

임상시험의 표본크기 계산

박선일¹ · 오태호*

강원대학교 수의과대학 및 동물의학종합연구소, *경북대학교 수의과대학

(게재승인: 2014년 6월 18일)

Sample Size Calculation for Cluster Randomized Trials

Son-Il Pak¹ 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 : A critical assumption of the standard sample size calculation is that the response (outcome) for an individual patient is completely independent to that for any other patient. However, this assumption no longer holds when there is a lack of statistical independence across subjects seen in cluster randomized designs. In this setting, patients within a cluster are more likely to respond in a similar manner; patient outcomes may correlate strongly within clusters. Thus, direct use of standard sample size formulae for cluster design, ignoring the clustering effect, may result in sample size that are too small, resulting in a study that is under-powered for detecting the desired level of difference between groups. This paper revisit worked examples for sample size calculation provided in a previous paper using nomogram to easy to access. Then we present the concept of cluster design illustrated with worked examples, and introduce design effect that is a factor to inflate the standard sample size estimates.

Key words : sample size, nomogram, cluster design, design effect.

서 론

지난 호(16)에서는 표본크기 선정과 관련된 통계적 이론을 임상에서 흔히 접하는 비율과 평균 비교를 예시하여 설명하였다. 표본크기를 계산하는데 사용되는 공식이 매우 다양하기 때문에 통계적 이론에 익숙하지 않은 연구자의 입장에서는 접근하기가 용이하지 않을 수 있다. 이 경우 표본크기 계산표(sample size table)나 그래프(1,12,26)를 활용할 수 있다. 그러나 연구자가 원하는 모든 모수에 대한 표를 작성하기 어렵고 정확한 계산을 위해서는 분석용 패키지를 사용하는 것이 최선이다.

표본크기는 연구계획의 종류(study design), 자료의 형태, 연구목적 등에 따라 다양한 방법으로 계산되며 표본추출방법도 고려해야 한다. 예를 들어 약제 A의 창상치유 효과에 대한 연구를 계획한다고 가정하자. 개의 피부에 실험적으로 피부 결손창을 유발한 후 특정 약제를 투여하고 2주 후 창상부위의 면적이 60%에서 90% 이상 변화할 때 임상적으로 의미가 있는 효과로 판단하고 이 실험을 4곳의 동물병원을 대상으로 각각 20마리를 할당하여 평가할 때 연구목적 달성을 위해서는 실험견 몇 마리가 필요한가? 본 예제의 특징은 실험군과 대조군 간 비율의 변화를 추정하는 연구를 4

곳의 병원에서 수행한다는 점이다. 즉 연구 대상 개체인 환자를 무작위 배정(patient-based, individual-based randomized trial)하는 것이 아니라 동물병원(집락, 군집)을 무작위로 배정하는 임상시험이다(2,4,5,20,21,24).

이러한 집락 무작위배정 임상시험(cluster randomized trial, CRT)은 임상의학에서 널리 사용되는 전통적인 방법으로 연구계획, 수행 및 분석에서 차이가 있다(11,15,17,25). 가장 중요한 것은 집락성(clustering)이 연구의 효율(efficiency)을 감소시키기 때문에 표본크기 계산에서 반드시 이를 고려해야 한다는 점이다(3,10,19). 즉 patient-based 시험에서는 표본크기 계산에서 개체 간 변동만을 고려하면 되지만 CRT에서는 집락 내 개체 간 변동과 집락 간 변동을 동시에 고려해야 한다. 즉 집락 내 개체 간 의존성의 정도를 나타내는 지표로 집락 내 상관계수를 적용하여 표본크기를 증가시켜 주어야 한다.

본 연구에서는 지난 호(16)에서 설명한 예제에 대하여 표본크기 계산을 간편하게 수행할 수 있도록 그래프를 활용하는 방법과 CRT 연구에 필요한 표본크기를 계산하는 방법을 제시한다.

결 론

Nomogram 활용

표본크기를 계산할 때 위와 같이 계산하는 과정이 번거롭

¹Corresponding author.
E-mail : paksi@kangwon.ac.kr

기 때문에 분석용 소프트웨어[PASS (NC, USA), nQuery Advisor (MA, USA), Power and Precision (NJ, USA), Epi-Info (CDC, USA), STATA (TX, USA), SAS (NC, USA), SPSS (NY, USA)]나 이미 계산된 표(1,12,25)를 활용하는 방법이 있다. 본 절에서는 후자에 대하여 간단히 설명한다.

비율 비교

표준화 차이(d)와 실험군 당 표본크기(n)는 다음과 같이 계산한다. 표준편차는 두 비율(p_1, p_2)의 평균($\bar{p} = (p_1 + p_2)/2$)을 이용한 표준편차($\sqrt{\bar{p}(1-\bar{p})}$)를 사용한다.

$$d = \frac{p_1 - p_2}{\sqrt{\bar{p}(1-\bar{p})}}, n = \frac{p_1(1-p_1) + p_2(1-p_2)}{(p_1 - p_2)^2} \times C_{p, power}$$

$C_{p, power}$ 는 주어진 유의수준과 검정력에 대한 상수다. 예를 들어 유의수준 5%일 때 $Z_{0.05} = 1.96$, 검정력 80%일 때 $Z_{0.2} = 0.84$ 이므로 $(1.96 + 0.84)^2 = 7.9$ 가 된다. 흔히 사용되는 유의수준과 검정력에 대한 $C_{p, power}$ 를 요약하면 Table 1과 같다.

예를 들어 지난 호(16)에서 설명한 예제5에 대하여 계산하여 보자. 개의 질병 'x'에 대한 새로운 치료제의 효과를 폐사율로 측정하여 비교하는 연구에서 기존 약제는 40%, 신약은 31%의 폐사율을 보인다고 할 때(즉 9%의 폐사율 차이를 유의하다고 판단할 때) 유의수준 5%에서 검정력 90%를 달성하는 표본크기를 계산하는 것이다. 이 경우 $\bar{p} = (0.4 + 0.31)/2 = 0.355$, 표준편차 0.479이므로 표준화 차이는 0.188이 된다. 즉 $(0.4 - 0.31)/\sqrt{0.355(1-0.355)} = 0.188$. 따라서 nom-

Table 1. Commonly used values for $C_{p, power}$

p	Power (%)						
	50	60	70	80	90	95	99
0.05	3.8	4.9	6.2	7.9	10.5	13.0	18.4
0.01	6.6	8.0	9.6	11.7	14.9	17.8	24.0

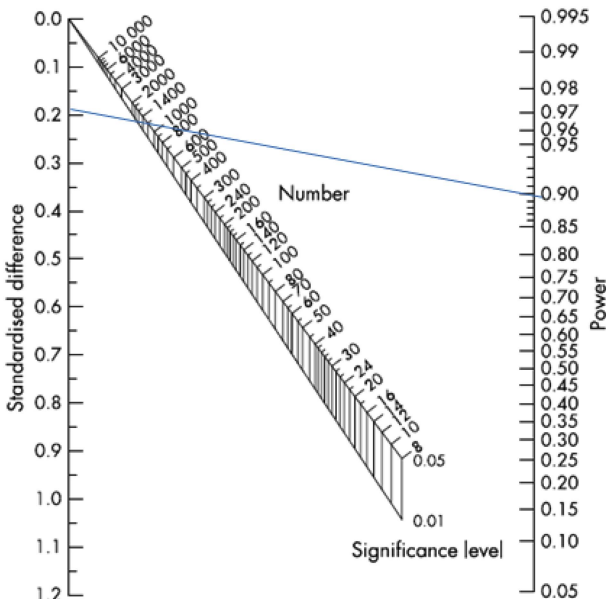


Fig 1. Nomogram for the calculation of sample size for comparison of proportions.

ogram으로 추정하면 유의수준 5%, 검정력 90%일 때 $C_{0.05, 90\%} = 10.5$ 이므로 약 1,200두가 되며 공식을 이용하여 계산된 1,180두와 유사한 결과를 보인다(Fig 1). 이 과정을 직접 계산하면 각 군당 약 590두로 총 1,180두가 되어 동일한 결과를 얻는다.

$$n = \frac{0.4_1(1-0.4) + 0.31(1-0.31)}{(0.4 - 0.31)^2} \times 10.5 \approx 590$$

평균 비교

Nomogram을 활용하기 위해서는 연구자가 검출하고자 하는 임상적으로 중요하다고 판단되는 유효크기와 측정변수의 표준편차를 이용하여 표준화 차이(standardized difference, d)를 계산해야 한다. 평균 비교에 대한 표준화 차이와 실험군 당 표본크기(n)는 다음과 같이 계산한다.

$$d = \frac{\delta}{\sigma} = \frac{\mu_1 - \mu_2}{\sigma}, n = \frac{2}{d^2} \times C_{p, power}$$

여기에서 δ = 두 집단 간 평균 차이 ($\mu_1 - \mu_2$) 즉 유효크기, σ = 모집단에서 측정변수의 표준편차이다. 예를 들어 지난 호(16)의 예제6에서 모집단의 표준편차를 16.8로 가정하면 표준화 차이는 0.32가 된다. 즉 $(132.86 - 127.44)/16.8 = 0.32$. 따라서 nomogram으로 추정하면 유의수준 5%, 검정력 80%일 때 $C_{0.05, 80\%} = 7.9$ 이므로 약 300두가 되며 공식을 이용하여 계산된 304두와 유사하며(Fig 2), 또한 표준화 차이를 이용하여 표본크기를 계산하면 각 군당 155두, 총 310두가 되어 유사한 결과를 얻는다.

$$n = \frac{2}{0.32^2} \times 7.9 \approx 155$$

모집단에서 측정변수의 표준편차에 대한 정보를 알기 어렵지만 기존의 연구결과, 전문가의견, 예비실험 등을 통하여 대략적인 추정치를 알 수 있다. 이 값을 지나치게 보수적으

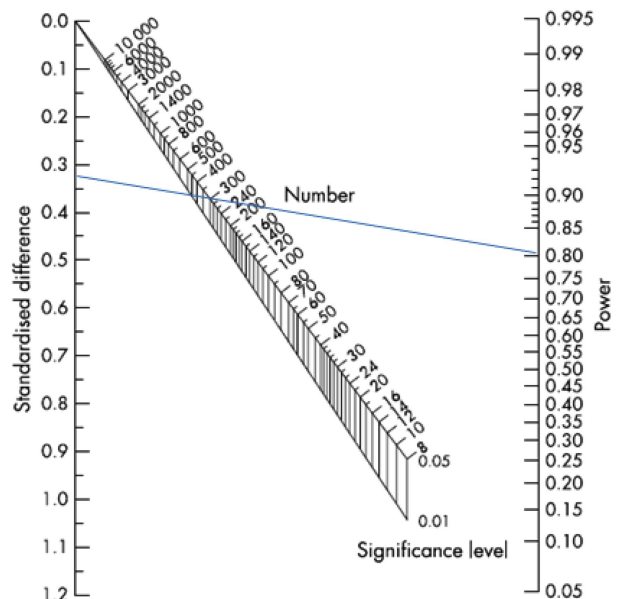


Fig 2. Nomogram for the calculation of sample size for comparison of means.

Table 2. Sample size required in each experimental group for comparing two group means at the two-sided 5% significance level for different standardized difference (d) and power

d	Power (%)					
	0.99	0.95	0.90	0.80	0.70	0.60
0.10	3,675	2,600	2,102	1,571	1,235	981
0.20	920	651	526	393	310	246
0.30	409	290	234	175	138	110
0.40	231	163	132	99	78	62
0.50	148	105	85	64	50	40
0.60	103	73	59	45	35	28
0.70	76	54	44	33	26	21
0.80	58	42	34	26	20	16
0.90	46	33	27	20	16	13
1.00	38	27	22	17	13	11
1.10	31	22	18	14	11	9
1.20	27	19	16	12	10	8
1.30	23	16	13	10	8	7
1.40	20	14	12	9	7	6
1.50	17	13	10	8	6	5
1.60	15	11	9	7	6	5
1.70	14	10	8	6	5	4
1.80	12	9	7	6	5	4
1.90	11	8	7	5	4	4
2.00	10	7	6	5	4	3

d = (effect size)/(standard deviation)

로 설정하면 표준화차이가 증가하여 표본크기가 감소하게 되어 연구의 검정력이 저하될 수 있음을 주의해야 한다. Table 2와 3은 두 집단 평균 비교를 위하여 표준화 차이를 이용할 때 지정한 유의수준과 검정력에 따른 실험군 당 표본크기를 요약한 것으로 표준화 차이를 *d*라 할 때 공식 $n = 1 + 2C_{p,power}(1/d)^2$ 을 사용하여 계산한 결과다(23).

Cluster design 표본크기

독립성 가정

환자(patient-based) 무작위배정 임상시험에 필요한 표본크기를 계산할 때 모집단의 구성원 즉 표본추출 단위(sampling unit)가 상호 독립적이라는 가정을 전제로 한다. 독립성(independence)이란 이를테면 환자 A의 치료결과는 환자 B의 치료결과와 전혀 관련이 없다는 것을 의미한다. 연구 계획이 집단 무작위배정 임상시험(cluster randomized trial, CRT)인 경우에는 특정 집단 내(within cluster)의 환자들이 서로 유사한 반응을 보이기 때문에 독립성 가정을 위반한다. 예를 들어 특정 병원에 내원하는 환자들은 해당 병원의 환자관리나 기술수준 등이 동일하기 때문에 동일한 질병에 대하여 서로 유사한 치료결과를 보이며 이러한 결과는 병원 간에는 큰 차이를 보이는 경우다. 치료결과의 유사성은 ICC로 평가하고, ICC가 클수록 집단(병원) 내 변동에 비하여 병원 간 변동이 더 크다는 것을 의미하므로(16) 단순무작위추출에 비하여 CRT 연구에서는 표본크기를 보정해주어야 한다.

Table 3. Sample size required in each experimental group for comparing two group means at the two-sided 1% significance level for different standardized difference (d) and power

d	Power (%)					
	0.99	0.95	0.90	0.80	0.70	0.60
0.10	4,807	3,564	2,977	2,337	1,923	1,602
0.20	1,203	892	745	585	482	401
0.30	535	397	332	261	215	179
0.40	301	224	187	147	121	101
0.50	193	144	120	94	78	65
0.60	135	100	84	66	54	45
0.70	99	74	62	49	40	34
0.80	76	57	47	37	31	26
0.90	60	45	38	30	25	21
1.00	49	37	31	24	20	17
1.10	41	30	26	20	17	14
1.20	34	26	22	17	14	12
1.30	29	22	19	15	12	10
1.40	26	19	16	13	11	9
1.50	22	17	14	11	10	8
1.60	20	15	13	10	9	7
1.70	18	13	11	9	8	7
1.80	16	12	10	8	7	6
1.90	14	11	9	7	6	5
2.00	13	10	8	7	6	5

d = (effect size)/(standard deviation)

Design effect

전술한 상호 독립성 가정은 이를테면 집락추출에서와 같이 측정결과(유병률)가 집락의 수준과 연관되어 있다면 분산 추정치가 과소평가되기 때문에 표본크기를 계산할 때 분산을 조정 즉 팽창시킬 필요가 있다. 즉 집락 내 개체 간 독립성이 충족되지 못할 때 분산을 팽창시킬 목적으로 집락 내 상관계수를 사용하여 보정한다는 의미에서 design effect(*D*)를 분산 팽창보정계수(variance inflation correction factor)라고도 한다(13,14). 수학적으로 design effect는 집락 당 표본크기와 집락 내 상관성을 보정하여 계산하며 집락추출에서 발생하는 오차의 왜곡 정도를 나타낸다(6,7,8).

$$D = 1 + \rho(k - 1)$$

여기에서

ρ = 집락 내 상관계수(ICC)

k = 집락크기 (집락 당 선발두수)

$D > 1$ 은 단순무작위추출과 비교할 때 사용한 표본추출계획(집락추출)의 정밀도가 감소하고(표본오차 증가), $D < 1$ 은 사용한 표본추출계획(층화추출)의 정밀도가 증가(표본오차 감소)함을 의미한다. 예를 들어 $D = 3$ 은 동일한 표본크기를 단순무작위추출로 선발할 때와 비교하여 집락추출의 분산이 3배 더 크다고 할 수 있다. 이를 달리 해석하면 $D = 3$ 인 집락추출법에 비하여 단순무작위 추출을 이용하면 표본크기가 1/

3로 감소함을 의미한다. 이 식에서 보듯이 design effect는 ρ 와 k 로 결정되며, 집락 내 상관계수가 매우 작다고 하더라도 집락크기가 크면 design effect에 상당히 큰 영향을 미친다. 집락 내 상관계수(ρ)는 design effect에 대한 정보가 있다면 전술한 공식 $\rho = (D - 1)/(k - 1)$ 을 이용하여 직접 계산하고, 관련 정보가 없는 경우에는 분산분석으로 추정한다(8,9,18,22).

집락추출에서 집락수 C , 집락 당 선발두수(집락크기, cluster size)를 k 라 할 때 총 표본크기 (n)는 $n = Ck$ 가 되며 집락성을 고려한 총 표본크기는 다음과 같이 계산된다.

$$ESS = \frac{Ck}{D} \quad [\text{단, } D = 1 + \rho(k - 1)]$$

$$n = Ck = ESS \times D$$

이 식에서 보듯이 총 표본크기는 집락표본의 표본크기 즉 유효 표본크기(effective sample size, ESS)에 design effect를 곱하여 계산된다. 이러한 관계에 의해 지난 호(16)에서 설명한 두 독립표본 비율비교에서 design effect를 고려한 표본크기는 다음의 공식으로 변형된다(3,10).

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

또는

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 [p_1(1 - p_1) + p_2(1 - p_2)]}{(p_1 - p_2)^2} \times D$$

연구 상황

개의 만성 신장염에 대한 새로운 내과적 처치 방법에 따른 치료효과를 평가하기 위하여 집락 무작위배정 임상시험을 계획하고 있다. 병원에 내원하는 환자의 특성(품고, 유전적, 환경적 등)은 병원의 소재지에 따라 차이가 있다고 판단되어 2곳의 동물병원을 대상으로 각각 20두의 환자를 치료하였다. 만성 신장염 환자의 치료 성공률이 약 40%로 알려져 있다고 할 때 새로운 치료법을 적용할 때 성공률이 적어도 20% 이상 증가하는 것을 기대하는 경우 유의수준 5%에서 80%의 검정력을 유지하기 위해서는 몇 두가 필요한가? 이 예제는 만성 신장염 환자의 치료 성공률이 병원에 내원하는 환자의 특성에 따라 다를 것으로 예상하여 2곳의 병원을 대상으로 집락표본을 이용한 연구계획이다.

본 예제에서 독립성을 가정할 때 표본크기는 각 군 당 97두로 총 194두가 필요하다.

$k = 1$ ($n_1 = n_2$), $p_1 = 0.4$, $q_1 = 0.6$, $p_2 = 0.6$, $q_2 = 0.4$, $\bar{p} = (0.4 + 0.6)/2 = 0.5$

$$n = \frac{[1.96\sqrt{2 \times 0.5 \times 0.5} + 0.84\sqrt{0.4 \times 0.6 + 0.6 \times 0.4}]^2}{(0.4 - 0.6)^2} \approx 97$$

여기에서 집락 내 상관계수를 이론테면 $ICC = 0.06$ 로 매우 작게 가정하여도 $D = 2.14[1 + 0.06 \times (20 - 1)]$ 이므로 총 표본크기는 415두(194×2.14)가 되어 독립성 가정(194두)에 비하여 표본크기가 2배로 증가한다.

전술한 연구 상황에서는 각 병원에 동일한 환자수를 배정

하였다($k = 1$). 예를 들어 동물병원(집락)에 내원하는 환자수에 차이가 있어 각 병원에 동일한 환자 수(집락크기)를 배정하기 곤란한 경우 지정된 유의수준에서 연구의 검정력을 달성할 수 있는 동물병원의 개수(집락수)와 집락크기를 적절히 조정해야 한다(3,10). 이 경우 집락수를 증가시키고 집락크기를 줄이는 방법과 집락수를 줄이고 집락크기를 증가시키는 방법이 있는데 일반적으로 집락수가 증가하면 연구에 필요한 비용이 증가하기 때문에 집락수와 집락크기의 다양한 조합을 고려하여 선택할 필요가 있다. 또한 ρ 가 클수록 조사비용이 증가하므로 ρ 가 1에 근사할 때 비용을 최소화하기 위해서는 집락수를 증가시키고 집락크기를 줄이는 전략이 바람직하다. 본 예제의 경우 유의수준 5%, 검정력 80%에서 $ICC = 0.06$ 을 가정할 때 분석용 패키지를 사용하면 집락크기 2, 3, 4, 5, 6, 7, 8, 9, 10에 대하여 집락수는 각각 104, 74, 58, 50, 44, 38, 36, 32, 30두로 계산된다. 즉 2개의 집락과 집락 당 104두를 선발하거나 혹은 3개의 집락과 집락 당 74두를 모집해도 동일한 검정력을 달성한다. 한편 동일한 가정에서 $ICC = 0.01$ 일 때 집락수는 98, 66, 50, 42, 34, 30, 26, 24, 22두로 계산된다.

임상의학에서는 환자의 질병 상태를 신속하게 진단하는 목적으로 진단 키트를 사용한다. 다음과 같은 상황을 가정하자.

연구상황

개의 질병 X를 진단하기 위해 새로 개발된 진단키트의 민감도가 90%로 기대되고, 유병률이 10%인 모집단에서 이 키트의 민감도가 95% 신뢰수준에서 적어도 75% 이상이라는 것을 90% 확신하기 위해서 감염 개체와 건강한 대조군 몇 두를 대상으로 평가해야 하는가?

시판 중인 많은 키트 중에서 어떠한 제품이 상대적으로 우수한지를 판단할 때 임상가에 따라 결정기준이 다르겠지만 키트의 능력을 평가하는 지표인 민감도와 특이도를 검증하는 방법을 이해한다면 결정기준을 설정하는데 매우 유용할 것이다. 다음 호에서는 진단키트의 능력을 검증하는데 필요한 표본크기를 계산하는 통계적 이론과 사례에 대하여 설명한다.

감사의 글

본 연구는 2012년도 농림수산검역본부의 용역연구개발사업(Z-1541779-2012-12-03)과 강원대학교 동물의학종합연구소의 지원에 의해 이루어졌으며 이에 감사드립니다.

참고문헌

1. Altman DG. How large a sample? In: Gore SM, Altman DG (eds). Statistics in practice. London, UK: British Medical Association 1982.
2. Bernay F, Bland JM, Häggström J, Baduel L, Combes B, Lopez A, Kaltsatos V. Efficacy of spironolactone on survival in dogs with naturally occurring mitral regurgitation caused by myxomatous mitral valve disease. J Vet Intern Med 2010; 24: 331-341.
3. Campbell MK, Thomson S, Ramsay CR, MacLennan GS, Grimshaw JM. Sample size calculator for cluster randomized

- trials. *Comput Biol Med* 2004; 34: 113-125.
4. Chandler HL, Gemensky-Metzler AJ, Bras ID, Robbin-Webb TE, Saville WJ, Colitz CM. In vivo effects of adjunctive tetracycline treatment on refractory corneal ulcers in dogs. *J Am Vet Med Assoc* 2010; 237: 378-386.
 5. Coskun A, Sen I, Guzelbektes H, Ok M, Turgut K, Canikli S. Comparison of the effects of intravenous administration of isotonic and hypertonic sodium bicarbonate solutions on venous acid-base status in dehydrated calves with strong ion acidosis. *J Am Vet Med Assoc* 2010; 236: 1098-1103.
 6. Donald A. Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm. *Prev Vet Med* 1993; 15: 125-145.
 7. Donald AW, Gardner IA, Wiggins AD. Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. *Prev Vet Med* 1994; 19: 167-187.
 8. Donner A. Statistical methodology for paired cluster designs. *Am J Epidemiol* 1987; 126: 972-979.
 9. Elbers ARW, Stegman JA, de Jong MF et al. Estimating sample sizes for a two-stage sampling survey of seroprevalence of pseudorabies virus (PRV)-infected swine at a regional level in the Netherlands. *Vet Q* 1995; 17: 92-95.
 10. Killip S, Mabfoud Z, Pearce K. What is an intracluster correlation coefficient? crucial concepts for primary care researchers. *Ann Fam Pract* 2004; 2: 204-208.
 11. Levine JM, Scrivani PV, Divers TJ, Furr M, Mayhew IJ, Reed S, Levine GJ, Foreman JH, Boudreau C, Credille BC, Tennent-Brown B, Cohen ND. Multicenter case-control study of signalment, diagnostic features, and outcome associated with cervical vertebral malformation-malarticulation in horses. *J Am Vet Med Assoc* 2010; 237: 812-822.
 12. Machin D, Campbell MJ, Tan SB, Tan SH. Sample size tables for clinical studies. 3rd ed, Wiley-Blackwell, 2009.
 13. McDermott JJ, Schukken YH, Shoukri MM. Study design and analytic methods for data collected from clusters of animals. *Prev Vet Med* 1994; 18: 175-191.
 14. McDermott JJ, Kadohira M, O'Callaghan CJ, Shoukri MM. A comparison of different models for assessing variations in the sero-prevalence of infectious bovine rhinotracheitis by farm, area and district in Keyna. *Prev Vet Med* 1997; 32: 219-234.
 15. Neumann S, Stolt P, Braun G, Hellmann K, Reinhart E. Effectiveness of the homeopathic preparation Zeel compared with carprofen in dogs with osteoarthritis. *J Am Anim Hosp Assoc* 2011; 47: 12-20.
 16. Pak S, Oh T. Sample size calculation in medical research. *J Vet Clin* 2012; 29: 68-7.
 17. Prange T, Holcombe SJ, Brown JA, Dechant JE, Fubini SL, Embertson RM, Peroni J, Rakestraw PC, Hauptman JG. Resection and anastomosis of the descending colon in 43 horses. *Vet Surg* 2010; 39: 748-753.
 18. Revie CW, Gettingby G, Treasurer JW, Wallace C. Evaluating the effect of clustering when monitoring the abundance of sea lice populations in farmed Atlantic salmon. *J Fish Biol* 2005; 66: 773-783.
 19. Roudsari B, Fowler R, Nathens A. Intracluster correlation coefficient in multicenter childhood trauma studies. *Inj Prev* 13: 344-347, 2007.
 20. Saarto EE, Hielm-Björkman AK, Hette K, Kuusela EK, Brandão CV, Luna SP. Effect of a single acupuncture treatment on surgical wound healing in dogs: a randomized, single blinded, controlled pilot study. *Acta Vet Scand* 2010; 15: 52-57.
 21. Schilte AF, Portegijs PJ, Blankenstein AH, van Der Horst HE, Latour MB, van Eijk JT, Knottnerus JA. Randomised controlled trial of disclosure of emotionally important events in somatisation in primary care. *BMJ* 2001; 323: 86-89.
 22. Solis-Calderon JJ, Segura-Correa VM, Segura-Correa JC. Bovine viral diarrhoea virus in beef cattle herds of Yucatan, Mexico: seroprevalence and risk factors. *Prev Vet Med* 2005; 72: 253-262.
 23. Stursberg U, Zenker I, Hecht S, Hartmann K, Schulz BS. Use of propentofylline in feline bronchial disease: prospective, randomized, positive-controlled study. *J Am Anim Hosp Assoc* 2010; 46: 318-326.
 24. Willesen JL, Kristensen AT, Jensen AL, Heine J, Koch J. Efficacy and safety of imidacloprid/moxidectin spot-on solution and fenbendazole in the treatment of dogs naturally infected with *Angiostrongylus vasorum* (Baillet, 1866). *Vet Parasitol* 2007; 147: 258-264.
 25. Young MJ, Bresnitz EA, Strom BL. Sample size nomograms for interpreting negative clinical studies. *Ann Intern Med* 1983; 99: 248-251.