Table 2. Anatomic distribution and incidence of BCCs in PUVA cohort 1990–1999 and Queensland (Australia) 1997–2006

	PUVA cohort		Australia[1]	
Anatomical site	No. of tumors (%)	Incidence rate[2]	No. of tumors (%)	Incidence rate[2]
Head and neck	488 (51)	122	379 (57)	29
Trunk	284 (30)	71	111 (17)	9
Upper extremities	77 (8)	19	118 (18)	9
Lower extremities[3]	108 (11)	27	54 (8)	4
All sites	959	239	662	51

BCC, basal cell cancer; PUVA, psoralens plus UVA.
[1]Richmond-Sinclair *et al.* (2008).
[2]BCCs per 1,000 person-years.
[3]Includes buttocks and hip.

started PUVA treatment by the age of 25 years and were <40 years in the 1990s (mean age starting PUVA = 19 years) had a significantly higher risk of BCC than Nambour patients of comparable age (IRR (incidence rate ration) = 2.64, 95% CI (95% confidence interval) = 1.04–6.70).

When each tumor is counted, the incidence of tumors was far higher in the PUVA cohort than in the Nambour cohort (age-adjusted incidence 239 vs 51 per 1,000 patient-years, $P<0.001$). The incidence of BCC (tumor counts) was significantly higher in patients with more than 200 PUVA treatments than those with fewer treatments (IRR = 3.1, 95% CI = 2.74–3.54).

Table 2 provides the anatomic distribution of tumors in both cohorts. The incidence of BCC on the trunk and lower extremities was more than six times higher for PUVA patients than for Australians.

In contrast to the far lower incidence of BCC in the United States' general population, North American patients who use PUVA and live in temperate climates have at least as high a risk of developing at least one BCC as seen in a fairer-skinned population living in a subtropical Australia. The greatest increase in risk was seen among those who started PUVA treatment before age 25 years. The average number of tumors per patient who developed at least one BCC was about three times higher for PUVA patients than for Australians (Tables 1 and 2). Earlier reports suggest that the increase in BCC among Australians is nearly 10 times that of residents in United States, a finding markedly different from the experience of the PUVA cohort reported here (Stern, 1999). A comparison of our data and population-based data from Australia provides a context for informing clinicians and advising patients about the long-term risk of BCC with PUVA, particularly for those exposed when young.

CONFLICT OF INTEREST

The author state no conflict of interest.

Robert S. Stern[1]

[1]*Department of Dermatology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA*
E-mail: rstern@bidmc.harvard.edu

REFERENCES

Ahmad O, Boschi-Pinto C, Lopez AD, Murray CJL, Lozano R, Inoue M *et al.* (2001) *Age standardization of rates: a new WHO standard.* Geneva: World Health Organization

Gallagher RP, Hill GB, Bajdik CD, Fincham S, Coldman AJ, McLean DI *et al.* (1995) Sunlight exposure, pigmentary factors and risk of nonmelanocytic skin cancer. I. Basal cell carcinoma. *Arch Dermatol* 131:157–63

Nijsten TE, Stern RS (2003) The increased risk of skin cancer is persistent after discontinuation of psoralen+ultraviolet A: a cohort study. *J Invest Dermatol* 121(2):252–8

Richmond-Sinclair N, Pandeya N, Ware RS, Neale RE, Williams GM, van der Pols JC *et al.* (2008) Incidence of basal cell carcinoma multiplicity and detailed anatomic distribution: longitudinal study of an Australian population. *J Invest Dermatol* 129(2):323–8

Stern R (1999) The mysteries of geographic variability in nonmelanoma skin cancer incidence. *Arch Dermatol* 135:843–4

Tumorigenic Effect of Moisturizing Creams in UVB-Pretreated High-Risk Mice

Journal of Investigative Dermatology (2009) **129,** 2316–2318; doi:10.1038/jid.2009.158

TO THE EDITOR

As a biostatistician, I read the article by Lu *et al.* (2008) with concern. I wish to raise several issues, some echoed by other commentators (Ellefson, 2009; Staeb *et al.*, 2009).

ISSUE 1

I have questions about the design of the study. Was randomization used to select treatment groups? If so, it is important that randomization was carried out at the end—not the beginning—of the pre-treatment period. If mice were separated into groups before this time, it is possible that this part of the experiment induced differences before treatment application.

ISSUE 2
Were the investigators, who adjudicated the occurrence, number, and size of tumors, blinded? This is particularly important given apparent variation across identical experiments.

ISSUE 3
Apparently not all mice were available for assessment (Tables 2–5). For example, only 25 mice had results in the Custom Blend group, suggesting that 17% of the mice either never started on treatment or were lost to follow-up. What was the reason for exclusions? (These mice were assumed to have died by Ellefson, 2009). Removing mice because of tumor development and/or death during pre-treatment biases subsequent comparisons if exclusions occurred unequally across groups. Similar bias arises if mice were removed after treatment because of adverse events possibly related to the risk of tumor formation. In either case, groups with larger number of removals might then show artificially low rates of tumor formation with high-risk mice selectively excluded. This is a concern in Experiment 2, where the largest number of removals came from the Untreated and Custom blend groups, on the lower end of outcome measures.

ISSUE 4
It is not clear to me why the cumulative incidence curves for the Untreated and Dermovan groups (Figures 1a and d) decline from weeks 9–13, impossible for Kaplan–Meier estimators. Further, plotted numbers in Figure 1c do not always agree with Table 4, the latter reporting cumulative incidence of 77% for the Water control group where the plotted point (Figure 1c) is above 80%.

ISSUE 5
There are no apparent differences in percent of mice developing tumors, except for Dermabase (Figure 1c) and Dermovan (Figure 1d) comparisons with Water control (Experiment 2). The *P*-value 0.003 associated with Figure 1c, presumably from a log-rank test, is surprising, given cumulative incidences in the Dermabase, and Water control, groups of 90 and 77%, respectively (Table 4). A simple 2×2 table comparison of these incidences among 29 and 30 mice, respectively, produces a *P*-value 0.2 using the standard χ^2-test. This asymptotic test depends on large samples (does as log-rank testing); the exact test, not dependent on such approximations, yields a *P*-value 0.3. However, neither of these tests uses occurrence time information used by log-rank calculations. Approximately reconstructing tumor incidence over time (Figure 1c), a log-rank test yields a *P*-value 0.06. Ignoring the variation among these various *P*-values, none of them are compatible with 0.003.

ISSUE 6
The variables, number of tumors, and tumor volume per mouse are not explicitly defined. Apparently, ''tumors per mouse'' includes mice with no tumors (Table 4), as now confirmed by the authors (Conney *et al.*, 2009). Comparison of the mean of this random variable across groups is therefore influenced by differences in percentages of mice that ever develop tumors, rendering comparative interpretation of ''tumors per mouse'' problematic. For example, the ''tumors per mouse'' comparison between the Dermabase and Custom blend groups of Experiment 2 (Table 4) uses averages 7.52 and 4.88, respectively. The average number of tumors per mouse, given that at least one tumor developed, is actually 8.4 for Dermabase, and 6.8 for Custom Blend. These numbers (with standard deviations approximately 1.2) are substantially closer, suggesting that comparisons of ''tumors per mouse'' are overstated. An identical comment applies to ''tumor volume per mouse'' (Table 5).

ISSUE 7
Further, can the authors discuss the adequacy of the Poisson model for the number of tumors per mouse (Conney *et al.*, 2009), as this is a very restrictive statistical assumption (more restrictive than the two-sample *t*-test used by Staeb *et al.*, 2009), particularly an issue when structural zeros (the mice with no tumors) are automatically included with positive counts.

ISSUE 8
In Tables 4–5, the Water and Untreated ''control'' groups are combined, presumably to allow for additional precision. However, other commentators have raised the possibility of similarly combining the Dermabase groups (and, similarly, the untreated groups) from Experiments 1 and 2, apparently identical experiments. Comparing the results, the average number of tumors per mouse under Dermabase is 12.39 (SE = 1.53) in Experiment 1 (Table 2), but only 7.52 (SE = 1.18) in Experiment 2 (Table 4), a difference of about 4 SEs. This between-experiment variation is much greater than the variation between the Dermabase and Water (or Untreated) groups in Experiment 2, and is statistically significant: two apparently equivalent experiments produce noticeably different results. This source of variation is not accounted for by the separate analyses of Experiments 1 and 2.

ISSUE 9
Incorporating between-experiment variation is desirable, likely substantially increasing comparative *P*-values. In one response, the authors note that the differences between experiments were attributable to them being conducted ''more than a year apart'' (Conney *et al.*, 2009). Unfortunately, this does not directly address the reproducibility of Experiment 2; without further details, one must assume that differences between the experiments are because of random variation.

ISSUE 10
It would be helpful to know more explicitly the random effects model used for analysis and further assumptions about additional within-mouse correlation assumptions. As no intercepts were used, what random slopes were considered? A specific choice of random effects necessarily induces a particular within-mouse correlation structure. I would have preferred standard regression techniques with a ''working'' within-mouse correlation structure, with estimation using generalized estimating equations, as coefficient estimates are more robust than those from mixed effects models, where unverified assumptions can substantially influence results.

ISSUE 11

Estimated percent increases (in parentheses, Tables 2–5) are highly variable representing ratios of quantities that themselves vary substantially. As such, it would be helpful to see confidence intervals reported (including allowance for between-experiment variation). This issue arises in the Abstract, where estimated percent increases are reported without uncertainty information. This is particularly crucial as quite different increases (or decreases) arise with a different ''control'' comparison. For example, the appropriate average number of tumors per mouse (if tumors occur—see above) for the Eucerin group (Experiment 2) is 7.43, and 8.04 for the Untreated group of Experiment 1. This ''just-as-appropriate'' comparison yields a decrease in number of tumors per mouse of 8%, not the increase of 24% reported in the Abstract.

ISSUE 12

Finally, is it not possible that any apparent difference between treatment groups in the number (volume) of tumors observed might be because of effects after at least one tumor has occurred—the latter attribute barely varying across groups—rather than arising from any mutagenic or carcinogenic effects? This concern reinforces what other commentators have noted about the misleading nature of the article's title (Ellefson, 2009; Staeb *et al.*, 2009).

There are many fine experimental papers in the literature that generate similar data and that successfully handle most of the issues raised here. See, for example, the study by Lerche *et al.*, 2008. Unfortunately, given the concerns raised above, I join previous correspondents in asserting that the authors' interpretations of their results are not scientifically justifiable.

CONFLICT OF INTEREST

The author has previously consulted with Nixon Peabody LLP (less than 10 hours) who, he believes, represent Beiersdorf Inc. in the United States.

***Nicholas P. Jewell*[1]**

[1]*Division of Biostatistics, University of California, School of Public Health, Berkeley, California, USA*
E-mail: jewell@berkeley.edu

REFERENCES

Conney AH, Lu Y-P, Shih WJ (2009) Response to Staeb *et al. J Invest Dermatol* 129:517–8

Ellefson RD (2009) Comment on ''Tumorigenic effect of some commonly usedmoisturizing creams when applied topically to UVB-pretreated high-risk mice''. *J Invest Dermatol* 129:513–4

Lerche CM, Philipsen PA, Sehested M, Wulf HC (2008) Photocarcinogenesis of topical tazarotene and isotretinoin alone and in combination with valproic acidin hairless mice. *Experiment Dermatol* 17:972–4

Lu YP, Lou YR, Xie JG, Peng Q, Shih WJ, Lin Y *et al.* (2008) Tumorigeniceffect of some commonly used moisturizing creams when applied topically toUVB-pretreated high-risk mice. *J Invest Dermatol* 129: 468–75

Staeb F, Gerber E, Kolbe L (2009) Comment on ''Tumorigenic effect of somecommonly used moisturizing creams when applied topically to UVB-pretreated high-risk mice''. *J Invest Dermatol* 129:515–6

Response to Jewell

Journal of Investigative Dermatology (2009) **129,** 2318–2321; doi:10.1038/jid.2009.160

TO THE EDITOR

Dr Nicholas P. Jewell expresses several concerns about our publication ''Tumorigenic effect of some commonly used moisturizing creams when applied topically to UVB-pretreated high risk mice'' (Lu *et al.*, 2008). Many of these concerns were raised earlier (Ellefson, 2009; Staeb *et al.*, 2009) and answered (Conney *et al.*, 2009a; Conney *et al.*, 2009b). We answer Dr Jewell's concerns as follows:

RESPONSE TO ISSUE 1

All randomizations were carried out after the completion of the UVB pretreatment. It might not be clear or explicit enough, but we did imply this order in our paper (Materials and Methods, descriptions of Experiments 1 and 2) by stating ''In experiment 1, we treated 60 female SKH-1 mice with UVB … for 20 weeks. … UVB irradiation was stopped, and half of the mice were treated with 100 mg Dermabase once a day, …for 17 weeks, and the control group was untreated.'' A similar description was given for experiment 2. We add that this is carried out routinely in our laboratory, as indicated in our earlier publications. See *Proc Natl Acad Sci USA* 99:12455–12460, 2002 and *Carcinogenesis* 28:199–206, 2007 (cited in our manuscript) as well as other papers from our laboratory. For experiment 1, we randomized 58 UVB-pretreated tumor-free mice (29 in each group). In experiment 2, we randomized all 210 UVB-pretreated tumor-free mice (30 per group).

RESPONSE TO ISSUE 2

Although the topical treatment of mice and the collection of data for Figure 1 were not blinded, they were performed by a lab technician (J.-G. Xie) who had no scientific interest in the results. The histological evaluation of tumors (data in Tables 2–5) was blinded so that the histologist did not know which group or mouse was being evaluated, as we have performed in earlier studies (see *Carcinogenesis* 28:199–206, 2007; cited in our manuscript).

We should point out that experiments 1 and 2 were similar in their procedures, but they were not identical. In fact, they were separated by more than a year and one was a pilot study, whereas the other was confirmatory (see additional response later). Although Dr Jewell believes there was variation between our two experiments, both experiments demonstrated a tumorigenic effect of Dermabase in UVB-pretreated mice (discussed later).