Up-and-Down Procedure:
Brief description of the method and results of a study of some statistical properties

Elizabeth H. Margosches, Ph.D., USEPA/OPPTS/OPPT
with programming assistance from Timothy Barry, Sc.D., USEPA/OP

One of the alternatives offered as a replacement for the Acute Oral Toxicity Assay (OEDC 401) is a specific form of an Up-and-Down method (OEDC 425), as specified by the ASTM in Standard E1 163-87 (note this standard has been reissued in 1997 as E1163-90). This alternative offers the opportunity to reduce the total number of animals used for the toxicity test itself, when that test is used for identifying the LD50, provided certain requirements are met. It has the prospect, however, of utilizing many more animals than the OECD 401 if, for instance, it is used to estimate a percentile considerably distant from the median or the spacing of doses is inefficient. Since each animal can only be dosed after the outcome of the previous one is known, there can be problems in identifying in advance a cadre for testing where weights and other measures are comparable so that randomization is not in question.

Background on the Method

This test calls for dosing individual animals in sequence singly at 24-hour intervals, with the initial dose set at "the toxicologist's best estimate of the LD50." Following each death (or moribund state) the dose is lowered; following each survival, it is increased, according to a prespecified dose progression factor. If a death follows an initial direction of increasing doses, or a survival follows an initial direction of decreasing dose, four additional animals are tested following the same dose adjustment pattern and then testing is ended. The OECD 425 protocol calls for a default dose progression factor of 1.3 and default sigma for maximum likelihood calculations of 0.12, i.e., log(1.3).

The method has been described over the years in the statistical literature. An Up-and-Down Procedure (sometimes called a Staircase Design) was first proposed in the 1 940's by Wilfrid Dixon and Alexander Mood; there have been papers on such issues as its use with small samples (Brownlee, K.A, J. L. Hodges, Jr., & M. Rosenblatt, 1953, J Amer Stat Assoc 48:262-277) and its use with multiple animals per dose (Hsi, B.P, 1969, J Amer Stat Assoc 64:147-162). One of the most extensive discussions appears in a draft monograph entitled Design and Analysis of Quantal DoseResponse Experiments (with Emphasis on Staircase Designs) prepared by W. Dixon and Dixon Statistical Associates for a U.S. National Institutes of Health [[NIH]] Phase I Final Report, Reduction in Vertebrate Animal Use in Research, produced under SBIR Grant No. 1-R43-RR06151-01, on April 19, 1991. This draft monograph, available from its author for a fee or from the National Center for Research Resources of the NIH to individuals under the Freedom of Information Act, will be the Dixon source quoted below.

Most of the statistical treatment has assumed that there will be some form of prior or historical information available on the tested compound. This means, for instance, that Brownlee et al. write "We have not considered the problem of estimating the scale parameter σ [sigma]. The reason for this is...primarily that with small samples no estimate for σ [sigma] can be accurate enough to have much value. Even if μ [mu] were known, and even if the trials are conducted at stimuli giving the most efficient estimation, over 200 trials would be required to estimate [sigma] within 20 per cent with confidence of 95 per cent. Our experience is that in most experimental situations, the scale parameter is sufficiently stable that the experimenter can guess its value in advance from past experience more accurately than he can estimate it from a small sample. Fortunately, our procedures require only that σ be known within rough limits, and the performance of the estimates for μ [mu] are not sensitive to errors in the guessed value of σ [sigma]."

[σ = sigma, μ = mu]

Because testing submitted to the member nations of the OECD may be the first ever done on compounds of a given family, it may be that σ will not be known even so well as Brownlee assumes. In addition to relying on the monograph of Dr. Dixon, EPA has carried some simulations out based on theoretical distributions, where the underlying μ (LD50 in base 10 logarithmic units) and σ (standard deviation in base 10 logarithmic units) are known, and the Up-and-Down Procedure is performed with the default values identified in the DECO 425 method. These simulations indicate that there can be considerable bias in the estimates when the starting value for testing is distant from the LD50 and, when the starting value is considerably above the LD50, the consequent estimate would have a high probability of overestimating the safety of the compound. That is, the estimated LD50 can be considerably greater than the true one (in the case of the computer runs, the starting LD50 for the simulations) with a potential to place a compound in a less severe hazard classification, depending on the size of the classes and the location of the LD50. As Dixon points out, based on Hsi's results, bias is influenced by the initial test level, the step size, the stopping rule, the number of trials, the number of organisms per trial and the phasing factor [the distance from the true LD50 to the nearest test level].

Simulation trials

To carry out the simulations, with 1000 trials each, the EPA assumed lognormality with 3 possible magnitudes of LD50 (1.5, 50, 1500), 3 possible log sigmas (including the one specified by the Up-and-Down protocol, 0.12; the dosing interval, 1.3; 2.5), and 3 possible starting points (LD10, LD50, LD80), along with routines to estimate only the LD50 with an assumed log sigma of 0.12 and to estimate both parameters. For the most part the two estimation procedures plot on the 45deg. line; namely, their estimated LD50 values are essentially equal.

Although some of these results are rather higher than would probably be tested in a laboratory (owing to limit tests and the ability of real live animals to absorb some doses that are very large), the general tendency seems to be counter-conservative (i.e., to say one has a larger LD50 than is the case). For log sigma the same as the assumption, while there is quite a spread of estimates, they're pretty balanced about the "true" LD50 regardless of starting value (although the spread can be pretty wide), but as log sigma increases to the dosing interval (Dixon suggests that a dose progression factor equal to sigma will improve design) and above, there is a pronounced tendency to overestimate the LD50 (i.e., underestimate hazard) with increasing starting value. These results are shown via a table with the percentiles of the UDP-estimated LD50 (Table 1). The spread of values can be seen by reading the median estimated LD50 value and observing how high the 75th and 90th percentile and how low the 25th and 10th percentile are. The underlining in the table indicates the interval which covers the "true" LD50. The simulation parameters (i.e., LD50 magnitude, log sigma) were chosen to reflect a gamut of possible compounds; six actual studies selected by the Office of Pesticide Programs show these values are not unreasonable, and there can be quite a bit of variability between tests on the same compound.

It is quite likely these results reflect the poor information going into the default design. That means, however, some form of adjustment to the starting dose and dose progression factor must be possible. That could be based on a sighting study for the compound or several related compounds together with quantitative information on structure activity relationships. Another possibility is to carry out several short sequences to estimate the standard error of the ED50. (This, by the way, is consistent with Dixon's and Brownlee et al.'s assertion, and the EPA simulations' suggestion, that single short series of trials provide limited information concerning the variance of the ED50 and thus it's not useful to get an MLE from such a single series). Performing such repeated testing will, of course, increase the number of animals used. It will not, however, be sufficient to discriminate the type of dose response -- all shapes being presumed one of a particular family of symmetric distributions. That means, all the testing methods for examining dose response or related parameters are based on a symmetric distribution, typically a normal or Gaussian one which assumes two parameters (the mean and variance or functions of

them) are needed to define its shape. There are not enough observations (and, hence, degrees of freedom) in many studies to add estimation of the shape to the list of statistical tests. That's part of why the Up-and-Down method requires a historical sigma be provided when the LD50 is estimated. A sighting study with one animal at each of several doses is equally subject to the variability of small samples, but with two or more animals per dose it can give a crude estimate of the LD50 location for starting an Up-and-Down test intended to estimate the LD50.

In particular, if the underlying shape in log dose can reasonably be assumed normal, Dixon provides a table (Dixon, Table 4.2) for use in estimating the LD50. He bases this on the following strategy:

> "A series of test levels is chosen with equal spacing between doses (usually in log units) and encompassing a starting level located at the initial estimate of the [LD50]. The spacing is equal to the initial estimate of σ.
>
> "A nominal sample size is selected. [This is done based on a desired standard error of the LD50 in σ-units, from his Table 4.1.]
>
> "A series of trials is carried out following the rule of a decrease in level following a response and an increase in level following a non-response. The initial level should be close to the [LD50].
>
> "Testing continues until the desired sample size is reached. [This nominal sample size, denoted N by Dixon, appears to correspond to the number of trails in addition to the trials in the initial run of constant sign, plus one, Brownlee et al.'s n. For OECD 425 that would appear to be 5: 4 additional animals, plus one. Dixon, however, interprets the stopping rule as described in Bruce (1985), which seems to be the same as OECD 425, to be a nominal sample size of six.]

"This strategy is based on the assumption that the response curve fits a normal model... and thus is not good for estimating small or large percentage points unless normality of the distribution throughout a wide range is assured. It is also assumed that the interval between testing levels is approximately equal to the standard deviation. This assumption will be well enough satisfied if the interval used is less than twice the standard deviation. [Note that the variety of sigmas used for sensitivity testing in Lipnick, R.L., J.A. Cotruvo, R.N. Hill, et al., 1995, Fd Chem Toxic 33:223-231 falls in a range that meets this assumption (e.g., 0.05 x 2 = 0.1 compared to 0.12, the interval of testing in log dose units), unlike the variety of sigmas considered in the EPA simulations. Thus it could be expected that Lipnick et al. would not necessarily have seen the anomalies shown in the EPA simulations.]

"...To obtain an estimate of [LD50 in log units] for the results of an up-and-down sequence, look up the configuration of responses and nonresponses in Table 4.2 and compute

$$[\text{LD50}] = X_f + kd$$

where X_f, = last dose administered; k = value from Table 4.2; d = interval between dose levels [difference in log units]." Because the EPA has not automated the look-up into this table, the EPA has not examined how this procedure compares in its simulations. It is, however, based on maximum likelihood solutions and should compare well to the solutions from the computer runs.

In his correspondence with the EPA regarding his monograph and EPA's simulations, Dr. Dixon has suggested:

> "If you are concerned that the method should be cautious toward testing at levels too high for the biology of the animal, one can use shorter steps up than down after reversal and then use a ML estimate. However, in my experience, concern is apt to arise about large doses since the investigator does not really believe the fog normal character of the biological response even when it actually is true. Another safety approach is to use smaller spacing and start at a conservative initial value. Loss of efficiency will not be great."

Additional possible uses following from method adaptations

The Dixon monograph also summarizes several modifications in the procedure that would permit estimation of other percentiles. One estimates a discrete set of percentage points p, that may be other than p = 50%. This modification, based on the logistic model (by contrast to the normal or Gaussian, for the standard method), was proposed by Wetherill et al. (Wetherill, G.B., H. Chen, & R.B. Vasudeva, 1966, Biometrika 53:439-454). From a preliminary estimate of the LDp with equally spaced dose levels centered about it, apply the usual procedure, until a nonresponse is observed. After each subsequent trial, estimate the proportion p' of positive responses (if $p > 0.5$) or zero responses (if $p < 0.5$) at the level used for the current trial, counting only those trials used since the last change of level. The dose progression rule requires specification of the minimum number of trials required for a change in response type and the relation of p' to p in deciding whether to change dosage levels.

Wetherill proposes stopping after a specified number of changes in response type. Dixon shows the Average Sample Number estimates (expected sample size) for several percentiles and two stopping rules. Estimation of the 80th percentile with as few as 2 changes of response type can take 8 animals, or as many as 32 if 8 changes of response type are required for stopping. For percentiles other than the median, Dixon believes the estimates from this Up-and-Down transformed response rule are likely to be better than extrapolating from an LD50 with an assumed standard deviation, particularly if little is known about the underlying standard

deviation or distributional form. Note that the sample size will increase rapidly as the percentile desired moves away from the 50th. It may still be worthwhile, however, to carry out such a test or some other test designed for dose response estimation as an adjunct for specific instances where a specific other percentile is needed.

Conclusions and summary

Performing toxicity testing sequentially can introduce some additional considerations in implementation. For instance, compared to OECD 401, while all animals that MIGHT start on test will be identified at the outset, their dosing regimens will not start for them at the same age. Although use of a bodyweight-adjusted concentration may roughly account for size differences, the potential effects of weight and other growth changes on response should be considered in such choices as rodent strain, starting age, litter mate usage, etc.

The Up-and-Down method has been suggested as a generally useful alternative to the OECD 401. The EPA results, however, suggest that the Up-and-Down Method may have serious problems with under or over estimation of LD50's, depending on how well the starting value and progression factor are chosen and how well the assumed sigma reflects the true variability of response across doses. Adjunct studies (e.g., sighting and structure activity relationship work) are needed to improve its performance.

Table 1

Up- and-Down Procedure
PERCENTILES of the estimated LD50
by "true" LD50, sigma, starting point
1000 simulated sets each row

'True' LD50	'True' Slope	Starting Dose	10% of results were this value or less	25% of results were this value or less	50% of results were this value or less	75% of results were this value or less	90% of results were this value or less
1.5	8.33	LD10	1.2003	1.3485	1.4596	1.6697	1.8087
		LD50	1.2408	1.3308	1.4641	1.5678	1.8134
		LD80	1.2606	1.3651	1.5217	1.6600	1.8109
	0.80	LD10	0.0515	0.0809	0.1367	0.2489	0.5074
		LD50	0.9428	1.1443	1.5678	1.9828	2.4444
		LD80	3.1598	5.1987	7.9219	12.839	16.339
	0.40	LD10	1.907E-03	2.896E-03	5.530E-03	0.0142	0.0323
		LD50	0.7773	1.1347	1.4641	2.0791	2.7127
		LD80	20.547	41.889	76.291	120.25	167.18
50	8.33	LD10	40.009	45.117	50.569	55.784	60.291
		LD50	41.359	44.943	48.805	54.822	60.446
		LD80	42.020	45.503	50.725	55.334	60.362
	0.80	LD10	1.6849	2.6954	4.5553	7.6984	14.321
		LD50	27.648	37.825	47.838	64.049	83.744
		LD80	113.13	187.90	277.87	430.90	544.64
	0.40	LD10	0.0496	0.0785	0.1716	0.3771	1.0531
		LD50	27.648	37.825	48.805	66.094	90.423
		LD80	807.03	1504.7	2543.0	4408.9	5711.1
1500	8.33	LD10	1200.3	1348.5	1488.3	1669.7	1763.1
		LD50	1206.0	1315.6	1464.1	1690.8	1813.4
		LD80	1260.6	1365.1	1521.7	1660.0	1810.9
	0.80	LD10	51.492	80.863	136.66	248.68	420.62
		LD50	942.82	1171.0	1567.8	1982.8	2505.5
		LD80	3150.3	5322.3	8336.2	1.284E+04	1.634E+04
	0.40	LD10	1.4924	2.7252	5.1489	14.380	32.309
		LD50	829.50	1141.2	1567.8	1982.8	2895.0
		LD80	2.297E+04	4.514E+04	7.629E+04	1.323E+05	1.713E+05

Each table entry represents the percentile LD50 value estimated by the single-parameter maximum likelihood method and assuming a sigma of 0.12, from an up-and-down procedure starting at the specified "start" with observations from a lognormal distribution with LD50 as shown by "True LD50" and "True Slope". Slope = 1 /sigma. Underlining is explained in the accompanying text.

Table 2
"Central" Starting Points
PERCENTILES of the estimated LD50
by "true" LD50, sigma, starting point
1000 simulated sets each row

'True' LD50	'True' Slope	Starting Dose	10% of results were this value or less	25% of results were this value or less	50% of results were this value or less	75% of results were this value or less	90% of results were this value or less
1.5	2.00	LD30	0.7193	0.8572	1.1371	1.4091	1.7868
		LD40	0.8747	1.0721	1.2776	1.6148	1.925
		LD60	1.1989	1.3934	1.7611	2.0988	2.5722
	0.80	LD30	0.2738	0.3473	0.4529	0.6755	1.0139
		LD40	0.5316	0.6703	0.8495	1.1138	1.6522
		LD60	1.3617	2.0538	2.6488	3.6510	4.6462
50	2.00	LD30	23.977	30.414	37.892	46.972	61.094
		LD40	28.256	35.735	45.154	54.194	67.547
		LD60	37.041	46.446	58.705	69.959	87.981
	0.80	LD30	9.2311	11.864	15.097	24.464	35.555
		LD40	17.718	22.409	28.315	37.263	55.079
		LD60	47.763	67.090	88.292	111.89	153.24
1500	2.00	LD30	719.32	857.22	1084.1	1409.1	1917.6
		LD40	874.73	1069.0	1277.6	1614.8	2026.4
		LD60	1182.7	1393.4	1761.1	2098.8	2654.3
	0.80	LD30	273.78	347.28	452.92	646.48	1013.9
		LD40	487.58	623.37	849.45	1109.2	1652.4
		LD60	1361.7	2018.9	2648.8	3356.6	4439.8

Each table entry represents the percentile LD50 value estimated by the single-parameter maximum likelihood method and assuming a sigma of 0.12, from an up-and-down procedure starting at the specified "start" with observations from a lognormal distribution with LD50 as shown by "True LD50" and "True Slope". Slope = 1 /sigma. Underlining identifies the range of estimated LD50 values that includes the "true" one.

Table 3
Up-and-Down Procedure
PERCENTILES of the estimated LD50
by "true" LD50, sigma, starting point
1000 simulated sets each row

'True' LD50	'True' Slope	Starting Dose	10% of results were this value or less	25% of results were this value or less	50% of results were this value or less	75% of results were this value or less	90% of results were this value or less
1.5	8.33	LD10	1.2003	1.3485	1.4596	1.6697	1.8087
		LD50	1.2408	1.3308	1.4641	1.5678	1.8134
		LD80	1.2606	1.3651	1.5217	1.6600	1.8109
	2.00	LD10	0.4756	0.6203	0.8720	1.2010	1.5980
		LD50	1.0120	1.2400	1.5678	1.8657	2.2521
		LD80	1.2930	1.6809	2.3600	2.9903	3.5530
	0.80	LD10	0.0515	0.0809	0.1367	0.2489	0.5074
		LD50	0.9428	1.1443	1.5678	1.9828	2.4444
		LD80	3.1598	5.1987	7.9219	12.839	16.339
	0.50	LD10	6.526E-03	0.0110	0.0220	0.0495	0.1091
		LD50	0.8294	1.1347	1.4641	1.9717	2.5773
		LD80	9.4059	17.131	28.951	50.192	69.184
50	8.33	LD10	40.009	45.117	50.569	55.784	60.291
		LD50	41.359	44.943	48.805	54.822	60.446
		LD80	42.020	45.503	50.725	55.334	60.362
	2.00	LD10	16.478	21.483	28.567	39.888	52.028
		LD50	33.302	40.200	48.805	62.189	75.072
		LD80	43.099	53.933	76.686	99.675	115.56
	0.80	LD10	1.6849	2.6954	4.5553	7.6984	14.321
		LD50	27.648	37.825	47.838	64.049	83.744
		LD80	113.13	187.90	277.87	430.90	544.64
	0.50	LD10	0.2290	0.3681	0.6713	1.4749	3.6227
		LD50	29.101	39.032	52.260	65.726	90.423
		LD80	298.06	561.21	965.03	1661.7	2136.6

Table 3 (continued)
Up-and-Down Procedure
PERCENTILES of the estimated LD50
by "true" LD50, sigma, starting point
1000 simulated sets each row

'True' LD50	'True' Slope	Starting Dose	10% of results were this value or less	25% of results were this value or less	50% of results were this value or less	75% of results were this value or less	90% of results were this value or less
1500	8.33	LD10	1200.3	1348.5	1488.3	1669.7	1763.1
		LD50	1206.0	1315.6	1464.1	1690.8	1813.4
		LD80	1260.6	1365.1	1521.7	1660.0	1810.9
	2.00	LD10	494.33	644.49	871.99	1200.7	1554.3
		LD50	999.05	1206.0	1500.4	1865.7	2330.0
		LD80	1376.9	1768.8	2425.6	3007.2	3553.0
	0.80	LD10	51.492	80.863	136.66	248.68	420.62
		LD50	942.82	1171.0	1567.8	1982.8	2505.5
		LD80	3150.3	5322.3	8336.2	1.284E+04	1.634E+04
	0.50	LD10	6.6846	11.045	22.516	43.969	108.68
		LD50	829.50	1134.7	1567.8	1982.8	2712.7
		LD80	9.600E+04	1.769E+04	2.961E+04	5.019E+04	6.502E+04
3000	8.33	LD10	2400.5	2697.0	3034.1	3337.2	3526.3
		LD50	2481.5	2737.9	3135.6	3337.5	3626.8
		LD80	2521.2	2730.2	3043.5	3320.0	3621.7
	2.00	LD10	906.86	1289.0	1839.5	2458.4	3274.9
		LD50	1998.1	2412.0	2928.3	3731.3	4677.3
		LD80	2585.9	3361.7	4601.1	5980.5	6933.6
	0.80	LD10	102.98	161.73	273.32	461.91	861.24
		LD50	1840.9	2282.3	2928.3	3943.4	4888.9
		LD80	6679.9	1.040E+04	1.667E+04	2.687E+04	3.268E+04
	0.50	LD10	13.012	20.497	44.033	98.936	234.24
		LD50	1746.0	2288.7	3073.5	3965.7	5425.4
		LD80	1.882E+04	3.830E + 04	5.922E + 04	1.004E + 04	1.300E + 04

Each table entry represents the percentile LD50 value estimated by the single-parameter maximum likelihood method and assuming a sigma of 0.12, from an up-and-down procedure starting at the specified "start" with observations from a lognormal distribution with LD50 as shown by "True LD59" and "True Slope". Slope = 1/sigma. Underlining identifies the range of estimated LD50 values that includes the "true" one.

Table 4
Up-and-Down Procedure
Number of Animals Used
by "true" LD50, sigma, starting point
1000 simulated sets each row

'True' LD50	'True' Slope	Starting Dose	mean no. of animals (s.d..)	median no. of animals	maximum no. of animals	% using 6 animals	% using 7 animals
1.5	2.00	LD10	8.6(1.95)	8	15	16	18
		LD50	6.6(0.82)	6	11	55	32
		LD80	7.5(1.48)	7	14	33	26
	0.50	LD10	11.3(4.21)	10	28	9	11
		LD50	6.9(1.23)	6	14	52	26
		LD80	8.7(2.72)	8	20	24	20
50	2.00	LD10	8.6(1.91)	8	15	15	19
		LD50	6.5(0.80)	6	11	61	28
		LD80	7.5(1.46)	7	14	35	24
	0.50	LD10	11.2(4.07)	10	30	8	11
		LD50	6.8(1.17)	6	13	53	25
		LD80	8.7(2.76)	8	23	24	19
1500	2.00	LD10	8.6(1.85)	9	16	14	17
		LD50	6.6(0.87)	6	11	59	28
		LD80	7.4(1.45)	7	13	36	26
	0.50	LD10	11.3(4.04)	11	28	8	11
		LD50	6.9(1.23)	7	14	50	27
		LD80	8.6(2.75)	8	20	27	19
3000	8.3	LD10	6.8(0.74)	7	9	41	41
		LD50	6.2(0.38)	6	8	85	15
		LD80	6.4(0.60)	6	8	64	31
	2.00	LD10	8.6(1.93)	8	15	16	16
		LD50	6.6(0.82)	6	10	58	28
		LD80	7.5(1.52)	7	13	33	24
	0.80	LD10	10.4(3.17)	10	22	9	12
		LD50	6.8(1.02)	6	12	53	28
		LD80	8.4(2.31)	8	18	27	18
	0.50	LD10	11.3(4.21)	11	27	10	11
		LD50	7.0(1.29)	7	15	49	28
		LD80	8.6(2.68)	8	21	25	20

Slope = 1/sigma