BIOMETRICS 55, 1236–1240
December 1999

Nonparametric Identification of the Minimum Effective Dose

Yuh-Ing Chen

Institute of Statistics, National Central University, Chung-Li, 32054, Taiwan
email: ychen@stat.ncu.edu.tw

SUMMARY. We consider identifying the minimum effective dose (MED) in a dose–response study, where the MED is defined to be the lowest dose level producing an effect over that of the zero-dose control. Proposed herein is a nonparametric procedure based on the Mann–Whitney statistic incorporated with the step-down closed testing scheme. A numerical example demonstrates the feasibility of the proposed nonparametric procedure. Finally, the comparative results of a Monte Carlo level and power study for small sample sizes are presented and discussed.

KEY WORDS: Dose–response study; Monte Carlo study; One-way layout; Step-down closed test.

1. Introduction

In toxicological and drug development studies, several increasing dose levels of a substance are usually compared with the zero-dose control to investigate the effect of the substance. For this purpose, a dose–response experiment is often conducted in a one-way layout in which the doses of the substance under consideration are administered to separate groups of subjects. There are different concerns in these studies. In toxicological studies, the major concern is the safety of the toxin under consideration. Therefore, the goal is to estimate the highest dose that shows no significant difference from the zero-dose control, which is generally called the no statistical significance of trend (NOSTASOT; Tukey, Ciminera, and Heyse, 1985) or no observed adverse event level (NOAEL; Ryan, 1992) dose. In drug development studies, however, the primary interest is identifying the lowest dose level producing a desirable effect over that of the zero-dose control, which is commonly referred as the minimum effective dose (MED; Ruberg, 1989).

The conventional approach in toxicological studies is to identify the NOSTASOT or NOAEL dose and apply appropriate safety factors to it to reach a safe dose level. Since this approach tends to overestimate the safe dose level in smaller and less sensitive experiments, an alternative approach based on estimation of the benchmark dose (Crump, 1984) from a suitable dose–response model seems to be more preferable in recent years. The alternative approach involves fitting data with a dose–response curve and estimating the dose level corresponding to a specified risk level (e.g., ED_{01}, which causes a 1% increase in risk over the zero-dose control). Based on the upper confidence limit on the risk level at the estimated ED_{01}, a safety factor is obtained to arrive at the safe dose level.

The regression-based quantitative approach is not commonly used in drug development studies since it is impractical to specify such an amount of increase in effect over the zero-dose control so that the corresponding dose level causes a desirable effect. Moreover, no extrapolation from the experimental data is involved. Therefore, this paper mainly discusses a test-based approach to identifying the MED in drug development studies. Note that, in these drug studies, increasing dose levels are frequently expected to produce stronger or at least equal treatment effects. However, it also happens often that, due to the toxic effects at high doses, an ordering in the treatment effects is anticipated that is monotonically increasing up to a point, followed by a monotonic decrease. Since this corresponds to an up-down ordering of the treatment effects, they are said to follow an umbrella pattern. The point that separates the treatment effects into the two different ordering groups is called the peak of the umbrella (Mack and Wolfe, 1981).

The problem of identifying the MED has been investigated by several authors for normally distributed responses with a common variance. For example, Williams (1971) considered a closed testing procedure based on the isotonic regression of the sample means for a monotonic dose–response relationship. Ruberg (1989) suggested tests based on different contrasts of sample means to identify the MED. Tamhane, Hochberg, and Dunnett (1996) further proposed contrast-based closed testing procedures for identifying the MED. In dose–response studies, however, it occurs frequently that the normal assumption is not tenable or the observations are too few to rely on the central limit theorem for normality. In these cases, nonparametric procedures providing practical alternatives for identifying the MED are then needed. For example, Shirley (1977) considered a nonparametric equivalent of Williams' (1971) test for contrasting increasing dose levels. Williams (1986) further suggested a modification of Shirley's (1977) test. Chen and Wolfe (1993) proposed nonparametric procedures for comparing umbrella pattern treatment effects with a control for cases both when the peak of the umbrella is known or unknown. Moreover, Chen (1993) suggested a modified Chen–Wolfe test for peak-known umbrella setting. In fact,

the modified Chen–Wolfe test is identical to Williams' (1986) closed test when treatment effects are monotonically ordered. However, all of these nonparametric procedures are based on the isotonic regressions of the average ranks under appropriate order restrictions, which need quite a lot of computational effort. Therefore, we consider in this paper employing the step-down closed testing scheme suggested by Tamhane et al. (1996), but we utilize the Mann–Whitney (Mann and Whitney, 1947) statistic for identifying the MED.

In Section 2, we propose a nonparametric closed testing procedure to identify the MED. In Section 3, the use of the proposed procedures is demonstrated with the numerical example involving Ames salmonella/microsome test data previously analyzed in Chen and Wolfe (1993). Section 4 presents the comparative results of a Monte Carlo study investigation of the relative level and power performances of several competing procedures for a variety of patterns of treatment effects configurations. The final section contains some conclusions.

2. The Proposed Testing Procedure

For the ith sample ($i = 0, 1, \ldots, k$), let $Y_{i1}, \ldots, Y_{in_i}$ be independent and identically distributed random variables, each with a continuous distribution function F_i. Suppose that the zero population ($i = 0$) is the zero-dose control and the other k populations correspond to increasing dose treatments. Furthermore, assume that the $k + 1$ samples are independent of each other. In this paper, specifically, we consider estimation of the MED, which is the smallest i so that the response in the ith population is stochastically larger than that in the control, namely, $F_i < F_0$, $i = 1, 2, \ldots, k$, when the dose–response relationship is either monotonic (ordered; $F_0 \geq F_1 \geq \cdots \geq F_k$) or nonmonotonic with a down turn (umbrella patterned; $F_0 \geq F_1 \geq \cdots \geq F_p \leq \cdots \leq F_k$ for some p, $1 < p < k$).

As noted in Tamhane et al. (1996), the family of null hypotheses $\mathrm{H} = \{\mathrm{H}_{0i}\}$, where H_{0i}: $(F_0 = F_1 = \cdots = F_{i-1} = F_i)$ for $i = 1, 2, \ldots, k$, is closed under intersection in the sense that $\mathrm{H}_{0i} \in \mathrm{H}$ and $\mathrm{H}_{0j} \in \mathrm{H}$ imply $\mathrm{H}_{0i} \cap \mathrm{H}_{0j} \in \mathrm{H}$. Hence, a level-$\alpha$ closed procedure that includes separate level-α tests of individual H_{0i} applied in a step-down manner can be employed in finding the MED. Moreover, the closed testing scheme strongly controls the familywise error rate (FWE), which is the probability that at least one true H_{0i} is rejected. Therefore, we consider using the Mann–Whitney statistic incorporated into the step-down closed testing scheme to estimate the MED.

The two-sample Mann–Whitney statistic comparing the ith dose group with the combined groups of all the lower dose levels (including the control) is

$$T_i = \sum_{j=0}^{i-1} \sum_{u=1}^{n_i} \sum_{v=1}^{n_j} I(Y_{iu} - Y_{jv}), \qquad i = 1, 2, \ldots, k,$$

where $I(a) = 1$ if $a > 0$ and 0 otherwise. Let

$$T_i^* = [T_i - \mu(T_i)]/\sqrt{\sigma^2(T_i)}, \qquad i = 1, 2, \ldots, k,$$

where $\mu(T_i) = n_i N_{i-1}/2$ and $\sigma^2(T_i) = n_i N_{i-1}(N_i + 1)/12$, with $N_i = \Sigma_{j=0}^{i} n_j$, are the null ($\mathrm{H}_{0i}$) mean and variance of T_i, respectively. Then the test based on T_i^* is appropriate for testing against the alternative hypothesis H_{1i}: $(F_0 = F_1 = \cdots = F_{i-1} > F_i)$, $i = 1, 2, \ldots, k$. Note that, if there are ties among the N_i observations, a modification of T_i^* is obtained by replacing the $N_i + 1$ in $\sigma^2(T_i)$ with $N_i + 1 - \Sigma_{j=1}^{g} t_j(t_j^2 - 1)/[N_i(N_i - 1)]$, where g is the number of tied groups and t_j is the size of tied group j. Moreover, the results in Terpstra (1952) and the projection theorem (cf., Randles and Wolfe, 1979) imply that, under the null hypothesis $\mathrm{H}_{0k}, T_1^*, T_2^*, \ldots, T_k^*$ are asymptotically independent and identically distributed (i.i.d.) standard normal.

We describe the step-down closed testing scheme proposed by Tamhane et al. (1996) together with the test statistics T_i^* as follows: To estimate the MED, we first let $k_1 = k$ and find $T_{(k_1)}^*$, where $T_{(k_1)}^*$ is the maximum of $T_1^*, T_2^*, \ldots, T_{k_1}^*$. Since the statistics $T_1^*, T_2^*, \ldots, T_{k_1}^*$ are asymptotically i.i.d. standard normal under the null hypothesis H_{0k_1}, we observe that $P\{T_{(k_1)}^* \leq z(a) \mid \mathrm{H}_{0k_1}\} = [P\{T_1^* \leq z(a) \mid \mathrm{H}_{0k_1}\}]^{k_1} \approx (1 - \alpha)^{k_1}$, where $z(a)$ is the upper ath percentile of the standard normal distribution. Let $\alpha(k_1) = 1 - (1 - \alpha)^{1/k_1}$. Define $d(k_1)$ to be the antirank of $T_{(k_1)}^*$, i.e., $T_{(k_1)}^* = T_{d(k_1)}^*$. Then, if $T_{(k_1)}^* \geq z(\alpha(k_1))$, reject H_{0j}, $j = d(k_1), \ldots, k_1$, and go to the second step with $k_2 = d(k_1) - 1$; otherwise, stop testing and accept all hypotheses. In general, at the ith step, set $k_i = d(k_{i-1}) - 1$ and $\alpha(k_i) = 1 - (1 - \alpha)^{1/k_i}$. Let $d(k_i)$ be the antirank of $T_{(k_i)}^*$, where $T_{(k_i)}^*$ is the maximum of $T_1^*, T_2^*, \ldots, T_{k_i}^*$. If $T_{(k_i)}^*$ or $T_{d(k_1)}^* \geq z(\alpha(k_i))$, then reject H_{0j}, $j = d(k_i), \ldots, k_i$; otherwise, stop testing. When testing stops at, say, the mth step, estimate the MED as $k_m + 1$ or $d(k_{m-1})$.

3. An Example

Consider the data set in Table 1 analyzed in Chen and Wolfe (1993), which contains five dose levels and a zero-dose control. There are three observations in each group. The observations are numbers of visible revertant colonies observed on plates containing Salmonella bacteria of strain TA98 and exposed to different doses of Acid Red 114. The Mann–Whitney statistics, their corresponding means and ties-adjusted variances, and the modified Mann–Whitney statistics are obtained in the following: $T_1 = 7, T_2 = 18, T_3 = 27, T_4 = 16, T_5 = 2, \mu(T_1) = 4.5, \mu(T_2) = 9, \mu(T_3) = 13.5, \mu(T_4) = 18, \mu(T_5) = 22.5, \sigma^2(T_1) = 5.10, \sigma^2(T_2) = 14.88, \sigma^2(T_3) = 29.08, \sigma^2(T_4) = 47.66, \sigma^2(T_5) = 70.96, T_1^* = 1.11, T_2^* = 2.33, T_3^* = 2.50, T_4^* = -.29, T_5^* = -2.43$. Note that the largest statistic among the five T_i^*'s is T_3^*, so $d(5) = 3$. Since, at the level $\alpha = .05$, $T_3^* = 2.50 > z(.010) = 2.326$ ($.010 \approx 1 - (.95)^{1/5}$), we go to the second step with $k_2 = 2$. We observe that $d(2) = 2$ and $T_2^* = 2.33 > z(.025) = 1.96$ ($.025 \approx 1 - (.95)^{1/2}$), but $T_1^* = 1.11 < z(.05) = 1.645$. Therefore, we estimate that, at the 5% significant level, the MED is the second dose level.

Table 1
Revertant colonies for Acid Red 114, TA98, hamster liver activation

Dose (μg/ml)					
0	100	333	1000	3333	10,000
23	27	28	41	28	16
22	23	37	37	21	19
14	21	35	43	30	13

Table 2
Estimated FWE and power for $\alpha = .05$, $k = 3$, and $n_0 = n_1 = n_2 = n_3 = 5$

			FWE				Power			
θ_{10}	θ_{20}	θ_{30}	WILM	CHEN	CW	SDT	WILM	CHEN	CW	SDT
Normal Distribution										
0	0	3	.042	.042	.014	.046	.501	.501	.412	.544
0	3	3	.046	.046	.008	.033	.421	.421	.368	.483
3	3	3	—	—	—	—	.446	.446	.365	.380
0	2	3	.037	.037	.007	.027	.255	.255	.283	.246
1	2	3	—	—	—	—	.110	.110	.041	.079
0	3	2	.041	.044	.009	.034	.323	.521	.382	.462
0	3	0	.027	.043	.016	.027	.101	.473	.406	.464
2	3	2	—	—	—	—	.251	.297	.193	.198
2	3	0	—	—	—	—	.136	.280	.206	.198
Average power							.283	.367	.295	.339
Exponential Distribution										
1	1	4	.043	.043	.014	.045	.432	.432	.359	.455
1	4	4	.039	.039	.003	.032	.328	.328	.309	.402
4	4	4	—	—	—	—	.329	.329	.295	.296
1	3	4	.038	.038	.004	.028	.239	.239	.199	.281
2	3	4	—	—	—	—	.133	.133	.059	.112
1	4	3	.037	.043	.006	.028	.279	.453	.312	.384
1	4	1	.026	.038	.011	.027	.081	.408	.346	.386
3	4	3	—	—	—	—	.230	.288	.202	.210
3	4	1	—	—	—	—	.117	.273	.245	.216
Average power							.241	.320	.258	.305

Note that the Chen–Wolfe (Chen and Wolfe, 1993) test with an estimation of the umbrella peak at high dose levels concludes that, at the 5% significant level, the third dose level is the only one more effective than the zero-dose control.

4. Monte Carlo Study

We conducted a Monte Carlo study to examine the relative level and power performances of the competing tests, including the one considered in Williams (WILM) for a monotonic dose–response relationship, the one proposed in Chen and Wolfe (CW) for unknown-peak umbrella pattern treatment effects with a down-turn at high dose levels, the modified Chen–Wolfe peak-known test suggested in Chen (CHEN), and the step-down closed procedure (SDT) proposed in this paper for identifying the MED. The study was performed for comparing $k = 3$ and 5 treatments with a control, with $n_0 = n_1 = \cdots = n_k = n = 5$ observations per sample in each case, and for a variety of dose–response relationships.

For each of these settings, appropriate normal and exponential deviates were derived by the IMSL routines RNNOR and RNEXP, respectively. The normal distributions under consideration have the same variance (five) but different means (θ_i's), and the exponential distributions have various scale parameters (θ_i's). The designated alternative configurations correspond to values of $\theta_{i0} = \theta_i - \theta_0$ for normal distributions and $\theta_{i0} = \theta_i/\theta_0$ for exponential distributions, $i = 1, 2, \ldots, k$, which include step- and linear-type ordered treatment effects and umbrella patterned treatment effects. The FWE and powers for the four tests are simulated with 10,000 replications. These estimators and the average powers, which are employed for assessing the power performances of the four tests over all the situations considered in the study, are then presented in Tables 2 and 3. Note that the configurations with true MED = 1 involve no type I errors, so the entry of estimated FWE = .000 is omitted for all procedures.

We observe from the simulation results that the FWEs of all the four procedures are not significantly higher than the nominal level .05 since they are all less than .054 ($= .05 + 2[(.05)(.95)/10{,}000]^{1/2}$). In fact, the CW test tends to be conservative in controlling its FWE.

The simulation results indicate that the WILM test has excellent power when the treatment effects have a monotonic ordering. Likewise, the CHEN test provides excellent power against umbrella pattern treatment effects when the peak is correctly chosen. This is not surprising since both tests are designed to estimate the MED for their respective special classes of alternatives. The proposed SDT test is, in general, superior to the CW test and outperforms the WILM test for umbrella treatment effects configurations, especially, with a sharp downturn. The SDT test is even better than the WILM or CHEN test for step-type ordered configurations except for the case where all the treatments are better than the control. Moreover, for identifying the MED with the umbrella patterned configurations, the power of the CHEN test is the highest one when $k = 3$, while the SDT test has the best power performance when $k = 5$. On average, although the power of the SDT test is second to the CHEN test for $k = 3$, the SDT test is the best one for $k = 5$.

Table 3
Estimated FWE and power for $\alpha = .05$, $k = 5$, and $n_0 = n_1 = \cdots = n_5 = 5$

					FWE				Power			
θ_{10}	θ_{20}	θ_{30}	θ_{40}	θ_{50}	WILM	CHEN	CW	SDT	WILM	CHEN	CW	SDT
Normal Distribution												
0	0	0	0	5	.049	.049	.011	.040	.796	.796	.710	.911
0	0	0	5	5	.046	.046	.003	.042	.787	.787	.677	.910
0	0	5	5	5	.046	.046	.000	.050	.801	.801	.681	.888
0	5	5	5	5	.050	.050	.000	.043	.806	.806	.657	.861
5	5	5	5	5	—	—	—	—	.891	.891	.634	.682
0	0	0	4	5	.054	.054	.004	.046	.669	.669	.522	.769
0	0	3	4	5	.046	.046	.001	.042	.469	.469	.258	.495
0	2	3	4	5	.041	.041	.000	.021	.241	.241	.067	.211
1	2	3	4	5	—	—	—	—	.109	.109	.006	.060
0	0	4	5	4	.049	.049	.001	.051	.662	.683	.505	.751
0	0	4	5	0	.049	.053	.005	.045	.296	.668	.517	.757
0	3	4	5	4	.047	.047	.001	.032	.460	.465	.222	.472
0	3	4	5	0	.043	.046	.002	.034	.312	.466	.250	.479
0	4	5	4	3	.049	.049	.000	.039	.648	.700	.538	.724
0	4	5	0	0	.032	.048	.001	.037	.115	.680	.510	.718
Average power									.537	.616	.457	.646
Exponential Distribution												
1	1	1	1	6	.050	.050	.012	.037	.578	.578	.476	.586
1	1	1	6	6	.050	.050	.004	.041	.487	.487	.383	.586
1	1	6	6	6	.044	.044	.001	.039	.455	.455	.333	.562
1	6	6	6	6	.045	.045	.000	.031	.434	.434	.297	.527
6	6	6	6	6	—	—	—	—	.486	.486	.279	.355
1	1	1	5	6	.043	.043	.004	.038	.439	.439	.324	.514
1	1	4	5	6	.042	.042	.001	.035	.329	.329	.203	.380
1	3	4	5	6	.035	.035	.001	.020	.202	.202	.082	.241
2	3	4	5	6	—	—	—	—	.128	.128	.012	.077
1	1	5	6	5	.047	.048	.001	.040	.397	.457	.290	.489
1	1	5	6	1	.038	.044	.005	.037	.176	.428	.329	.473
1	4	5	6	5	.386	.039	.000	.026	.314	.337	.169	.373
1	4	5	6	1	.035	.040	.001	.027	.188	.336	.216	.358
1	5	6	5	4	.041	.042	.001	.032	.369	.483	.273	.463
1	5	6	1	1	.026	.043	.005	.029	.069	.432	.324	.449
Average power									.367	.401	.266	.429

5. Conclusions

In conclusion, the use of the proposed nonparametric test to identify the MED is recommended for two reasons. First, the proposed test involves only the two-sample Mann–Whitney statistics, which are very easy to compute relative to the established nonparametric procedures. The proposed test is also very easy to implement since the necessary critical values can be found from a standard normal table. Second, the proposed test controls familywise error rate well and has an appreciable power performance compared to competing tests.

Acknowledgements

This research was supported in part by the National Science Council of Taiwan under the contract NSC87-2121-M-008-006. The author wishes to thank the editor and two referees for their careful reading of the manuscript and their suggestions, which lead to an improved presentation of the results.

Résumé

On cherche à identifier la dose efficace minimale (DEM) dans une étude dose-réponse où la DEM est définie comme la dose la plus faible produisant un effet surpassant celui du groupe contrôle soumis à dose zéro. La méthode proposée est une procédure non paramétrique basée sur la statistique de Mann-Whitney (1947) associée à la procédure de tests, fermée, pas à pas descendante, suggérée par Tamhane et al. (1996). Un exemple numérique supporte la faisabilité de cette procédure non paramétrique. Pour conclure, les résultats comparatifs avec une procédure de Monte Carlo et une analyse de la puissance statistique dans de petits échantillons sont présentés et discutés.

References

Chen, Y. I. (1993). Nonparametric comparisons of umbrella pattern treatment effects with a control in a one-way

layout. *Communications in Statistics—Simulation and Computation* **22,** 749–764.

Chen, Y. I. and Wolfe, D. A. (1993). Nonparametric procedures for comparing umbrella pattern treatment effects with a control in a one-way layout. *Biometrics* **49,** 455–465.

Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology* **4,** 854–871.

Mack, G. A. and Wolfe, D. A. (1981). K-sample rank tests for umbrella alternatives. *Journal of the American Statistical Association* **76,** 175–181.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18,** 50–60.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics.* New York: John Wiley.

Ruberg, S. J. (1989). Contrasts for identifying the minimum effective dose. *Journal of the American Statistical Association* **84,** 816–822.

Ryan, L. (1992). Quantitative risk assessment for developmental toxicity. *Biometrics* **48,** 163–174.

Shirley, E. (1977). A nonparametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. *Biometrics* **27,** 103–117.

Tamhane, A. C., Hochberg, Y., and Dunnett, C. (1996). Multiple test procedures for dose finding. *Biometrics* **52,** 21–37.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking. *Indagationes Mathematica* **14,** 327–333.

Tukey, J. W., Ciminera, J. L., and Heyse, J. F. (1985). Testing the statistical certainty of a response with increasing doses of a compound. *Biometrics* **41,** 295–301.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose level. *Biometrics* **27,** 103–117.

Williams, D. A. (1986). A note on Shirley's nonparametric test for comparing several dose levels with a zero-dose control. *Biometrics* **42,** 183–186.

Received March 1998.
Revised August 1998 *and January* 1999.
Accepted January 1999.