

Clinical aspects of allergic disease

Risk assessment in determining systemic reactivity to honeybee stings in sting-threatened individuals

James H. Day, MD, David L. Buckeridge, BSc, and April C. Welsh, BSc
Kingston, Ontario, Canada

We evaluated the diagnostic parameters (sting rate, venom-specific RAST IgE, venom-specific RAST IgG, venom-specific IgG₄, venom-specific skin test) of 224 individuals at risk of honeybee sting, assigned them a level of risk on the basis of a hypothetical risk model, and then sting challenged each subject with a live honeybee. Of the 70 subjects at the lowest risk level, only three (4.3%) experienced equivocal or mild systemic reactions, whereas eight (72.7%) of the 11 at the highest risk level experienced systemic reactions, of which five (45.5%) were severe. Increase in risk level was associated with an increased proportion of systemic reactions ($p < 0.001$). Of the individual diagnostic parameters, venom-specific RAST IgE was the best single predictor of reactivity (likelihood ratio = 0.759, $p < 0.01$). Venom-specific skin testing at a concentration of 0.1 µg/ml was a better discriminator than a concentration of 1.0 µg/ml but did not predict one severe systemic reaction. Three methods of prediction of risk and severity of reaction were examined; the one that assessed risk in the most clinically useful manner was similar to the hypothetical risk model. This model compared well with classical methods of risk assessment and may be used to assess high-risk individuals exposed to honeybees when management could include prophylactic immunotherapy. (J ALLERGY CLIN IMMUNOL 1994;93:691-705.)

Key words: Anaphylaxis, bees, bites and stings, risk, skin tests, IgE, IgG

Bee, wasp, and hornet stings are believed to cause over 30 deaths annually in the United States,¹ whereas in Canada they cause approximately one per year.² In addition, there is considerable morbidity in the form of systemic (anaphylactic) reactions, especially in family members of beekeepers who are only occasionally stung. Approximately 20% of these individuals have systemic reactions, which are indicative of a predisposition to potentially severe anaphylactic responses.³ This incidence of systemic reactivity and its possible consequences are a cause of anxiety among the nearly 20,000 individuals employed in the honey industry in Canada who together produce an annual crop of honey worth over 50 million dollars.⁴

Abbreviations used

CRM:	Classical risk model
HRM:	Hypothetical risk model
LR:	Likelihood ratio
PRM:	Proposed risk model
RM:	Risk model
SC:	Sting challenge
VS:	Venom-specific

Numerous studies debating the use of immunologic and clinical parameters in the prediction of the severity of reaction to a sting challenge (SC) have been published,⁵⁻⁹ but despite the large volume of material produced, there is no clear consensus on the issue. Accordingly, we designed a large-scale study, which incorporated a controlled SC to evaluate the use of individual and combined parameters in the prediction of both the severity and risk of reaction to a bee sting. It was the intention of this research to determine the most accurate method (prediction of severity of reaction or prediction of risk of reaction) and combination of parameters that would allow proper treatment of persons likely to be stung.

From Queen's University, Kingston; and Division of Allergy and Immunology, Kingston General Hospital.

Supported by the Agriculture Research Council of Alberta, Alberta Occupational Health and Safety (a Heritage Grant Program), and the Canadian Honey Council.

Reprint requests: James H. Day, MD, Division of Allergy and Immunology, Kingston General Hospital, 76 Stuart St., Kingston, Ontario, Canada K7L 2V7.

Copyright © 1994 by Mosby-Year Book, Inc.
0091-6749/94 \$3.00 + 0 1/1/51930

METHODS

Subjects

Human subject procedures used in this study were approved by the Ethics Committee of the Department of Medicine at Queen's University and by the hospital boards where sting challenges were performed. All subjects gave informed consent. Two hundred twenty-four subjects were recruited by advertising the project through the Alberta Beekeeper's Association with the assistance of Alberta Agriculture. The individuals accepted into the study were all associated with the honeybee industry in some capacity but had differing degrees of exposure to honeybees and differing frequencies of being stung. Beekeeper's family members usually had occasional exposure and may never have been stung, whereas other subjects, such as some beekeepers and their employees, received numerous stings daily. All were threatened by stings because of involvement with the industry.

Hypothetical risk model

A hypothetical risk model (HRM) was proposed before research began. This model took into account evidence that reduced or absent venom-specific (VS) RAST IgE¹⁰ combined with reduced skin test reactivity^{11, 12} and elevated levels of VS RAST IgG¹³ were associated with a low level of reactivity (low risk), whereas elevated VS RAST IgE,¹⁴ positive skin test results,^{15, 16} and low levels of VS RAST IgG¹⁷ were associated with a high level of systemic reactivity (high risk). On the basis of this information, five separate categories of risk were derived, with risk being interpreted as the likelihood of systemic, possibly fatal, response to a sting. The categories ranged from risk level 0 (nonresponsiveness) to risk level 4 (distinct likelihood of systemic reaction). It was believed that categories between these extremes would reflect possible trends in either direction. The structure of the HRM is presented in Fig. 1.

To make this model useful for screening the majority of exposed individuals, only parameters that were objective and potentially available for all exposed individuals were used. History of reactivity to sting was not necessarily available or predictive in exposed individuals, so it was not included in the model. Furthermore, some data suggest that individuals who are sensitized to honeybee venom¹⁸ and other allergens, including penicillin,¹⁹ are only transiently sensitized and spontaneously lose their sensitivity over time.

Procedure

Each subject completed a questionnaire regarding exposure to honeybees and previous reactions to honeybee stings. Each subject also gave a blood sample, which was examined for VS IgE, VS IgG, and VS IgG₄, and underwent skin testing with honeybee venom at concentrations of 0.1 µg/ml and 1.0 µg/ml. Based on the results of these tests, the subjects were assigned a level of risk according to the HRM. Finally, each subject was

stung challenged by a honeybee in the emergency room of a hospital and the response was recorded.

The levels of VS IgE and VS IgG in the serum of each subject were measured by the RAST method with a commercial kit (Phadebas; Pharmacia, Uppsala, Sweden) according to the manufacturer's instructions. The VS RAST IgE was reported in classes 0 to 4, and VS RAST IgG was reported as units of IgG per milliliter.

IgG₄ was measured with the RAST technique developed by Dr. L. Perlmutter, in which specially labeled anti IgG₄ is used. Briefly, 50 µl of each diluted sample (1/100) was added to the discs (in duplicate), and the tubes were incubated at room temperature for 3 hours. The discs were washed with a water-Tween NaCl solution three times (10 minutes between each wash). Fifty microliters of iodine 125-labeled anti-IgG₄ was added to each tube. The tubes were then incubated overnight, washed as in the previous step, and counted for 2 minutes. VS RAST IgG₄ was expressed as a percent of the total counts of the isotope applied. Two times background was considered positive.

All subjects were requested not to take antihistamines for at least 2 weeks before VS skin testing, and all concurrent medications were reviewed at the time of testing. Serial end point intradermal titration tests were performed with 0.02 ml of freeze-dried Hymenoptera venom protein allergenic extracts (Pharmalgen) supplied by Pharmacia at concentrations of 0.1 µg/ml and 1.0 µg/ml. Negative diluent and positive histamine controls were tested concurrently with venom. Responses were scored 15 minutes after administration. The *in vitro* diagnostic products and the *in vivo* products supplied by Pharmacia were antigenically comparable to assure conformity.

For the SC, each subject was stung on the medial aspect of the forearm by a single honeybee, and the stinger was left in place for at least 2 minutes in order to represent the worst possible field sting. These challenges were performed in an emergency room setting in different hospitals with appropriate medical staff and treatment facilities present. Individuals identified as being at risk of systemic reaction by virtue of the HRM or previous reactions were monitored closely, and an intravenous line was established before the sting challenge. Reactions were managed by a team of specially instructed physicians from each facility under the direction of the senior author. All subjects were monitored after SC for 30 to 60 minutes. Those with no response or a painless local response after this time were discharged without further treatment. On discharge all subjects were instructed to return to the emergency room if they experienced a late systemic reaction or persistent uncomfortable local reaction. Subjects with a large or painful local response were treated with ice, analgesics, and one 50 mg dose (25 mg for children) of orally administered diphenhydramine before discharge. Subjects showing signs of a mild systemic reaction were monitored for 30 to 60 minutes after the onset of symptoms to determine whether they

		VS IgE Level				
VS	VS IgG U/ml	0	1	2	3	4
Skin Test Result	<20	Risk Level 1		Risk Level 2		
	20-<100					
	100-<500	Risk Level 0				
	500-<1000					
	=>1000	Risk Level 1				
	<20					
	20-<100					
	100-<500	Risk Level 0				
	500-<1000					
	=>1000	Risk Level 2		Risk Level 3		
	<20					
	20-<100					
	100-<500					
	500-<1000					
	=>1000	Risk Level 3		Risk Level 4		
	<20					
	20-<100					
	100-<500					
	500-<1000					
	=>1000	Risk Level 4				

FIG. 1. Hypothetical risk model (HRM).

would progress to a severe systemic reaction. If there was no progression after this time, they were treated with diphenhydramine 50 mg (25 mg for children) four times a day for 1 day and as required afterward. Any sign of progression to a severe systemic reaction was immediately treated with 0.3 to 0.5 ml (0.01 ml/kg for children) subcutaneously administered aqueous epinephrine 1:1000 dilution, which was repeated as required. Intravenous administration of epinephrine (1:10,000 dilution) was reserved for shock situations when intravenous colloidal solution was administered as well.

Statistical analyses

Initially, the results for each of the measures were considered individually so that their independent association with SC response could be assessed. Then the relationship between the HRM and actual SC responses was examined. Finally, all of the factors were considered together in a series of regression analyses to determine whether a different combination of measures would be a better predictor of response to a sting than the HRM.

Both parametric and nonparametric tests were used

as indicated. These tests included Kruskal-Wallis non-parametric analysis of variance, Kendall's rank correlation, and Pearson's goodness of fit chi square. Finally, least-squares and logit regression models were fitted to the data. For the chi square tests and the regression analyses, the data for each measure were classified according to the criteria in Table I. Analyses were performed with the SAS (SAS Institute, Cary, N.C.) and BMDP (BMDP Statistical Software, Inc., Los Angeles, Calif.) statistical software packages.

RESULTS
Demographics

The 128 male and 96 female subjects ranged in age from 6 to 79 years with a mean age of 34.7 years. Table II shows the relationship between occupation and sting frequency during the bee-keeping season. One hundred ninety-two of the subjects (85.7%) had had the same sting frequency for more than 2 years, five (2.2%) had the same frequency for between 1 and 2 years, and 27 (12.1%) had experienced a change in frequency within the last year.

TABLE I. Classification of data for statistical analyses

VS RAST IgE			VS skin tests		VS RAST IgG	
Class	Level (PRU/ml)	Grade	Erythema (mm)	Wheal (mm)	Class	Level (U/ml)
0	<0.35	0	<5	<5	1	<20
1	≥0.35 and <0.70	±	5 to 10	5 to 10	2	≥20 and <100
2	≥0.70 and <3.50	1	11 to 20	5 to 10	3	≥100 and <500
3	≥3.50 and <17.5	2	21 to 30	5 to 10	4	≥500
4	≥17.5	3	31 to 40	10 to 15 or pseudopods		
		4	>40	>15 or many pseudopods		

TABLE II. Relationship between sting frequency and occupation

	Sting frequency (per yr)		Sting frequency (per day)					Totals
	Never	<1	<1	1-4	5-10	11-25	26-100	
Occupation								
Full-time beekeeper	0	0	1	5	12	10	2	30
Family member	2	30	24	21	3	0	0	80
Employee	0	0	2	9	1	2	0	14
Hobbyist	0	10	7	28	4	0	1	50
Retired beekeeper	0	7	1	0	0	0	0	8
Entomologist	0	0	2	3	1	0	0	6
Other	6	24	4	2	0	0	0	36
Totals	8	71	41	68	21	12	3	224

TABLE III. Relationship of reactivity to SC to VS skin test results with 0.1 µg/ml and 1.0 µg/ml honeybee venom

Challenge reactivity	Venom concentration (µg/ml)	Grade of reaction				
		0	±	1	2	3
Local	0.1	133	22	23	7	0
	1.0	95	35	39	14	2
Mild systemic	0.1	3	10	13	3	1
	1.0	3	4	14	8	1
Severe systemic	0.1	1	0	3	5	0
	1.0	2	0	3	2	2

Occupation was found to be associated with time from last sting ($\chi^2[6] = 46.75, p < 0.001$); beekeepers and employees were stung most recently, whereas family members tended to have a greater interval since their last sting. A relationship was also noted between occupation and response to SC ($\chi^2[3] = 25.52, p < 0.001$). This relationship was most apparent for beekeepers and family members, with the former experiencing a low number of systemic reactions and the

latter experiencing a high number of systemic reactions.

Among the subjects who experienced a systemic reaction from SC, the average time between last sting and reporting for the SC was 705 days, with one of the 39 (2.6%) having no history of a previous sting and one (2.6%) receiving his last sting on the same day he reported for the SC. The latter subject had an equivocal mild systemic reaction at SC, which was not out of character

VS RAST IgG ₄		Sting response and history of reaction		
Class	Level (%)	Reaction category	Class	Criteria
1	< 4	None	0	No local swelling
2	≥ 4 and < 9	Normal local	1	Local swelling < 5 cm in diameter
3	≥ 9 and < 13	Large local	2	Local swelling ≥ 5 cm in diameter
4	≥ 13	Mild systemic	3	Not life-threatening; generalized reaction confined to the skin and consisting of urticaria and/or angioedema
		Severe systemic	4	Potentially life-threatening; may include cutaneous symptoms but has laryngeal, and/or respiratory, and/or cardiovascular symptoms

with the reaction earlier in the day. Seven (3.8%) of the 185 subjects who did not have a systemic reaction to SC had no history of sting, and 37 (20.0%) had their last sting on the same day they reported for SC. The average time between last sting and reporting for SC was 615 days for the nonreactors. A nonparametric analysis of variance showed the difference between these two means to be significant ($p < 0.001$). However, this observation is confounded by the association between occupation and time from last sting and the association between occupation and response to SC.

VS skin tests

Of the 224 subjects tested, 55 (24.6%) showed a reaction of 1 or greater on skin testing at a venom concentration of 0.1 µg/ml, and 16 (7.1%) of these were level 2 or greater. When the same subjects were tested at a venom concentration of 1.0 µg/ml, 85 (37.9%) showed a reaction of 1 or greater, and 29 (12.9%) of these were level 2 or greater.

It appeared that subjects who were stung less frequently were more likely to respond positively to skin testing, and this trend was found to be significant after data were condensed to fit the requirements of the chi square goodness of fit test ($\chi^2[9] = 17.39, p < 0.05$).

Table III shows the response to SC in relationship to skin test values. Skin testing with 0.1 µg/ml honeybee venom had a higher level of specificity than with 1.0 µg/ml, while still having a comparable level of sensitivity. Using 2 or greater as a positive skin test response, we found that skin testing with 1.0 µg/ml honeybee venom more than doubled the number of subjects with clinically irrelevant positive responses (i.e., those with skin tests ≥ 2 who did not have a systemic reaction to SC) relative to 0.1 µg/ml, from seven (3.8%) to 16

(8.6%). This reduction in specificity was accompanied by only a minor increase in subjects with clinically relevant positive responses (i.e., those with skin tests ≥ 2 who did have a systemic reaction to SC), from nine (23.1%) to 13 (33.3%), and a decrease in sensitivity at the severe systemic level. The strength of the relationship between sting reactivity and skin test results at 0.1 µg/ml was assessed with a chi square goodness of fit test. The results of this test showed that the skin test results were associated with the severity of the reaction to stinging ($\chi^2[16] = 96.42, p < 0.001$). Even though the VS skin test response was a significant predictor, it did not identify all reactors. One subject experienced a severe systemic reaction with a negative VS skin test result but had a positive histamine control and an elevated VS RAST IgE level (2.4 Phadebas RAST units/ml, level 2). The association between skin test results and severity of reaction to sting is demonstrated in Fig. 2 by the increase in systemic reactions across levels of skin test response. A similar increase across response levels can be seen for the severe systemic reactions, even though the trend appears to be affected at the highest response level because of the sample size there.

With a chi square goodness of fit test, a positive association of skin test responses with VS RAST IgE was found ($\chi^2[16] = 154.95, p < 0.001$). The negative association of skin test results with VS RAST IgG was also significant ($\chi^2[12] = 37.76, p < 0.001$).

VS RAST IgE

A total of 32 subjects had substantially elevated levels of VS RAST IgE (≥ 3); 24 (75%) of these were stung less than once per day during the beekeeping season, whereas the remaining eight were stung between one and four times per day.

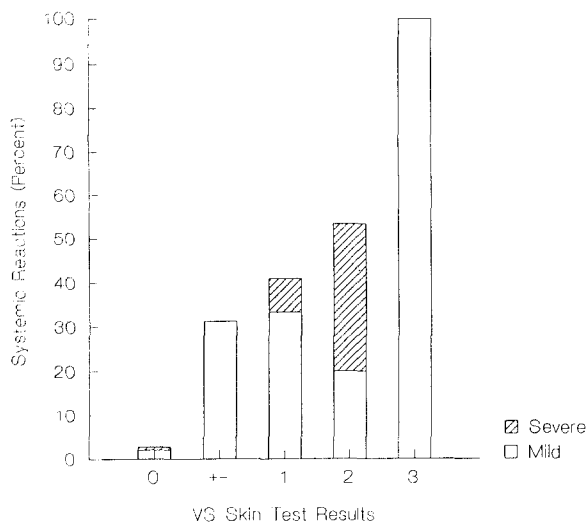


FIG. 2. Proportion of systemic reactions by VS skin test results.

None of the 36 subjects who had been stung more than four times a day during the season had an elevated VS RAST IgE level, whereas 30 (83.3%) had class 0 levels. Thus for the most part, the subjects who had been stung more frequently had lower VS RAST IgE levels than those who had been stung less frequently. This association was confirmed by a Pearson's chi square goodness of fit test ($\chi^2[24] = 44.47, p < 0.01$).

The association between VS RAST IgE levels and response to SC is shown in Fig. 3. A trend of increasing systemic reactions across the VS RAST IgE levels is evident and is more pronounced for the severe systemic reactions alone. Of the 32 subjects with elevated VS RAST IgE levels (≥ 3), 20 (62.5%) sustained a systemic reaction, including eight of the nine subjects who had severe reactions. Only six (4.5%) of the 134 subjects with undetectable levels of VS RAST IgE had a systemic reaction to the SC, none of which were severe; whereas 120 (90.0%) had no response or a mild local response. Thus VS RAST IgE levels were a sensitive predictor of risk of a reaction to honeybee stings ($\chi^2[16] = 128.01, p < 0.001$). The only systemic reaction that was not associated with a VS RAST IgE level of 3 or greater (2.4 PRU/ml) was also not predicted by VS skin test results. This example demonstrates the usefulness of VS RAST IgE levels in the prediction of risk because even though the subject's level in this case was not 3 or greater, it was still elevated enough to indicate allergic reactivity. The VS skin test result was negative with normal histamine response and was of no predictive value.

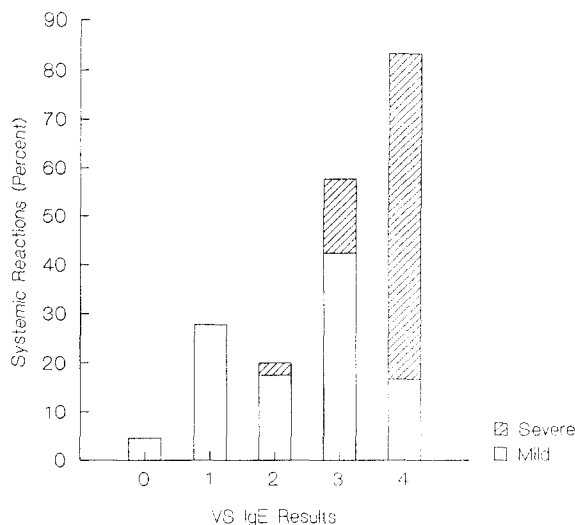


FIG. 3. Proportion of systemic reactions by VS RAST IgE results.

VS RAST IgG

In general, levels of VS RAST IgG increased with an increased frequency of stinging, but there was still a wide range of levels among subjects who reported similar sting rates. The association between VS RAST IgG levels and frequency of stinging was confirmed with a Kruskal-Wallis rank sum test ($\chi^2[2] = 86.00, p < 0.001$).

When the VS RAST IgG levels were classified into four categories, the number of subjects with elevated VS RAST IgG levels (> 500 U/ml, class 4) increased substantially as sting frequency increased ($\chi^2[18] = 156.47, p < 0.001$). Of the 104 persons who had been stung at least once per day during the beekeeping season, 58 (55.8%) had VS RAST IgG levels above 500 U/ml, whereas four (3.8%) had no measurable level. Of the 112 persons who had been stung less than once per day, four (3.6%) had levels greater than 500 U/ml, whereas 45 (40.2%) had undetectable levels of VS RAST IgG. Six of the eight subjects who had never been stung had no measurable VS RAST IgG, and two had a measurable but low level (36 and 44 U/ml, respectively).

When the VS RAST IgG levels were related to the subjects' subsequent reactions to SC, there was no significant association ($\chi^2[2] = 2.36, p > 0.30$). However, it is noteworthy that none of the 62 subjects with a VS RAST IgG level greater than 500 U/ml had a severe systemic reaction to SC, and only three (4.8%) had a mild systemic reaction (two of these three had a VS RAST IgE class of ≥ 3).

The relationship of VS RAST IgG₄ levels with

those of VS RAST IgG and the ability of VS RAST IgG₄ to predict the response to the SC was explored. The absolute values of VS RAST IgG and VS RAST IgG₄ were substantially correlated ($r = 0.77$, $p < 0.01$). For further analysis of the data, the subjects' VS RAST IgG₄ values were classified into four levels. As would be expected, there was considerable agreement between VS RAST IgG and VS RAST IgG₄ levels ($\chi^2[9] = 235.99$, $p < 0.001$), and the association between VS RAST IgG₄ and sting rate was significant ($\chi^2[18] = 177.10$, $p < 0.001$). When VS RAST IgG₄ was used as a predictor of the response to the SC, the association was again significant ($\chi^2[2] = 12.11$, $p < 0.01$).

With the chi square goodness of fit test, it was found that for the subjects in this study as a whole, VS RAST IgE was negatively associated with both VS RAST IgG ($\chi^2[12] = 33.92$, $p < 0.001$) and IgG₄ ($\chi^2[12] = 28.192$, $p < 0.01$), but on an individual scale, five (2.2%) subjects had both a VS RAST IgE level of 3 or greater and a VS RAST IgG level of greater than 500 U/ml.

Sting challenge

Thirty-nine subjects had systemic reactions (9 severe and 30 mild), 157 subjects had local responses, and 28 had negative responses to SC.

One child with no history of sting but with daily exposure to honeybee products reacted systemically to SC. Other systemic reactions occurred in 31 subjects (79.5%) who were stung less than once per day during the beekeeping season and in seven (17.9%) who were stung between one and four times per day during the beekeeping season. None were stung more than four times a day during the beekeeping season. On the history of reaction scale shown in Table I, the mean reaction level for the 178 subjects who had been stung previously was 1.76, whereas for the 38 subjects who had systemic reactions and had been stung previously, the mean reaction level was 3.34 ($p < 0.001$). Of the 38 previously stung systemic reactors, seven (18.4%) reported having only a local reaction, whereas 31 (79.5%) reported having systemic reactivity to a honeybee sting at some time in the past. However, 54 (63.5%) of the 85 who reported a history of systemic reaction did not experience a systemic reaction to SC.

HRM

Of the 70 subjects placed in the lowest risk group with the use of the HRM only three (4.3%) had reactions that could not be considered local.

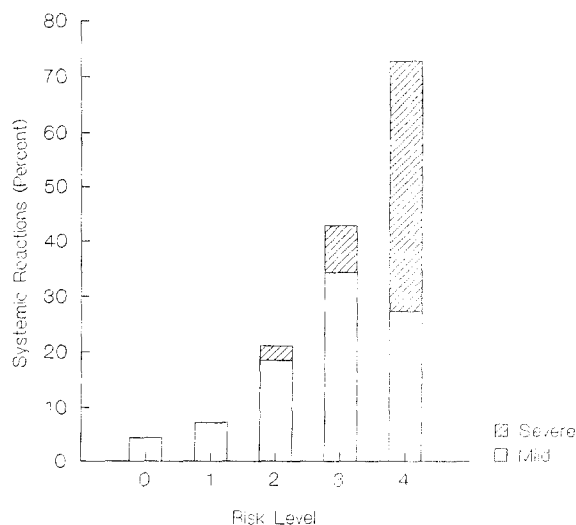


FIG. 4. Proportion of systemic reactions by risk level according to the HRM.

Two of these three subjects had mild equivocal reactions. Nasal congestion was seen in one, whereas the other seemed to have symptoms related to a preexisting pharyngitis. The third in this category had some coughing and lacrimation. In the second risk group five (7.1%) of the 70 subjects experienced a mild systemic reaction, with three of the five having reactions that did not progress beyond mild cutaneous expression. No subjects in this risk group had a severe systemic reaction. Eight (21.1%) of the 38 subjects in the middle risk group had a systemic reaction, and one (2.6%) had a severe systemic reaction. In the second highest risk group there were 15 (42.9%) systemic reactions among the 35 subjects, but the number of severe reactions increased to three (8.6%). Of the 11 subjects predicted to be at the highest risk of having systemic reactivity, five (45.5%) had severe systemic reactions and three (27.3%) had mild systemic reactions, resulting in a total of eight (72.7%) systemic reactions overall. A stepwise tendency to systemic reactivity with a distinct order of risk among the levels can be seen in Fig. 4. A similar increase is also apparent for the severe reactions, which increase in proportion dramatically between risk levels 3 and 4.

The apparent association between the proportion of systemic reactions (both total and severe) and risk level was tested by performing a logit analysis on the data; an increase in risk level was significantly associated with an increased proportion of systemic reactions ($p < 0.001$). In other words, the odds of having a systemic reaction increased as the risk level increased.

TABLE IV. Chi square values and LR for a number of logit models tested on the study data

Model No.	VS RAST IgE	VS RAST IgG	VS RAST IgG ₄	VS skin test	Intercept	LR
Three-parameter models						
1	20.71	3.92 (NS)		8.65 (NS)	31.68	0.999
2	19.26		9.76	7.64 (NS)	12.43	0.999
Two-parameter models						
3	41.04	4.85 (NS)			30.38	0.600
4	38.60		11.28		12.15	0.834
5	19.13			9.64*	118.95	0.819
6		1.63 (NS)		38.82	31.83	0.236
7			6.88 (NS)	37.07	12.07	0.264
One-parameter models						
8	40.15				122.40	0.759
9		1.43 (NS)			26.31	0.003
10			9.30 (NS)		12.06	0.033
11				38.97	143.02	0.050

Values are significant at $p < 0.01$ unless otherwise noted.

NS, Not significant.

* $p < 0.05$.

HRM refinement

The preceding section has shown that the HRM provides an accurate method for predicting an individual's risk of systemic reaction to SC. However, because the goal of this study was to determine the most accurate combination of parameters for the prediction of risk, we have in this section looked for ways to improve on the predictive power offered by the original HRM. To accomplish this we used two statistical methods to examine the predictive power of all possible combinations of the study parameters.

The first statistical analysis was performed with the logit or logistic regression method. This analysis determined the ability of each model to predict the occurrence of a systemic reaction after a SC. For the purpose of this analysis, subjects who had a mild systemic or severe systemic reaction were considered as having a systemic reaction, and all other subjects were considered as having as nonsystemic reaction. Table IV shows that VS RAST IgE is the best single predictor (model no. 8) and provides a significant contribution to the predictive power of every model in which it is included.* The VS skin test result was also found to be a good predictor on its own, but

it was not totally reliable and its significance greatly decreased whenever it was included in a model that also contained VS RAST IgE. When examined by itself, the predictive power of VS RAST IgG₄ was not significant (model no. 10), but it did seem to have complementary function when combined with VS RAST IgE (models nos. 2 and 4). VS RAST IgG was not significant by itself, nor in any combination with other parameters.

Although the three-parameter models appear to be the best predictors of systemic reaction, they are of limited use, as demonstrated by the non-significant parameters.† A less powerful but more widely applicable model is the two-parameter model with VS RAST IgE and VS RAST IgG₄ (model no. 4). The likelihood ratio (LR) for this model was high (0.834), and both parameters were significant with a p value of less than 0.01. The predictive power of another two-parameter model in which VS RAST IgE and VS skin tests are used (model no. 5) was a close second (LR = 0.819). When this model is compared with VS RAST IgE on its own (model no. 8), it is apparent how little the combination of VS skin test results and VS RAST IgE improves on the

*In a four-parameter model (VS RAST IgE, VS RAST IgG, VS RAST IgG₄, VS skin test) none of the parameters were significant.

†Nonsignificant parameters in a model indicate that the model is generally not applicable even though it may be a good model for the study population.

TABLE V. Results of the regression analyses performed on the study data

Model No.	VS RAST IgE	VS RAST IgG	VS RAST IgG ₄	VS skin test	IgE · IgG	Intercept	R ²	Sting rate
1	0.504	−0.03		0.404	−0.092	1.96	0.379	
2	0.716		−0.084	0.342	−0.154	2.11	0.410	
3	0.322		−0.169	0.410		2.32	0.389	
4	0.326	−0.07		0.438		2.05	0.371	
5	0.801			0.343	−0.188	1.90	0.406	
6			−0.155	0.846		2.40	0.303	
7	0.458		−0.192			2.42	0.355	
8	0.315			0.451		1.88	0.365	
9*	0.329			0.387		2.22	0.418	−0.174
10*				0.832		2.32	0.389	−0.162
11*	0.456					2.29	0.388	−0.191

All models are significant at $p < 0.01$.

IgE · IgG is a parameter that takes into account interaction between VS RAST IgE and VS RAST IgG.

*Includes sting rate (last column).

predictive power of VS RAST IgE alone. Sting frequency was also included in a series of analyses along with other parameters, and it improved the fit of the model when included with VS RAST IgE.

The second statistical analysis performed was linear regression. This analysis determined the ability of each model to predict the severity of a reaction to SC as opposed to the occurrence of reaction. This analysis predicted the severity of each subject's reaction to SC on a continuous scale. The R^2 values* in Table V demonstrate that no combination of parameters is very effective in predicting the severity of a response to SC. The best combination was a three-parameter model with VS RAST IgE, sting rate, and VS skin test (model no. 9; $R^2 = 0.418$). Similar predictive power was seen in model no. 2 (VS RAST IgE, VS RAST IgG₄, VS skin test, and IgE · IgG) with an R^2 value of 0.410. Further examination showed that model no. 5 (VS RAST IgE, VS skin test, IgE · IgG) provides equivalent predictive power ($R^2 = 0.406$) to model no. 2 with one less parameter.

DISCUSSION

On the basis of the results of this study, it appears that skin testing with 0.1 $\mu\text{g/ml}$ venom, as opposed to 1.0 $\mu\text{g/ml}$, markedly increases specificity while having little effect on overall sensitivity

and improving sensitivity at the severe systemic level. This increase in specificity is in contrast to the statement by Golden and Schwartz²⁰ that clinically irrelevant VS skin test results are quite unusual with concentrations of venom of 1.0 $\mu\text{g/ml}$ or less. We conclude that for skin testing, 0.1 $\mu\text{g/ml}$ venom is the best concentration to predict systemic reactivity to honeybee stings. Our results in this area support findings of some¹² but differ with those of others who claim that skin testing with venom at a concentration of 1.0 $\mu\text{g/ml}$ provides better results.^{21, 22} Although a skin test at venom concentration of 0.1 $\mu\text{g/ml}$ was found to be a good predictor of reaction to honeybee stings, it was not as effective as VS RAST IgE. There also appeared to be a relationship between VS skin test results and sting frequency. In another study²³ the authors found that neither the size of the skin test reaction to venom nor the lowest concentration yielding a positive venom skin test result correlated with the most severe systemic reaction during venom immunotherapy. This observation is generally in agreement with the results of our regression analysis, which showed that no single parameter or combination of parameters was strongly correlated with severity of response to SC. However, results of this nature must be interpreted within the boundaries of the tests applied to the data, which some authors have failed to do.²⁴ The results do not show an absence of relationship between VS skin test results and response to SC, but rather they show that the variability of response to SC is not satisfactorily explained by a linear relationship with VS skin test results. Other methods of analysis have

*A large R^2 value indicates that the model has strong predictive power. An R^2 value of 1.0 indicates perfect prediction and an R^2 value of 0 indicates no relationship between the parameters in the model and response to SC.

TABLE VI. Relative predictive power and clinical applicability of evaluated methods of prediction

Method	Groups	Predictive power	Clinical applicability
Linear regression	∞	Low	Medium
HRM	5	Medium	High
Logistic regression (logit)	2	High	Low

clearly shown that the relationship between VS skin test results and severity of response to SC is not a random one.

In this study the VS RAST IgE level was found to be a sensitive indicator of response to a honeybee sting. However, other studies have produced conflicting results.^{25, 26} In addition, the VS RAST IgE level was found to have a strong positive association with VS skin test response, which is in agreement with results of previous studies.^{16, 17, 27, 28} A negative association between VS RAST IgE levels and both VS RAST IgG and VS RAST IgG₄ levels was also noted; this expected finding has been observed by others.²⁶

In confirmation of findings by other investigators,²⁹ little association was found between levels of VS RAST IgG and adverse response to honeybee sting even though it was significantly associated with sting rate. However, it appears that VS RAST IgG at a sufficient level may be protective against a severe systemic reaction to a honeybee sting because no person in this study with a VS RAST IgG level greater than 500 U/ml had such a reaction, and only three had a mild systemic reaction in spite of coexisting VS RAST IgE observed in two subjects. Even though a level of significance cannot be attached to this observation because of the small number of subjects with high levels of VS RAST IgG and contradictory elevated indices of risk, this observation should lend support to previous attempts to define a protective level of VS RAST IgG.^{30, 31} In particular, it supports the observation by Urbanek et al.³² that the majority of patients are protected by a VS RAST IgG level greater than 400 U/ml. Because there is considerable agreement between levels of VS RAST IgG and VS RAST IgG₄, VS RAST IgG₄ offers no practical advantage over VS RAST IgG in the prediction of response to SC.

We believe our results to be accurate and representative of the population from which they were sampled, but there are confounding factors inherent in a study of this nature. For example, subjects were directed to not take medications that could interfere with skin test results for 2 weeks before SC, and these medications were

reviewed again immediately before challenge. Nevertheless, it is possible that a minority of subjects were noncompliant in this regard. A more likely source of bias is that some of the reactions to SC may have progressed in severity if we had waited longer before administering treatment. None of the subjects who were untreated (those with minor local reactivity only) progressed to systemic reactivity over the subsequent 24 hours. Other subjects who had symptoms that required management with antihistamines, but not epinephrine, may have had progressive reactivity if left untreated. However, the incidence of systemic reactivity to SC was relatively high in this study (17.4%), so it appears that this was not a large source of bias. It should also be noted that the subjects in this study were recruited from a population of individuals at higher risk of being stung by honeybees than the general population, so caution should be used when applying the results of this study to the general population.

The ethical issue regarding SC of subjects who may have had adverse reactions in the past centers on the recognition that previous reactions are not necessarily associated with subsequent reactivity, which is also observed in penicillin sensitivity. It was the objective of this study to determine factors, other than history of reactivity, which would be reliable indicators of risk because "the first reaction" may be severe but predictable if certain tests were applied. Fundamental to the ethical considerations was that all of those who were sting challenged were connected with the beekeeping industry and thus at risk of being stung on a daily basis. Given this information and the remoteness of some of the sites of beekeeping operations, it was deemed more acceptable for subjects to be stung under controlled circumstances than where medical services were not readily available. Common sense prevailed, and no subject who claimed a history of very severe reactions to stings either expressed a desire or was accepted for an SC.

Three methods of predicting an outcome of SC were individually evaluated. A comparison of these methods shows a definite order among them

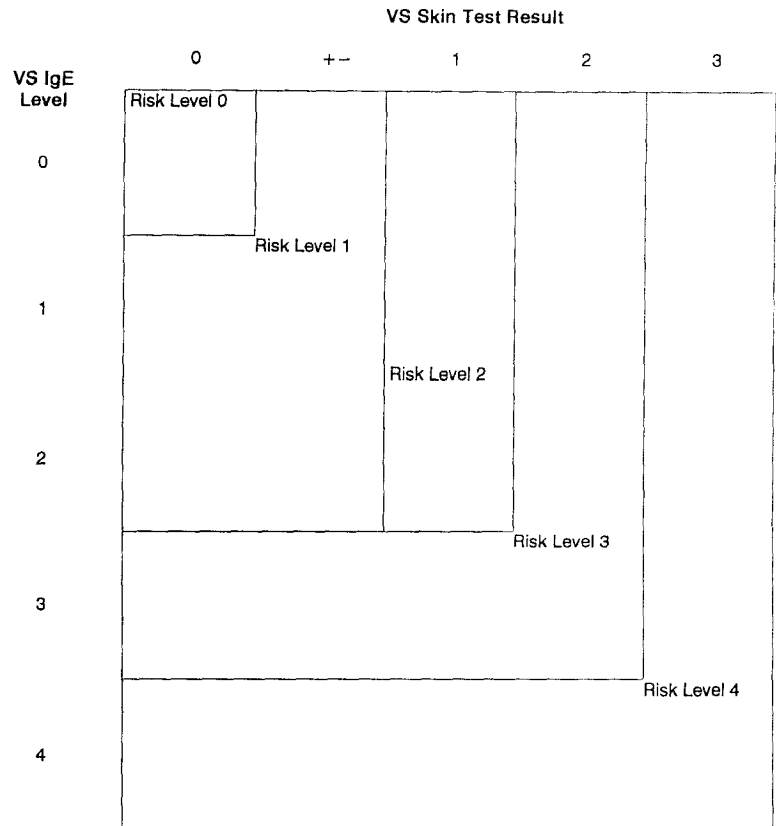


FIG. 5. Proposed risk model.

in certain areas. As one would expect, it appears that the methods that attempt to make more precise predictions of outcome have lower predictive power, whereas those that attempt to make more general predictions have higher predictive power. For example, the linear regression method attempts to predict the severity of a subject's response to SC on a continuous scale, which effectively involves placing each of the subjects into one of an infinite number of groups. When this method provides a correct prediction, it is very descriptive and clinically useful, but as evidenced by the R^2 values in Table V, it has a low predictive power and does not regularly provide a correct prediction. At the other extreme is the logit method, which attempts to predict a subject's response to SC by placing each subject in one of two groups: systemic reaction or nonsystemic reaction. Because of its simplified approach, this method has a much higher predictive power than the linear regression method. However, it has a drawback in that it is making an exclusive prediction with less than 100% accuracy. In other words, this model will invariably place some subjects in the nonsystemic reaction risk group when they

are liable to have a systemic reaction, possibly severe.

Thus a balance must be struck between predictive power and clinical applicability. The HRM achieves this balance by assigning each subject a level of risk for having a systemic reaction to a SC. This is accomplished by placing each subject in one of five discrete, ordinal groups from low risk (level 0) to high risk (level 4). This method does not have the predictive power of the logit method, but it allows for less accurate prediction by having a number of groups with sequential levels of risk. Similarly, it does not provide as descriptive a prediction as the linear regression method, but it has a much higher predictive power. These observations are summarized in Table VI.

The preceding discussion has shown that the best balance between predictive power and clinical applicability is achieved by using a risk model (RM) method, but the choice of parameters for the best model is not clear. Analysis of other methods suggests that an RM with greater predictive power could be constructed by varying the parameters from those used in the HRM.

An improved HRM would still have to include

TABLE VII. Response to sting challenge and history of reaction

	Response to sting challenge		
	Systemic	Nonsystemic	Total
History of reaction			
Systemic	31	54	85
Nonsystemic	7	124	131
Total	38	178	216

VS RAST IgE because our results indicate that it is the best single predictor, and it was an integral part of every model in which it was included. On the other hand, the predictive ability of VS RAST IgG was found to be minimal. The role of VS skin test in an improved RM is not as clear as the other parameters. VS skin test results were highly correlated with and complementary to VS RAST IgE levels, so one must logically question what they could offer an RM that VS RAST IgE levels do not, aside from the obvious clinical convenience. However, skin test results did make a significant addition to some of the best models, even when included with VS RAST IgE.

Two parameters not used in the HRM, which could add to the predictive power of an RM, are VS RAST IgG₄ and sting rate. Unfortunately, these parameters are not as widely applicable as the ones already mentioned. Sting rate would not be very useful because the average individual will have a sting rate that is limited and of questionable accuracy. The other parameter, RAST IgG₄, made a significant addition to the predictive power of some of the models but was not a useful predictor in its own right, and methods for its detection are not widely available.

Consideration of all the parameters suggests that the best model would rely mainly on VS RAST IgE levels for its prediction of risk but would also include some input from VS skin test results. Sting rate could also add to an RM but only if the subject has an extensive and reliable sting rate. Finally, VS RAST IgG₄ level would be of limited use because of its weak predictive power and the difficulty involved in its determination.

Our research suggests that in clinical practice, patients suspected of being at risk for a life-threatening reaction to honeybee stings, as determined by either history of systemic reactivity or exposure to honeybee stings and bee products, may be accurately assessed by undertaking VS RAST IgE testing and VS skin testing and then applying the results to an RM. For practical purposes, we have structured the RM along the

same lines as the HRM (Fig. 1), with a slight revision of parameters and a restructuring of risk groups. The structure of the proposed model can be seen in Fig. 5.

To evaluate the performance of the proposed RM (PRM) we have compared it with the classical manner of determining recommendations for immunotherapy. However, it must be emphasized that the PRM is designed to assess risk for those in danger of being stung and not to recommend which individuals in the general population require immunotherapy. For the comparison, data regarding the subjects' most severe reaction to honeybee sting before the SC were analyzed, and a classical RM (CRM) in which VS skin test and reaction history were used was constructed.

History of reaction was a significant predictor of reactivity to SC as an individual parameter ($\chi^2[2] = 34.87, p < 0.001$) and was found to be a better predictor of decreased reactivity than increased reactivity. Table VII shows how history of reaction and response to sting challenge are related. From Table 7 it can be seen that the predictive value of a history of nonsystemic reaction is very high (124 of 131 or 94.7%). It is also apparent that a history of systemic reaction is of little predictive value (31 of 85 or 36.5%) in determining who will have a systemic reaction to SC. Accordingly, the classical RM relies mainly on VS skin test results to identify the reactors and on reaction history to identify the nonreactors. This model is shown in Fig. 6.

The two RMs were compared by calculating the odds ratio* of systemic reaction at each risk level. This value expresses the odds that a given risk level would be expected in a patient with (as opposed to

*The *odds ratio* is usually referred to as the *likelihood ratio* when diagnostic tests are evaluated in medical literature. However, the term *likelihood ratio* has already been used in this article in the true statistical sense with respect to the logit analysis of the individual and combined parameters. To avoid confusion the term *likelihood ratio* will be used in reference to the logit analysis and the term *odds ratio* will be used in reference to the evaluation of diagnostic tests.

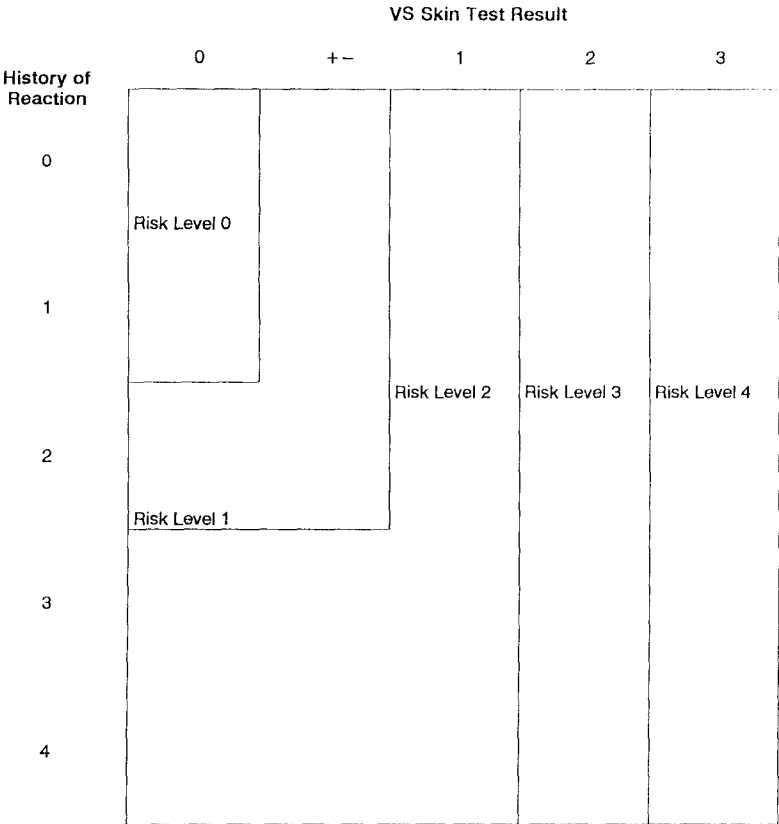


FIG. 6. Classical risk model.

TABLE VIII. Odds ratios and diagnostic levels for the five risk levels of the proposed and classical risk models

Risk level	Subjects with systemic reaction to SC		Subjects without systemic reaction to SC		Odds ratio	Diagnostic level
	No.	Proportion	No.	Proportion		
PRM						
4	6	0.1538	1	0.0054	28	Rule-in
3	16	0.4103	15	0.0811	5.1	Intermediate high
2	7	0.1795	18	0.0973	1.8	Indeterminate
1	7	0.1795	36	0.0162	0.92	Intermediate low
0	3	0.0769	115	0.6216	0.12	Rule-out
CRM						
4	2	0.0513	0	0	*	Rule-in
3	25	0.6410	33	0.1784	3.6	Intermediate high
2	10	0.2564	35	0.1892	1.4	Indeterminate
1	1	0.0256	30	0.1622	0.16	Intermediate low
0	1	0.0256	87	0.4703	0.05	Rule-out

*The odds ratio is undefined here because both the subjects at this risk level experienced a systemic reaction, which places a 0 in the denominator of the odds ratio.

one without) a systemic reaction.³³ A clinically useful RM must have a high odds ratio at the highest risk level, a low odds ratio at the lowest risk level, and intermediate levels in between.

The odds ratio and diagnostic levels³⁴ displayed in Table VIII indicate that both the PRM and CRM satisfy the criteria for a clinically useful RM. Direct comparison is hampered because the

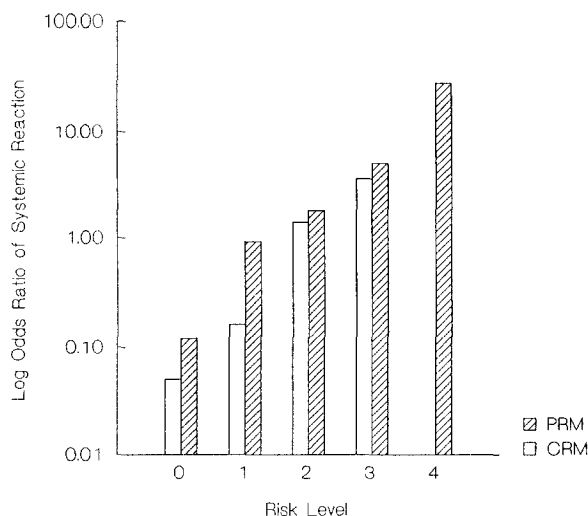


FIG. 7. Log odds ratio of systemic reaction by risk level according to the proposed and classical risk models.

odds ratio for risk level 4 is undefined in the CRM. This is a function of small sample size because both of the two subjects at this risk level experienced a systemic reaction, which places a zero in the denominator of the odds ratio. However, the odds ratio for risk level 4 in the PRM is quite high, which indicates that individuals who are in this risk level are likely to experience a systemic reaction when stung by a honeybee. As a means of comparison, the diagnostic level associated with an odds ratio of this magnitude is similar to that seen with a positive sputum smear for tuberculosis as confirmed by culture.³⁵ At the lower end of the risk scale, the RMs have comparable odds ratio, with the CRM identifying fewer nonreactors, but with more accuracy. The similarity of the intermediate odds ratios is evident in Fig. 7.

The thrust of this article has been to use laboratory criteria in determining risk of systemic reaction to honeybee stings, because it has been observed that prior clinical reactivity to venoms and drugs (e.g., penicillin) does not in itself predict future responsiveness to these antigens. This study was designed to determine a test or a combination of readily available tests, which would have a high level of predictive value. Previous reactions to sting, although recorded, were not a part of the RM. A secondary objective was to determine what test or group of tests could be applied to assist in treatment of potentially sting-reactive persons at risk from stings, recognizing that a life-threatening reaction could occur without warning.

Clearly anyone who had anaphylaxis in the past would be considered to be at risk, but the presence or absence of immunologic risk indicators, which have been shown to project reactivity, would be the best predictor of such reactivity. A child who has had mild dermal reactions and who has indicators of anaphylactic reactivity would be treated in part for the potential to be stung. For example, management could depend on whether he or she is in any particular risk situation such as being a member of a beekeeping family. A person who has had "anaphylaxis" to honeybee sting and demonstrates no immunologic reactivity to honeybee venom would not be expected to have IgE-mediated anaphylaxis to honeybee sting and would be treated accordingly. The ability of both the HRM and PRM to accurately assign risk is impressive, but the absence of strong correlation between any indicator of risk and the severity of reaction to SC, which is not unusual in evaluating risk in a variety of medical conditions, ultimately requires medical judgment.

Immunologic criteria have been shown in this study to be very important, and the combination of VS skin test reactivity and VS RAST IgE, which is used in the HRM and improved on in the PRM, is possibly the best indicator of allergic reactivity. A non-IgE-mediated reaction, such as an acute vasovagal episode, could not be predicted by the test. Indeed, the absence of allergic correlates after a collapse or other forms of response to a sting would be helpful in determining the nature of such a reaction.

We acknowledge the assistance of the STATLAB at Queen's University for their advice on the application and interpretation of statistical tests.

REFERENCES

1. Barnard JH. Studies of 400 Hymenoptera sting deaths in the United States. *J ALLERGY CLIN IMMUNOL* 1973;52:259-64.
2. Statistics Canada. Causes of death 1989. Health Reports 1991;3(suppl 11):166.
3. David MF. The incidence of anaphylactic sensitivity to sting among beekeepers in Manitoba [Abstract]. *Clin Invest Med* 1987;10(B suppl):B48.
4. Statistics Canada, Agriculture Division. Production and value of honey and maple products 1989. Ottawa: Minister of Supply and Services Canada, 1990:2-3.
5. Mosbech H. Insect allergy: a comparative study including case histories and immunological parameters. *Allergy* 1984;39:543-9.
6. Zora JA, Swanson MC, Yunginger JW. A study of the prevalence and clinical significance of venom-specific IgE. *J ALLERGY CLIN IMMUNOL* 1988;81:77-82.
7. Heinig JH, Engel T, Weeke ER. Allergy to venom from

- bee or wasp: the relation between clinical and immunological reactions to insect stings. *Clin Allergy* 1988;18:71-8.
8. Devey ME, Lee SR, Richards D, Kemeny DM. Serial studies of the functional affinity and heterogeneity of antibodies of different IgG subclasses to phospholipase A₂ produced in response to bee-venom immunotherapy. *J ALLERGY CLIN IMMUNOL* 1989;84:326-30.
9. Müller U, Berchtold E, Helbling A. Honeybee venom allergy: results of a sting challenge 1 year after stopping immunotherapy in 86 patients. *J ALLERGY CLIN IMMUNOL* 1991;87:702-9.
10. Nordvall SL, Uhlin T, Einarsson R, Johansson SGO, Öhman S. Bee keepers' IgG and IgE antibody responses to bee venom studied by means of crossed radioimmuno-electrophoresis. *Clin Allergy* 1984;14:341-50.
11. Bar-Sela S, Shalit M, Kalbfleisch JH, Fink JN. The relative value of skin tests and RAST tests in the diagnosis of bee sting hypersensitivity. *J ALLERGY CLIN IMMUNOL* 1983;72:690-4.
12. Georgitis JW, Reisman RE. Venom skin tests in insect-allergic and insect non-allergic populations. *J ALLERGY CLIN IMMUNOL* 1985;76:803-7.
13. Aalberse RC, van der Gaag R, Leeuwen J. Serological aspects of IgG antibody I: prolonged immunisation results in an IgG₄-restricted response. *J Immunol* 1983;130:722-6.
14. Bousquet J, Coulomb Y, Robinet-Levy M, Michel FB. Clinical and immunological surveys in bee keepers. *Clin Allergy* 1982;12:331-42.
15. Santrach PJ, Peterson LG, Yunginger JW. Comparison of diagnostic tests for Hymenoptera sting allergy. *Ann Allergy* 1980;45:130-6.
16. Patrizzi R, Mueller U, Yman L, Hoigne R. Comparison of skin tests and RAST for the diagnosis of bee sting allergy. *Allergy* 1979;34:249-56.
17. Miyachi S, Lessof MH, Kemeny DM. Evaluation of bee sting allergy and antibody assays. *Int Arch Allergy Applied Immunol* 1979;60:148-53.
18. Golden DBK, Marsh DG, Kagey-Sobotka A, et al. Epidemiology of insect venom sensitivity. *JAMA* 1989;262:240-4.
19. Sullivan TJ, Wedner HJ, Shatz GS, Yecies LD, Parker CW. Skin testing to detect penicillin allergy. *J ALLERGY CLIN IMMUNOL* 1981;68:171-80.
20. Golden DBK, Schwartz HJ. Guidelines for venom immunotherapy. *J ALLERGY CLIN IMMUNOL* 1985;77:727-8.
21. Hunt KJ, Valentine MD, Sobotka AK, et al. Diagnosis of allergy to stinging insects by skin testing with Hymenoptera venoms. *Ann Intern Med* 1976;85:56-9.
22. Golden DBK. Diagnosis and prevalence of stinging insect allergy. *Clin Rev Allergy* 1987;5:119-36.
23. Lockey RF, Turkeltaub PC, Olive ES, Hubbard JM, Baird-Warren IA, Bukantz SC. The Hymenoptera venom study III: safety of venom immunotherapy. *J ALLERGY CLIN IMMUNOL* 1990;86:775-80.
24. van der Zwan JK, Hack CE, van der Linden PWG, Struyvenberg A. Skin tests and specific immunoglobulin E are not related to the severity of insect-sting anaphylaxis in 324 patients [Abstract]. *J ALLERGY CLIN IMMUNOL* 1993;91:282.
25. Blaauw PJ, Smithuis LOMJ. The evaluation of the common diagnostic methods of hypersensitivity for bee and yellow jacket venoms by means of an in-hospital insect sting. *J ALLERGY CLIN IMMUNOL* 1985;75:556-62.
26. Kampelmacher MJ, Van der Zwan JC. Provocation tests with a living insect as a diagnostic tool in systemic reactions to bee and wasp venom: a prospective study with emphasis on the clinical aspects. *Clin Allergy* 1987;17:317-27.
27. Hoffman DR. Comparison of the RAST test to intradermal skin testing in the diagnosis of stinging insect venom allergy. *Ann Allergy* 1979;43:211-3.
28. Bousquet J, Menardo JL, Aznar R, Robinet-Levy M, Michel FB. Clinical and immunological survey in beekeepers in relation to their sensitization. *J ALLERGY CLIN IMMUNOL* 1984;73:322-40.
29. Parker JL, Santrach PJ, Dahlberg MJE, Yunginger JW. Evaluation of Hymenoptera-sting sensitivity with deliberate sting challenges: inadequacy of present diagnostic methods. *J ALLERGY CLIN IMMUNOL* 1982;69:200-7.
30. Golden DBK, Meyers DA, Kagey-Sobotka A, Valentine MD, Lichtenstein LM. Clinical relevance of the venom-specific IgG antibody level during immunotherapy. *J ALLERGY CLIN IMMUNOL* 1982;69:489-93.
31. Urbanek R, Forster J, Kuhn W, Ziupa J. Discontinuation of bee immunotherapy in children and adolescents. *J Pediatr* 1985;107:367-71.
32. Urbanek R, Krauss U, Ziupa J, Smedegård G. Venom-specific IgE and IgG antibodies as a measure of the degree of protection in insect-sting-sensitive patients. *Clin Allergy* 1983;13:229-34.
33. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown and Company, 1991:120.
34. Simel DL, Feussner JR, Delong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med Decis Making* 1987;7:107-14.
35. Boyd JC, Marr JJ. Decreasing reliability of acid-fast smear techniques for detection of tuberculosis. *Ann Intern Med* 1975;82:489-92.