Ann. occup. Hyg., Vol. 48, No. 7, pp. 601–606, 2004

Published by Oxford University Press
doi:10.1093/annhyg/meh052

Reliability of a Visual Scoring System with Fluorescent Tracers to Assess Dermal Pesticide Exposure

AURORA ARAGÓN[1,2]*, LUIS BLANCO[1,2], LYLLIAM LÓPEZ[1], CAROLA LIDÉN[3], GUN NISE[2] and CATHARINA WESSELING[2,4]

[1]*Department of Preventive Medicine and Public Health, Faculty of Medicine, National Autonomous University of Nicaragua, León, Nicaragua;* [2]*Department of Public Health Sciences, Division of Occupational Health, Karolinska Institutet, Stockholm, Sweden;* [3]*Department of Medicine, Occupational and Environmental Dermatology, Karolinska Institutet and Stockholm County Council, Sweden;* [4]*Central American Institute for Studies on Toxic Substances, Universidad Nacional, Heredia, Costa Rica*

Received 2 January 2004; in final form 16 April 2004; Published online 20 September 2004

Background and objectives: We modified Fenske's semi-quantitative 'visual scoring system' of fluorescent tracer deposited on the skin of pesticide applicators and evaluated its reproducibility in the Nicaraguan setting.

Methods: The body surface of 33 farmers, divided into 31 segments, was videotaped in the field after spraying with a pesticide solution containing a fluorescent tracer. A portable UV lamp was used for illumination in a foldaway dark room. The videos of five farmers were randomly selected. The scoring was based on a matrix with extension of fluorescent patterns (scale 0–5) on the ordinate and intensity (scale 0–5) on the abscissa, with the product of these two ranks as the final score for each body segment (0–25). Five medical students rated and evaluated the quality of 155 video images having undergone 4 h of training. Cronbach alpha coefficients and two-way random effects intraclass correlation coefficients (ICC) with absolute agreement were computed to assess inter-rater reliability.

Results: Consistency was high (Cronbach alpha = 0.96), but the scores differed substantially between raters. The overall ICC was satisfactory [0.75; 95% confidence interval (CI) = 0.62–0.83], but it was lower for intensity (0.54; 95% CI = 0.40–0.66) and higher for extension (0.80; 95% CI = 0.71–0.86). ICCs were lowest for images with low scores and evaluated as low quality, and highest for images with high scores and high quality.

Conclusions: Inter-rater reliability coefficients indicate repeatability of the scoring system. However, field conditions for recording fluorescence should be improved to achieve higher quality images, and training should emphasize a better mechanism for the reading of body areas with low contamination.

Keywords: fluorescent tracer; Nicaragua; pesticides; reliability; skin exposure; visual scoring system

INTRODUCTION

Fluorescent tracers have been used since the 1980s to visualize dermal exposure of pesticides with ultraviolet light (Franklin *et al*., 1981; Fenske *et al*., 1986). The use of the tracer has helped to demonstrate the non-uniformity of exposure, which is a substantial source of uncertainty in dermal exposure and risk assessment (Fenske, 1990). It has also been used for evaluating performance of protective devices, and for educational purposes (Houghton *et al*., 1999). A further development of the method was the Video Imaging Technique to Assess Exposure (VITAE) (Fenske *et al*., 1986), which transformed qualitative observations of fluorescent skin images into quantitative values by means of a computer program. VITAE has been shown to perform well with substances such as malathion (Fenske, 1988a) and pirimicarb (Archibald *et al*., 1994).

However, since computerized quantification by image analysis would be impractical in many field

*Author to whom correspondence should be addressed.
e-mail: aragon@unanleon.edu.ni

situations, Fenske developed the 'Visual Scoring System', a more accessible semi-quantitative method based on visual observation and scoring of both the extension and the intensity of fluorescent tracer on the skin. This Visual Scoring System was validated for the hands and face of applicators and correlated well with the computerized VITAE (Fenske, 1988b). Fenske validated the Visual Scoring System reading photographs, but he also suggested the possibility of using videotapes by freezing the images. Fenske's validated Visual Scoring System was recommended for field studies, but, to our knowledge, it has been put into practice in the field only once to test the efficiency of protective equipment simulating the use of granulated pesticide with graphite powder (Machado-Neto *et al.*, 1996).

The fluorescent tracer technique has been occasionally used in developing countries as part of educational programs for agricultural workers. The skin is the major absorption route for pesticides (Bos *et al.*, 1998). In Nicaragua, skin contact with the toxic agent was identified in the majority of work-related pesticide poisonings (Corriols *et al.*, 2001). In the early 1990s, we built up some educational experience with this technique with the intent of expanding this experience towards a robust but valid alternative method for semi-quantitative exposure assessment in a Third World context, where human, technical and financial resources for occupational hygiene are almost nonexistent. Although Fenske's visual scoring system seemed affordable and feasible, it was necessary to adapt it to the particular working conditions in Nicaragua. In particular, Fenske's scoring system only evaluated face and hands, the body surface segments considered to be most prone to skin contamination in his study subjects, working with protective devices and under controlled conditions, whereas Nicaraguan farmers usually do not wear protective clothing and experience more extensive pesticide skin contamination (Aragón *et al.*, 2001). Therefore, we extended Fenske's scoring system to the entire body surface of the applicator and we scored the dermal exposure of 33 farmers from the surroundings of León, Nicaragua, who participated in a larger study looking into a number of pesticide exposure assessment methods. Here we report the feasibility and reliability assessment of the modified visual scoring system by estimating the inter-rater agreement among five raters for a subset of five farmers.

MATERIAL AND METHODS

The exposure of total body surface was assessed using 31 body surface segments for scoring (Table 1). Another adaptation was the use of videotapes instead of photographs to record the fluorescence of the skin, a more suitable method to assess exposure of the entire body. Furthermore, we prepared a guide for characterization of the fluorescent images, to assist in the identification of the most common patterns of skin contamination during scoring. A final minor modification was the use of a minimum score of zero, as opposed to the 1 in Fenske's system, to denote skin surfaces without any fluorescence, or 'clean areas'.

Fluorescent tracer and videorecording

At the start of the workday, the fluorescent tracer Tinopal CBS-X® was placed into the tank of the backpack sprayer (260 mg/l) by one of the researchers, before the farmer poured the concentrated pesticide. Toxicological data of Tinopal CBS-X® do not

Table 1. Body segments defined for the Visual Scoring System

Body area	No.	Body segment	Body area	No.	Body segment
Face	1	Right	Hands	16	Right palm
	2	Front		17	Right dorsal
	3	Left		18	Left palm
Neck	4	Front		19	Left dorsal
	5	Back	Thighs	20	Right front
Thorax	6	Front		21	Right back
	7	Back		22	Left front
Arm	8	Right front		23	Left back
	9	Right back	Legs	24	Right front
Forearm	10	Right front		25	Right back
	11	Right back		26	Left front
Arm	12	Left front		27	Left back
	13	Left back	Feet	28	Right dorsal
Forearm	14	Left front		29	Right plantar
	15	Left back		30	Left dorsal
				31	Left plantar

indicate any adverse health effects (Houghton *et al*., 1999). The farmer applied the pesticide as usual. Immediately after finishing the application, the deposition of fluorescent images on the skin of the farmer was videotaped in the field, with a Hitachi® VMH-640A Hi8 camcorder with a minimum illumination of 1.0 lux, a 16× optical zoom and a 300× digital zoom. Videotaping took place inside a 1.2 × 1.2 × 2 m foldaway darkened room especially built for the purpose.

Farmers were observed wearing underpants. A hand held UV lamp (UVP®, model UVSL-26P) operated with long wave (365 nm) was used for the illumination. To simplify the simultaneous observations and recording, the lamp was attached to the video camera and held together by the investigators. The camera-and lamp-to-subject distance ranged from 30–50 cm. No zooming was applied. We usually started with a contaminated area, and then continued towards clean areas. Although care was taken to have the optimal degree of darkness, there were problems in filming low fluorescent skin surfaces in the dark room, which resulted in missing data (range 1–9). The video recording took between 30 and 60 min. Several breaks of a few minutes each were scheduled during the recording process because of increases in the ambient temperature inside the room. All videotapes were scored by the main investigator (A.A.). For the present report, the videotapes of five farmers were randomly selected from the 24 applicators with the most complete data (≥28 scored body segments).

Raters

Five students of the fifth year of medical school, who had previously shown interest in occupational health, were invited as raters (R1–R5). They were not previously connected with the study nor with the farmers, and did not participate in the data collection. Two of the authors (L.B., L.L.) conducted the training, which took 4 h and included theoretical background of the use of fluorescent tracer and practical training. The rating of the main investigator (A.A.), who had previously scored the farmers, was included for comparison (R6). She did not participate in the training and was not present during the rating process.

The scoring system

Each of the 31 body surface segments was scored separately. The scoring was based on a two-dimensional matrix where the rows denote the degree of extension of the fluorescence on the body part and the columns the intensity of the fluorescence, as proxies for extension and intensity of pesticide exposure. The extension of the fluorescence is classified into five categories: clean surface (0), 1–20% (1), 21–40% (2), 41–60% (3), 61–80% (4) and 81–100% (5), and intensity of fluorescence is represented by a scale of none (0), low (1), moderate low (2) moderate (3), moderate high (4) and high (5). The product of these two variables results in an image score ranging from 0 to 25 for each of body part. The total score for an individual was computed as the sum of the scores of the 31 body surface segments.

Reading the videotapes

The readings were carried out in a classroom simultaneously by the five raters. The two trainers assisted only by running and stopping the videos to allow the scoring of each body segment. The raters were not allowed to communicate during the reading except to ask for the tape to be rewound for reviewing. The next image was not shown until every rater had completed reading the current one. The scores were recorded in a previously designed form, which contained the list of the 31 body surface segments. Raters were also asked to annotate the pattern of fluorescence as mist, splash or smear (Figure 1), and to judge the quality of each image (poor, average, good). When the quality of the image was considered poor, the rater had the choice to not read it, resulting in a missing value. The reading of the 31 body areas of a farmer took ≤30 min.

Data analysis

Statistical analyses were performed using SPSS version 10.1. The total visual scores were plotted. The Cronbach alpha coefficient of reliability was calculated to assess internal consistency of the rating. To evaluate inter-rater agreement, we used the two-way random model of intraclass correlation coefficients (ICC), with measures of absolute agreement. In addition, we selected the option for single measure intraclass correlation to estimate the reliability of a single rating instead of a mean of several ratings (Armstrong *et al*., 1992; Yaffee, 1998). The formula used is as follows:

$$ICC = \frac{MSP - MSE}{MSP + (k-1)MSE + k(MSR - MSE)/n}$$

where n is the total number of scored body surface segments by the raters, k is the number of raters, MSP is the between body surface segments mean square, MSE is the residual mean square and MSR is the between-rater mean square from a two-way analysis of variance.

RESULTS

Figure 2 shows the total visual score of each farmer, as assessed by each rater. Farmer 1 was evaluated as the least contaminated and Farmer 3 as the most contaminated by all raters. This was also in agreement with the scoring of R6, the main investigator. Considering all body segments scored, Cronbach alpha

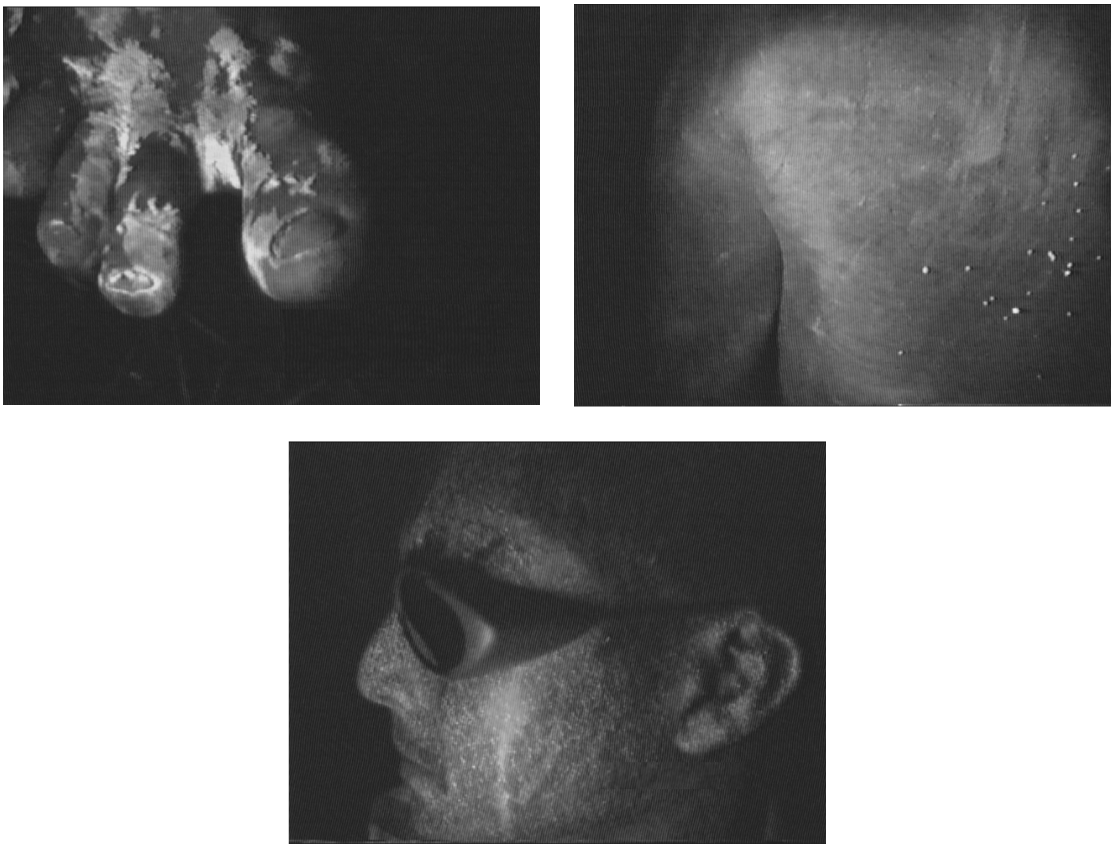

Fig. 1. Images of splash on the right foot (A); smear on the right side of the back (B) and mist on the left side of the face (C).

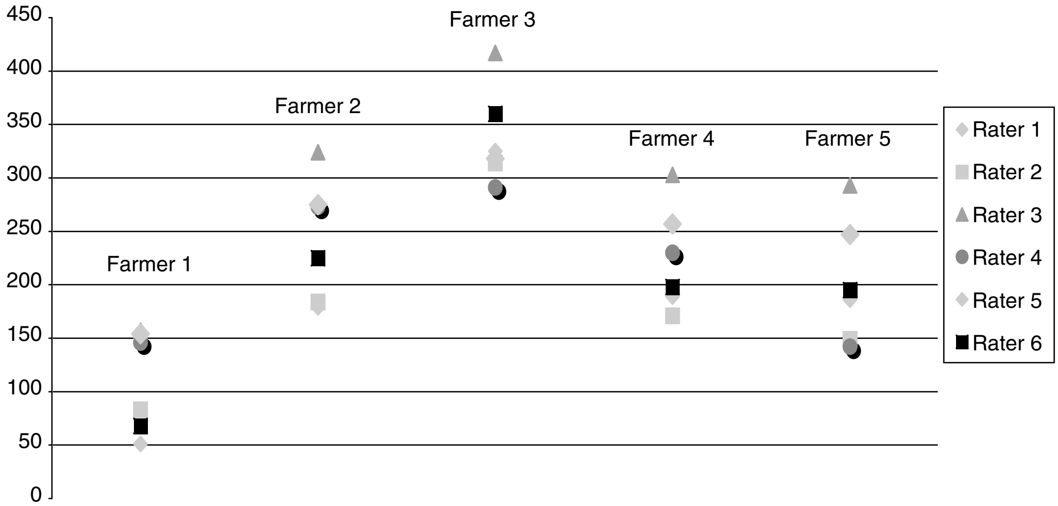

Fig. 2. Total visual score of fluorescent tracer on the body surface of five sprayers, according to five raters after 4 h training (R1–R5) and the main investigator (R6). This figure is based on the 21 body areas that had no missing data for any of the raters.

for the five raters was 0.96, and 0.97 including R6, indicating high internal consistency independent of the differences in individual score levels of the raters. R3 tended to give high scores, and R1 and R2 gave low scores. The mean (± SD) score of the images read by the individual raters ranged from 7.9 ± 7.7 to 13.7 ± 8.5. The mean score for the main investigator, R6, was 8.5 ± 8.7.

The ICC of the visual score for the five raters combined was 0.75 [95% confidence interval (CI) = 0.62–0.83], ranging between rater pairs from 0.63 (95% CI = 0.10–0.83) to 0.88 (95% CI = 0.85–0.92). The ICC of the pairs of the main investigator, R6, and the newly trained raters varied between 0.70 and 0.87. When separating the visual score into its two components of extension and intensity, the ICC for extension was 0.80, considerably higher than 0.54 for intensity (Table 2).

Table 3 shows the relative frequencies of the quality of images according to the raters. The five raters were consistent with regard to the poor images, the proportion of bad quality and illegible images ranging between 6 and 10%. After classifying the quality of all images for which three or more raters coincided in their judgement (96% of all images), ICC was 0.80, 0.47 and 0.32 for images of good, average and poor quality, respectively. Mean scores decreased with decreasing image quality (range = 10.5–14.2 for good quality images, 4.2–8.0 for average quality images and 1.2–3.4 for poor quality images). The image most identified was smear (41%), followed by mist (20%) and the combination of smear and mist (23%). Mists were scored lowest and smears, alone or combined, highest. The mean scores (SD) were 4.8 (4.1) for mist, 5.2 (3.5) for splash, 11.9 (8.4) for smear, 13.3 (8.7) for the combination of mist and smear, 14.8 (7.7) for splash and smear, and 16.7 (8.5) for mist, splash and smear combined.

DISCUSSION

We evaluated the feasibility and reliability of a modified version of Fenske's Visual Scoring System of pesticide exposure (scoring of video images of fluorescence extension and intensity on the entire body, divided into 31 parts) under real-life circumstances in a developing country. Although Fenske's semi-quantitative Visual Scoring Method for the hands and the face showed a good correlation with the computerized VITAE system (Fenske, 1988b), this was first time that the Fenske's Visual Scoring System was applied to the whole body and in field circumstances. Only one study, in Brazil (Machado *et al*., 1996), used a similar two-dimensional scoring matrix, but that study used a different agent (graphite powder) and for a different purpose (to evaluate potential exposure in relation to the efficiency of protective clothing in a controlled setting).

Table 2. Intraclass correlation coefficients (ICC) and the Cronbach alpha coefficient of the Visual Scoring System by distribution and intensity: a two-way random effect model with absolute agreement

	ICC	95% CI	Cronbach alpha
Visual score (five raters)	0.75	0.62–0.83	0.96
Extension (five raters)	0.80	0.71–0.86	0.96
Intensity (five raters)	0.54	0.40–0.66	0.90

We found a very high between-rater consistency for ranking. However, absolute differences in scores were considerable. The inter-rater agreements accounting for variability between raters (ICCs) were less satisfactory than consistency of ranking. The overall ICC of 0.75 (95% CI = 0.62–0.83) indicates quite good agreement. There is, however, considerable subjectivity in the assessment of the combined intensity–extent dimension of exposure. When stratifying by extension and intensity of exposure, and by poor and good quality images, the discrepancies seemed to be confined primarily to a low agreement for the intensity component of fluorescence, and for the scoring of poor quality images.

After the scoring session, the raters were unanimous in that low intensity images were most difficult to read. Sources of discrepancies in the reading were also associated with poor quality of the image. Both factors are related. To obtain a well-focused image during the videotaping, the camera should be fixed during several seconds on the body part. Highly intense images were easy to focus but low intensities took more time to capture and sustain. Nicaragua is a tropical country with ambient temperatures often exceeding 35°C with a high humidity. The portable room was comfortable early in the morning, but if the farmer ended the workshift late, the heat inside the room forced the investigators to reduce the time for videotaping of each body part as a precaution. This partly explains the link between low intensity and poor quality of the images. However, difficulty in filming body surface segments with low or no fluorescence in a dark room is inherent in the method, and probably suggests that many missing data correspond to zero or very low scores.

Table 3. Proportion of images by quality

Quality	Rater 1 (*n* = 155)	Rater 2 (*n* = 155)	Rater 3 (*n* = 155)	Rater 4 (*n* = 155)	Rater 5 (*n* = 155)
High	0.63	0.46	0.79	0.59	0.76
Moderate	0.31	0.48	0.15	0.31	0.17
Low	0.01	0.01	0.02	0.04	0.04
Not read	0.05	0.05	0.04	0.06	0.03

Cherrie *et al*. (2000) pointed out that it is important to categorize the mechanisms by which the contaminant is emitted (splashing, spilling, etc.) to elucidate the exposure determinants. Our raters used a guide for categorization, which was intended for facilitation of the learning process of scoring. This characterization gave some clues about the mechanism of the contamination, splashes being related to spilling, smears to friction or wiping of surfaces, and mist to transport by air and deposition on the skin in the form of aerosol. Although the raters considered the image characterization a useful tool for scoring, their judgment did not always agree. This could be the result of the short training. In any case, characterization of the images has to be further explored.

After this first experience, a number of improvements can be suggested. The design of the foldaway dark room needs to be elaborated further through engineering techniques in order to reduce filming time, aiming at increased space, tolerable ambient temperature and proper illumination. Also, further improving the videotaping technique with methods to stabilize the hand-held camera and achieve steady focus at a specified distance from the skin, as well as the use of filters for the video camera and for the source of the UV illumination, would contribute to better quality of the images. As a next step, it is important to standardize the scoring of the intensity of the fluorescent images. Training should emphasize the reading of low contamination parts, and the inclusion of illustrating pictures for this purpose can enhance the scoring guide.

In conclusion, inter-rater reliability coefficients are acceptable and indicate repeatability of the scoring system in general terms. However, the low ICCs for the intensity component must be addressed. With improvements for intensity scoring, this method can be applied in developing countries, as a tool for exposure assessment in epidemiologic studies, in occupational hygiene evaluations and as an educational tool.

Acknowledgements—We thank Dr Richard Fenske of the University of Washington for his initial advice to perform the Visual Scoring System. Thanks go also to Indiana López and Rafael Espinoza, UNAN León, for their advice on the statistical analyses. We especially thank Dr Timo Partanen, Universidad Nacional in Costa Rica, for his comments and fruitful discussion. Our gratitude is also due to the medical students who participated as raters. This study was carried out with the financial support of the Swedish Agency for Research Cooperation in Developing Countries (SAREC/Sida).

REFERENCES

Aragón A, Aragón C, Thörn A. (2001) Pest, peasants and pesticides in the Pacific Plain of Nicaragua. Int J Occup Environ Health; 7: 295–302.

Archibald BA, Solomon KR, Stephenson G. (1994) Estimating pirimicarb exposure to greenhouse workers using video imaging. Arch Environ Contam Toxicol; 27: 126–9.

Armstrong B, White E, Saracci R. (1992) Validity and reliability studies. In Armstrong BK, White E, Saracci R, editors. Principles of exposure measurement in epidemiology. New York: Oxford University Press. pp. 79–114.

Bos PM, Brouwer DH, Stevenson H, Boogaard PJ, de Kort WL, van Hemmen JJ. (1998) Proposal for the assessment of quantitative dermal exposure limits in occupational environments: Part 1. Development of a concept to derive a quantitative dermal occupational exposure limit. Occup Environ Med; 55: 795–804.

Cherrie J, Brouwer D, Roff M, Vermeulen R, Kromhout H. (2000) Use of qualitative and quantitative fluorescence techniques to assess dermal exposure Ann Occup Hyg; 44: 519–22.

Corriols M, Silva D, Marin J, Berroteran J, Lozano LM, Martinez J. (2001) Incidencia de intoxicaciones agudas por plaguicidas y estimación del subregistro en Nicaragua. Research Report. Managua, Nicaragua: OPS/OMS.

Fenske RA, Leffingwell JT, Spear RC. (1986) A video imaging technique for assessing dermal exposure. I. Fluorescent tracer testing. Am Ind Hyg Assoc J; 47: 764–70.

Fenske RA. (1988a) Correlation of fluorescent tracer measurements of dermal exposure and urinary metabolite excretion during occupational exposure to malathion. Am Ind Hyg Assoc J; 49: 438–44.

Fenske RA. (1988b) Visual scoring system for fluorescent tracer evaluation of dermal exposure to pesticides. Bull Environ Contam Toxicol; 41: 727–36.

Fenske RA. (1990) Nonuniform dermal deposition patterns during occupational exposure to pesticides. Arch. Environ Contam. Toxicol; 19: 332–7.

Franklin CA, Fenske RA, Greenhalgh R *et al*. (1981) Correlation of urinary pesticide metabolite excretion with estimated dermal contact in the course of occupational exposure to Guthion. J Toxicol Environ Health; 7: 715–31.

Houghton D, Archibald B, Solomon K. (1999) Review of video imaging techniques for estimating dermal exposure to pesticides. In Ecobichon D, editor. Occupational hazards of pesticide exposure. Sampling, monitoring, measuring. Philadelphia, PA: Taylor & Francis. pp. 135–82.

Machado-Neto JG, Matuo T, Matuo YK. (1996) Semiquantitative evaluation of dermal exposure to granulated insecticides in coffee (*Coffea arabica* L.) crop and efficiency of individual protective equipment. Bull Environ Contam Toxicol; 57: 946–51.

Yaffee R. (1998), Enhancement of reliability analysis: application of intraclass correlations with SPSS/Windows v.8. Available from: URL: http//www.nyu.edu/its/socsci/docs/intracls.html.