TREND ESTIMATION USING SUMMARIZED DOSE-RESPONSE DATA

Junshan Qiu
Office of New Animal Drug, Center for Veterinary Medicine, FDA, 7500 Standish Place, Rockville, MD 20855

Abstract
Meta-analysis of trend estimation based upon summarized data has been widely used to obtain a robust estimate of dose-response relationship. In practice, exposure level for each subject or study unit is reported as an interval rather than as a single value. This raises a well-recognized statistical problem: How can variation of exposure levels be integrated into trend estimation in a statistically sound manner? Shi and Copas (2004) recommended using observed frequency of subjects in each exposure category, reported bounds of exposure intervals and even prior knowledge from historical studies to estimate the distribution of exposure. In the light of the method by Shi and Copas (2004), we developed a new strategy for trend estimation using summarized information from literature. The new method features estimating the distributions of both exposure and probability of being a case in each exposure category.

Key Words: trend estimation, dose-response, summarized data

1. Introduction

Epidemiological studies on the association of a particular exposure and a disease outcome often group the exposure levels into several categories and report the relative risks for each exposure category compared with a single reference category. Current methods for estimation of the dose response relation based upon the reported exposure categories and relative risks are derived from the original method by Greenland and Longnecker (1992). This method advanced the weighted least square regression by adjusting the covariance among the relative risks. However, this classical method assigned a point estimate for each exposure category and ignored the imbedded variation. In addition, the number of reported exposure intervals is relatively small such that the trend estimation is based upon paucity data.

Longnecker (1988) proposed to use the historical data to obtain the population distribution of exposure levels. However, the historical data may be collected from a population which is different from the current population. Shi and Copas (2004) suggested estimating the distribution of latent doses based upon the observed frequencies (n_{ij}) of dose group (G_{ij}) with lower bound (LG_{ij}) and upper bound (UG_{ij}) for dose group j in study i. Further, the estimated exposure distribution was incorporated into estimation of relationship between reported relative risks and exposure levels. Here, instead, we developed a strategy to estimate the dose trend between directly observed outcomes and exposure levels to overcome potential limitations of the aforementioned methods and facilitate extrapolation at low exposure levels.

1.1 Aims
We used simulation studies to demonstrate three characteristics of our proposed methods in terms of sensitivity to the number of exposure intervals and the range and distribution of exposure.

1.2 Hypotheses
Assume that response variable is a binary outcome and a linear dose effect (β) function is connected to the response variable via a logit link. Further, assume that there is no effect if dose level is zero. We are interested in the null hypothesis that the empirical estimate of dose effect (β_k) is equal to the true dose effect (β) and the alternative hypothesis that the empirical estimate of dose effect (β_k) is not equal to the true dose effect (β) given simulation scenario k.

2. Methods

2.1 Estimation of Exposure Distribution
Shi and Copas (2004) assumed that the dose levels have a normal distribution, N (u, γ^2). Under this assumption, the log likelihood is

$$\sum_{i,j} n_{ij} \log P(x \in G_{ij}) = \sum_{i,j} n_{ij} \log\left\{ \Phi\left(\frac{UG_{ij} - u}{\gamma}\right) - \Phi\left(\frac{LG_{ij} - u}{\gamma}\right) \right\}$$

Where Φ is the standard normal distribution. We used the same method to estimate the distribution of exposure and explored the potential distributions: Log-Normal and Gamma.

.

2.2 Estimation of Distribution of Probability of Being a Case
Probability of being a case within each dose group varies across the dose groups. The reported number of cases and total number of subjects for each dose group can be used to estimate the distribution of probability of being a case. Assume the probability of being a case (p_j) within dose group j follows a Beta distribution, B(a_j,b_j). Then use the total number of subjects (n_j) and the number of cases (c_j) in dose group j to calculate empirical mean ($\hat{\mu}_j$) and variance ($\hat{\sigma_j}^2$) for p_j. The denominator of empirical variance was adjusted by a power parameter (r) which ranges from 0 to 1. The beta distributions were estimated via minimizing the distance (Δ) which is defined as sum of squares of the differences between the empirical means and variances ($\hat{\mu}_j$, $\hat{\sigma_j}^2$) and the true ones (μ_j,σ_j^2).

2.3 Empirical Estimation Using Simulated Data
We simulated two types of data: exposure and case/control using the estimated distributions of exposure and probability of being a case. The exposure data were simulated via random sampling from the estimated distribution ($\hat{\Phi}$) for each dose group and the sample size is proportional to n_j. The case/control data were simulated by random sampling from the estimated beta distribution ($\hat{B}$) for each dose group and the sample size is proportional to n_j.

A logistic regression was used to estimate the dose trend per each simulated data set. Empirical mean and variance of the dose trends estimated across the simulated data sets

were summarized and compared with the true mean and variance. Biases were calculated for each simulation scenario.

3. Simulations

3.1 Simulation Design

We simulated six scenarios to demonstrate some characteristics of our method preliminarily. The simulation design was summarized in Table 1.

Table 1: Summary of Simulation Design

	Exposure Parameters			**Dose-response Parameters**		
Scenario	Distributions	Ranges	# of Categories (percentiles)	Distributions	Slopes (βs)	Links
1	Gamma (2.1,1)	0.02-10.5	4 (<25%, [25%, 50%), [50%, 75%), ≥75%)	Binomial	0.35	logit
2	Gamma (2.1,1)	0.02-10.5	5 (<20%, [20%, 40%), [40%, 60%), [60%, 80%), ≥80%)	Binomial	0.35	logit
3	Gamma (2.1,1)	0.02-10.5	6 (<12.5%, [12.5%, 25%), [25%, 37.5%), [37.5%, 50%), [50%, 75%), ≥75%)	Binomial	0.35	logit
4	Log-Normal (mean=4.48, Var=34.51)	0.11-72.96	4 (<25%, [25%, 50%), [50%, 75%), ≥75%)	Binomial	0.35	logit
5	Log-Normal (mean=4.48, Var=34.51)	0.11-72.96	5 (<20%, [20%, 40%), [40%, 60%), [60%, 80%), ≥80%)	Binomial	0.35	logit
6	Log-Normal (mean=4.48, Var=34.51)	0.11-72.96	6 (<12.5%, [12.5%, 25%), [25%, 37.5%), [37.5%, 50%), [50%, 75%), ≥75%)	Binomial	0.35	logit

4. Results and Conclusions

4.1 Results

Simulation results were summarized in Tables 2. Summary statistics for the estimated dose trends (β_ks) based upon 1,000 samples include empirical mean, standard deviation, minimum, maximum, 95% lower confidence bound, 95% upper confidence bound and bias (%).

Table 2: Summary of the estimated dose trends (β_ks) based upon 1,000 samples

Scenario	Empirical Mean	Empirical Standard Deviation	Minimum	Maximum	95% Lower CI	95% Upper CI	Bias (%)
1	0.3553	0.03248	0.2690	0.4582	0.3533	0.3573	1.5%
2	0.3473	0.03414	0.2382	0.4607	0.3452	0.3495	-0.8%
3	0.3572	0.03137	0.2594	0.4527	0.3552	0.3591	2.0%
4	0.3502	0.02741	0.2612	0.4362	0.3485	0.3519	0.06%
5	0.350002	0.02797	0.2558	0.4385	0.3483	0.3517	0.0006%
6	0.3499	0.02397	0.2812	0.4316	0.3484	0.3514	-0.03%

4.2 Conclusions

As demonstrated in Table 2, we have the following conclusions:

- The proposed method for the dose effect estimation is not sensitive to the number of exposure categories. There is no obvious trend across the scenarios with different number of exposure categories.

- The biases range from -0.8% to 2% given the distribution of exposure is Gamma; the biases range from -0.03% to 0.06% given the distribution is Log-Normal.

- The proposed method yields relatively smaller biases given the distribution is Log-Normal and ranges from 0.11-72.96.

Acknowledgements

I would like to thank Dr. Laura Hungerford and Dr. Anna Nevius for their comments and suggestions.

References

Greenland S, Longnecker MP. 1992. Methods for trend estimation from summarized dose–response data, with applications to meta-analysis. American Journal of Epidemiology. 135:1301–1309.

Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. 1988. A meta-analysis of alcohol consumption in relation to risk of breast cancer. Journal of the American Medical Association. 260:652-656.

Shi JQ, Copas JB. 2004. Meta-analysis for trend estimation. Statistics in Medicine. 23(1):3–19, discussion 159–162.