Sample size estimation using negative binomial distribution for determining the risk of adverse drug reaction in the post-marketing study.

2018年6月15日 統計数理研究所 オープンハウス

伊藤 陽一 データ科学研究系 教授

INTRODUCTION:

In Good Vigilance Practice (GVP), safety manager must make a Risk Management Plan (RMP) for safety use of medical products. In RMP, identified risks are specified. An identified risk is an untoward occurrence for which there is adequate evidence of an association with the medicinal product of interest, for example, an adverse reaction observed in well-designed clinical trials or epidemiological studies for which the magnitude of the difference compared with the comparator group, on a parameter of interest suggests a causal relationship.

In Good Post-marketing Study Practice (GPSP), post-marketing safety control manager must conduct drug use-results surveys for evaluating the magnitude of the risk according to the post-marketing survey protocol with the pre-planned sample size. In almost all post-marketing survey protocols in Japan, the sample size was determined by the well known rule-of-three approach; N=3/λ for an anticipated risk λ.[1)] However, this approach accommodates for detection of unknown ADR (Adverse Drug Reaction). Thus it is not useful for evaluating the magnitude of risk of known ADR.

Negative binomial distribution is the distribution of random variable X, which is the total number of patients needed to get r ADRs. Zhu and Lakkis proposed sample the size estimation formula for negative binomial regression for accommodating over-dispersion of Poisson regression for rates[2)]. In this study, we propose sample size estimation approach using negative binomial distribution for determining the risk of ADR in the post-marketing study.

Proposed Sample Size Estimation Approach:

On the label of drugs, the risk of ADR was often classified into three categories; (a) <1%, 1% - 5%, >= 5% or (b) <0.1%, 0.1% - 1%, >=1%. It is clinically useful if the risk of ADR was clearly classified according to the above categories.

In our approach, two sample sizes (n_1, n_2) and two numbers of ADR (k_1, k_2) are estimated according to the case of categories such as (a) or (b) and a decision table is made by these figures. Table 1 shows an example of decision table for the case of categories (a).

Table 1: An example of decision table for the case of categories (a)

	$N<=n_1$	$n_1<N<n_2$	$N>=n_2$
$k<k_1$	Unknown	probably <1%	definitely <1%
$k_1<=k<k_2$	probably >=5%	probably 1%-5%	probably 1%-5%
$k>=k_2$	definitely >=5%	probably 1%-5%	definitely 1%-5%

The estimation algorithm is constructed by the following five steps.

Step 1; determine the lower number of ADR (k_1) we will observe.

Step 2; estimate the lower sample size (n_1) based on the cumulative probability of negative binomial distributions to meet the condition: $\Pr(k_1 \mid \lambda=$the upper threshold of the risk of ADR$(\lambda_2)) > 95$ %. Cumulative probability $\Pr(r \mid \lambda)$ for the sample size n is defined as the following formula.

$$\Pr(k|\lambda) = \lambda^k \sum_{j=0}^{n-k} \binom{k+j-1}{k-1} (1-\lambda)^j$$

Step 3; estimate the upper number of ADR (k_2) based on the sample size n_1 and the upper threshold of the risk of ADR(λ_2) to meet the condition: $\Pr(k_2 \mid \lambda=$the upper threshold of the risk of ADR(λ_2) , $n_1) < 5$ %.

Step 4; estimate the upper sample size (n_2) to meet the condition: $\Pr(k_1 \mid \lambda=$the lower threshold of the risk of ADR(λ_1)) > 95 %.

Step 5; estimate the cumulative probability (p) based on the lower threshold of the risk of ADR(λ_1) and the upper number of ADR (k_2) and the upper sample size (n_2): $\Pr(k_2 \mid \lambda=$the lower threshold of the risk of ADR$(\lambda_1))$.

Applied example:

For applied example, we estimate the parameters of decision tables for the two cases of categories; (a) <1%, 1% - 5%, >= 5% and (b) <0.1%, 0.1% - 1%, >=1%. Figure 1 showed the cumulative probabilities against sample sizes for the case of categories (a); λ_1=1%, λ_2=5% and k=2,3,4. Figure 2 showed the cumulative probabilities against sample sizes for the case of categories (b); λ_1=0.1%, λ_2=1% and k=1,2,3. Table 2 showed the parameters of decision tables for these cases of categories.

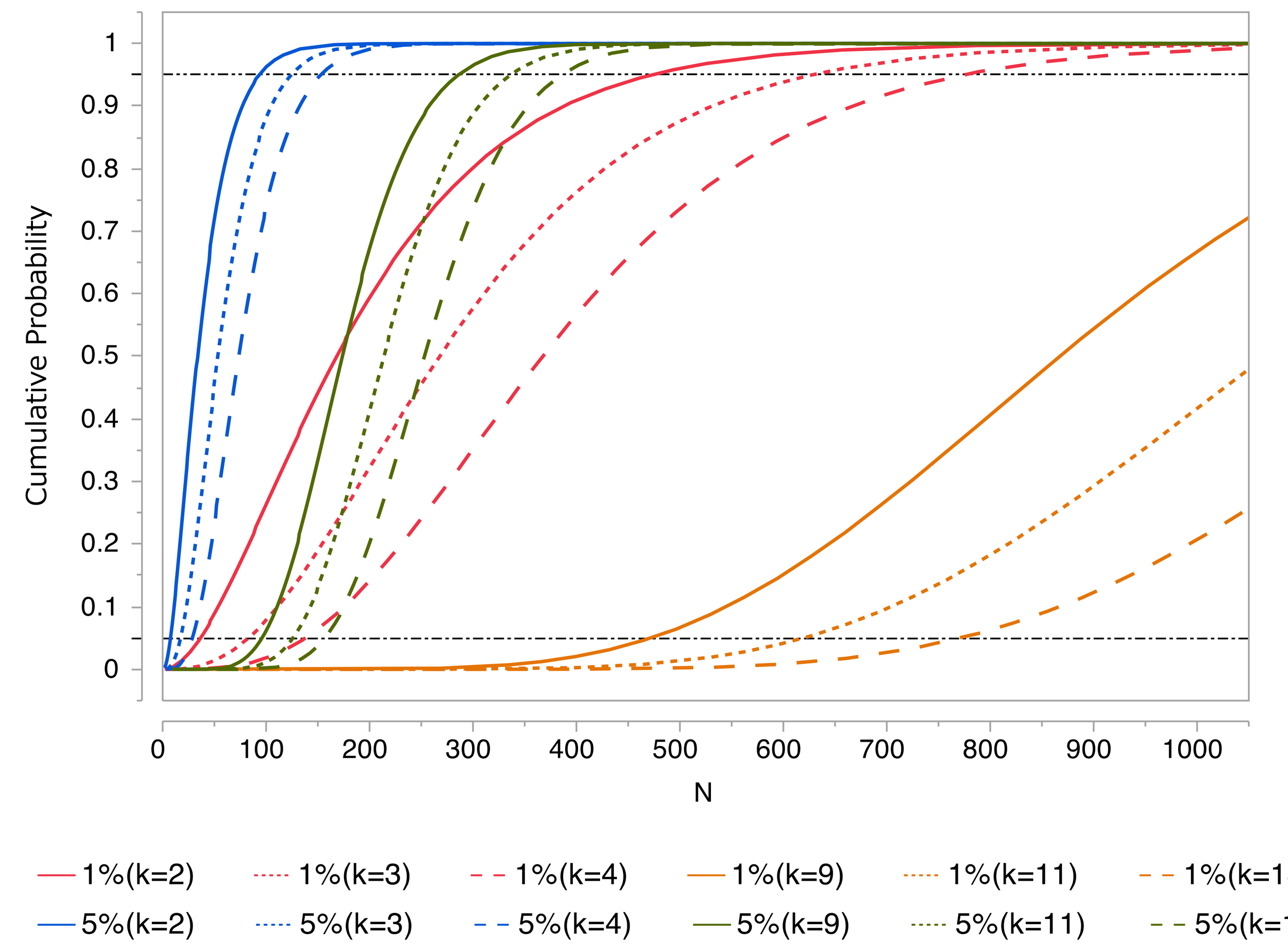

Figure 1: Cumulative probability for the case of categories (a) <1%, 1%-5%, >5% and k=2,3,4.

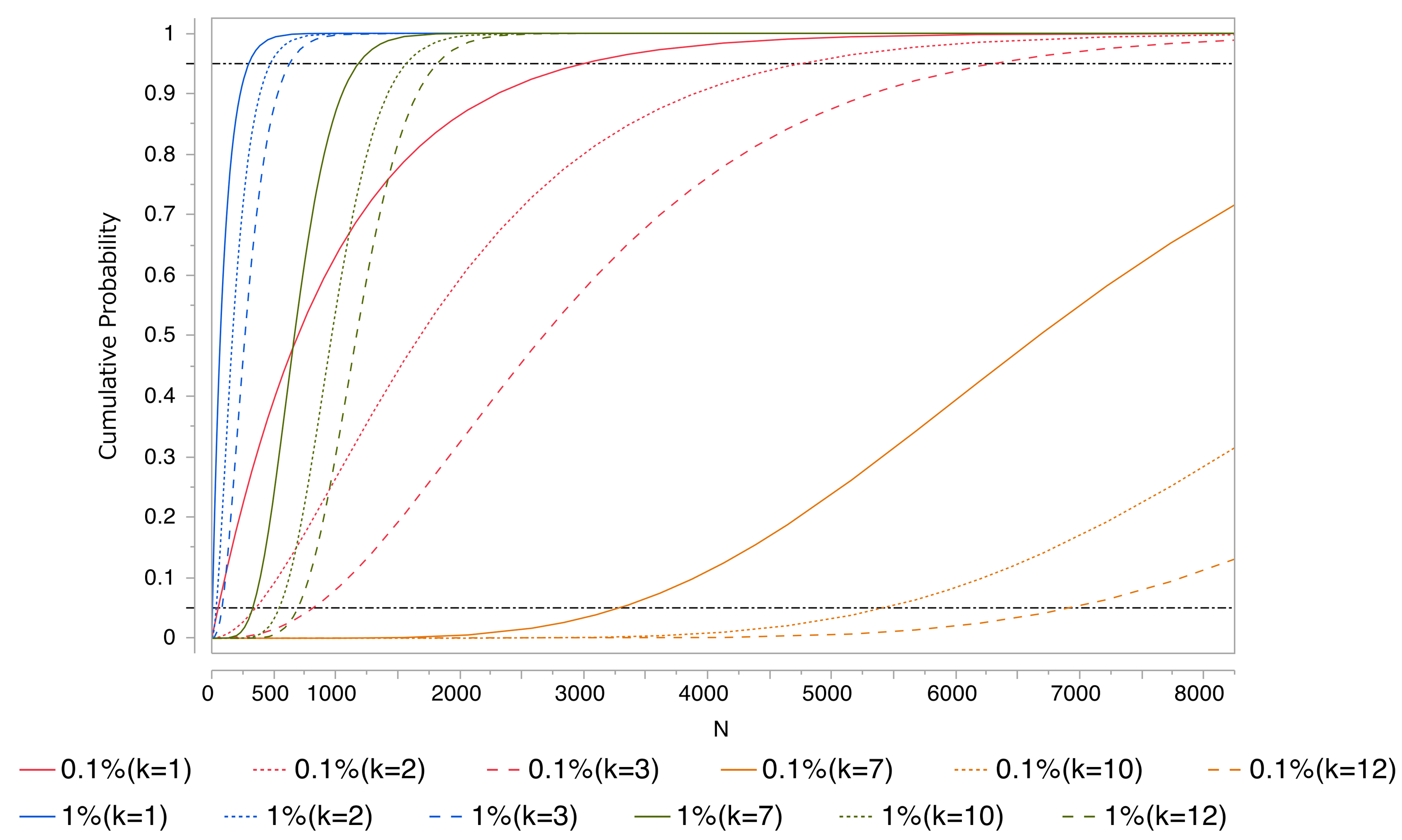

Figure 2: Cumulative probability for the case of categories (b) <0.1%, 0.1%-1%, >1% and k=1,2,3.

Table 2: Estimated parameters of decision tables

	k_1	k_2	n_1	n_2	p
(a)	2	9	93	473	0.051
	3	11	124	628	0.054
	4	13	153	773	0.051
(b)	1	7	299	2995	0.033
	2	10	473	4772	0.023
	3	12	628	6294	0.028

Discussion:

Our approach was focused on constructing the decision table for classification of the risk of ADRs based on the negative binomial distribution. On the post-marketing survey, the real-time assessment of the risk of ADRs is needed, so we considered that the negative binomial distribution was appropriate for modeling the observed number of ADRs.

The cumulative probabilities (p) in the decision tables can be interpreted with the probability that more than k2 ADRs observe when the lowest category of the risk of ADRs is true. In Table 2, all cumulative probabilities (p) were sufficiently low, so we can clearly declare the risk of ADRs was the middle category of the risk.

If we choose k=1, the cumulative probability was equivalent with usual binomial distribution. So our approach to the sample size estimation was equivalent with the rule-of-three approach[2)] and our approach was a natural extension of the rule-of-three approach to sample size estimation using more than one ADRs.

Estimated sample sizes for (b) were huge especially on n_2, so this revealed that rare events such as less than 0.1% were difficult to be classified clearly. However our approach was useful for classification of the moderate risk because sample sizes for the risks with more than 1% were feasible. In addition if the risk of ADRs was very high more than expected, the number of ADRs exceed k_2, so we can detect it as early as possible.

Referece:

(1) Eypasch, E, Lefering R, Kum C.K, Troidl H. Probability of adverse events that have not yet occurred: A statistical reminder. BMJ 1995;311(7005):619–620.

(2) Zhu H, Lakkis H. Sample size calculation for comparing two negative binomial rates. Stat Med. 2014;33(3):376-87.