

## 진단검사의 특성 추정을 위한 표본크기

박선일<sup>1</sup> · 오태호\*

강원대학교 수의과대학 및 동물의학종합연구소, \*경북대학교 수의과대학

(Accepted: December 08, 2014)

### Sample Size Requirements in Diagnostic Test Performance Studies

Son-Il Pak<sup>1</sup> and Tae-Ho Oh\*

College of Veterinary Medicine and Institute of Veterinary Science, Kangwon National University, Chuncheon 200-701, Korea

\*College of Veterinary Medicine, Kyungpook National University, Daegu 702-701, Korea

**Abstract:** There has been increasing attention on sample size requirements in peer reviewed medical literatures. Accordingly, a statistically-valid sample size determination has been described for a variety of medical situations including diagnostic test accuracy studies. If the sample is too small, the estimate is too inaccurate to be useful. On the other hand, a very large sample size would yield the estimate with more accurate than required but may be costly and inefficient. Choosing the optimal sample size depends on statistical considerations, such as the desired precision, statistical power, confidence level and prevalence of disease, and non-statistical considerations, such as resources, cost and sample availability. In a previous paper (J Vet Clin 2012; 29: 68-77) we briefly described the statistical theory behind sample size calculations and provided practical methods of calculating sample size in different situations for different research purposes. This review describes how to calculate sample sizes when assessing diagnostic test performance such as sensitivity and specificity alone. Also included in this paper are tables and formulae to help researchers for designing diagnostic test studies and calculating sample size in studies evaluating test performance. For complex studies clinicians are encouraged to consult a statistician to help in the design and analysis for an accurate determination of the sample size.

**Key words :** sample size, diagnostic test, accuracy, sensitivity, specificity.

## 서 론

의학연구에서 통계적으로 유효한 표본크기(sample size)를 산정하는 것은 연구결과의 유용성을 높이고 연구의 궁극적인 목적을 달성하는데 매우 중요한 부분이다. 표본크기가 너무 적으면 표본에서 얻은 통계량에 대한 신뢰성이 훼손되거나 실제로 존재하는 차이를 검출하지 못하게 되고, 반대로 너무 많으면 유용한 자원과 시간을 불필요하게 낭비하는 결과를 초래한다(17). 표본크기는 임상의학뿐만 아니라 유병률(prevalence) 추정(1,18), 질병검출(4,7), 질병 비발생(disease freedom) 증명(3,10), 백신검증(6) 등 다양한 분야에서 활용되고 있다. 표본크기는 임상의학에서 환자의 감염 상태를 판별하기 위해 사용되는 다양한 진단검사의 특성을 평가하거나 2개 이상의 검사를 상호 비교하는 연구는 치료 방법과 방향을 설정하는데 매우 중요하다. 만일 적정 수준 이하의 표본크기를 사용한다면 추정치의 정확도가 낮아 유용성이 저하되며, 반대로 과도한 표본크기를 사용하면 추정치의 정확도는 향상되지만 자원과 시간이 낭비되는 문제가 초래된다.

한편 이용 가능한 표본(표본크기)을 대상으로 이미 시험을 수행하였다면 연구자는 이 시험의 검정력(statistical power)이 어느 정도인지 평가하게 되는데 이 경우에도 표본크기는 중요한 평가요소가 된다.

진단검사법의 정확도(accuracy)를 반영하는 속성으로 민감도(sensitivity)와 특이도(specificity)를 검증하기 위한 표본크기를 결정할 때 추정치의 정밀도(precision) 뿐만 아니라 진단검사의 특성에 영향을 미치는 요인으로 표본 집단의 유병률을 반드시 고려해야 한다. 적정 수준 보다 적은 수의 표본을 대상으로 하는 경우 민감도나 특이도가 정확하게 추정되지 못하여 임상적으로 중요한 정보를 제공하는데 실패할 수 있다(5,8,9,12,14,15,19). 또한 새로운 진단검사법을 개발하는 과정에서 적용한 표본 집단의 유병률과 개발한 이후 검증과정에서 선정된 표본의 유병률이 상당한 차이를 보이는 경우 진단검사법의 예측도는 다르게 추정될 수 있다. 일반적으로 새로운 진단검사법은 비용, 신속성, 정확성, 실용성, 간편성 측면에서 기존의 방법에 비하여 적어도 동등한 수준 이상이라고 판단될 때 임상적으로 사용하게 된다(16). 새로 개발된 검사법(진단키트)을 상용화하기 위해서는 키트의 민감도와 특이도를 정확하게 평가하고 이러한 정보를 사용자에게 제공해주어야 한다. 진단검사의 특성을 평가하기 위해서는 추

<sup>1</sup>Corresponding author.  
E-mail : paksi@kangwon.ac.kr

정하고자 하는 지표에 대하여 임상적 혹은 과학적으로 수용할 수 있는 정밀도의 수준, 가설상의 민감도와 특이도 추정치, 연구 대상 모집단에서 질병의 유병률 등 다양한 정보를 필요로 한다(13). 국내 수의학 분야의 연구에서 표본크기와 검정력(statistical power)에 대한 중요성은 충분히 인식되지 않고 있는 실정을 감안하여 본 연구에서는 진단검사의 특성을 평가하는 연구에 필요한 표본크기를 계산하는 방법을 설명한다.

## 결론

### 통계적 고려사항

표본크기 산정에서 유병률을 고려한다는 것은 이들 추정치에 새로운 성분을 추가함으로써 정확도를 높일 수 있다. 진단검사의 특성을 추정하는데 필요한 표본크기는 첫째, 신뢰구간의 폭(즉 임상적으로 수용할 수 있는 수준의 정밀도) 둘째, 모집단에서 관심을 두고 있는 질병의 유병률 셋째, 민감도와 특이도에 대한 가설상의 값을 필요로 한다. 각 항목에 대하여 접근하는 방법은 다음과 같다.

### 정밀도

연구자는 민감도와 특이도에 대하여 임상적으로 수용할 수 있는 수준의 95% 신뢰구간으로 이를테면  $\pm 10\%$ 와 같이 결정한다. 요구되는 신뢰구간의 폭이 좁을수록(즉 정밀도가 높을수록) 표본크기는 증가하므로 적정수준을 연구자가 지정한다.

### 유병률

모집단에서의 유병률은  $N$  두 중 질병을 가지고 있는 개체수( $TP + FN$ )와 질병이 없는 개체수( $FP + TN$ )를 결정한다.  $TP + FN$ 과  $FP + TN$ 는 각각 민감도와 특이도의 분모가 되기 때문에 이들 두 군의 크기는 신뢰구간의 폭에 영향을 미치고 결과적으로 총 표본크기에도 영향을 미친다. 민감도의 경우 유병률이 높을수록 요구되는 표본크기는 감소하고 특이도의 경우 유병률이 낮을수록 요구되는 표본크기는 감소한다.

### 가설상의 추정치

신뢰구간의 폭은 민감도와 특이도에 대한 가설상의 값에 좌우된다. 표본크기를 고정시킬 때 민감도(혹은 특이도)가 50%일 때 신뢰구간의 폭이 최대가 된다. 즉 민감도나 특이도가 클수록 신뢰구간의 폭은 좁아진다. 만일 연구자가 민감도

나 특이도의 기댓값을 알지 못하는 경우에는 보수적으로 50%를 가정하여 최대의 표본크기를 결정하는 것이 바람직하다.

### 검정의 형태

많은 가설검정에서는 일반적으로 양측검정을 다루지만 경우에 따라서는 단측검정을 사용할 수 있으며, 동일한 검정력에서 단측검정을 사용할 때 표본크기가 감소한다(11,12). 단측검정은 치료효과나 진단검사의 특성을 비교할 때 어느 한 쪽 방향에만 관심을 두는 경우다. 이를테면 기존 키트 A와 새로 개발한 키트 B의 민감도(특이도)를 비교하는 연구에서 키트 B가 A와 차이가 있는지에 관심을 두기도 하지만, 키트 B의 민감도가 키트 A의 민감도와 적어도 동등한 수준이라는 것을 증명하는 것에 관심을 둘 수 있다. 전자는 전형적인 양측검정이지만 후자는 동등성(equivalence) 검정으로 단측검정에 해당한다. 따라서 연구를 계획할 때 어떠한 검정을 사용할 것인지를 분명히 설정하고 연구를 진행하는 것이 바람직하다(2).

### 이분형 결과

일반 모집단에서 질병 'x'의 유병률이 20%로 알려져 있다고 하자. 어느 연구자가 이 모집단으로부터 무작위로 선발된  $N$  두를 대상으로 질병 'x'를 진단하기 위한 진단검사법의 민감도와 특이도를 추정한다고 가정하자. 이는 질병 'x'의 유병률이 20%인 모집단에서 무작위로 선발된 표본에서도 질병 'x'의 유병률이 동일하다고 가정이 성립된다면 표본에서 얻은 연구결과를 모집단으로 일반화할 수 있다.

진단검사 결과(test result)와 환자의 질병상태(disease status)가 이분형(binary outcome)이라고 할 때(Table 1), 민감도, 특이도, 예측도(predictive value)는 비율(proportion) 자료이므로 추정치와 표준오차는 Table 2와 같이 계산된다. 지난호(17)에서 설명하였듯이 비율(민감도  $se$ , 특이도  $sp$ )의 신뢰구간을 추정할 때 최대 가능한 오차(maximum marginal

**Table 1.** Two-by-two table for determining test performance characteristics

Test result	Disease state	
	Present (D+)	Absent (D-)
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

**Table 2.** Definition of performance characteristics of clinical laboratory tests

Characteristics	Definition	Formula	standard error
Sensitivity (Se)	Probability that test will be positive, given the subject is diseased	$Se = \frac{TP}{TP + FN}$	$\sqrt{\frac{Se(1-Se)}{TP + FN}}$
Specificity (Sp)	Probability that test will be negative, given the subject is disease-free	$Sp = \frac{TN}{TN + FP}$	$\sqrt{\frac{Sp(1-Sp)}{TN + FP}}$
Positive predictive value (PPV)	Probability that the subject is diseased, given a positive test	$PPV = \frac{TP}{TP + FP}$	$\sqrt{\frac{PPV(1-PPV)}{TP + FP}}$
Negative predictive value (NPV)	Probability that the subject is disease-free, given a negative test	$NPV = \frac{TN}{TN + FN}$	$\sqrt{\frac{NPV(1-NPV)}{TN + FN}}$

TP = true positive, TN = true negative, FP = false positive, FN = false negative

error,  $d$ )와 신뢰수준 $[(1 - \alpha)\%]$ 을 고려한 표본크기는 다음과 같이 계산된다.

$$n = \frac{z_{1-\alpha/2}^2 se(1-se)}{d^2} \quad \text{또는} \quad n = \frac{z_{1-\alpha/2}^2 sp(1-sp)}{d^2}$$

상기 두 공식에서 우측 항을 유병률( $p$ )과  $1-p$ 로 나누어 주면 민감도와 특이도 추정에 필요한 표본크기( $n_{se}$ ,  $n_{sp}$ )가 된다.

$$n_{se} = \frac{z_{1-\alpha/2}^2 se(1-se)}{d^2 \times p} \quad \text{또는} \quad n_{sp} = \frac{z_{1-\alpha/2}^2 sp(1-sp)}{d^2 \times (1-p)}$$

표준오차( $SE$ )를 이용한 비율 추정치에 대한 95% 신뢰구간은 추정치  $\pm 1.96 \times SE$ 로 추정할 수 있으며 계산 결과는 95% 신뢰구간 내에 참값이 포함되는 것을 95% 확신한다고 해석

한다. 이 방법은 표본크기가 매우 크다는 가정을 전제로 하며 만일 표본크기가 적은 경우에는 이항확률(binomial probability)을 이용한 정확계산법(exact method)으로 계산해야 한다. 위의 식에서 95% 신뢰구간의 폭(width)을 나타내는  $1.96 \times SE$ 는 민감도, 특이도, 예측도 추정치의 정밀도가 되며 이 폭이 좁을수록 정밀도가 높고, 넓을수록 정밀도가 낮다는 것을 의미한다.

Table 3-1과 Table 3-2는 다양한 유병률과 민감도에서 양측검정의 95% 신뢰구간 폭을 각각  $\pm 10\%$ 와  $\pm 5\%$ 로 추정하는데 필요한 표본크기를 요약한 것이며, Table 4-1과 Table 4-2는 동일한 조건에서 특이도를 추정하는데 필요한 표본크기이다. 이 표에서 유병률의 수준에 따라 표본크기가 매우 작게 계산되는데 이 경우 정확계산법을 이용하는 것이 바람

**Table 3-1.** Sample size requirement to yield a 10% width of a 2-sided 95% confidence interval for various values of sensitivity, with differing prevalences of disease

Prevalence of disease (%)	Expected sensitivity (%)						
	50	60	70	80	90	95	99
1	9,604	9,220	8,068	6,147	3,458	1,825	381
5	1,921	1,844	1,614	1,230	692	365	77
10	961	922	807	615	346	183	39
20	481	461	404	308	173	92	20
30	321	308	269	205	116	61	13
40	241	231	202	154	87	46	10
50	193	185	162	123	70	37	8
60	161	154	135	103	58	31	7
70	138	132	116	88	50	27	6
80	121	116	101	77	44	23	5
90	107	103	90	69	39	21	5

**Table 3-2.** Sample size requirement to yield a 5% width of a 2-sided 95% confidence interval for various values of sensitivity, with differing prevalences of disease

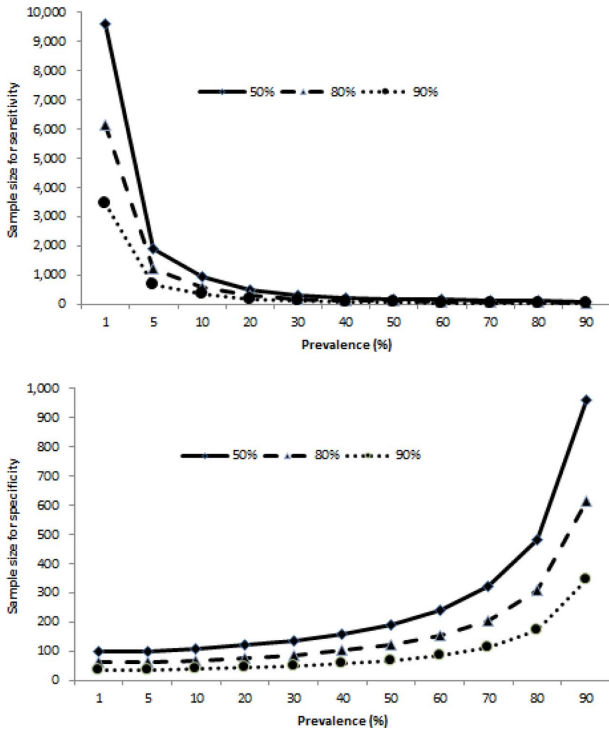
Prevalence of disease (%)	Expected sensitivity (%)						
	50	60	70	80	90	95	99
1	38,416	36,880	32,270	24,587	13,830	7,300	1,522
5	7,684	7,376	6,454	4,918	2,766	1,460	305
10	3,842	3,688	3,227	2,459	1,383	730	153
20	1,921	1,844	1,614	1,230	692	365	77
30	1,281	1,230	1,076	820	461	244	51
40	961	922	807	615	346	183	39
50	769	738	646	492	277	146	31
60	641	615	538	410	231	122	26
70	549	527	461	352	198	105	22
80	481	461	404	308	173	92	20
90	427	410	359	274	154	82	17

**Table 4-1.** Sample size requirement to yield a 10% width of a 2-sided 95% confidence interval for various values of specificity, with differing prevalences of disease

Prevalence of disease (%)	Expected specificity (%)						
	50	60	70	80	90	95	99
1	98	94	82	63	35	19	4
5	102	98	85	65	37	20	5
10	107	103	90	69	39	21	5
20	121	116	101	77	44	23	5
30	138	132	116	88	50	27	6
40	161	154	135	103	58	31	7
50	193	185	162	123	70	37	8
60	241	231	202	154	87	46	10
70	321	308	269	205	116	61	13
80	481	461	404	308	173	92	20
90	961	922	807	615	346	183	39

**Table 4-2.** Sample size requirement to yield a 5% width of a 2-sided 95% confidence interval for various values of specificity, with differing prevalences of disease

Prevalence of disease (%)	Expected specificity (%)						
	50	60	70	80	90	95	99
1	389	373	326	249	140	74	16
5	405	389	340	259	146	77	17
10	427	410	359	274	154	82	17
20	481	461	404	308	173	92	20
30	549	527	461	352	198	105	22
40	641	615	538	410	231	122	26
50	769	738	646	492	277	146	31
60	961	922	807	615	346	183	39
70	1,281	1,230	1,076	820	461	244	51
80	1,921	1,844	1,614	1,230	692	365	77
90	3,842	3,688	3,227	2,459	1,383	730	153



**Fig 1.** Sample size requirement for estimating (a) sensitivity and (b) specificity at various prevalences of disease.

직하다. 유병률과 민감도(특이도)에 따른 표본크기의 변화를 보면(Fig 1), 첫째, 민감도는 유병률이 높을수록 표본크기가 감소하고 특이도는 유병률이 낮을수록 표본크기는 감소하며 둘째, 민감도나 특이도가 50%일 때 표본크기는 최대가 되고 셋째, 정밀도가 높을수록 표본크기는 증가한다는 것을 알 수 있다.

**단일 진단검사의 민감도(특이도) 평가에 필요한 표본크기**

지정한 진단검사의 민감도( $p_0$ ) 혹은 특이도에 대하여 새로운 검사법의 민감도(특이도)를 비교하는 경우 귀무가설과 대립가설은  $H_0: Se = p_0, H_a: Se \neq p_0$  (or  $Se = p_1$ )로 정의할 수 있다. 검정력( $1 - \beta$ )을 고려할 때 단일 집단 비율 검정에 필요한 표본크기 계산 공식(17)은 다음과 같다.

$$n = \frac{[z_{1-\alpha/2}\sqrt{p_0q_0} + z_{1-\beta}\sqrt{p_1q_1}]^2}{(p_1 - p_0)^2}$$

**두 진단검사의 민감도(특이도) 비교에 필요한 표본크기**

2개의 진단검사에 대한 민감도(특이도)를 비교하는데 필요한 표본크기는 두 집단 비율( $p_1, p_2$ )의 차이 검정에 필요한 표본크기를 계산하는 공식(17)을 사용한다. 이 경우 검정해야 할 가설은  $H_0: p_1 = p_2, H_a: p_1 \neq p_2$ 가 되며, 검정력을  $1 - \beta$ , 두 집단의 할당비(allocation ratio,  $k$ )가 동일한 경우( $k = 1$ ) 표본크기( $n$ )는 다음의 공식으로 계산한다.

$$n = \frac{[z_{1-\alpha/2}\sqrt{2 \times \bar{p}\bar{q}} + z_{1-\beta}\sqrt{p_1q_1 + p_2q_2}]^2}{(p_1 - p_2)^2}$$

$$[q_1 = 1 - p_1, q_2 = 1 - p_2, \bar{p} = \frac{p_1 + kp_2}{1 + k}, \bar{q} = 1 - \bar{p}]$$

**유병률을 고려한 진단검사의 정확도 평가**

민감도를 추정하는데 필요한 표본크기( $N1$ )는 특이도를 추정하는데 필요한 표본크기( $N2$ )와 다르다. 민감도의 경우 연구자가 지정한 신뢰구간의 폭과 민감도의 기댓값을 사용하여 표본크기( $TP + FN$ )를 계산한 후 이 값과 주어진 유병률에서 민감도를 추정하는데 필요한 표본크기( $N1$ )를 계산한다. 마찬가지로 특이도의 경우 먼저 질병이 없는 개체수( $FP + TN$ )를 계산한 후 특이도 추정에 필요한 표본크기( $N2$ )를 계산한다. 총 표본크기( $N$ )는  $N1$ 과  $N2$  중 더 큰 값이 된다. 만일 예측도를 추정하는 것에 관심을 둔다면 표본크기는 양성과 음성 개체수의 기대 비율, PPV 및 NPV에 대한 기대 추정치를 사용하여 계산한다. 총 표본크기( $N$ )를 계산하는 절차를 요약하면 다음과 같다.

(1) 모수 지정

$W = 95\%$  신뢰구간의 폭

$P =$  모집단에서 유병률 추정치

$Se =$  진단검사의 민감도

$Sp =$  진단검사의 특이도

(2) 질병 양성 개체 수( $TP + FN$ ) 계산

$$TP + FN = Z_{\alpha/2}^2 \frac{Se(1 - Se)}{W^2}$$

여기에서  $Z_{\alpha/2}$ 는 유의수준  $\alpha$ 에서 표준정규분포의 값이다. 예를 들어 양측검정에서 95% 신뢰구간을 지정하는 경우  $\alpha = 0.05$ 이므로  $Z_{0.05/2} = 1.96$ 이 되며, 90% 신뢰구간을 지정하는 경우  $\alpha = 0.1$  이므로  $Z_{0.1/2} = 1.645$ 가 된다.

(3) 민감도( $N1$ ) 추정에 필요한 표본크기 계산

$$N1 = \frac{TP + FN}{P}$$

(4) 질병 음성 개체 수( $FP + TN$ ) 계산

$$FP + TN = Z_{\alpha/2}^2 \frac{Sp(1 - Sp)}{W^2}$$

(5) 특이도( $N2$ ) 추정에 필요한 표본크기 계산

$$N2 = \frac{FP + TN}{1 - P}$$

(6)  $N1$ 과  $N2$  중 더 큰 값을 총 표본크기( $N$ )로 결정

총 표본크기는 유병률이  $P$ 인 모집단에서 민감도와 특이도를  $\pm W$  이내로 추정하기 위하여 무작위로 선발해야할 개체수이다.

**연구 상황 1**

예를 들어 민감도가 95%인 진단검사를 유병률이 5%인 모집단을 대상으로 조사하여 민감도 추정치의 최대 오차가 5%를 초과하지 못하는 것을 95% 신뢰하기 위해서는 Table 3-2에 근거하면 1,460두가 필요하다.

**연구 상황 2**

새로 개발한 진단검사의 민감도 추정치 85%( $p_0$ )에 대하여

5%의 차이를 검출하는 것을 검정력 90%에서 99% 신뢰하기 위해서는 681두가 필요하다.

$$H_0: Se = 85\%, H_a: Se \neq 85\%$$

$$n = \frac{[2.58 \times \sqrt{0.85 \times 0.15} + 1.282 \times \sqrt{0.9 \times 0.1}]^2}{0.05^2} = 681$$

### 연구 상황 3

두 진단검사의 민감도가 각각 75%, 85%라고 할 때 99% 신뢰수준에서 90%의 검정력으로 10%의 차이를 검출하기 위해서는 474두가 필요하다.

$$n = \frac{[2.58 \times \sqrt{2 \times 0.8 \times 0.2} + 1.282 \times \sqrt{0.75 \times 0.25 + 0.85 \times 0.15}]^2}{(0.1)^2} = 474$$

### 연구 상황 4

소의 브루셀라병을 신속진단하기 위하여 개발한 진단키트의 민감도와 특이도를 평가한다고 가정하자. 이 키트의 정확도는 현재 시판되고 있는 제품과 적어도 동등할 것이라고 판단하여 민감도와 특이도의 기댓값(expected value)을 각각 93%와 98%로 지정한다. 정부의 강력한 방역정책으로 브루셀라병의 유병률이 현저히 감소하였지만 분석의 목적으로 1%, 5%, 10%로 가정하고, 민감도와 특이도의 95% 신뢰구간 폭이 5%를 초과하지 않기를 원한다고 가정할 때 정확도를 평가하는데 필요한 표본크기를 계산하여 보자.

$$(i) P = 1\%, Se = 93\%, Sp = 98\%, W = 0.05$$

$$TP + FN = Z_{\alpha/2}^2 \frac{Se(1-Se)}{W^2} = 1.96^2 \times \frac{0.93(1-0.93)}{0.05^2} = 100.035 \text{ 이므로}$$

$$N1 = \frac{TP + FN}{P} = 100.035 / 0.01 \approx 10,004$$

$$FP + TN = Z_{\alpha/2}^2 \frac{Sp(1-Sp)}{W^2} = 1.96^2 \times \frac{0.98(1-0.98)}{0.05^2} = 30.11 \text{ 이므로}$$

$$N2 = \frac{FP + TN}{1-P} = \frac{30.1}{1-0.01} \approx 31$$

따라서  $N1 > N2$  이므로 10,004두를 선발해야 한다.

$$(ii) P = 5\%, Se = 93\%, Sp = 98\%, W = 0.05$$

$$N1 = \frac{TP + FN}{P} = 100.035 / 0.05 \approx 2,001$$

$$N2 = \frac{FP + TN}{1-P} = \frac{30.1}{1-0.05} \approx 32$$

따라서  $N1 > N2$ 이므로 2,001두를 선발해야 한다.

$$(iii) P = 10\%, Se = 93\%, Sp = 98\%, W = 0.05$$

$$N1 = \frac{TP + FN}{P} = 100.035 / 0.1 \approx 1,001$$

$$N2 = \frac{FP + TN}{1-P} = \frac{30.1}{1-0.1} \approx 34$$

따라서  $N1 > N2$ 이므로 1,001두를 선발해야 한다.

## 감사의 글

본 연구는 2014년도 농림축산검역본부 용역연구사업 (조류 인플루엔자 유입감시 체계 구축, Project code: Z-1543068-

2014-14-01)과 강원대학교 동물의학종합연구소의 지원에 의해 이루어졌으며 이에 감사드립니다.

## 참고문헌

- Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med* 2002; 21: 835-852.
- Altman DG. *Practical statistics for medical research*. London, England: Chapman and Hall, 1993.
- Cameron AR, Baldock FC. A new probability formula for surveys to substantiate freedom from disease. *Prev Vet Med* 1998; 34: 1-17.
- Cannon RM. Sense and sensitivity: designing surveys based on an imperfect test. *Prev Vet Med* 2001; 49: 141-163.
- Carley S, Dosman S, Jones SR, Harrison M. Simple nomograms to calculate sample size in diagnostic studies. *Emerg Med J* 2005; 22: 180-181.
- Carpenter TE. Use of sample size for estimating efficacy of a vaccine against an infectious disease. *Am J Vet Res* 2001; 62: 1582-1584.
- Christensen J, Gardner IA. Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Prev Vet Med* 2000; 45: 83-106.
- Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005; 58: 859-862.
- Georgiadis MP, Johnson WO, Gardner IA. Sample size determination for estimation of the accuracy of two conditionally independent tests in the absence of a gold standard. *Prev Vet Med* 2005; 71: 1-10.
- Hardon DC, Rüfenacht J, Hauser R, Stärk KDC. Risk-based design of repeated surveys for the documentation of freedom from non-highly contagious diseases. *Prev Vet Med* 2002; 56: 179-192.
- Karl E. Peace, The alternative hypothesis: one-sided or two-sided? *J Clin Epidemiol* 1989; 42: 473-476.
- Knottnerus JA, Bouter LM. The ethics of sample size: two-sided testing and one-sided thinking. *J Clin Epidemiol* 2001; 54: 109-110.
- Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol* 2003; 56: 1118-1128.
- Li J, Fine J. On sample size for sensitivity and specificity in prospective diagnostic accuracy studies. *Stat Med* 2004; 23: 2537-2550.
- Obuchowski NA. Sample size calculations in studies of test accuracy. *Stat Methods Med Res* 1998; 7: 371-392.
- Obuchowski NA, Graham RJ, Baker ME, Powell KA. Ten criteria for effective screening: their application to multislice CT screening for pulmonary and colorectal cancers. *Am J Roentgenol* 2001; 176: 1357-1362.
- Pak SI, Oh TH. *Sample Size Calculation in Medical Research*. *J Vet Clin* 2012; 29: 68-77.
- Wagner B, Salman MD. Strategies for two-stage sampling designs for estimating herd-level prevalence. *Prev Vet Med* 2004; 66: 1-17.
- Young MJ, Bresnitz EA, Strom BL. Sample size nomograms for interpreting negative clinical studies. *Ann Intern Med* 1983; 99: 248-251.