Risk Analysis, Vol. 21, No. 1, 2001

A Randomization Test-Based Method for Risk Assessment in Neurotoxicology

Michael A. Bogdan,[1] Robert C. MacPhail[2] and John R. Glowa[3]*

A current trend in risk assessment for systemic toxicity (noncancer) endpoints is to utilize the observable range of the dose-effect curve in order to estimate the likelihood of obtaining effects at lower concentrations. Methods to accomplish this endeavor are typically based on variability in either the effects of fixed doses (benchmark approaches), or on variability in the doses producing a fixed effect (probabilistic or tolerance-distribution approaches). The latter method may be particularly desirable because it can be used to determine variability in the effect of an agent in a population, which is an important goal of risk assessment. This method of analysis, however, has typically been accomplished using dose-effect data from individual subjects, which can be impractical in toxicology. A new method is therefore presented that can use traditional groups-design data to generate a set of dose-effect functions. Population tolerances for a specific effect can then be estimated from these model dose-effect functions. It is based on the randomization test, which assesses the generality of a data set by comparing it to a data set constructed from randomized combinations of single point estimates. The present article describes an iterative line-fitting program that generates such a data set and then uses it to provide risk assessments for two pesticides, triadimefon and carbaryl. The effects of these pesticides were studied on the locomotor activity of laboratory rats, a common neurobehavioral end point. Triadimefon produced dose-dependent increases in activity, while carbaryl produced dose-dependent decreases in activity. Risk figures derived from the empirical distribution of individual dose-effect functions were compared to those from the iterative line-fitting program. The results indicate that the method generates comparable risk figures, although potential limitations are also described.

KEY WORDS: Risk assessment; randomization; dose-response models; neurotoxicity; pesticides; carbaryl; triadimefon

1. INTRODUCTION

Toxicological risk assessment attempts to characterize the probability of obtaining an adverse effect in humans exposed to an agent. Typically, this estimate is based on effects of the agent in laboratory animals, since both practical and ethical issues can constrain the collection of data from exposed humans. There have been several different approaches to using data obtained in animals to assess chemical risks in humans. One of the earliest approaches, often used with carcinogenic end points, has been the attempt to predict the effects of extremely low doses by extrapolating from effects seen at much higher doses.[1] Low-dose extrapolation results in risk estimates that are model dependent, often lack statistical

[1] Stanford University Medical Center, Stanford, CA.
[2] Neurotoxicology Division, National Health and Environmental Effects Research Laboratory, Research Triangle Park, NC.
[3] Louisiana State University Medical Center at Shreveport, Shreveport, LA.
*Address correspondence to John R. Glowa, Department of Pharmacology and Therapeutics, Louisiana State University Medical School in Shreveport, 1501 Kings Highway, P.O. Box 33932, Shreveport, LA 71103-3932; jglowa@lsumc.edu.

0272-4332/01/0200-0107$16.00/1

or biological justification, and are impossible to validate for most end points. Another early risk assessment method was the safety-factor (SF), or more recently the reference dose (RfD), approach. It first attempts to experimentally determine a dose that is the lowest observable adverse effect level (LOAEL) or the no observable adverse effect level (NOAEL). It then divides that dose by a series of uncertainty factors (UFs) to arrive at an exposure level that is presumed to be either devoid of risk or will allow an acceptably safe daily exposure to the agent.[2] UFs can be based on suspected differences between animal and human susceptibility, individual differences, the nature of exposure (i.e., less than lifetime versus chronic), or "confidence" in the data base. The attempt to establish "safe" doses has been criticized for a lack of precision, a dependence on dose spacing, and a failure to take into account the slope and variability in effect of the agent. Perhaps the greatest concern, however, is that this approach does not fully utilize the available dose-effect data.

Newer methods of risk assessment, based on effects obtained in the measurable portion of the dose-effect function, have been proposed.[3–7] Most of these methods precisely (mathematically) fit the dose-effect function, using both the slope and variability in effect to produce risk figures, and do not extrapolate far beyond the low end of the dose-effect function. Two general classes of these quantitative risk assessment methods have been identified.[6] One is based on variability in the effect of fixed doses (benchmark approach) and the other is based on variability in the dose producing fixed effects (tolerance-distribution approach). One difference between these approaches is that benchmark assessments are based on confidence intervals about a maximum-likelihood estimate of the mean dose-effect function, while tolerance-distribution assessments are based on probabilities generated from a distribution of effects of an agent in a sample. The latter may be preferable because risk assessments should include a factor that is based on differences in sensitivity within the population, for example, the young or elderly, rather than simply estimating the likelihood of a mean effect. Operationally, tolerance-distribution approaches fit a function to dose-effect data from individual organisms. The distribution of some small, but measurable effect (e.g., ED_{10}) over these functions is then used to estimate the likelihood of obtaining that effect at still lower doses. As single-subject experimental designs are rarely used in toxicology, the current method was developed to adapt the dose-tolerance approach for use with data obtained from more traditional between-groups experimental designs.

The basis for the current method was the randomization test,[8] which compares empirically obtained outcomes to all possible outcomes, determined by exhaustive permutation of the data set. In order to extend the randomization test method to the tolerance-distribution approach, a single data point from each dose studied was used to produce an "individual" dose-effect function. An iterative software program computed all possible permutations of these individual dose-effect functions. For example, in the current studies three doses of an agent were studied using nine animals per dose. For first exposures, this resulted in 729 ($9 \times 9 \times 9$) possible permutations of an individual dose-effect function. Each dose-effect function was then treated in a manner similar to the traditional tolerance-distribution approach; that is, an ED_{10} was calculated from each dose-effect function and the distribution of these ED_{10}s was used to estimate the likelihood of obtaining that effect at lower doses. Additional formulations of these data allowed several different comparisons to evaluate the efficacy of the approach.

Two pesticides, triadimefon and carbaryl, were used to generate data. Their effects were assessed on the motor activity of laboratory rats because each pesticide produces a different effect on motor activity (i.e., an increase or decrease) and because of the widespread use of this measure in the assessment of neurotoxic chemicals.[9–11]

2. METHODS

2.1. Procedure

Adult male Long-Evans rats (Charles River, Raleigh, NC) weighing about 300 g were used ($n = 36$ per experiment). Rats were placed in a photocell-based motor activity monitor[10] for 30-min sessions that were conducted 5 days/week. Sessions occurred for 1 to 2 weeks prior to dosing in order to establish stable baselines of activity. A modified groups-design was used: following initial training, animals were arbitrarily assigned to one of four groups ($n = 9$ per group) that received either vehicle or one of three doses of triadimefon (Experiment 1) or carbaryl (Experiment 2). Rats were given the pesticide a suitable amount of time before the session to allow absorption (see below) and then their motor activity was tested as on preceding days. On subsequent days the rats were tested without injection for 4 days to allow reassess-

ment of their baseline. Animals were then retested with one of the other doses or vehicle. This procedure was followed until all animals received all doses and vehicle. Order of dosing was counterbalanced so that each group of 9 rats received one of the three doses (or vehicle) on the first, and each subsequent, occasion.

2.2. Agents

Triadimefon (10, 56, or 100 mg/kg; Rhone-Poulenc) and carbaryl (3, 10, or 30 mg/kg; Miles, Inc.) were given (IP) in a volume of 1 ml/kg body weight. Triadimefon was given 30 min before a session and carbaryl was given 20 min before a session. The vehicle for both pesticides was 5% Emulphor and 5% ethanol in saline.

2.3. Analysis

A computer program was developed to fit a linear function (using least squares regression) to each possible permutation of an effect at each dose. *Effect* was defined as percent of control (mean vehicle effect) activity at a particular dose for each rat, with dose converted to a natural logarithm (ln).[12] For example, 9 animals per dose with three doses tested resulted in 729 ($9 \times 9 \times 9$) possible functions for the first exposure. The dose estimated to produce a 10% effect—10% decrease in activity for carbaryl or a 10% increase in activity for triadimefon—was then determined for each of these functions. The mean and standard deviation (*SD*) of these ED_{10} estimates were then adjusted by the appropriate z value (one tailed) to produce estimates of the dose that would affect selected percentiles of the sample distribution. The z values used to estimate doses expected to affect 10% ($p = 0.10, z = 1.28$), 1% ($p = 0.01, z = 2.33$), and 0.1% ($p = 0.001, z = 3.16$) of the population were obtained from standard normal probability tables. The program automatically excluded lines with a slope of zero (a rare occasion), as an ED_{10} could not be determined. Two sets of randomization analyses were performed for each pesticide. The first involved the initial exposure data, that is, the effect of the first dose tested in each of three groups of 9 animals each, compared to the vehicle control group, resulting in 729 possible lines. The second involved analysis of the total exposure data, that is, the effect of all three doses in each of the 36 animals, compared to the vehicle control group, resulting in $36 \times 36 \times 36$, or 46,656 possible lines. Doses of the pesticide estimated to affect smaller proportions of the population were truncated at three decimal places.

Two methods of constraining the randomization analyses were used to characterize the results. One method simply excluded data points that lay outside the mean of the distribution of effects for any particular dose by a factor of z or more (outliers). In order to demonstrate the effect of this constraint, a range of different z values was used. The range included those from the largest z value to produce a difference in any estimated parameter from the nonconstrained results to a fixed z of 1. Those z values within that range were studied in increments of 0.25, but results are only reported for those increments that produced a change in estimates. The loss of a fixed number of data points could have different effects. For example, the loss of two data points could result in permutations of $9 \times 9 \times 7$ or $9 \times 8 \times 8$, resulting in 567 or 576 lines, respectively. As such, both the number of data points used ($n = 25$ in this example) and number of calculated lines (# lines) are reported for each level of constraint. The second method excluded lines with slopes that were different in valence from the mean function (i.e., slope reversals). This could occur if doses were spaced too closely, resulting in an overlap in the variances of the high and low doses. A line could be lost without exclusion of any data points.

The results from the randomization program were compared to those of the "true" or empirical distribution of results determined by the method described by Dews.[4,12,13] Dose-effect data for each rat were converted to a percent of the effect of vehicle in that rat. Linear functions were then fit by regression to relate the ln of dose to locomotor-increasing effects of triadimefon or locomotor-decreasing effects of carbaryl. The dose producing a 10% change in activity was determined for each function (rat). The distribution of these point estimates and t distribution statistics were then used to estimate doses expected to produce a 10% change in activity in successively smaller proportions of the population: for example, 10% ($p = 0.10, t = 1.307$), 1% ($p = 0.01, t = 2.437$), and 0.1% ($p = 0.001, t = 3.332$).

3. RESULTS

3.1. Triadimefon

3.1.1. Overall Dose Effects

Table I shows motor activity data for the triadimefon experiment. With vehicle injections, motor activity varied from 1,697–5,621 counts per session, with a grand mean and *SD* of 3,676 and 932 counts per session, respectively. The coefficient of variation

Table I. Session Motor Activity Counts for Triadimefon

Animal #	Vehicle	10 mg/kg	56 mg/kg	100 mg/kg
Group 1				
1	**4,394**	5,282	7,351	10,215
5	**2,754**	5,059	6,239	11,823
9	**2,298**	3,724	8,029	9,411
13	**3,800**	5,823	12,929	11,711
17	**3,625**	5,351	8,957	12,441
21	**4,097**	3,596	9,280	11,011
25	**3,701**	5,879	8,573	11,948
29	**2,016**	3,484	8,666	7,510
33	**2,280**	4,239	10,353	10,138
Mean	3,218	4,715	8,931	10,690
SD	886	962	1,900	1,555
Group 2				
2	3,953	**4,539**	5,942	8,582
6	3,012	**3,951**	10,068	12,302
10	3,442	**2,668**	10,437	12,586
14	3,191	**3,339**	6,355	11,524
18	3,552	**3,142**	5,698	10,132
22	3,575	**3,495**	10,074	10,652
26	3,147	**3,436**	6,696	13,806
30	3,308	**4,510**	6,263	11,142
34	3,742	**4,154**	7,688	11,640
Mean	3,436	3,693	7,691	11,374
SD	303	637	1,959	1,510
Group 3				
3	4,379	3,727	**5,732**	8,657
7	4,660	5,895	**9,943**	15,129
11	3,838	4,106	**7,405**	10,805
15	5,621	4,347	**7,740**	10,203
19	3,584	3,530	**6,817**	10,868
23	3,660	4,014	**7,346**	11,302
27	3,398	3,701	**6,112**	8,365
31	1,697	2,242	**7,483**	11,712
35	3,742	3,517	**4,667**	9,214
Mean	3,842	3,898	7,027	10,695
SD	1,062	958	1,483	2,035
Group 4				
4	4,653	5,912	11,089	**14,157**
8	5,538	5,251	9,274	**7,659**
12	4,965	6,065	8,339	**8,878**
16	5,520	7,094	12,379	**12,449**
20	4,150	4,916	8,668	**8,799**
24	3,723	5,975	10,807	**9,815**
28	3,873	4,917	9,794	**8,049**
32	2,459	3,994	9,609	**8,995**
36	3,006	4,265	8,036	**9,134**
Mean	4,210	5,377	9,777	9,771
SD	1,067	978	1,421	2,139
Grand mean	3,676	4,421	8,357	10,632
SD	932	1,094	1,959	1,844
Coefficient var.	25.4%	24.7%	23.4%	17.3%

Note: Bold numbers represent first exposures.

for vehicle data was approximately 25%. Triadimefon increased motor activity in a dose-dependent manner. Although the absolute variability in effect (*SD*) was greatest at the two highest doses, coefficients of variation at these doses did not exceed that for vehicle. Effects of the first exposure to doses of triadimefon or vehicle are in bold print. The mean effects of the first exposures, as well as their variability, were similar to those for total exposures.

3.1.2. Randomized First Exposure

Table II shows that when the data for the first exposures were assessed using the iterative line-fitting program, and no exclusionary criteria were applied, all 729 possible lines were computed from the 27 (3 doses × 9 animals) independent data points (n). From these data, the mean dose of triadimefon to increase motor activity by 10% was 12.104 mg/kg, and doses of 8.563, 6.296, and 4.884 mg/kg were estimated to produce this effect in 10%, 1%, and 0.1% of the sample distribution, respectively. No combination of data points resulted in a line with a negative slope. When the data were assessed with increasingly conservative limits on "outliers" (i.e., using a smaller z), the mean dose to produce a 10% increase in motor activity was very similar to that found with no limitations. The effects of using various values for z on the ED_{10} and estimates of that effect in 10%, 1%, and 0.1% percent of the population are also presented in Table II. The table shows that all the data lay within ±2.25 *SD* of the mean, and the use of different values for z had little effect on risk estimates.

3.1.3. Randomized Total Exposure

Table III shows that when all the exposure data were assessed with the iterative line-fitting program with no exclusionary criteria applied, all of the 46,656 possible lines were obtained. Under these conditions, the mean dose to increase activity by 10% was 8.184 mg/kg. However, the large variability in effect (*SD* = $\sim 26 \times 10^6$ mg/kg), precluded the determination of low-dose estimates (i.e., to three significant figures). This was due to 7 lines with slope reversals, resulting in biologically implausible intercepts. Limiting risk calculations to iterations with positive slopes decreased the number of lines to 46,649. This reduced variability (*SD* = ~150 mg/kg) to the point where a dose estimated to increase activity in 10% of the sample (0.011 mg/kg) could be determined, but doses affecting smaller proportions of the population could

Table II. Parameters of the Iterative Line-Fitting Program for First Exposure to Triadimefon

z	m	ln ED_{10}	ln SD	ED_{10}	10%	1%	0.1%	No. lines	n
1.00	+	2.571	0.1759	13.074	10.35	8.363	6.965	336	21
1.50	+	2.466	0.2469	11.777	8.499	6.34	4.956	512	24
1.75	+	2.510	0.2664	12.310	8.663	6.324	4.864	576	25
2.00	+	2.485	0.2731	12.000	8.372	6.076	4.656	648	26
2.25	+	2.494	0.2631	12.104	8.563	6.296	4.884	729	27
None	All	2.494	0.2631	12.104	8.563	6.296	4.884	729	27

not be estimated. As all the data lay within ±2.5 *SD* of the mean, using higher values of z did not affect risk estimates. Decreasing z to 2.25 eliminated three outlying data points, decreasing the number of lines calculated to 42,875. This resulted in a mean dose of 8.308 mg/kg to increase activity by 10%, and doses of 4.025, 2.183, and 1.363 mg/kg were estimated to produce that effect in 10%, 1%, and 0.1% of the population, respectively. As increased restrictions (i.e., decreases in z from 2.25 to 1.0) were placed on the data, the number of excluded outliers increased and the number of lines decreased. In addition, the mean ED_{10} slightly increased and *SD*s decreased (as expected). As a consequence, risk estimates increased. The mean ED_{10}, however, remained 7.4–9.7 mg/kg regardless of exclusionary criteria. For example, the most conservative approach used to fit lines to these data excluded outliers greater than ±1 *SD* from the mean and only accepted lines with a positive slope. This resulted in the loss of 34 data points and iterated only 14,872 out of 46,656 possible lines. From these data, doses expected to produce effects in 10%, 1%, and 0.1% of the population were determined to be 6.784, 5.073, and 4.057 mg/kg, respectively.

3.1.4. Single-Subject Exposures

The empirical distribution of ED_{10}s was determined by fitting a linear regression function to the dose-effect data for each animal. The mean ln ED_{10} for a 10% increase in activity was 2.176 (9.74 mg/kg) with an *SD* of 0.457. When these figures were then used to estimate doses of triadimefon to produce a 10% increase in activity in successively smaller proportions of the sample population, doses of 4.853, 2.897, and 1.925 mg/kg were estimated to affect 10%, 1%, and 0.1% of the sample population, respectively. No animal produced a dose-effect function with a negative slope.

3.1.5. Comparison of Approaches

The ED_{10}s and risk figures (in mg/kg) resulting from each approach are listed below to allow direct comparisons. The risk figures from the first-exposure

	ED_{10}	10%	1%	0.1%
Randomized first exposure	12.1	8.56	6.30	4.88
Randomized total exposure	8.31	4.03	2.18	1.36
Single-subject exposures	9.74	4.85	2.90	1.93

data are unconstrained, and those from the total exposure utilize the least possible constraint to produce 0.1% risk figures (z = 2.25). The risk figures from the first-exposure and total-exposure data spanned those derived from the single-subject data. Since the utility of the randomization approach places a focus on the first-exposure data, a relevant comparison is between the randomized first-exposure data and the single-subject exposure data. From this summary it can be seen that the randomized first-exposure approach produced less

Table III. Parameters of the Iterative Line-Fitting Program for Total Exposure to Triadimefon

z	m	ln ED_{10}	ln SD	ED_{10}	10%	1%	0.1%	No. lines	n
1.00	+	2.259	0.2655	9.570	6.784	5.073	4.057	14,872	74
1.25	+	2.277	0.2897	9.746	6.695	4.877	3.821	23,490	86
1.50	+	2.120	0.4742	8.328	4.505	2.682	1.799	32,736	96
1.75	+	2.089	0.5537	8.077	3.941	2.151	1.349	38,148	101
2.00	+	2.122	0.5615	8.348	4.032	2.182	1.360	41,650	104
2.25	+	2.117	0.5592	8.308	4.025	2.183	1.363	42,875	105
2.50	+	1.998	5.0200	7.374	0.011	0.000	0.000	46,649	108
None	All	2.102	17.1097	8.184	0.000	0.000	0.000	46,656	108

conservative risk estimates, both because the mean ED_{10} was higher and the *SD* was lower than those obtained when all the data were formulated in a single-subject exposures. In contrast, randomizing the entire data set produced estimates that were very similar to those produced by the single-subject analysis.

3.2. Carbaryl

3.2.1. Overall Dose Effects

Table IV shows motor activity data for the carbaryl experiment, as well as the effects of the first (bold) and subsequent exposures to doses of this pesticide and vehicle. Activity following vehicle varied from 1,921–4,530 counts per session, with a grand mean and *SD* of 3,458 and 612 counts per session, respectively. This yielded a coefficient of variation of approximately 18%. The two highest doses of carbaryl decreased activity in a dose-dependent manner, resulting in a negative slope for the mean dose-effect function. Variability in activity counts was greatest at the intermediate dose of carbaryl. The decrease in variability at the highest dose compared to the intermediate dose did not appear to result from a floor effect, as that dose did not totally eliminate activity counts. The mean effects of carbaryl and their variability were similar for the first exposure and total exposures.

3.2.2. Randomized First Exposure

Table V shows that when the data for the first exposure to carbaryl were used with the iterative line-fitting program and no exclusionary criteria were applied, all 729 of the possible combinations were iterated. No slope reversals were obtained, and all of the ED_{10}s lay within 2 *SD*s of the mean. The mean ED_{10} was 2.491 mg/kg, and 1.195, 0.622, and 0.363 mg/kg were estimated to produce a comparable effect in 10%, 1%, and 0.1% of the population, respectively. When the data were analyzed by progressively excluding outliers, the mean ED_{10} increased slightly and risk estimates increased. For example, eliminating data points more than 1 *SD* from the mean resulted in the loss of five data points and 378 lines. This resulted in a mean ED_{10} of 2.791 mg/kg, and doses of 1.862, 1.290, and 0.943 mg/kg were estimated to decrease activity in 10%, 1%, and 0.1% of the population, respectively.

3.2.3. Randomized Total Exposure

Table VI shows that when all the carbaryl data were used with the iterative line-fitting program, with

Table IV. Session Motor Activity Counts for Carbaryl

Animal no.	Vehicle	3 mg/kg	10 mg/kg	30 mg/kg
Group 1				
1	**4,481**	4,101	2,251	339
5	**3,804**	2,638	471	449
9	**3,219**	3,203	496	533
13	**2,771**	3,290	481	289
17	**1,921**	2,749	727	611
21	**2,571**	2,727	1,112	504
25	**3,719**	3,710	588	311
29	**2,995**	3,791	2,118	436
33	**2,206**	957	1,337	381
Mean	3,076	3,018	1,065	428
SD	820	929	703	108
Group 2				
2	4,166	**3,182**	769	336
6	3,085	**2,136**	751	173
10	3,861	**3,662**	1,328	389
14	4,305	**3,951**	1,021	252
18	3,021	**3,134**	1,193	359
22	3,594	**2,810**	933	869
26	3,208	**3,432**	1,150	367
30	3,594	**2,882**	2,518	229
34	4,289	**3,646**	4,763	395
Mean	3,680	3,204	1,603	374
SD	507	550	1,298	201
Group 3				
3	3,372	3,915	**1,027**	474
7	3,577	3,540	**1,650**	148
11	3,255	3,333	**455**	98
15	3,828	3,809	**362**	440
19	3,472	2,336	**211**	271
23	3,763	3,242	**921**	416
27	3,378	2,888	**392**	219
31	3,712	2,890	**546**	349
35	2,770	3,190	**2,765**	330
Mean	3,459	3,238	925	305
SD	323	493	821	131
Group 4				
4	4,530	4,162	1,312	**298**
8	3,500	4,009	2,160	**90**
12	3,350	3,579	863	**206**
16	4,141	4,391	988	**634**
20	3,975	4,894	262	**739**
24	2,508	3,699	507	**477**
28	3,392	4,156	753	**236**
32	3,926	2,623	833	**880**
36	3,226	3,773	893	**434**
Mean	3,616	3,921	952	444
SD	597	629	540	265
Grand Mean	3,458	3,345	1,136	388
SD	612	730	892	187
Coefficient var.	17.7%	21.8%	78.5%	48.2%

Note: These data previously appeared in Glowa and MacPhail,[6] in a preliminary introduction to the method. Bold numbers represent first exposures.

Table V. Parameters of the Iterative Line-Fitting Program for First Exposure to Carbaryl

Q6

z	m	ln ED_{10}	ln SD	ED_{10}	10%	1%	0.1%	No. lines	n
1.00	–	1.027	0.3054	2.791	1.862	1.290	0.943	378	22
1.25	–	1.017	0.3135	2.765	1.826	1.256	0.915	441	23
1.50	–	1.076	0.3262	2.934	1.908	1.298	0.941	576	25
1.75	–	1.067	0.3358	2.906	1.867	1.259	0.907	648	26
2.00	–	0.913	0.5581	2.491	1.195	0.622	0.363	729	27
None	All	0.913	0.5581	2.491	1.195	0.622	0.363	729	27

no exclusionary criteria applied, the program calculated 46,654 lines. The maximum number of iterations did not obtain because 2 lines with a zero slope were eliminated. No slope reversals occurred and all the data lay within ±4.25 *SD*s of the mean. Under these conditions the mean ED_{10} was 2.168 mg/kg, but the high variability precluded the determination of doses (to three decimal places) estimated to decrease activity in less than 10% of the population. When the data were assessed using progressively more conservative constraints, the mean ED_{10} increased (and *SD* decreased tenfold) when two outliers exceeding ±2.75 *SD*s were eliminated. There was little additional change with the elimination of more outliers. Eliminating the two most extreme outliers also allowed estimates of doses expected to produce effects in less than 10% of the population. When data exceeding ±1 *SD* were excluded, 25 outliers were eliminated resulting in the iteration of 21,141 lines. Under these conditions the mean ED_{10} was 2.9 mg/kg and doses of 1.976, 1.430, and 1.114 mg/kg were estimated to decrease activity in 10%, 1%, and 0.1% of the population, respectively.

3.2.4. *Single Subject Exposures*

When the effects of carbaryl were assessed for individual animals, the mean ln ED_{10} was 0.403 (2.030 mg/kg) with an *SD* of 1.053. The estimates of carbaryl doses to produce this effect in 10%, 1%, and 0.1% of the sample population were 0.378, 0.115, and 0.045 mg/kg, respectively. No animal produced a dose-effect function with a positive slope.

3.2.5. *Comparison of Approaches*

The mean carbaryl ED_{10}s and risk figures (in mg/kg) using each approach are compared below. The carbaryl data are summarized in a manner similar to

	ED_{10}	10%	1%	0.1%
Randomized first exposure	2.49	1.20	0.62	0.36
Randomized total exposure	3.00	1.65	0.99	0.67
Single-subject exposures	2.03	0.38	0.12	0.05

that described for triadimefon, except the least possible constraint to produce 0.1% risk figures for total exposures was $z = 2.75$.

From these comparisons it can be seen that the ED_{10} for the first-exposure data for carbaryl was higher than that for the single-subject analysis, and the *SD* was smaller. This resulted in less conservative risk figures for the first-exposure analysis. Randomizing the total-exposure data produced the highest ED_{10} of the three methods, but it also resulted in the lowest variability. Consequently, it produced the least conserva-

Table VI. Parameters of the Iterative Line-Fitting Program for Total Exposure to Carbaryl

Q6

z	m	ln ED_{10}	ln SD	ED_{10}	10%	1%	0.1%	No. lines	n
1.00	–	1.065	0.2960	2.900	1.976	1.430	1.114	21,141	83
1.25	–	1.113	0.3385	3.043	1.962	1.355	1.019	29,667	93
1.50	–	1.098	0.3825	2.998	1.826	1.202	0.871	33,759	97
1.75	–	1.074	0.4380	2.927	1.659	1.027	0.711	38,148	101
2.00	–	1.085	0.4450	2.959	1.662	1.021	0.702	40,460	103
2.25	–	1.105	0.4546	3.019	1.675	1.019	0.695	41,650	104
2.75	–	1.099	0.4641	3.002	1.645	0.990	0.670	44,100	106
3.50	–	0.739	4.1803	2.094	0.009	0.000	0.000	45,360	107
4.25	–	0.772	4.1308	2.164	0.010	0.000	0.000	46,654	108
None	All	0.774	4.1392	2.168	0.010	0.000	0.000	46,654	108

tive risk estimates. There was about an order of magnitude difference in risk figures produced by the empirical distribution and that produced by randomizing the total sample.

4. DISCUSSION

There is currently considerable interest in quantitative approaches to risk assessment for systemic toxicity produced by environmental agents. Collectively, these approaches mathematically describe the dose-effect data and then focus on different aspects of data variability to estimate risks.[6] Effect-tolerance models focus on variability in the effect of fixed dose or doses,[3] while the dose-tolerance approach focuses on variability in the dose producing a fixed effect.[12,13] The dose-tolerance model was previously used to estimate risks of performance impairment for a number of solvents based upon laboratory research with mice.[14,15] Concentration-effect functions were obtained for individual mice trained to emit operant responses for food reinforcement, by exposing each to successively increasing concentrations of solvents. Linear functions were fit to the lns of concentrations that decreased rates of responding, and the concentration to produce a 10% decrease was determined for each mouse. Individual differences in these ED_{10}s produced a tolerance distribution of doses. Normal probability statistics were then used to estimate concentrations expected to produce a 10% performance decrement in successively smaller portions of the population (1 in 10, 1 in 100, 1 in 1,000, etc.). Although this approach to assessing risks from tolerance distributions of within-subject effects can be used with a number of exposure scenarios, agents, and effects, it may be difficult or impossible to provide point estimates of small effects for some other types of agents, exposure regimens, or effects. For example, single-subject designs are difficult (if not impossible) to use with developmental exposures, chronic low-level exposures, and persistent agent-induced effects. As a consequence, the dose-tolerance approach was modified using randomization techniques to allow point estimates based on all possible combinations of effects. This modification allows the dose-tolerance approach to be applied to more traditional between-subject toxicological designs.

The current study allowed direct comparisons between doses estimated using the single-subject, dose-tolerance approach and those derived from the randomization approach. For triadimefon the randomization approach produced risk estimates that spanned those produced by the single-subject design, although the estimates from all three data sets were very similar. For carbaryl the randomization approach produced risk estimates that were less conservative than those produced using the single-subject design, primarily because of the smaller variability in the point estimates produced by randomizing the sample. If randomization inherently lowered variability, it could represent a serious limitation to the approach. That was not the case, however, with triadimefon (see below).

These data also allowed comparisons between the first-exposure risk figures and the total-exposure risk figures, in order to evaluate the effect of increasing n on the estimates produced by the randomization approach. For many types of exposures and end points—lethality, developmental, cancer—this comparison would be impossible because subjects are assigned to only one dose group. In the present experiment, however, the acute effects of both triadimefon and carbaryl were reversible and the baselines were stable. This suggests that early exposures did not influence the effects of subsequent exposures in the same animal. For triadimefon, the mean ED_{10} was smaller and the *SD* was larger for the randomized total-exposure data compared to the first-exposure data, producing more conservative risk figures. For carbaryl, the mean ED_{10} for the total-exposure data was larger and the *SD* was smaller than that of the first-exposure data, producing less conservative risk figures. These comparisons illustrate that increasing the number of iterations from 729 to 46,656 did not necessarily decrease risk estimates. Comparing the results of randomization analysis of the first- and total-exposure data for both agents, however, carbaryl's effects appeared to be more consistent than those of triadimefon. This observation should also take into consideration that the coefficient of variation for vehicle-control data for carbaryl was lower than that for triadimefon. Nevertheless, the consistency between first- and total-exposure data suggests that using first (only) exposures to characterize agents that may have irreversible effects is an acceptable practice.

Finally, analyses of the total-exposure data and the single-subject-design data are perhaps the most appropriate comparison since they are based on identical data sets. For triadimefon, the means, *SD*s, and risk figures were remarkably similar. For carbaryl, the single-subject design produced more conservative risk figures than the total-exposure data, because the mean ED_{10} was smaller and the *SD* was larger. Despite these differences, the agreement between the

numbers from all three assessments was high. This suggests that the randomization procedure produces risk estimates that are similar to those based on the empirical (or "true") distribution of ED_{10}s.

There are several other reasons to consider the use of the randomization approach in risk assessment. One advantage of a randomization-based dose-tolerance model is that it uses all (or nearly all) the experimental data to create tolerance distributions. As a result, it tends to capture the full breadth of variability in the sample. More variability in the effects of each dose will obviously broaden the distribution. In this regard, the between-subjects dose-tolerance model will be affected by variability in dose effects in the same manner as is the within-subject dose-tolerance model.[6] There are several other advantages to the randomization approach. For one thing, it is ideal for analyzing pilot data for dose-spacing purposes. Two or more small groups can be used to determine which doses should be used for a larger study. Dose spacings that do not result in slope reversals are obviously preferred. The approach also places incentives on well-controlled experiments, since the greater the spread in the ED_{10} distribution the lower will be the estimated effective dose(s) and vice versa.[6] Additionally, as on average there was no improvement in risk figures with the repeated experimental design ($n = 36$) compared to the first-exposure data set ($n = 9$), the method appears appropriate for use with relatively small sample sizes.

The randomization approach has some limitations. For example, it tends to produce a "worst-case" scenario by combining low point estimates (large effects) in "sensitive" animals with high point estimates in relatively insensitive animals. This creates a shallow dose-effect function that may not be representative of any one organism. The opposite also occurs, that is, high point estimates at low doses are combined with low point estimates at high doses. Together, this may provide a greater range of ED_{10}s than actually exists. The impact that a "sensitive" or "insensitive" animal may have on the risk assessment process, however, is probably diluted by combining its effects with other, more "normal" animals. For the same reasons, the approach can also suffer from poor dose spacing. In the worst case, too many closely spaced doses, with high variability, would generate an unacceptably large number of slope reversals and excessively broaden the tolerance distribution. This handicap could easily be overcome by more appropriate dose spacing, as the farther apart the doses are spaced the lower is the likelihood of a "reversal." This possibility can be confirmed by analysis of effects produced by only the lowest and the highest doses. If this subset of data produces reversals, broader dose spacing would be indicated. Alternatively, rules can be developed for censuring data. For the present data, a restriction limiting the data set to individual points that did not exceed the *SD* of the dose-group mean by some factor (z) eliminated relatively few data points and greatly reduced the variance in the tolerance distribution. The additional restriction of no slope reversals in this data set had little effect, confirming well-selected doses and dose spacing. The appropriateness of these and other censoring techniques can be easily evaluated by comparing the resulting randomized tolerance distribution with that actually obtained in the animals using the present experimental design.

The increase in attention being given to assessment of the neurotoxic potential of chemicals, along with several other related systemic effects, raises the likelihood of regulating chemical exposures based on neurotoxicity. Low-dose extrapolation models have been used in estimating risks due to carcinogens while the SF/RfD approach has traditionally been used to assess the risks of noncancer effects. Both approaches have been routinely criticized as a basis for risk estimation. Dose-effect models offer many potential advantages over these approaches to risk estimation. To date, there have been few systematic approaches to data collection and analysis that have allowed direct comparisons of the resulting risk estimates based on these different methods, but see Glowa and MacPhail[6] for a contrast. The current study outlines a new approach to risk assessment that avoids many of the criticisms of older approaches. It produced estimates that were close to those that were produced when individual animal data were used to assess risks. Using the approach on a reversible end point produced reasonable risk estimates, which helps validate the method. As this approach may be applicable to end points for which repeated exposures (i.e., a single-subject design) are not possible, it has much to recommend itself. Further studies should evaluate other types of end points and agents.

REFERENCES

1. Cranmer, M. F. (1981). Extrapolation from long term low dose animal studies. In C. G. Berg & H. D. Maille (Eds.), *Measurement of risks* (pp. 415–441). New York: Plenum Press.
2. Barnes, D. G., & Dourson, M. (1988). Reference dose (RfD): Description and use in health risk assessments. *Reg. Toxicol. Pharmacol., 8*, 471–486. Q4
3. Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fund. Appl. Toxicol., 4*, 854–871. Q4

4. Dews, P. B. (1980). Estimation of low risks. *The Pharmacologist, 22*, 159.
5. Gaylor, D. W., & Slikker, W. (1990). Risk assessment for neurotoxic effects. *Neurotoxi., 11*, 211–218.
6. Glowa, J. R., & MacPhail, R. C. (1995). Quantitative approaches to risk assessment in neurotoxicology. In L. Chang & W. Slikker (Eds.), *Handbook of neurotoxicology: Vol. 3. Approaches and methods* (pp. 777–787). New York: Academic Press.
7. MacPhail, R. C., & Glowa, J. R. (1999). Quantitative risk assessment in neurotoxicology: Past, present and future. In H. A. Tilson & G. J. Harry (Eds.), *Target organ toxicology series: Neurotoxicology* (pp. 367–382). Philadelphia: Taylor & Francis.
8. Edgington, E. S. (1987). *Randomization tests.* New York: Marcel Dekker, Inc.
9. Reiter, L. W., & MacPhail, R. C. (1979). Motor activity: A survey of methods with potential use in toxicity testing. *Neurobehav. Toxicol., 1*(Suppl. 1), 53–66.
10. Crofton, K. M., Howard, J. L., Moser, V. C., Gill, M. W., Reiter, L. W., Tilson, H. A., & MacPhail, R. C. (1991). Interlaboratory comparison of motor activity experiments: Implications for neurotoxicological assessments. *Neurotoxicol. Teratol. 13*, 599–609.
11. Crofton, K. M., & MacPhail, R. C. (1996). Reliability of motor activity assessments. In K. P. Ossenkopp, M. Kavaliers, & P. R. Sanberg, (Eds.), *Measuring movement and locomotion: From invertebrates to humans* (pp. 227–252). Amsterdam: Elsevier Publishers.
12. Dews, P. B. (1986). On the assessment of risk. In N. Krasnegor, J. Gray, & T. Thompson (Eds.), *Developmental behavioral pharmacology* (pp. 53–65). Hillsdale, NJ: Lawrence Erlbaum Associates.
13. Dews, P. B. (1986). Some general problems in behavioral toxicology. In Z. Annau (Ed.), *Neurobehavioral toxicology* (pp. 424–434). Baltimore: The Johns Hopkins University Press.
14. Glowa, J. R., DeWeese, J., Natale, M. E., Holland, J. J., & Dews, P. B. (1986). Behavioral toxicology of volatile organic solvents I. Methods: Acute effects of toluene. *J. Environ. Pathol. Toxicol. Oncol., 6*, 153–168.
15. Glowa, J. R., & Dews, P. B. (1987). Behavioral toxicology of volatile organic solvents IV. Comparisons of the behavioral effects of acetone, methyl ethyl ketone, ethyl acetate, carbon disulfide, and toluene of the responding of mice. *J. Amer. Coll. Toxicol., 6*, 461–469.

Risk Analysis ms #093 Query Sheet

Author: Please print this file, and respond to the following queries. Insert additions/corrections in the exact location on the page where changes need to be made; and return the complete set of proofs to the address given on the cover sheet.

Q1: Please provide location (city and state) for for "Rhone-Poulenc" and "Miles Inc."

Q2: Please spell out "IP." Does this stand for "intraperitoneally"?

Q3: Please rephrase last part of of sentence "From this summary it can be seen . . ." Meaning is unclear.

Q4: Please provide complete (unabbreviated) journal title.

Q5: Article title OK as set?

Q6: Is "m" supposed to be the "mean"? If not please explain.