,  $B$ , into this model. The baseline parameter allows some vertical translocation of the resultant fit, and thus improves the fitting in some cases (subjects 3a, 5b and 6).

Fig. 1 shows two sets of data 1a and 2b which were well fitted by all the three models. The estimated total SC thickness for subject 1a was  $12.9 \pm 0.19 \mu\text{m}$  (LM),  $13.1 \pm 0.07 \mu\text{m}$  (S-NL) and  $13.2 \pm 0.16 \mu\text{m}$  (BC-NL). The estimated  $H$  for subject 2b was  $8.8 \pm 0.12$ ,  $8.5 \pm 0.10$  and  $8.0 \pm 0.17 \mu\text{m}$  according to the LM, the S-NL and the BC-NL models, respectively. Thus, when there is no initial plateau, namely for subjects having little SC disjunctum, the three models fit well the data and the values of  $H$  obtained are very similar. Other sets of data that showed this behaviour are highlighted in Table 1. Next, we should consider the data shown in Fig. 2 or the cases showing an apparent initial plateau. Contrary to the LM, both the non-linear models fit the data very well. Both the non-linear models resulted in very close values of  $H$  which were approximately  $1 \mu\text{m}$  smaller than those estimated by the linear model. Interestingly enough, when values are excluded from the linear fitting the differences become smaller. For example, the estimated values of  $H$  for subject 5a were  $8.5 \pm 0.04 \mu\text{m}$  (S-NL) and  $8.4 \pm 0.05 \mu\text{m}$  (BC-NL); clearly lower than that estimated by linear regression  $9.9 \pm 0.45 \mu\text{m}$ . However, when 12 points were excluded from the linear regression, the value of  $H$  decreased to  $8.26 \pm 0.07 \mu\text{m}$ , very similar to the non-linear estimates. A similar trend is observed for all the cases shown in Fig. 2 and for a total of 16 of the 31 cases analyzed. This suggests that the three models would agree on a common value for  $H$  if the linear regression is applied only to the latter linear portion of the data. However, as discussed before, excluding the initial plateau values, and deciding on linearity is a process subjected to researchers' subjectivity. On the contrary, the two non-linear models fit all the experimental data letting the model to correct for the baseline in an objective way.

The values of  $H$  estimated by the S-NL and BC-NL models were very similar in cases 3a, 5a, 5b and 7 (Fig. 2). Some differences between the S-NL and the BC-NL models are illustrated by subject 6: the S-NL prediction misses six experimental points, while the BC-NL misses only two; the latter model is probably aided by the vertical

Table 1

Estimation of stratum corneum thickness ( $H$ ) by the linear model (LM) and by the simple (S-NL) and baseline-corrected (BC-NL) non-linear models for 18 subjects (31 experiments)

| Subject <sup>a</sup> | LM                    |      |        | S-NL                  |      |        | BC-NL                 |      |        | $\Delta AIC$ <sup>b</sup> | Probability (%) BC-NL is preferred | $F^*/F_{table}$ |
|----------------------|-----------------------|------|--------|-----------------------|------|--------|-----------------------|------|--------|---------------------------|------------------------------------|-----------------|
|                      | $H$ ( $\mu\text{m}$ ) | SER  | CV (%) | $H$ ( $\mu\text{m}$ ) | SER  | CV (%) | $H$ ( $\mu\text{m}$ ) | SER  | CV (%) |                           |                                    |                 |
| 1a                   | 12.9                  | 0.19 | 1.5    | 13.1                  | 0.07 | 0.6    | 13.2                  | 0.16 | 1.2    | 0.39                      | 45                                 | 1.4/6.2         |
| 1b*                  | 20.1                  | 0.53 | 2.7    | 17.2                  | 0.33 | 1.9    | 15.4                  | 0.09 | 0.6    | −37.74                    | 100                                | 101.5/56.0      |
| 2a                   | 11.5                  | 0.25 | 2.2    | 11.4                  | 0.15 | 1.3    | 11.5                  | 0.34 | 3.0    | 1.84                      | 29                                 | 0.1/5.8         |
| 2b                   | 8.8                   | 0.12 | 1.4    | 8.5                   | 0.10 | 1.2    | 8.0                   | 0.17 | 2.2    | −3.90                     | 88                                 | 5.8/6.0         |
| 3a*                  | 10.8                  | 0.41 | 3.8    | 9.6                   | 0.06 | 0.6    | 9.3                   | 0.08 | 0.8    | −10.93                    | 100                                | 14.7/5.7        |
| 3b                   | 6.4                   | 0.19 | 2.9    | 5.8                   | 0.03 | 0.5    | 5.7                   | 0.05 | 0.8    | −2.43                     | 77                                 | 4.2/6.0         |
| 4                    | 10.5                  | 0.25 | 2.4    | 10.1                  | 0.06 | 0.6    | 9.9                   | 0.13 | 1.3    | −0.88                     | 61                                 | 2.6/6.0         |
| 5a*                  | 9.9                   | 0.45 | 4.5    | 8.5                   | 0.04 | 0.5    | 8.4                   | 0.05 | 0.6    | −11.50                    | 100                                | 15.85/5.8       |
| 5b*                  | 10.2                  | 0.51 | 5.0    | 8.6                   | 0.10 | 1.2    | 8.1                   | 0.08 | 1.0    | −18.97                    | 100                                | 33.1/6.2        |
| 6*                   | 10.0                  | 0.44 | 4.4    | 8.6                   | 0.20 | 2.3    | 7.4                   | 0.07 | 1.0    | −32.39                    | 100                                | 91.8/6.3        |
| 7*                   | 10.6                  | 0.41 | 3.8    | 9.6                   | 0.08 | 0.8    | 9.4                   | 0.11 | 1.2    | −4.31                     | 90                                 | 6.3/6.6         |
| 8a*                  | 7.4                   | 0.29 | 3.8    | 6.8                   | 0.07 | 1.0    | 6.5                   | 0.05 | 0.7    | −18.50                    | 100                                | 36.6/6.8        |
| 8b                   | 13.0                  | 0.37 | 2.8    | 11.8                  | 0.14 | 1.2    | 11.3                  | 0.16 | 1.4    | −7.19                     | 97                                 | 9.9/6.1         |
| 8c                   | 10.0                  | 0.20 | 2.0    | 9.5                   | 0.04 | 0.5    | 9.5                   | 0.08 | 0.8    | 1.36                      | 34                                 | 0.5/9.9         |
| 8d*                  | 7.5                   | 0.28 | 3.8    | 6.5                   | 0.17 | 2.7    | 5.7                   | 0.05 | 1.0    | −26.04                    | 100                                | 70.5/6.7        |
| 9a*                  | 13.0                  | 0.88 | 6.8    | 10.4                  | 0.14 | 1.4    | 10.0                  | 0.12 | 1.2    | −9.29                     | 99                                 | 13.2/6.3        |
| 9b                   | 8.8                   | 0.14 | 1.5    | 8.7                   | 0.05 | 0.6    | 8.8                   | 0.09 | 1.1    | −0.02                     | 50                                 | 1.8/6.3         |
| 9c                   | 12.4                  | 0.27 | 2.2    | 11.7                  | 0.11 | 0.9    | 11.2                  | 0.11 | 1.0    | −10.86                    | 100                                | 15.0/5.9        |
| 10a                  | 6.4                   | 0.17 | 2.6    | 5.9                   | 0.04 | 0.7    | 5.7                   | 0.05 | 0.9    | −7.40                     | 98                                 | 10.4/6.4        |
| 10b*                 | 8.6                   | 0.33 | 3.9    | 7.2                   | 0.06 | 0.9    | 7.0                   | 0.05 | 0.7    | −12.85                    | 100                                | 19.0/6.1        |
| 11                   | 6.2                   | 0.16 | 2.6    | 6.0                   | 0.04 | 0.8    | 6.0                   | 0.09 | 1.5    | 0.21                      | 47                                 | 1.5/6.4         |
| 12a*                 | 13.1                  | 0.38 | 2.9    | 11.3                  | 0.10 | 0.9    | 10.8                  | 0.07 | 0.7    | −24.33                    | 100                                | 38.5/5.7        |
| 12b                  | 14.5                  | 0.24 | 1.7    | 15.2                  | 0.06 | 0.4    | 15.4                  | 0.10 | 0.7    | −9.18                     | 99                                 | 12.6/5.9        |
| 13                   | 8.1                   | 0.25 | 3.1    | 7.6                   | 0.05 | 0.6    | 7.6                   | 0.09 | 1.2    | 1.76                      | 29                                 | 0.2/6.2         |
| 14a*                 | 16.2                  | 0.52 | 3.2    | 13.9                  | 0.32 | 2.3    | 12.0                  | 0.07 | 0.6    | −53.67                    | 100                                | 192.6/5.8       |
| 14b*                 | 18.1                  | 0.68 | 3.8    | 17.8                  | 0.12 | 0.7    | 17.2                  | 0.08 | 0.5    | −25.40                    | 100                                | 56.2/6.3        |
| 14c*                 | 9.4                   | 0.27 | 2.9    | 9.6                   | 0.16 | 1.7    | 10.6                  | 0.66 | 6.2    | −3.41                     | 85                                 | 5.3/6.0         |
| 15*                  | 12.7                  | 0.27 | 2.2    | 12.0                  | 0.11 | 0.9    | 11.4                  | 0.13 | 1.2    | −10.28                    | 99                                 | 14.2/5.9        |
| 16                   | 9.1                   | 0.20 | 2.2    | 8.5                   | 0.06 | 0.7    | 8.4                   | 0.09 | 1.1    | −2.77                     | 80                                 | 4.5/6.5         |
| 17*                  | 8.8                   | 0.34 | 3.9    | 7.5                   | 0.09 | 1.2    | 7.1                   | 0.05 | 0.7    | −23.63                    | 100                                | 51.5/6.4        |
| 18                   | 5.6                   | 0.22 | 4.0    | 5.0                   | 0.12 | 2.4    | 4.4                   | 0.13 | 2.8    | −7.81                     | 98                                 | 11.1/6.6        |
| Average              | 10.7                  | 0.30 | 3.1    | 9.8                   | 0.10 | 1.1    | 9.5                   | 0.10 | 1.3    |                           | 84                                 |                 |
| $H \pm SD^c$         | 10.7 $\pm$ 3.4        |      |        | 9.8 $\pm$ 3.2         |      |        | 9.5 $\pm$ 3.1         |      |        |                           |                                    |                 |

<sup>a</sup> The cases where a plateau was apparent to the authors are indicated with \*.

<sup>b</sup>  $\Delta AIC = AIC_{BC-NL} - AIC_{S-NL}$ .

<sup>c</sup> Average  $H$  ( $\mu\text{m}$ ) with standard deviation (SD) for 31 volunteers measured according to the three models.

translocation permitted by the baseline parameter. This results in a slightly larger (1.16  $\mu\text{m}$ ) difference in the values of  $H$  estimated by the two non-linear models in this case.

Before discussing the goodness-of-fit of the three models by statistical tools we should discuss whether  $H$ , the parameter of interest for dermato-pharmacokinetic studies, is significantly different when estimated by different models. A compilation of the  $H$  estimates together with the standard error of the regression (SER) and the coefficient of variation (CV(%)) for the 31 data sets is presented in Table 1. The  $H$  and corresponding SER may be compared graphically in Fig. 3. A matched-observations Friedman test (equivalent to a non-parametric 1-way matched ANOVA) followed by the corresponding Dunn's post-test was used to compare the values of  $H$  estimated by the LM (with no data excluded), the S-NL and the BC-NL models. The test concluded that the value of  $H$  estimated by the linear model was statistically significantly greater than that derived from the S-NL ( $p < 0.01$ ) and the BC-NL

( $p < 0.001$ ) models. In other words, the linear model tends to overestimate the thickness of the stratum corneum. On the other hand, although the S-NL model tends to estimate a higher value of  $H$  than the BC-NL model, the differences between the non-linear models did not reach the level of statistical significance ( $p < 0.05$ ).

Table 1 also shows the coefficient of variation (CV%) associated with each value of  $H$  and model. It is interesting to note that in 28 of the 31 cases considered, the highest relative error was associated with the linear method. The performance of the linear model can be improved but only via the exclusion of some data points. Overall, these results suggest that the two non-linear models proposed here offer a better-quality fit than the linear model; allowing a better estimation of the parameter  $H$  (lower CV%), in an objective way.

Finally, we should discuss the relative performance of the two non-linear models. The S-NL model fits directly the experimental values (TEWL and cumulative thickness

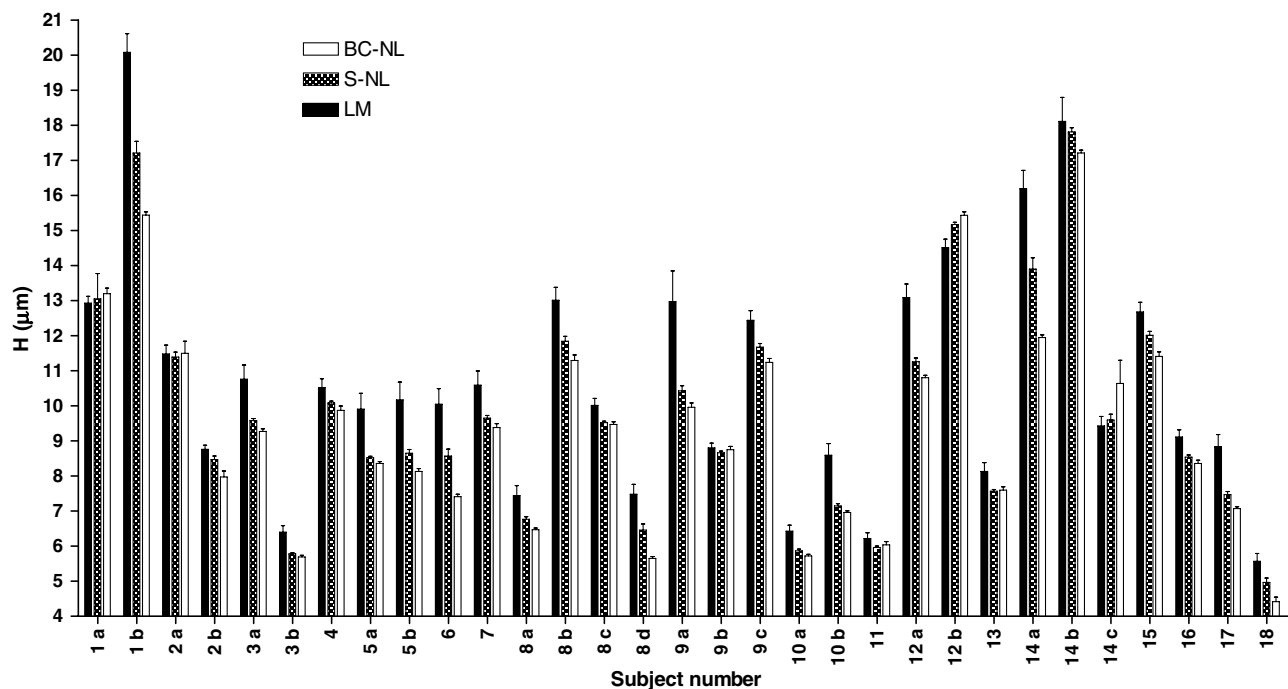


Fig. 3. Stratum corneum thickness,  $H$  ( $\mu\text{m}$ ) as determined by the LM, S-NL, and the BC-NL models, for 18 different volunteers at 31 different sites. The error bars show the standard error of the regression (SER) associated with each estimate.

removed) to Fick's first law. This basic difference already provides a certain advantage over linearly fitting the transformed values of water loss ( $1/\text{TEWL}$ ) [22]. The baseline parameter was built into the BC-NL model to describe the removal of the outer SC layers that do not significantly contribute to the barrier to water loss. However, the BC-NL model would be expected to fit the data better, and has smaller residual sum of squares than the S-NL model simply because it has one extra parameter. Therefore, to decide which model was superior, the use of some statistical tools that take into consideration the number of parameters used by each model was required, as outlined in Section 2. Two statistical tools: the AIC and an  $F$ -test (Table 1) (described in Section 2) were used to compare the non-linear models. The  $F$ -test performed here is only applicable for nested models (i.e., where one model (S-NL) can be considered a simplified version of a more complex one (BC-NL)). The value of  $F^*$  calculated is compared to a critical  $F$  value obtained from published  $F$  tables [26]. If the value of  $F^*$  is bigger than the critical value, the null hypothesis can be rejected and we can accept that the more complex model fits the data better. Table 1 shows that in 19 of the 31 cases, the BC-NL model is preferred ( $p < 0.05$ ).

The AIC offers complementary information as it tells us the probability of the preferred model being the correct one. The value  $\Delta\text{AIC} = \text{AIC}_{\text{BC-NL}} - \text{AIC}_{\text{S-NL}}$  was calculated as described in Section 2 and the values are shown in Table 1. The model with the lowest AIC is considered superior, therefore negative values of  $\Delta\text{AIC}$  indicate that the BC-NL is preferred. The magnitude of this difference can be used to determine the probability of this assumption

being correct via Eq. (7). For simplicity, Table 1 always shows the probability of the BC-NL model being the correct one. For example, for subject 1b, the negative value of  $\Delta\text{AIC}$  ( $-37.74$ ) indicates that the BC-NL model fits the data better than the S-NL model, and that there is a 100% chance that the BC-NL model is the correct one. On the contrary, in the case of subject 2a, the positive value of  $\Delta\text{AIC}$  (1.84) indicates that the S-NL model is superior in this case, and that there is a 71% chance of the S-NL model being correct (29% chance for the BC-NL model shown in the column). On the whole, the  $\Delta\text{AIC}$  of 25 of the 31 data sets were negative, indicating a preference for the BC-NL model (probability of BC-NL being correct  $>61\%$ ). In fact, in 20 of the 31 cases there was a probability  $\geq 90\%$  of the model BC-NL being correct.

The use of these two parameters can be illustrated by considering subject 6. The SC thickness for subject 6 is  $8.6 \pm 0.2 \mu\text{m}$  ( $\text{CV} = 2.3\%$ ) according to the S-NL model and  $7.4 \pm 0.07$  ( $\text{CV} = 1\%$ ) according to the BC-NL model. The difference in the  $\Delta\text{AIC}$  for this subject was  $-32.39$ , resulting in a probability of 100% the BC-NL model being correct; the  $F$ -test also shows that the BC-NL model fits better than the S-NL model. This is in good agreement with the graphic representation: the S-NL prediction misses 6 experimental points, while the BC-NL misses only two. As discussed before, approximately 25 of the 31 cases were better modelled by the BC-NL model, indicating the usefulness of the inclusion of the baseline parameter to fit the full breadth of the experimental data. In summary, the results in Table 1 suggest that the BC-NL has a better overall performance than the S-NL model.



In addition, Fig. 3 shows the data for some volunteers who participated on different occasions. It is worth noting that the  $H$  estimated on different occasions may differ markedly for the same volunteer, independent of the model used to fit the data. See for example, subjects 1–3, 5, 8, 9, 10, 12 and 14 (Fig. 3). These differences imply that the SC thickness changes with time, potentially due to environmental conditions and the use of drugs, cosmetics or exfoliating agents. It follows that a determination of the SC thickness is necessary every time a dermato-pharmacokinetic study is performed in a subject, so the correct thickness is used to normalise the drug-penetration profiles.

In summary, we believe that non-linear models perform better than the standard linear model typically used to estimate the stratum corneum thickness via tape-stripping experiments combined with TEWL measurements. When the LM performs well, that is in the absence of an initial plateau, the three models estimate comparable values of  $H$ . However, when an initial plateau appears, probably corresponding to the removal of the SC disjunctum, the LM model tends to overestimate the thickness of the SC unless some subjective exclusion of data is made. On the contrary, the non-linear models offer a robust procedure, incorporating all data points, reflecting better experimental observations, and estimating the thickness with a smaller coefficient of variation.

The incorporation, rather than the exclusion, of the initial plateau by the non-linear approach should be preferred as it follows more closely the SC physiology and its division into the SC compactum and disjunctum. The statistical comparison of the two non-linear models showed a higher probability for the baseline-corrected model being the preferred one. However, there were no statistical differences between the values of  $H$  estimated from either of these models.

## Acknowledgements

We thank Pharsight Corporation Inc. for a PAL WinNonlin<sup>®</sup> licence and Dr. Dan Weiner for scientific input during the development and assessment of the non-linear models. The financial support of the U.S. National Institutes of Health (EB-001420) is gratefully acknowledged. We thank Prof. R.H. Guy and other members of our group at the University of Bath for encouraging discussions.

## References

- [1] I. Alberti, Y.N. Kalia, A. Naik, R.H. Guy, Assessment and prediction of the cutaneous bioavailability of topical terbinafine, in vivo, in man, *Pharm. Res.* 18 (2001) 1472–1475.
- [2] FDA, Draft Guidance: Guidance for Industry: topical dermatological drug product NDAs and ANDAs-in vivo bioavailability, bioequivalence, in vitro release, and associated studies, U.S. Department of Health and Human Services, Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), 1998.
- [3] V.P. Shah, G.L. Flynn, A. Yacobi, H.I. Maibach, C. Bon, N.M. Fleischer, et al., Bioequivalence of topical dermatological dosage forms – methods of evaluation of bioequivalence, *Pharm. Res.* 15 (1998) 167–171.
- [4] K.A. Holbrook, G.F. Odland, Regional differences in the thickness (cell layers) of the human stratum corneum: an ultrastructural analysis, *J. Invest. Dermatol.* 62 (1974) 415–422.
- [5] M. Denda, L.C. Wood, S. Emami, C. Calhoun, B.E. Brown, P.M. Elias, K.R. Feingold, The epidermal hyperplasia associated with repeated barrier disruption by acetone treatment or tape stripping cannot be attributed to increased water loss, *Arch. Dermatol. Res.* 288 (1996) 230–238.
- [6] G. Eriksson, L.O. Lamke, Regeneration of human epidermal surface and water barrier function after stripping. A combined study with electron microscopy and measurement of evaporative loss, *Acta Derm. Venereol.* 51 (1971) 169–178.
- [7] T. Frodin, M. Skogh, Measurement of transepidermal water loss using an evaporimeter to follow the restitution of the barrier layer of human epidermis after stripping the stratum corneum, *Acta Derm. Venereol.* 64 (1984) 537–540.
- [8] I. Alberti, Y.N. Kalia, A. Naik, J. Bonny, 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 (2001) 11–19.
- [9] I. Alberti, Y.N. Kalia, A. Naik, J.D. Bonny, R.H. Guy, In vivo assessment of enhanced topical delivery of terbinafine to human stratum corneum, *J. Control. Release* 71 (2001) 319–327.
- [10] C. Herkenne, A. Naik, Y.N. Kalia, J. Hadgraft, R.H. Guy, Pig ear skin ex vivo as a model for in vivo dermatopharmacokinetic studies in man, *Pharm. Res.* 23 (2006) 1850–1856.
- [11] C. Herkenne, A. Naik, Y.N. Kalia, J. Hadgraft, R.H. Guy, Dermatopharmacokinetic prediction of topical drug bioavailability in vivo, *J. Invest. Dermatol.* 127 (2007) 887–894.
- [12] F. Pirot, Y.N. Kalia, A.L. Stinchcomb, G. Keating, A. Bunge, R.H. Guy, Characterization of the permeability barrier of human skin in vivo, *Proc. Natl. Acad. Sci. USA* 94 (1997) 1562–1567.
- [13] Y.N. Kalia, I. Alberti, N. Sekkat, C. Curdy, A. Naik, R.H. Guy, Normalization of stratum corneum barrier function and transepidermal water loss in vivo, *Pharm. Res.* 17 (2000) 1148–1150.
- [14] F. Pirot, E. Berardesca, Y.N. Kalia, M. Singh, H.I. Maibach, R.H. Guy, Stratum corneum thickness and apparent water diffusivity: facile and noninvasive quantitation in vivo, *Pharm. Res.* 15 (1998) 492–494.
- [15] J. Crank, *The Mathematics of Diffusion*, Clarendon Press, Oxford, 1975.
- [16] Y.N. Kalia, F. Pirot, R.H. Guy, Homogeneous transport in a heterogeneous membrane: water diffusion across human stratum corneum in vivo, *Biophys. J.* 71 (1996) 2692–2700.
- [17] R.L. Anderson, J.M. Cassidy, Variations in physical dimensions and chemical composition of human stratum corneum, *J. Invest. Dermatol.* 61 (1973) 30–32.
- [18] E.H. Choi, M.-Q. Man, F. Wang, X. Zhang, B.E. Brown, K.R. Feingold, P.M. Elias, Is endogenous glycerol a determinant of stratum corneum hydration in humans? *J. Invest. Dermatol.* 125 (2005) 288–293.
- [19] P. Corcuff, F. Fiat, A.M. Minondo, Ultrastructure of the human stratum corneum, *Skin Pharmacol. Appl. Skin Physiol.* 14 (Suppl. 1) (2001) 4–9.
- [20] M. Haftek, M. Simon, G. Serre, Corneodesmosomes: pivotal actors in the stratum corneum cohesion and desquamation, in: P.M. Elias, K.R. Feingold (Eds.), *Skin Barrier*, Taylor & Francis, New York, London, 2006, pp. 171–190.
- [21] R.I. Spearman, Some light microscopical observations on the stratum corneum of the guinea-pig, man and common seal, *Br. J. Dermatol.* 83 (1970) 582–590.
- [22] J. Gabrielsson, D. Weiner, Modelling strategies, in: J. Gabrielsson, D. Weiner (Eds.), *Pharmacokinetic/Pharmacodynamic Data Analysis: Concepts and Applications*, Apotekarsocieteten, Stockholm, 2006, pp. 389–463.
- [23] C. Herkenne, I. Alberti, A. Naik, Y.N. Kalia, F.X. Mathy, V. Preat, R.H. Guy, In vivo methods for the assessment of topical drug bioavailability, *Pharm. Res.* 25 (2008) 87–103.

- [24] H. Motulsky, A. Christopoulos, Comparing models using Akaike's Information Criterion (AIC), in: H. Motulsky, A. Christopoulos (Eds.), *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, Oxford University Press, Oxford, 2004, pp. 143–148.
- [25] H. Motulsky, A. Christopoulos, Comparing models using the extra sum-of-squares F test, in: H. Motulsky, A. Christopoulos (Eds.), *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, Oxford University Press, Oxford, 2004, pp. 138–142.
- [26] J.H. Zar, *Biostatistical Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984, Appendix B: Statistical Tables and Graphs.
- [27] J.S. Lo, H.A. Oriba, H.I. Maibach, P.L. Bailin, Transepidermal potassium ion, chloride ion, and water flux across delipidized and cellophane tape-stripped skin, *Dermatologica* 180 (1990) 66–68.
- [28] C. Herkenne, A. Naik, Y.N. Kalia, J. Hadgraft, R.H. Guy, Ibuprofen transport into and through skin from topical formulations: in vitro–in vivo comparison, *J. Invest. Dermatol.* 127 (2007) 135–142.