J. Cosmet. Sci., 51, 153–167 (May/June 2000)

Development of a modified forearm controlled application test method for evaluating the skin mildness of disposable wipe products

MIRANDA A. FARAGE, *Baby Care Products, The Procter & Gamble Company, 11450 Grooms Road, Cincinnati, OH 45242.*

Accepted for publication March 15, 2000.

Synopsis

A modified forearm controlled application technique (modified FCAT) was developed to evaluate the relative mildness of baby wipe products. The test uses a semi-occlusive patch system to mimic the hydrated conditions common to the skin in the diaper area, interspersed with repeated product applications. The skin condition is scored for four endpoints: visual scoring for erythema and dryness, and instrumental scoring for redness (using a chromameter) and transepidermal water loss (TEWL by evaporimeter). The variables evaluated in the course of developing the method include the optimum patch system to achieve hydration, the specific procedure for applying the test products, and the minimum test duration required to achieve significant differences for all four endpoints. The optimum study design produces statistically significant differences between baby wipe test products in all four endpoints. This method has applicability for wipes intended for other uses where mildness and irritation are critical considerations, such as wipes intended for cosmetic removal.

INTRODUCTION

An important part of the safety assessment for new consumer products is an evaluation of potential skin effects. Initial evaluations may include patch tests on ingredients or on complete formulations for the potential to cause skin irritation or contact sensitization. For some consumer products, no additional testing is required. However, for products where expected use involves prolonged or repeated contact with the skin, additional testing is necessary to evaluate more subtle skin effects that may contribute to consumer acceptance of the product. Several tests have been developed over the years to evaluate household and personal cleansing products for potential irritant effects under conditions that mimic product use. Examples include the hand immersion test (1), the repeated open application test (2), the flex wash test (3), and the forearm controlled application technique (FCAT) (4).

We were interested in evaluating the skin effects of a very specialized consumer product, baby wipes. The intended use of baby wipes offers a unique challenge. The skin underneath a baby's diaper is almost always in a state of hydration as a result of the presence

of moisture and the semi-occluded nature of a diaper. Transepidermal water loss (TEWL) readings for baby's skin in the diaper area are markedly higher than TEWL readings for undiapered skin. As a result, diapered skin demonstrates a background level of irritation that can mask slight differences in the effects produced by different products. Clinical studies, where a panel of consumers actually use the products under evaluation for an extended period of time, have been used to study mildness and other skin effects. However, such studies for baby wipe products are extremely costly and time-consuming. Further, there are a number of factors that can confound the results, such as the appearance of diaper rash, heat rash, yeast infections, bacterial infections and other skin diseases, or variations in the use habits and practices of parents.

A practical, cost-effective test method was needed that would evaluate the relative mildness of various product formulations on skin that was in a state of hydration. Other investigators have been successful using the FCAT to determine the relative mildness of a variety of personal cleansing products (5). Therefore, we chose to develop a modified FCAT specifically for evaluating baby wipes.

The developmental program consisted of two phases run concurrently. The patch system development phase involved four small-scale tests designed to devise a practical and effective system for producing hydrated skin in the test site area. The two main variables examined were: (a) the use of synthetic urine versus water for hydration, and (b) the patch system used to achieve a semi-occluded environment. The FCAT modification phase involved three tests using baby wipe products. The main variables examined in this phase were: (a) the specific duration of test product application, (b) the need for a conditioning or pretreatment period, and (c) the minimum overall test duration and number of subjects required to achieve significant differences between products.

MATERIALS AND METHODS

PATCH DEVELOPMENT PHASE

Materials. Modified diaper patches were 4" × 6" rectangles cut from the core area of commercially available disposable diapers (Pampers Baby Dry®). These were saturated with 60 ml of fluid and applied to test sites by wrapping the diaper patch with Kerlix® mesh and securing it with 3M Blenderm® tape. Alternatively, patches consisted of Johnson & Johnson Band-Aid® Brand large adhesive pads (2⅞" × 4") saturated with 5 ml of fluid. In some experiments, the adsorbent pad of the bandage was supplemented with a 3" × 3" sterile gauze pad folded once, and/or the bandage adhesive was reinforced with Blenderm® tape. When the gauze pad was used together with the bandage, 9–10 ml of fluid was used to saturate the patch.

Synthetic urine consisted of 0.2% KCl, 0.2% Na_2SO_4, 0.085% $(NH_4)H_2PO_4$, 0.05% $MgCl_2 \cdot 6H_2O$, 0.025% $CaCl_2 \cdot 2H_2O$, and 0.015% $(NH_4)_2HPO_4$ in deionized water.

FCAT MODIFICATION PHASE

Materials. Four different currently marketed baby wipes were used in these experiments. These are referred to as products A, B, C, and D. All four of these products have been safely marketed for many years. Products A and B were Procter & Gamble brands. Internal, unpublished data available on these products indicate they are "very mild" in

a 21-day cumulative irritation test (6). In a four-day patch test (internal company protocol using semi-occluded patches on four consecutive days), these products are comparable to the negative irritant controls (water and isotonic saline). The products are non-irritating and are negative for sensitization and in human repeat insult patch tests (7).

Forearm washing protocol. In tests involving comparisons of baby wipe products, baby wipe application consisted of repeated washing cycles of the defined test sites with the test products. Washing sessions were 1½ hours apart, and were conducted four times each day, with the exception of the final day of the test, when one or two washing sessions were conducted. For each washing cycle, a baby wipe was folded twice, and wiped in a back and forth motion with light pressure on the test site for either 15 or 30 seconds. In some studies, the baby wipe was refolded to expose an unused portion, and the back and forth wiping was repeated. In some studies, the duration of the washing procedure for the last washing session of each day was doubled by using a second baby wipe and repeating the two-step washing procedure. The technicians conducting the washing procedure were instructed to take care to apply equal pressure to all test sites. A study of specific differences in the washing protocol are given below in the Results section.

BOTH PHASES

Subjects. For each study, the protocol was approved by the test facility's Institutional Review Board. Participants in all studies were healthy adult volunteers, ages 18–65, who had signed an informed consent form. Subjects were excluded from participation if: (a) they had allergies to soaps or fragrances, (b) they were taking anti-inflammatory corticosteriods or other medications that may interfere with test results, (c) they had active dermatitis, (d) they were pregnant or lactating, or (e) they had sunburn, acne, scar tissue, or other skin abnormality at the test site. In addition, for experiments during the FCAT modification phase, subjects were excluded if they had participated in a forearm patch test within four weeks prior to the study.

Initial studies used both male and female subjects. However, the relatively high amount of body hair on the forearms of the male subjects made visual scoring for erythema and dryness difficult. Therefore, the final two studies used female panelists only.

For each panelist, the forearms were divided into upper and lower test sites. These sites were at least 2 cm apart, and 2 cm from either the elbow or the wrist. Test patch systems or products were assigned to test sites using a randomization schedule. Panelists in each study were given the same brand of bar soap to use for personal cleansing and instructed not to use lotions, cremes, or moisturizers on the test sites, nor to wash the test sites during bathing throughout the duration of the test. In those experiments where modified diaper patches were used, subjects were instructed to put a plastic bag on the arm to cover the modified diaper patch while bathing or showering.

Evaluation of skin condition. In these experiments, transepidermal water loss (TEWL) was measured using a ServoMed Evaporimeter EP1® employing standardized procedures (8). This instrument concurrently measures evaporation rate (range 0–300 g/m^2hr), relative air humidity, and water vapor partial pressure. Readings were taken in triplicate. All measurements were taken after the subjects had spent a 30-minute acclimation period in the test facility's controlled temperature and humidity room. Specific room settings

varied slightly, but all were between 66° and 72°F, and 30% and 50% relative humidity. In all studies, individuals with a TEWL reading of >7.5 g/m^{2}hr at baseline (i.e., prior to the first patch application) were excluded from the study.

A Minolta Chromameter CR-200® was used to score redness. A complete description of the use of this instrument for similar applications has been given previously (9). Briefly, this instrument reads color in a three-dimensional format along coordinates L*, a*, and b*. Coordinate L* indicates the level of brightness between black and white. Coordinate a* indicates the red-green balance (with 100 being red only and −100 being green only). Coordinate b* indicates the yellow-blue balance. For this program, the reading along the a* coordinate was of greatest interest since higher a* values indicate greater irritation. Measurements were obtained by gently resting the aperture of the measuring probe on the skin surface. The L*a*b* values were directly inputted to a personal computer. Triplicate a* readings were taken at each test site.

Visual scoring was conducted by expert graders under a 100-watt incandescent daylight bulb. For each study, a single expert grader was used throughout the study. Persons scoring the test sites were unaware of the specific treatment at each test site. In the first study in the FCAT modification phase, dryness and erythema were graded on scales of 0–4, with "0" indicating perfect skin and "4" indicating a severe reaction. In all other studies, these scales were expanded to the seven-point scales given below.

Erythema grading scale

0 = none
1 = barely detectable redness
2 = slight redness
3 = moderate redness
4 = heavy or substantial redness
5 = severe redness
6 = extreme redness

Dryness grading scale

0 = none
1 = patchy, slight powderiness, small scales
2 = general, slight powderiness, small lifting scales
3 = general, moderate powderiness, cracking and scales
4 = general, heavy powderiness and cracking, lifting scales
5 = heavy cracking (possibly bleeding) and lifting scales
6 = severe scaling, bleeding cracks, sloughing of large scales

Fractional increments were assigned to slight differences or intermediate responses when two test sites were compared at a single grading point.

Statistical analyses of data. In the patch development phase studies, the Student's t-test was used to compare two treatments. When three or more treatments were being evaluated, an analysis of variance (ANOVA) was used. If the ANOVA indicated that the groups were significantly different, a Student's t-test between treatment pairs was conducted.

In the first experiment in the FCAT modification phase, instrumental measurements were evaluated by a repeated ANOVA run on the values obtained from the treated sites

compared to the non-treated control. If significant differences existed, multiple comparisons were done using Tukey's protected t-test, which allows one to examine differences between treatment pairs. Scores for erythema and dryness were analyzed by a Wilcoxon's matched-pairs test.

In the second experiment in the FCAT modification phase, Wilcoxon analysis was used to compare products A and C with regard to erythema, dryness, chromameter readings (redness), and TEWL.

In the third experiment in the FCAT modification phase, a Wilcoxon's signed rank test was used to compare erythema, dryness, and TEWL. Chromameter readings (redness) were analyzed using a two-sided paired t-test. In addition, analysis of covariance using baseline readings as the covariate was used to evaluate TEWL and chromameter readings.

RESULTS

PATCH DEVELOPMENT PHASE

The first three experiments in this phase were designed to determine an effective and practical way to achieve a hydrated environment at the skin test sites. In these experiments, either synthetic urine or distilled water was used in combination with various semi-occlusive patches worn overnight (approximately 20 hours) to achieve a hydrated environment at the skin site. The TEWL was measured at the test sites prior to patch application (baseline), immediately upon removal of the patch system, and 90 minutes after patch removal. In these experiments, an increase in TEWL over baseline levels was interpreted as an indication of skin barrier damage.

Distilled water versus synthetic urine. In experiments 1 and 2, the effects of water were directly compared to the effects of synthetic urine using two different patch systems: the modified diaper patch containing 60 ml of fluid, and the J&J bandage alone containing 9 ml of fluid. Several panelists lost patches during the night. In addition, subjects who took a morning shower had patches that became saturated, indicating a lack of occlusion. These subjects were excluded from the final data. TEWL readings immediately after patch removal for those subjects whose patches remained intact indicated that water and synthetic urine produced similar results using either the modified diaper patch or the J&J bandage (data not shown).

Patch system. In experiment 2, the modified diaper patch was compared to the J&J bandage used alone. Immediately after patch removal, the modified diaper patch moistened with water showed a mean TEWL reading of 15.00 ± 5.94 g/m^2hr. This was considerably higher than the reading produced by the J&J bandage (4.65 ± 1.77 g/m^2hr), indicating that the J&J bandage was less effective at hydrating the skin. In experiment 3 four modifications of the J&J bandage patch were compared: the J&J bandage alone, the J&J bandage secured with reinforcing tape, the J&J bandage supplemented with gauze, and the J&J bandage with gauze and tape (referred to as the J&J bandage system). The J&J bandage system resulted in the highest TEWL readings of all four modifications (17.15 ± 0.64 g/m^2hr). These readings were similar to those produced by the modified diaper patch.

Pilot for modified FCAT. Experiment 4 was an abbreviated modified FCAT designed (a)

to confirm the effectiveness of the J&J bandage system (i.e., J&J bandage with gauze and reinforcing tape containing 10 ml distilled water) compared to the modified diaper patch (containing 60 ml distilled water) in creating a hydrated skin environment, and (b) to provide preliminary data that the test system will distinguish between two baby wipe products that differ in skin mildness. This three-day experiment involved four washing sessions with test products A and C on days 1 and 2, and a single washing session on day 3. All washing sessions were 30 seconds except the final session on each day, which was 60 seconds. After the final washing session on days 1 and 2, patches were applied to test sites and worn overnight. Erythema was scored prior to the first washing session on all days, prior to the fourth washing session on days 1 and 2, and at the completion of the study. TEWL readings were taken prior to the first washing session on day 1, after the second washing session on day 2, and at completion of the study.

Results of this experiment are presented in Table I. For Product C, the three-day duration of this experiment was sufficient to produce statistically significant differences in the TEWL readings at both the intermediate and final readings when compared to the baseline readings with the modified diaper patch. When the J&J bandage system was used, the TEWL readings increased with time, but the differences were not statistically significant. For Product A, there were no significant differences between intermediate or final readings compared to the baseline readings with either the modified diaper patch or the bandage patch system.

When the scores for product A were compared to those for product C, there were no significant differences between products at any of the three time points (baseline, intermediate, or final) with either patch system.

FCAT MODIFICATION PHASE

The experiments in this phase were designed to examine the specific duration of test

Table I
Comparison of Two Baby Wipe Products in the Three-Day FCAT Pilot With Hydration of Test Sites Using Two Patch Systems

	TEWL (mean g/m^2 hr ± SD)					
	Modified diaper patch with 60 ml water			J&J bandage system with 10 ml water		
	Baseline (n = 5)[a]	Intermediate (n = 5)	Final (n = 4)	Baseline (n = 5)	Intermediate (n = 5)	Final (n = 4)
Product A	6.2 ± 0.84	5.4 ± 1.67	6.75 ± 1.50	5.8 ± 0.84	6 ± 0.71	6.75 ± 0.50
Product C	4.8 ± 1.64	5.8 ± 0.45	7.25 ± 0.50[b]	4.8 ± 1.3	6.2 ± 1.48	6.5 ± 1.29

This three-day experiment compared products A and C using the modified diaper patch and the J&J bandage system. One test site on each arm was subjected to four washing sessions with either test product A or C on days 1 and 2, and a single washing session on day 3. After the final washing session on days 1 and 2, patches were applied to test sites and worn overnight. The modified diaper patch was applied to one test site for each product, and the J&J bandage system was applied to the other site. TEWL readings were taken prior to the first washing session on day 1 (baseline), after the second washing session on day 2 (intermediate), and at completion of the study (final).

[a] Number of panelists.

[b] Significantly different from baseline (p = 0.025) and intermediate (p = 0.005) using a Student's t-test.

product application in the FCAT, the need for a conditioning or pretreatment period, and the overall test duration required to achieve significant differences between products. In addition, the final experiment in this phase incorporated the optimum patch system (the J&J bandage system) developed in the previous studies.

Experiment 1—FCAT without hydration. The objective of this experiment was to confirm that the FCAT was sensitive enough to detect mildness differences in baby wipe products on normal (non-hydrated) skin. The effects of three different products were compared to an untreated control. In this study, panelists underwent a one-week conditioning period prior to the start of the test during which they were asked to stop using any lotions, cremes, moisturizers, soaps, or other products on the test sites. Following the one-week conditioning period, the FCAT ran for five days. Four washing sessions were conducted on days 1–4, and two were conducted on day 5. Each washing session was a total of 30 seconds, with the exception of the final cycle on each day, which was 60 seconds.

Instrumental measurements were conducted prior to the first washing session on day 1, and after the final washing session on day 5. Visual scoring for erythema and dryness was conducted prior to the first and third washing sessions on days 1–4, prior to the first washing session on day 5, and at the completion of the study (following the second washing session on day 5). Five-point scales (0–4) were used for erythema and dryness. The washing procedure was terminated at any test site with a visual score for either erythema or dryness of ≥ 3. If an early termination was necessary, instrumental measurements were conducted at that site at the time of termination. These measurements were carried through to the end of the study.

Table II gives the final visual and instrumental scores. None of the untreated control sites or sites treated with Product A required early termination. For sites treated with products C and D, early terminations were required for 14% and 46%, respectively, of the 31 panelists who completed the study. For all four endpoints, product C was significantly less mild than product A. In addition, product D was less mild than

Table II
Comparison of Three Baby Wipe Products on Non-Hydrated Skin in the Five-Day FCAT

Treatment	Early terminations[a]	Final visual scores		Final instrumental scores	
		Erythema	Dryness	Redness (a*)	TEWL (g/m^2 hr)
Untreated	0%	0.11 ± 0.24	0.15 ± 0.23	9.2 ± 1.2	3.16 ± 1.2
Product A	0%	0.3 ± 0.31	0.23 ± 0.26	9.4 ± 1.2	4.5 ± 1.6
Product C	14%	1.28 ± 0.89[b]	0.51 ± 0.29[c]	10.8 ± 1.6[d]	10.7 ± 4.8[d]
Product D	46%	2.13 ± 0.92	0.62 ± 0.34	12.5 ± 1.9[e]	15.9 ± 7.1[e]

This five-day study compared products A, C, and D on non-hydrated (normal) skin only. Four washing sessions were conducted on days 1–4, and two were conducted on day 5. Data for the final scores are given above (mean ± SD). 31 panelists completed the study.

[a] The washing procedure was terminated due to a visual score of ≥ 3 at the test site.

[b] Significantly different from Product A ($p < 0.0001$) using Wilcoxon's matched-pairs test.

[c] Significantly different from Product A ($p < 0.001$) using Wilcoxon's matched-pairs test.

[d] Significantly different from untreated control, and Products A and D ($p < 0.01$) using Tukey's protected t-test.

[e] Significantly different from untreated control, and Products A and C ($p < 0.01$) using Tukey's protected t-test.

Table III
Comparison of Two Baby Wipe Products in a Five-Day FCAT With and Without Hydration of the Test Sites

	Final visual scores				Final instrumental scores			
	Erythema		Dryness		Redness (a*)		TEWL (g/m^2 hr)	
	Normal	Hydrated	Normal	Hydrated	Normal	Hydrated	Normal	Hydrated
Product A	0.47 ± 0.11	1.42 ± 0.16	1.03 ± 0.15	1.91 ± 0.21	8.51 ± 0.16	9.06 ± 0.15	3.93 ± 0.25	3.42 ± 0.20
Product C	0.97 ± 0.16[a]	1.67 ± 0.17	1.70 ± 0.11[a]	2.23 ± 0.16	9.13 ± 0.18[a]	9.70 ± 0.20	5.10 ± 0.39[b]	6.23 ± 0.72[b]

This five-day study compared products A and C on normal skin, and on skin that had been hydrated during a four-day pretreatment period. During the four-day pretreatment, panelists wore patches consisting of the modified diaper patch on one arm only. Following the pretreatment phase, one test site on each arm was subjected to four washing sessions with either test product A or C on days 1–4, and a single washing session on day 5. After the final washing session on days 1–4, patches were applied to the pretreated arm only and worn overnight. Final scores are given above as the mean ± SD. 32 panelists completed the study.

[a] Significantly different from Product A ($p = 0.01$) using Wilcoxon analysis.

[b] Significantly different from Product A ($p < 0.001$) using Wilcoxon analysis.

Table IV
Comparison of Two Baby Wipe Products in a 14-Day FCAT With and Without Hydration of the Test Sites

	Final visual scores				Final instrumental scores			
	Erythema		Dryness		Redness (a*)		TEWL (g/m^2 hr)	
	Normal	Hydrated	Normal	Hydrated	Normal	Hydrated	Normal	Hydrated
Product B	0.48 ± 0.09	1.38 ± 0.13	0.39 ± 0.09	0.52 ± 0.10	8.02 ± 0.17	8.55 ± 0.20	6.76 ± 0.23	8.89 ± 0.34
Product C	1.28 ± 0.12[a]	2.13 ± 0.13[a]	0.80 ± 0.12[a]	0.99 ± 0.12[b]	8.89 ± 0.22[c]	9.30 ± 0.22[c]	8.70 ± 0.33[a]	9.60 ± 0.42[d]

This 14-day study compared products B and C on both normal and hydrated skin. One test site on each arm was subjected to four washing sessions with either test product B or C on days 1–13, and two washing session on day 14. After the final washing session on days 1–13, patches consisting of the J&J bandage system were applied to test sites on one arm only and worn overnight. Data for the final scores are given above as the mean ± SD. 82 panelists completed the study.

[a] Significantly different from Product B ($p = 0.0001$) using Wilcoxon analysis.

[b] Significantly different from Product B ($p < 0.0005$) using Wilcoxon analysis.

[c] Significantly different from Product B ($p = 0.0001$) using a two-sided paired t-test.

[d] Significantly different from Product B ($p < 0.005$) using Wilcoxon analysis.

product A for the instrumental measurements. Sites treated with product A were not significantly different from the untreated control sites.

Experiment 2—FCAT with and without hydration. The objective of this experiment was to determine the effect of hydration on the ability of the FCAT to detect product mildness differences. In this study, panelists underwent a four-day pretreatment period during which patches were applied to test sites designated for semi-occlusion. Patches consisted of the modified diaper patch with 60 ml distilled water. The washing cycles were conducted on days 1 through 5 in a manner identical to those in Experiment 1. Patches were applied nightly after the final wash cycle of each day on days 1 through 4. Panelists were instructed to remove the patches at least one hour prior to arriving for the first washing cycle of each day. Instrumental and visual scoring was conducted at the same time points as in the previous experiment. Additional scoring was conducted prior to the pretreatment period. Visual scoring for erythema and dryness were done using seven-point scales (0–6), as described in the Materials and Methods section. The washing procedure was terminated at any test site with a visual score for either erythema or dryness of >3. If an early termination was necessary, instrumental measurements were conducted at that site at the time of termination.

Results are shown in Table III. All effects were increased when skin was hydrated through semi-occlusion except water loss for test product A. As in the first experiment, Products A and C were significantly different in all four parameters on the normal (non-occluded) test sites. On the hydrated sites, only water loss was significantly different. However, Product C was directionally worse (i.e., less mild) than A for the remaining three parameters.

Experiment 3—Extended FCAT with and without hydration. The objective of this experiment was to extend the duration of the study in an attempt to achieve statistically significant differences in all parameters. There was no conditioning or pretreatment period in this experiment. In order to better reflect actual consumer habits and practices for use of baby wipe products, the washing cycles were reduced to 15 seconds each, except for the final washing cycle on each day, which was 40 seconds. Washing cycles were conducted four times daily for 13 consecutive days, and twice on day 14. Hydration was achieved with the J&J bandage system and 10 ml distilled water. The patch was applied nightly, and removed one hour prior to the first wash of the next day.

Instrumental measurements were conducted prior to the first washing session on day 1, after the second washing session on day 5, and after the final washing session on day 14. Visual scoring for erythema and dryness was conducted prior to the first and third washing sessions on days 1–13, and prior to both the first and second washing sessions on day 14. Seven-point scales (0–6) were used for erythema and dryness. The washing procedure was terminated at any test site with a visual score of ≥ 5. If an early termination was necessary, instrumental measurements were conducted at that site at the time of termination.

Results are shown in Table IV. All effects were increased when skin was hydrated through semi-occlusion. The two products were significantly different in all four parameters on both normal (non-occluded) and hydrated (semi-occluded) test sites, with Product C being less mild than Product B.

Substantial differences in intermediate scores were observed throughout this study. An example of the intermediate erythema scores is shown in Figure 1. The intermediate

scores for dryness showed a similar pattern. One goal of this program was to develop a practical, cost-effective test method for measuring mildness differences on hydrated skin. Therefore, statistical analyses were conducted on the intermediate scores to determine if an experiment of shorter duration would provide statistically significant results for all endpoints.

Recorded scores at all time points starting with day 5 (scoring time point 10) in the 14-day experiment were analyzed. A summary of the days that yielded statistically significant results is shown in Table V. As shown in this table, the study consistently yielded statistically significant results for all endpoints except dryness on both normal and hydrated skin at all time points starting at day 5. After nine days, the dryness endpoint was also significantly different. The instrumental measurements were conducted on days 5 and 14, and were significantly different on both days.

DISCUSSION

A Modified FCAT has been developed that is capable of producing statistically significant differences between different baby wipe products in erythema, dryness, redness, and TEWL on both hydrated and normal skin. A patch system was developed for hydrating the skin that was convenient, cost-effective, and well accepted by the panelists. The first three experiments in the patch system development phase demonstrated that distilled water was roughly equal in effectiveness to synthetic urine in producing a hydrated environment, and a patch consisting of the J&J bandage with gauze and Blenderm tape provided adequate hydration when compared to the other patch systems tested, including the modified diaper patch.

Panelist acceptance was an important consideration for the patch system. Comments were solicited from the panelists with regard to the comfort and wearability of the various patch systems. The modified diaper patch was not well received. It was considered too large and unsightly for a test requiring several days. The 60 ml of fluid required to saturate the patch made the patch too heavy to be worn comfortably. Bathing with the modified diaper patches resulted in patches soaked with water. Not only did this contribute to panelist discomfort, it resulted in spuriously high TEWL readings, necessitating exclusion of some subjects from the final analysis. In addition, the modified diaper patches had to be hand cut from commercial diapers, which added considerably to the overall study costs.

The patches using the J&J bandage were well received by the panelists; however, without reinforcing tape these did not stay in place overnight for some, resulting in several panelists being dropped from these study groups. The gauze was a necessary addition in order to achieve a skin environment similar to that achieved with the modified diaper patches.

In the FCAT modification phase, the study summarized in Table II confirms that the FCAT has enough sensitivity to detect differences in the skin effects of products on normal (non-hydrated) skin in an experiment of only five days' duration. However, we expected that differences in mildness would be more challenging to detect on hydrated skin due to the background level of irritation. This increased level of irritation is demonstrated in Tables III and IV. In every instance but one (TEWL scores for Product

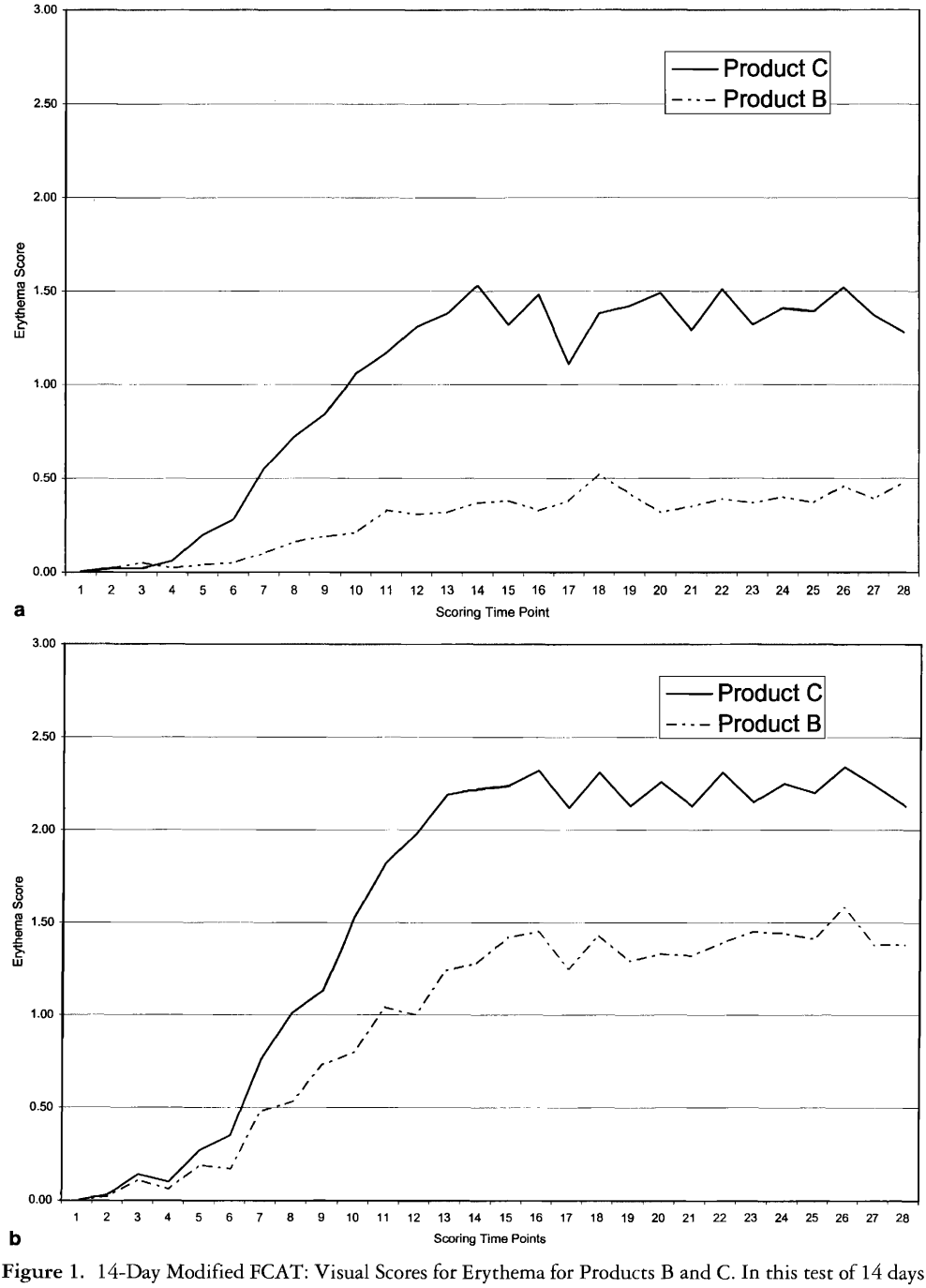

Figure 1. 14-Day Modified FCAT: Visual Scores for Erythema for Products B and C. In this test of 14 days' duration, products were compared on both normal (a) and hydrated (b) skin. Visual scoring was conducted twice daily (i.e., prior to the first and third washing cycle of each day). Therefore, scoring time points 1 and 2 occurred on day 1, scoring time points 3 and 4 occurred on day 2, etc. Data for the final scores for all four parameters (erythema, dryness, redness, and TEWL) are given in Table IV.

B in Table IV), hydrated skin showed higher scores for all four endpoints when treated in the same manner as normal skin.

The experiment summarized in Table III confirms that differences in skin effects are more difficult to detect on hydrated skin. In this experiment five days' duration, we were once again able to detect significant differences in all four endpoints on the normal skin sites. However, on the hydrated skin sites, only the TEWL measurements were significantly different. The other three endpoints were directionally different, but did not reach significance.

By increasing the overall duration of the test to 14 days, as in the experiment summarized in Table IV, we were able to achieve significant differences in all four endpoints on the hydrated skin sites. These differences were detected in spite of the fact that the duration of each product application was reduced to better reflect the actual consumer habits and practices of product use.

Eliminating the pretreatment period did not appear to compromise the ability to hydrate the skin sites, as demonstrated by comparing the normal to hydrated value for each product in the 14-day study presented in Table IV, where no pretreatment was used. For all four endpoints, the hydrated skin showed an increase in irritation.

A final step in development of this method was to determine the optimum test duration and number of panelists required to achieve significant results. It was possible that a test of 14 days using 80 subjects, such as the one presented in Table IV, was more than absolutely necessary.

Statistical analyses of the intermediate points from the experiment presented in Table V indicated that the duration of the test could be reduced to nine days and still produce significant differences between these two products for all endpoints. Subsequent studies have indicated that a test of seven days' duration involving approximately 60 subjects provides consistently reliable results.

Table V
Summary of Statistical Differences Between Products B and C At Different Time Points in the 14-Day Modified FCAT

	Day 5	Day 7	Day 8	Day 9	Day 14
Normal skin					
Erythema	Yes[a]	Yes	Yes	Yes	Yes
Dryness	No	Yes	Yes	Yes	Yes
Redness	Yes	—	—	—	Yes
TEWL	Yes	—	—	—	Yes
Hydrated skin					
Erythema	Yes	Yes	Yes	Yes	Yes
Dryness	Yes	No	No	Yes	Yes
Redness	Yes	—	—	—	Yes
TEWL	Yes	—	—	—	Yes

This is a summary of significant differences at the interim scoring time points for the 14-day study presented in Table IV. Statistical significance for erythema, dryness, and TEWL readings was determined using a Wilcoxon's signed rank test. Statistical significance for chromameter readings was determined using a two-sided paired t-test.

[a] "Yes" indicates a significant difference between Products B and C ($p < 0.05$). "No" indicates no significant difference.

Interestingly, a shorter duration test, such as the five-day test presented in Table III, may be adequate to provide a quick indication of mildness differences to aid in decisions during product formulation. Although the five-day test did not yield significant differences for all endpoints, the results demonstrated directional differences.

Most of these experiments used both visual and instrumental scoring methods. This provided an opportunity to compare the ability to detect differences when visual scoring was used versus instrumental scoring. In all cases, visual scoring was qualitatively similar to instrumental scoring, indicating that instrumental scoring may not be necessary to provide an indication of the relative mildness differences between products. This would be particularly useful for quick screening studies, where a complete understanding of mildness and irritating effects is not necessary.

CONCLUSIONS

- A modified FCAT has been developed that can consistently detect differences in mildness for baby wipe products under hydrated skin conditions such as those found in the diaper area. The test should be of nine days' duration, and use 60–80 female panelists. Washing cycles in the modified FCAT should be conducted four times daily for six to eight consecutive days, and twice on the final day. The washing cycles should be of 15-second duration except for the final washing cycle on each day, which should be 40 seconds. Hydration can be achieved with nightly patches consisting of the J&J bandage with gauze, reinforcing tape, and 10 ml of distilled water. No conditioning or pretreatment period is necessary. Four endpoints should be evaluated: TEWL, redness by chromameter, and erythema and dryness by visual scoring on seven-point scales.
- This method has applicability for wipes intended for other uses where mildness and irritation are critical considerations, such as wipes intended for cosmetic removal.
- Subjective measurements provide adequate endpoints in the absence of instrumental measurements. Significant differences among visual scores were very consistent with those observed with the instrumental scores.
- An abbreviated test design of five days' duration using visual scoring alone would be a cost-effective approach toward providing an indication of mildness differences to aid in formulation decisions. This abbreviated test can be followed by the more extensive study design given above for a complete understanding of mildness differences and for product claim support.

ACKNOWLEDGMENTS

The author is grateful to Ms. Sandy Meyer and Ms. Debbie Gilpin for their excellent technical support, to Sue Baldwin, PhD, for her scientific input into this program, to Celeste Kuta, PhD, for her technical input and review of the manuscript, and to Terresa L. Nusair, PhD, of Risk Assessment and Toxicology Services, Inc. (Cincinnati, Ohio) for the preparation of this manuscript.

REFERENCES

(1) D. J. Kooyman and F. H. Snyder, Tests for mildness of soap, *Arch. Dermatol.*, 46, 846–855 (1942).

(2) M. Hannuksela and H. Salo, The repeated open application test (ROAT), *Contact Dermatitis,* 14, 221–227 (1986).
(3) D. D. Stube, S. W. Koontz, R. I. Murahata, and R. F. Theiler, The flex wash test: A method for evaluating the mildness of personal washing products, *J. Soc. Cosmet. Chem.,* 40, 297–306 (1989).
(4) M. F. Lukacovic, F. E. Dunlap, D. E. Michaels, M. O. Visscher, and D. D. Watson, Forearm wash test to evaluate the clinical mildness of cleansing products, *J. Soc. Cosmet. Chem.,* 39, 355–366 (1988).
(5) K. D. Ertel, B. H. Keswick, and P. B. Bryant, A forearm controlled application technique for estimating the relative mildness of personal cleansing products, *J. Soc. Cosmet. Chem.,* 46, 67–76 (1995).
(6) L. Phillips, M. Steinberg, H. I. Maibach, and W. A. Akers, A comparison of rabbit and human skin response to certain irritants, *Toxicol. Appl. Pharmacol.,* 21, 369–382 (1972).
(7) J. Stotts, "Planning, Conduct and Interpretation: Human Predictive Sensitization Patch Tests," in *Current Concepts in Cutaneous Toxicity,* V. A. Drill and P. Lazar, Eds. (Academic Press, New York, 1980), pp. 41–53.
(8) J. Pinnagoda, R. A. Tupker, T. Agner, and J. Serup, Guidelines for transepidermal water loss (TEWL) measurements: A report from the standardization group of the European Society of Contact Dermatitis, *Contact Dermatitis,* 22, 164–178 (1990).
(9) S. W. Babulak, L. A. Rhein, D. D. Scala, F. A. Simion, and G. L. Grove, Quantitations of erythema in a soap chamber test using the Minolta Chroma (Reflectance) Meter: Comparisons of instrumental results with visual assessments, *J. Soc. Cosmet. Chem.,* 37, 475–479 (1986).