ANNEX 4
Biometrical Data Analysis (Phase III)

ECVAM Prevalidation Project:

EpiDerm™ Phototoxicity Test

Biometrical Data Analysis

Responsible Biostatistician:

PD Dr. H.G. Holzhütter
Humboldt-University Berlin, Medical School (Charité), Institut of Biochemistry
Monbijoustr. 2A, D-10117 Berlin
Phone: ++49 30 2802 6391, Fax: ++49 30 2802 6615

On Behalf of

Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin
ZEBET (Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch)
Diedersdorfer Weg 1
D-12277 Berlin

Berlin, 6 February 1998

1. Rationale and Intended Purpose of the Test

The *EpiDerm™ Phototoxicity Test* was established for predicting the phototoxic potential of chemicals by using a 3-dimensional human epidermis model. The test consists in the topical application of various concentrations of the test chemical onto the *EpiDerm™* tissue and determination of the cytotoxic effect in the absence (-UV) and presence (+UV) of UV and visible light. The maximum difference (Δ_{max}) between the (UV-) and (UV+) response (irrespective of the concentration at which it appears) is used as measure of the phototoxic potential. The prediction model of the test is a binary classification scheme which assigns a test chemical to one of the two possible toxicity classes non-phototoxic (npt) or phototoxic (pt) depending on whether the value of Δ_{max} falls below or above a predefined cut-off value.

2. Study Design

Participating laboratories: **3**
Beiersdorf, Procter & Gamble and ZEBET

Chemicals: **10** (5 in vivo phototoxic / 5 in vivo non-phototoxic)

Endpoints: Δ_{max} = Maximum(Response(-UV)-Response(+UV))

Measurements/Data:

- Each test chemical was applied in five different concentrations onto 2 different tissues per concentration (i.e. 1 vehicle control + 5 concentration = 12 tissues) (+UVA). A second set for the dark experiment (-UVA) is treated identically, Per chemical and **2 replicates** 24 tissues are needed.
- To reduce pipetting errors from each tissue **3 aliquots** of formazan extract were measured in the photometer and the mean of these 3 values was used for further calculations.
- This results in two dose-response relation comprising 20 + 4 data points per chemical and run
- Out of this, per run and chemical the mean dose difference between the (-UV) and the (+UV) responses was calculated at each test concentration and the maximum of this differences (Δ_{max}) was determined (= 1 value per chemical).
- Since each chemical were tested twice **(2 runs)** on seperate occasions, and 10 chemicals were tested, this results in 20 Δ_{max} values per laboratory as shown in summerising **Table 1.**
- All data (single values and tissue means) were transmitted on *Excel* files to the independent statistician.

3. Objectives of Biometrical Data analysis

The biometrical data analysis had three objectives:

1. Evaluation of the prediction model proposed in the SOP of this test
2. Evaluation of alternative prediction models developed for the 3T3-NRU phototoxicity assay
3. Assessment of data variability

4. Biometrical Methods Used

- Between-run variability of the endpoint value Δ_{max} within one laboratory was assessed by

$$V_{\Delta} = \frac{\left(\Delta_{max}(\text{run1}) - \Delta_{max}(\text{run2})\right)^2}{\left(\Delta_{max}(\text{run1}) - \text{cutoff}\right)\left(\Delta_{max}(\text{run2}) - \text{cutoff}\right)} \qquad (1)$$

where Δ_{max}(run1) and Δ_{max}(run2) denote the Δ_{max} values obtained in the first and second independent run. The quantity V_{Δ} takes into account that differences between the results of the two independent runs have to be evaluated according to the possible consequences for the classification process, i.e. as closer a single Δ_{max}

value comes to the cut-off value as more the discrepancy between the two runs may influence the decision process so that large V_Δ values indicate the necessity to carry out a third independent experiment.

- The possible presence of systematic inter-laboratory deviancies among the Δ_{max} values was checked visually by plotting the Δ_{max} values of two laboratories against each other in bivariate scattergrams and by applying the non-parametric *signs test*. This test is very robust against deviations of the residual distribution from the normal Gaussian distribution. The test statistic reads

$$\{\Delta_{max\ lab1}\} > \{\Delta_{max\ lab2}\} \text{ if } \frac{n_+}{n_- + 1} \geq F_{m_1,m_2;1-\alpha} \qquad (2a)$$

where n_+ and n_- count the number of cases where Δ_{max}(chem i, lab1) > Δ_{max}(chem i, lab2) or Δ_{max}(chem i, lab1) < Δ_{max}(chem i, lab2), respectively. $F_{m1,m2;1-\alpha}$ is the quantile of Firscher F-distribution with $m_1=2(n_+ +1)$ and $m_2=2n_-$ degrees of freedom. Analogously, one may test the hypothesis that the observations of laboratory 2 are systematically smaller than those of laboratory:

$$\{\Delta_{max\ lab1}\} < \{\Delta_{max\ lab2}\} \text{ if } \frac{n_-}{n_+ + 1} \geq F_{m_1,m_2;1-\alpha} \qquad (2b)$$

In the scattergram, fulfilment of conditions (2a) or (2b) means a significantly larger proportion of data points to be situated above or below the 45^o line.

- Evaluation of the prediction model proposed in the SOP was carried out in two variants: (a) classifying each test chemical twice per laboratory using theΔ_{max} values of the two independent runs as predictor variables and (b) classifying each test chemical only once per laboratory using the mean Δ_{max} value of the two runs as predictor variable. The predicted classifications were presented with the known *in vivo* phototoxic classes of the test chemicals in a summarising table. The degree of concordance between predicted and observed phototoxicity classifications was evaluated by 2 x 2 contingency tables:

2 x 2 contingency table

in vivo class	*in vitro* prediction phototoxic	non-phototoxic	row sum
phototoxic	a	b	a+b
non-phototoxic	c	d	c+d
column sum	a+c	b+d	N=a+b+c+d

Table statistics:

sensitivity: a/(a+c)

specificity: d/(b+d)

positive predictivity: a/(a+b)

negative predictivity: d/(c+d)

accuracy: (a+d)/(a+b+c+d)

The robustness of the prediction model against variations of the cut-off value (30%) was tested by computing the total number of misclassifications at varying the cut-off value between 1% and 50%.

- The two prediction models developed for the 3T3-NRU phototoxicity assay were also tested for their applicability to the *EpiDerm™ Phototoxicity Test.* The primary goal of this analysis was to check whether the range of applicability of the 3T3-NRU prediction models could be extended to different assays possessing a different biological background.

4. Results

1. The classification results obtained with the SOP prediction models and the V_Δ takes for each chemical and each laboratory are depicted in Table 1. Depending on whether the classifications were based on the Δ_{max} values of the two single runs or on mean Δ_{max} value, there were six or two misclassifications, respectively (cf. contingency tables in Table 2). In practice, the only reasonable way of classification should be based on the mean Δ_{max} values of the two experiments. Considering the quality of laboratory predictions derived on the basis of the between-run mean Δ_{max} values, the performance of this assay is indeed very encouraging.

2. As can be seen from Table 1, absolute V_Δ values larger than 3.0 resulted in all cases (except for tetracycline free base in laboratory P&G) where a chemical was misclassified according to the Δ_{max} value of at least one run, whereas for all correct classifications (except for neutral red in laboratory ZEBET) the V_Δ values were considerably smaller than 3.0. Thus, the V_Δ value can be regarded as a sensitive measure for the reproducibility of Δ_{max} measurements in independent experiments. Note that a negative V_Δ value is indicative for two controversy results where one Δ_{max} value was above and the second was below the cut-off value! Negative or/and large absolute V_Δ values (say larger than 3.0) indicate conflicting results of the two independent runs and , therefore, should be enough reason for the laboratory to carry out a third independent experiment.

3. Fig. 1 shows that the overall classification quality (i.e. the total number of misclassifications across the three participating laboratories) would not change in a relatively broad interval of cut-off values ranging from 20% and 38%. Thus, the cut-off value of 30% as proposed in the SOP seems to be a satisfactory choice.

4. The lab-to-lab plots (see Fig. 2) of the Δ_{max} values did not reveal any significant systematic between-lab differences.

5. Application of the PIF- and MPE based prediction model developed for the 3T3-NRU phototoxicity assay (for the definition of these two measures and their use in the prediction model cf. the SOP of the 3T3-NRU phototoxicity assay) provided the classifications depicted in Table 3 and Table 4. There were five misclassifications associated with the PIF-based prediction model with a clear tendency towards false-negatives and two misclassifications associated with the MPE-based prediction model. Taking into consideration that the two prediction models have been applied with the same cut-off values as optimised for the 3T3-NRU assay (5.0 for PIF and 0.1 for MPE), the MPE-based prediction model (which originally has been developed and proved to provide reliable predictions of the phototoxic potential in an assay using human

keratinocytes with the same cut-off value of 0.1) should be considered as a serious alternative to the SOP proposed prediction model.

Table 1: Summary of Laboratory Results

no	chemical	in vivo class	run	Beiersdorf				Procter & Gamble				ZEBET			
				Δmax	V_Δ	pred.	pred. mean	Δmax	V_Δ	pred. run	pred. mean	Δmax	V_Δ	pred. run	pred. mean
1	Chlorpromazine / W	**pt**	1	44.7	0.11	pt	**pt**	44.9	0.23	pt	**pt**	45	1.30	pt	**pt**
			2	40.6		pt		39.3		pt		74.4		pt	
2	Acridine	**pt**	1	67.3	0.17	pt	**pt**	91.6		pt	**pt**	91	0.00	pt	**pt**
	hydrochloride / O*		2	86.1		pt		92.2		pt		92.8		pt	
3	Bergamot oil / O	**pt**	1	60.4	0.14	pt	**pt**	50.1	0.01	pt	**pt**	79.6	0.00	pt	**pt**
			2	74.1		pt		52.1		pt		78.4		pt	
4	Neutral red / W	**pt**	1	69.8	0.01	pt	**pt**	54.8	0.46	pt	**pt**	39.9	3.73	pt	**pt**
			2	74.2		pt		78.1		pt		84.9		pt	
5	Tetracycline	**pt**	1	25.5	3.14	npt	**npt**	10.4	0.02	npt	**npt**	39.1	0.10	pt	**pt**
	free base / O		2	7.8		npt		7.2		npt		36.6		pt	
6	Penicillin G / W	**npt**	1	7.6	0.03	npt	**npt**	-4.3	-5.96	npt	**npt**	0.9	0.51	npt	**npt**
			2	11.2		npt		39.3		pt		15.5		npt	
7	Lauryl sulfate	**npt**	1	19.1	1.01	npt	**npt**	-1.7	0.15	npt	**npt**	6.7	0.05	npt	**npt**
	sodium / W		2	1.3		npt		8.3		npt		11.2		npt	
8	Octyl salicylate / O	**npt**	1	13.3	0.69	npt	**npt**	-2.3	0.00	npt	**npt**	3.1	0.00	npt	**npt**
			2	-7.4		npt		-0.3		npt		2.7		npt	
9	4-Methylbenzylidene	**npt**	1	13.4	0.59	npt	**npt**	-8.5	0.04	npt	**npt**	-2.8	0.03	npt	**npt**
	camphor / O		2	-5.2		npt		-1.6		npt		2.3		npt	
10	Octyl methoxy-	**npt**	1	10.7	0.25	npt	**npt**	3.7	0.23	npt	**npt**	7.4	-115.01	npt	**npt**
	cinnamate / O		2	-1.7		npt		13.7		npt		30.2		pt	

W = solvent: H_2O
O = vehicle/solvent: sesame oil
* = Beiersdorf: H_2O / Procter & Gamble, ZEBET: sesame oil

Δ_{max}: maximal difference between UV(-) and UV (+) response (tissue mean)
pred. run predicted class based on the Δ_{max} value of the run (> or < 30%)
pred. mean: predicted class based on the mean Δ_{max} value across the two independent runs (> or < 30%)
V_Δ: $(run1 - run2)^2 / ((run1 - cutoff)(run2 - cutoff))$

Table 2: Contingency tables for laboratory classifications

based on Δ_{max} values of single runs:

Beiersdorf

		in vivo pt	***in vivo*** npt
in vitro	pt	8	0
	npt	2	10

sensitifity:	80.00%
specificity:	100.00%
positive prediction:	100.00%
negative prediction:	83.33%
accuracy:	**90.00%**

Procter & Gamble

		in vivo pt	***in vivo*** npt
in vitro	pt	8	1
	npt	2	9

sensitifity:	80.00%
specificity:	90.00%
positive prediction:	89.00%
negative prediction:	82.00%
accuracy:	**85.00%**

ZEBET

		in vivo pt	***in vivo*** npt
in vitro	pt	10	1
	npt	0	9

sensitifity:	100.00%
specificity:	90.00%
positive prediction:	90.91%
negative prediction:	100.00%
accuracy:	**95.00%**

based on mean Δ_{max} values of two runs:

Beiersdorf

		in vivo pt	***in vivo*** npt
in vitro	pt	4	0
	npt	1	5

sensitivity:	80.00%
specificity:	100.00%
positive prediction:	100.00%
negative prediction:	83.33%
accuracy:	**90.00%**

Procter & Gamble

		in vivo pt	***in vivo*** npt
in vitro	pt	4	0
	npt	1	5

sensitivity:	80.00%
specificity:	100.00%
positive prediction:	100.00%
negative prediction:	83.33%
accuracy:	**90.00%**

ZEBET

		in vivo pt	***in vivo*** npt
in vitro	pt	5	0
	npt	0	5

sensitivity:	100.00%
specificity:	100.00%
positive prediction:	100.00%
negative prediction:	100.00%
accuracy:	**100.00%**

Table 3: Summary of Laboratory Results (PIF prediction model)

no	chemical	in vivo class	Beiersdorf Mean	Beiersdorf variance	Beiersdorf Type 1.	Beiersdorf class	Procter & Gamble Mean	Procter & Gamble variance	Procter & Gamble Type 1	Procter & Gamble class	ZEBET Mean	ZEBET variance	ZEBET Type 1	ZEBET class
1	Chlorpromazine	**pt**	6.21	0.51		**pt**	2.85	0.28		**npt**	3.77	2.13		**npt**
2	Acridine hydrochloride	**pt**	57.60	26.10		**pt**	59.32	58.35	>(1.2)*	**pt**	311.11	91.99	>(2:2)	**pt**
3	Bergamot oil	**pt**	6.27	3.30	>(2:2)	**pt**	2.78	0.30	>(2.2)	**pt**	6.08	1.79	>(2:2)	**pt**
4	Neutral red	**pt**	118.16			**pt**	51.32	50.33	=(1:2)	**pt**	27.73			**pt**
5	Tetracycline free base	**pt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**
6	Penicillin G	**npt**	1.00	0.00	=(2:2)	**npt**	1.79	0.79	=(1:2)	**npt**	1.00	0.00	=(2:2)	**npt**
7	Lauryl sulfate sodium	**npt**	0.91	0.15		**npt**	0.76	0.12		**npt**	1.12	0.06		**npt**
8	Octyl salicylate	**npt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**
9	4-Methylbenzylidene camphor	**npt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**
10	Octyl methoxy-cinnamate	**npt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**	1.00	0.00	=(2:2)	**npt**

* = Procter & Gamble: Type 2 =(1:2)

Beiersdorf

in vitro		***in vivo*** pt	***in vivo*** npt
	pt	4	0
	npt	1	5

sensitifity: 80.00%
specificity: 100.00%
positive prediction: 100.00%
negative prediction: 83.33%
accuracy: **90.00%**

Procter & Gamble

in vitro		***in vivo*** pt	***in vivo*** npt
	pt	3	0
	npt	2	5

sensitivity: 60.00%
specificity: 100.00%
positive prediction: 100.00%
negative prediction: 71.43%
accuracy: **80.00%**

ZEBET

in vitro		***in vivo*** pt	***in vivo*** npt
	pt	3	0
	npt	2	5

sensitifity: 60.00%
specificity: 100.00%
positive prediction: 100.00%
negative prediction: 71.43%
accuracy: **80.00%**

Table 4: Summary of Laboratory Results (MPE prediction model)

no	chemical	in vivo class	Beiersdorf Mean	Beiersdorf variance	Beiersdorf class	Procter & Gamble Mean	Procter & Gamble variance	Procter & Gamble class	ZEBET Mean	ZEBET variance	ZEBET class
1	Chlorpromazine	**pt**	0.43	0.08	**pt**	0.22	0.03	**pt**	0.35	0.31	**pt**
2	Acridine hydrochloride	**pt**	0.47	0.27	**pt**	0.41	0.40	**pt**	0.77	0.02	**pt**
3	Bergamot oil	**pt**	0.41	0.08	**pt**	0.33	0.04	**pt**	0.60	0.04	**pt**
4	Neutral red	**pt**	0.39	0.11	**pt**	0.55	0.06	**pt**	0.42	0.25	**pt**
5	Tetracycline free base	**pt**	0.05	0.03	**npt**	-0.01	0.02	**npt**	0.12	0.00	**pt**
6	Penicillin G	**npt**	0.01	0.01	**npt**	-0.08	0.33	**npt**	-0.01	0.11	**npt**
7	Lauryl sulfate sodium	**npt**	0.02	0.01	**npt**	0.02	0.01	**npt**	0.01	0.01	**npt**
8	Octyl salicylate	**npt**	-0.05	0.03	**npt**	-0.07	0.05	**npt**	-0.04	0.03	**npt**
9	4-Methylbenzylidene camphor	**npt**	0.10	0.10	**npt**	-0.08	0.04	**npt**	-0.03	0.02	**npt**
10	Octyl methoxy-cinnamate	**npt**	0.04	0.05	**npt**	0.03	0.02	**npt**	0.00	0.06	**npt**

Beiersdorf

		in vivo pt	***in vivo*** npt
in vitro	pt	4	0
	npt	1	5

sensitifity: 80.00%
specificity: 100.00%
positive prediction: 100.00%
negative prediction: 83.33%
accuracy: **90.00%**

Procter & Gamble

		in vivo pt	***in vivo*** npt
in vitro	pt	4	0
	npt	1	5

sensitivity: 80.00%
specificity: 100.00%
positive prediction: 100.00%
negative prediction: 83.33%
accuracy: **90.00%**

ZEBET

		in vivo pt	***in vivo*** npt
in vitro	pt	5	0
	npt	0	5

sensitifity: 100.00%
specificity: 100.00%
positive prediction: 100.00%
negative prediction: 100.00%
accuracy: **100.00%**

Figure 1: **Total number of misclassifications across 3 laboratories at varying cut-off value**

(based on Δ_{max} values)

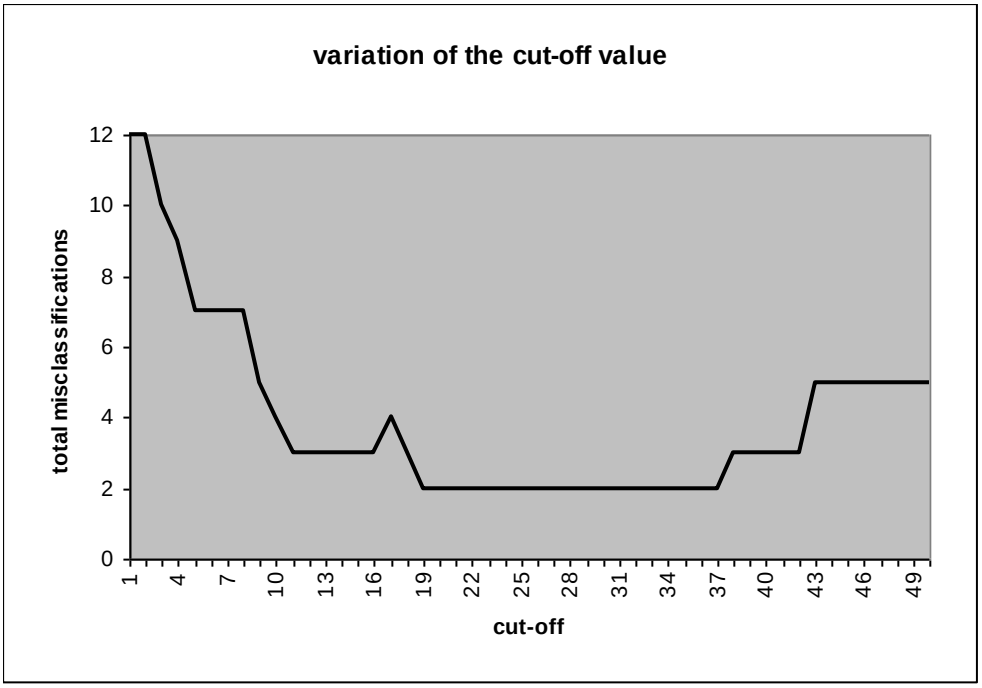
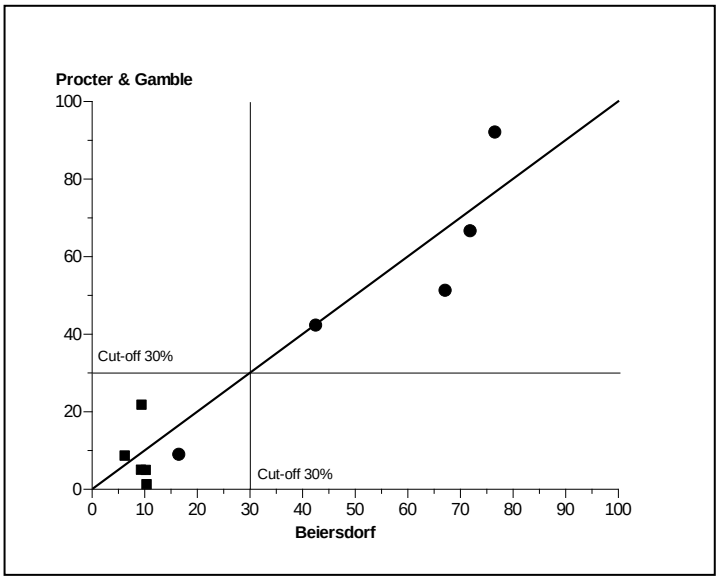
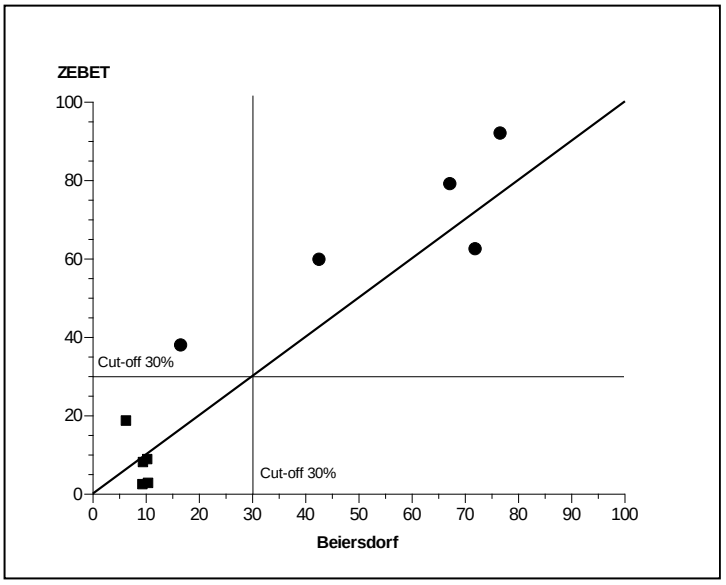
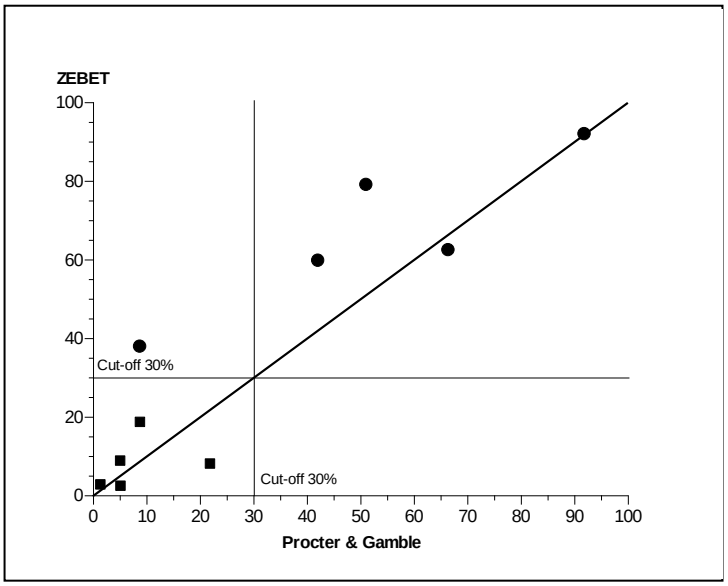


Figure 2: Between-Laboratory Comparison of Mean Δ_{max} Values







✧ = phototoxic
✪ = non phototoxic

Signs Tests:
no difference