Report

Induction of epidermal damage by tape stripping to evaluate skin mildness of cleansing regimens for the premature epidermal barrier

Robert J. O’Connor¹, MS, Julie Ogle¹, BS, and Mauricio Odio², PhD

¹Department of Global Clinical Sciences
The Procter & Gamble Company, Cincinnati, OH, USA, and
²Retired, The Procter & Gamble Company, Cincinnati, OH, USA

Correspondence
Robert J. O’Connor, MS
Global Clinical Sciences Department
The Procter & Gamble Company
11450 Grooms Road
Cincinnati, OH 45242, USA
E-mail: oconnor.rj@pg.com

Abstract

Background Previous works indicate a compromised skin model may be a possible surrogate for premature, undeveloped skin.

Objectives This study was performed to investigate the use of a current model as a surrogate test design.

Methods Serial tape stripping on the volar forearms of adult female volunteers (n = 36) was used to “thin” the stratum corneum. The forearm sites received intensive applications of different wiping options that exaggerated the exposures to cleansers that might be experienced by an infant in a neonatal intensive care unit. The recovery of skin barrier function during the wiping regimen, measured by transepidermal water loss (TEWL), was used as the primary indicator of the mildness of the cleansing options. Measurements of TEWL were made prior to the first wash on days 1–4, on day 5 and on day 8. Erythema was graded as a secondary endpoint.

Results There was an early and sustained distinction in rates of skin barrier recovery between the washcloth and water treatment, and the two wipes treatments (P < 0.05). Areas submitted to the wipes treatments showed recovery rates similar to that of the tape-stripped untreated site, indicating that the use of disposable wipes led to minimal perturbation of the recovery process. By contrast, cleansing with a cotton washcloth and water markedly perturbed the repair process compared with all other conditions (P < 0.05).

Conclusions This model shows promise as a possible surrogate model for assessing the mildness of skin cleansing products for the care of premature infants.

Introduction

Fragility of the outer layer of the skin, the stratum corneum, places premature infants at high risk for sepsis, irritation and dehydration. Thus, skin care is a special concern in the neonatal intensive care unit (NICU). Maintaining diaper skin hygiene poses additional challenges because the removal of feces and urine requires more vigorous cleansing. Tape stripping has been used as a surrogate for damaged skin in neurosensory evaluations of topical product stinging,¹ and as a model for the evaluation of the potential impact of topical treatments on barrier repair.² Both tape-stripped porcine skin and excised adult skin have been previously evaluated as potential in vitro surrogates for premature skin in topical drug absorption and chemical skin penetration studies.³⁻⁴ Barker et al. hypothesized as far back as 1987 that the preterm infant’s epidermis resembles the stripped skin of an adult.⁵ Based on this precedent, tape-stripped adult forearm skin was explored as a potential surrogate model for the screening of skin cleansing regimens for use on highly compromised or premature skin.

Early research comparing cleansing options for compromised skin found disposable wipes to be milder than a washcloth and water.¹ Building on this previous work, the present research was conducted to evaluate the efficacy of this model in the testing and development of improved skin cleansing options for use in the NICU in order to diminish the need for testing in premature infants.¹⁻³,⁷,⁸

This test was designed to compare cleaning implements with one another and with a no-treatment control for their impact on the recovery of epidermal barrier function.
Materials and methods

Study subjects
Generally healthy (self-reported) women aged 20–55 years, with Fitzpatrick skin types I–IV, were recruited for the study. Subjects agreed to comply with all aspects of the study protocol, which included requirements to refrain from tanning and sunbathing, and not to use excluded medications and skin preparations during the test period. Subjects were excluded if they were participating in another clinical study or any type of research study involving the forearms currently or had done within the previous month. Additionally, subjects were required to have visual erythema grades of ≤0.5 (on a scale of 0–4) and a baseline transepidermal water loss (TEWL) measurement at all sites of <12 g/m²/h prior to baseline measures and procedures.

Subjects were not enrolled if they had known or documented allergies to adhesives or products containing perfumes, if they had been diagnosed with skin cancer within the previous 12 months, or if they had any condition on the inner forearm that might prevent a clear assessment of the skin. Additionally, subjects were proscribed from entry to the study if they had a chronic condition or were taking any prescription or over-the-counter medication(s) that, in the opinion of the investigator, might have been likely to affect their response to treatment.

Tape stripping and TEWL measures
To maximize the relevancy of the model to the premature skin condition, serial tape stripping was performed to achieve a level of skin barrier compromise similar to that among infants at approximately 26–28 weeks of gestational age. The starting level of skin damage for the clinical model was established in line with published data on TEWL and represents some of the most fragile skin conditions under which disposable wipes may be used. All TEWL measurements were taken with a VapoMeter Evaporimeter (Delphin Technologies Oy, Kuopio, Finland). The skin damage targets for the study (TEWL of 40–50 g/m²/h) were similar to values reported in premature infants. This range is also comparable with TEWL values observed in children presenting with relatively severe diaper rash. When the target TEWL range could not be reached, a TEWL value of at least four times that at baseline was used and/or stripping was stopped at a maximum of 60 tape strips. These criteria minimized the probability of causing severe skin damage by tape stripping. Once the target was reached, a second confirmatory TEWL measurement was taken after 1 minute to verify that TEWL was within the target range.

In advance of the study execution, it was determined that if a subject’s post-tape strip confirmatory TEWL value was outside the target TEWL range by ±20% (<32 g/m²/h or >60 g/m²/h), the site would be considered non-usable. This protocol was established to eliminate unnecessary variability in starting TEWL values.

Measurements of TEWL were taken at the time-points designated in the protocol (see below) for all test sites, including the control sites at which neither tape stripping nor test products were applied. This was intended to facilitate the tracking of variations in the subject’s normal skin condition.

Study design and flow
The study protocol was reviewed and approved by an accredited external institutional review board ethics committee. The study was run at a third-party laboratory with extensive experience in executing these types of skin measurement studies. For logistical reasons, subjects were enrolled and investigated in two groups in February and March, respectively, 2013. All subjects signed statements of informed consent prior to the start of any study-related procedures.

Figure 1 shows the study flow and a breakdown of visit procedures.

This was a randomized, assessor-blinded study in which 36 subjects were randomized to treatment. The treatment portion of the study took place over five consecutive days (days 1–5) and was succeeded by a follow-up visit at which measurements were obtained on day 8 to verify that skin barrier repair had continued and that the skin condition was returning to baseline levels. During a run-in period of 3–4 days prior to day 1 and throughout the study, subjects used a standard cleansing bar on their forearms.

The same expert skin grader was utilized for the duration of the study. The skin grader was blinded to all treatments throughout the study and grading was performed in a separate location from other procedures. The grader’s status as expert was based on multiple years of experience with the standard patch grading scale used in the study (Fig. 2).

On day 1, at visit 1, six test sites were identified (demarcated) for evaluation (three sites per forearm). All subjects were exposed to the following four treatments: (i) a currently marketed disposable baby wipes product (Wipe A); (ii) a modified version of the marketed wipes (Wipe B); (iii) a commercially available, 100% cotton infant washcloth used with lukewarm tap water, and (iv) no treatment other than tape stripping.

In each subject, the four treatments were randomly assigned to one of four designated treatment sites (sites proximal to wrist or elbow). The area between the test sites on each forearm served as the no-stripping, no-treatment control site to which no treatment option was assigned, but at which assessments were performed to quantify the “normal” skin condition at each time-point.
Baseline erythema grades and measurements of TEWL were recorded for all six sites. The four treatment sites were then physically compromised by serial tape stripping using 1-inch Blenderm tape. Immediately following tape stripping on day 1, post-strip TEWL measurements and erythema grades were recorded and the wiping sequences were initiated.

**P&G Uniform Laboratory Patch Test Grading Scale (ERYTHEMA)**

0.0 No apparent cutaneous involvement.

0.5 Faint, barely perceptible erythema

1.0 Faint but definite erythema, no eruptions or broken skin

1.5 Well-defined erythema

2.0 Moderate erythema, may have a few papules or deep fissures, moderate to severe erythema in the cracks

2.5 Moderate erythema with barely perceptible edema or severe erythema not involving a significant portion of the patch (halo effect around the edges), may have a few papules or moderate-to-severe erythema.

3.0 Severe erythema (beet redness), may have generalized papules or moderate-to-severe erythema with slight edema (edges well defined by raising).

3.5 Moderate-to-severe erythema with moderate edema (confined to patch area) or moderate-to-severe erythema with isolated eschar formations or vesicles.

4.0 Generalized vesicles or eschar formations or moderate-to-severe erythema and/or edema extending beyond the area of the patch.

**Note** The degree of reaction expressed by such descriptive terms as "moderate" and "severe" is, in itself, subjective. Such terminology can be accurately understood only through experience. Any reaction of greater severity than Grade 4.0 should be described in detail. Unusual reactions not described by the scale should also be described.
Measurements of TEWL, as the primary efficacy and safety assessment, were taken prior to the first wash on days 1–4, on day 5 at 1.5–3.0 h after the third wiping session, and at the final visit on day 8. Erythema was graded using the standard patch grading scale of 0–4 as an indicator of skin irritation. The skin of all test sites was assessed for erythema on days 1–4 prior to the first and fourth wash cycles, on the fifth consecutive day at 1.5–3.0 h after the third wiping session, and at the final visit on day 8. The second erythema score on each day was used as a safety check and was not used for interpretation.

The test sites on the forearms designated for the three wiping treatments were “wiped” by a technician with the respective test products four times per day for four consecutive days. Each wiping cycle entailed the application of three fresh wipes or washcloths to the test site. Washcloths were prepared prior to wiping by soaking them in room-temperature tap water and were rung out until damp. Wipes were also used at room temperature to minimize any temperature variations in the treatments. Wipes and washcloths were folded so that their application did not overlap other test sites on the subjects’ forearms. Each of the three wipes or washcloths was applied for a total of 20 wipes across the test site. This resulted in the application of 60 swipes to each test site at each of four daily wiping sessions (240 swipes/test site/day). A minimum period of 1.5 h and a maximum of 3 h was allowed to elapse between wiping cycles. The time between wiping sessions was based roughly on consumer diaper change frequencies and was incorporated to allow “rest” periods between treatment application sessions. Prior to the study, technicians were trained in the wiping technique in an effort to minimize variability between technicians and over time. Wiping was standardized by having the technicians use the weight of their hand to press the wiping implement against the skin without adding additional pressure.

The impact on skin barrier recovery in this study was assessed based on the inhibitory effect of the test products on skin barrier repair following multiple uses of these test products over the course of the study. A plot of mean TEWL values per treatment over time was visually inspected in order to clinically evaluate the rate of skin barrier repair.

The washcloth and water treatment was included in this study to determine the effect on skin barrier repair of the use of a durable implement that is also commonly used in the care of the diaper area (cotton washcloth) and which is frequently regarded as being very mild to the skin. The no-treatment (i.e. tape stripping without wiping) site was included as a reference to determine the natural course of barrier repair in the absence of any exogenously applied treatments.

Statistical methods
The study used a complete block design whereby the four treatments were randomly assigned to the four forearm sites in all subjects. The treatment comparisons of interest were based on the treatment least squares (LS) means and LS mean differences specific to study objectives and hypotheses. A two-sided 0.05 type I error level was used to determine statistical significance. Study conclusions were based on the statistical test for the difference in treatment means in combination with an evaluation of the two-sided 95% confidence limits for day 5 differences between treatments in mean TEWL.

For this study, the confirmatory TEWL measurement was the last TEWL value recorded prior to initiating the first wiping; therefore, the confirmatory TEWL measurement was the measurement used in the analysis model as a covariate to adjust for variability in TEWL values just prior to wiping.

A pilot study of these same treatments provided preliminary data indicating the smallest mean difference that could be detected in an analysis of results obtained in groups of various sizes and two different standard deviations with 90% power and a two-sided type I error level (i.e. alpha) of 0.05. It was assumed that the same correlation would be seen in the proposed study.

Results
Subject demographics and accountability
A total of 57 healthy women aged 20–55 years (mean age: 42.7 years) with Fitzpatrick skin types I–IV were screened. Of these, 36 subjects qualified for the study and were randomized to treatment. The remaining 21 subjects did not meet the study entry criteria. All of the subjects randomized to treatment completed the study. However, in five subjects one site was excluded and in one subject two sites were excluded from analyses because post-tape strip TEWL was below the minimum cut-off level of 32 g/m²/h. Post-damage TEWL was comparable at all sites (Table 1).

<p>| Table 1 Mean ± standard error of the mean baseline and post-damage transepidermal water loss (TEWL) and number of tape strips applied by treatment sitea |</p>
<table>
<thead>
<tr>
<th>Treatment site</th>
<th>Baseline</th>
<th>Post-damage</th>
<th>Tape strips, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wipe A</td>
<td>7.5 ± 0.42</td>
<td>47.2 ± 1.00</td>
<td>38.9 ± 2.12</td>
</tr>
<tr>
<td>Wipe B</td>
<td>7.4 ± 0.34</td>
<td>44.5 ± 0.85</td>
<td>36.6 ± 1.85</td>
</tr>
<tr>
<td>Cotton washcloth</td>
<td>7.4 ± 0.35</td>
<td>44.3 ± 0.76</td>
<td>37.7 ± 1.98</td>
</tr>
<tr>
<td>Tape strip, no treatment</td>
<td>7.3 ± 0.35</td>
<td>43.8 ± 0.67</td>
<td>37.6 ± 1.97</td>
</tr>
</tbody>
</table>

Data refer to 33–36 observations for each test product. Variations in sample size reflect data excluded for sites that did not reach or exceeded damage acceptance criteria.
Primary efficacy and safety assessment

All pairwise comparisons were statistically significant for LS mean TEWL at day 5 as a result of the high sensitivity and low variability of the test design (Table 2).

Wipe B resulted in a statistically significantly higher day 5 LS mean TEWL than Wipe A ($P = 0.006$). At day 5, LS mean TEWL values for both Wipes A and B were statistically significantly lower than that for washcloth and water ($P < 0.0001$). Finally, LS mean TEWL values at day 5 for both Wipe A and Wipe B were significantly higher than that for the tape-stripped untreated control area ($P = 0.0149$ and $P < 0.0001$, respectively).

Following initial tape strip damage, TEWL at the tape-stripped untreated control site recovered by approximately 69% from post-damage levels by study day 5. Areas treated with Wipe A and Wipe B showed rates of recovery similar to that at the untreated site (68% and 61%, respectively). The tape-stripped sites at which the wiping treatments were applied showed an early and sustained distinction in the rate of skin barrier recovery between sites treated with the washcloth and water treatment (41% recovery) and those treated with the wipes ($P < 0.05$) (Fig. 3).

Secondary efficacy and safety assessment

Data on grades of erythema were not submitted to analysis beyond those of summary statistics and plots (Fig. 4). However, post hoc analyses of these data showed the same trends as the TEWL results.

Discussion

Although previous investigations of premature neonatal skin have shown differences in structure compared with adult skin, the primary difference has concerned the thickness of the protective stratum corneum barrier. Diaper skin hygiene practices pose additional challenges to the compromised skin of premature infants as a result of the need for more vigorous cleansing to remove feces and urine, which can induce further skin barrier compromise.

The test design utilized in this study was based on previous modifications of standard forearm controlled application test (FCAT) designs or in vitro data as...
referred to above. As premature baby skin continues to develop after birth and forms a more mature stratum corneum rapidly, this model has been designed to follow the impact of cleaning implements on barrier repair (and/or development) rather than disrupting that process.

Epidermis in this state of development and/or repair is easily disrupted by friction and abrasion; hence this model has been fairly predictive of the potential of cleaning materials to disrupt the repair and development process in the context of continual insult. Studies performed in atopic babies have also demonstrated the suitability of disposable wipes for use on compromised infant skin. After 4 weeks of product use in the previous study, results showed the wipes were at least as mild as the water and implement tested. More recent testing carried out by Bartels et al. in newborn infants showed that disposable wipes seemed to stabilize the skin barrier (in terms of TEWL) better than the use of water alone.

The results of this current study are consistent with those in previously published work comparing skin mildness of wipes. In this study, as well as in the previous study, the sites of both wipes treatments showed recovery rates similar to that at the site that was allowed to recover in the absence of any wiping, indicating that the use of disposable wipes led to minimal perturbation of the skin recovery process. By contrast, the cotton washcloth and water markedly perturbed the skin repair process, resulting in a TEWL value at day 5 that was substantially higher than those in all other conditions. The present authors are confident that this demonstrates the replicability of the test design and sensitivity of the model for showing differences between cleaning methods.

All pairwise comparisons were statistically significant for LS mean TEWL at day 5. However, not all statistical differences were considered clinically meaningful. For the purposes of this study, TEWL differences (between LS means) were considered clinically meaningful in terms of a change in skin barrier status only when the magnitude of difference was greater than the mean baseline TEWL (~7.4 g/m²/h in this study) of the subjects in the study. Therefore, differences equal to or larger than ~7.4 g/m²/h would be considered clinically meaningful, whereas smaller differences would not. Previous unpublished studies (internal studies in infants and adult forearm studies) have shown this level of difference in TEWL to be meaningful in comparison with visible erythema changes for skin barrier status (O’Connor R., Odio M; unpublished data, 1999, 2002).

Hence, using these criteria, the only differences considered clinically meaningful were between Wipes A and B, and the washcloth and water. Both Wipe A and Wipe B

produced significantly lower day 5 LM mean TEWL differences (~11.01 g/m²/h and ~8.92 g/m²/h, respectively) than the washcloth and water, indicating that both wipes were meaningfully gentler on skin.

Upon discontinuing the wiping procedure in this study, the site treated with the cotton washcloth and water achieved recovery comparable with the sites of the other treatments within 3 days (day 8). This result supports the hypothesis that the removal of the wiping procedure itself is the primary contributor to skin barrier repair at this site. Subsequent to this study, the test design was repeated using two disposable wipes that differed from those tested in this study, with consistent results.

The study shows that this tape-stripped FCAT, which models perturbation of barrier repair, shows promise as a possible surrogate model for assessing skin mildness of cleansing products that may be used in the care of premature infants. This conclusion is supported by the finding that starting TEWL values that approximate those in babies with a gestational age of 26–28 weeks can be achieved. Additionally, the design was able to rank order cleaning regimens and distinguish small differences in TEWL, which is consistent with previous testing, indicating that it is a sensitive and robust test model. This is further supported by similarities in data for the two wipes tested in this study, their consistent and comparable performance in the compromised skin model, and the knowledge that similar wipes tested in the NICU clinical study were proven to be mild and suitable for use in premature infants. Further, the study concludes that the disposable wipes meet a rigorous criterion pertaining to minimal interference with barrier repair and that differences in data from the sites submitted to treatment with the wipes and those derived from sites submitted to the cotton washcloth and water treatment were statistically significant, as is consistent with previous testing.

## Acknowledgments

The authors are grateful to Dr Marty Visscher, Independent Consultant, for helpful discussions and her invaluable perspective. We would also like to thank Dr Tom Youket and Dr Roger Gibb, Quantitative Sciences Department, The Procter & Gamble Company, for their assistance with statistical input and analysis.

## References


