

Tamaños de muestra para estimar prevalencia animal que aseguran cortos intervalos de confianza

Sample size determination for estimating animal prevalence that assure narrow confidence intervals

Osvaal Antonio Montesinos López^a, Abelardo Montesinos López^b, Eric Eduardo Santos Fuentes^a, Patricia Edwigis Valladares Celis^c, Martha Alicia Magaña Echeverría^d

RESUMEN

El cálculo del tamaño de las muestras juega un rol importante en el diseño óptimo de experimentos veterinarios y agrícolas para la estimación de proporciones de una población, incluyendo la prevalencia de enfermedades. Esta investigación propone un método de tres pasos para determinar el tamaño de muestra exacto para datos binomiales que asegura precisión en la estimación de la proporción, y muestra numéricamente el grado de subestimación que produce el uso de la fórmula tradicional (aproximación normal) para el cálculo del tamaño de la muestra. El paso 1 obtiene un tamaño de muestra que garantiza que la anchura relativa completa del intervalo de confianza (\bar{w}_r) es más estrecha que la amplitud deseada (re); el paso 2 incrementa iterativamente el tamaño de muestra hasta que \bar{w}_r es más pequeño que la amplitud deseada (re) con un grado de certeza especificado (γ) y el paso 3 obtiene el número de conglomerados requeridos. Datos simulados fueron creados para ilustrar el método propuesto; además se presenta un cuadro con escenarios útiles para los investigadores. Un programa en el paquete estadístico R es dado y explicado, de tal manera que reproduce los resultados de una manera sencilla.

PALABRAS CLAVE: Tamaño de muestra, Distribución binomial, Cortos intervalos de confianza, ARCIC.

ABSTRACT

Sample size calculation plays an important role in the design of optimal veterinary and agricultural experiments for estimate population proportions including disease prevalence. This research proposes a method with three steps to determine the exact sample size for a binomial distribution that ensures precision in the estimated proportion and numerically displays the degree of underestimation that occurs using the traditional formula (normal approximation) for calculating the sample size. Step 1 obtain a sample size that guarantees that the relative mean width of the CI (\bar{w}_r) is narrower than the desired width (re); the step 2 iteratively increase the sample size until \bar{w}_r is smaller than the desired width (re) with a specified degree of certainty (γ) and step 3 get the required number of clusters. Simulated data were created and tables presented showing possible scenarios. An R program is given and explained that will reproduce the results in an easy way.

KEY WORDS: Sample size, Binomial distribution, Narrow confidence intervals, CIRW.

Las decisiones en salud animal para los programas de control de enfermedades, comercio de animales, o las certificaciones de estar libre de enfermedades, confían en los conocimientos precisos y con base científica de los estados de salud de animales en una población animal específica. La cuestión

Animal health decisions for disease control programs, animal trade, or disease freedom certification rely on accurate, scientifically based knowledge of the health status of animals within a targeted animal population. The underlying question is to what extend or magnitude (incidence/prevalence

Recibido el 22 de enero de 2010. Aceptado el 27 de septiembre de 2010.

^a Facultad de Telemática, Universidad de Colima, Colima, México. oamontes1@uocol.mx. Correspondencia al primer autor.

^b Departamento de Estadística, Centro de Investigación en Matemáticas (CIMAT).

^c Bachillerato 22. Universidad de Colima, Colima, México.

^d Dirección General de Planeación, Universidad de Colima.

fundamental es conocer en qué extensión o magnitud (incidencia/nivel de predominio) ocurre en realidad la enfermedad en la población⁽¹⁾. Por ello, los gobiernos de países desarrollados y en desarrollo le han dando mucha importancia a estos programas de monitoreo y sistemas de vigilancia animal⁽²⁾.

Cuando se intenta estimar la prevalencia animal o la ausencia de agentes infecciosos en poblaciones, los métodos de muestreo son de suma importancia⁽³⁾. Por esta razón, el cálculo del tamaño de muestra juegan un papel importante en el diseño óptimo de experimentos veterinarios y agrícolas para la estimación de proporciones en poblaciones, incluyendo la prevalencia de una enfermedad⁽⁴⁾. Esto debido a que, una muestra demasiado pequeña no puede asegurar la precisión suficiente en la estimación del parámetro de interés, mientras que una muestra muy grande es un desperdicio innecesario de recursos⁽⁵⁾.

En lo que respecta a determinar el tamaño de muestra, tradicionalmente, estadísticos han formulado sus requerimientos en términos de potencia. Este enfoque es consistente con las pruebas de hipótesis para inferencia y presentan los resultados en términos de valores p (p -values). Recientemente, ha crecido mucho el interés en el uso de intervalos de confianza (ICs) en lugar de pruebas de hipótesis para la realización de inferencias⁽⁶⁾. De hecho, algunas revistas, reconociendo posibles problemas con las pruebas de hipótesis, últimamente han adoptado políticas editoriales o emitido declaraciones editoriales alentando el uso de ICs en los artículos que en ellas se publican, cuando dichos intervalos son justificados. Una gran ventaja de los ICs en la presentación de resultados en comparación con los valores p (p values que proporcionan las pruebas de hipótesis) es que son relativamente cercanos a los datos, estando en la misma escala de medida, mientras que los valores p (p -values) son una probabilidad abstracta⁽⁷⁾. Además, los ICs ayudan a asegurarse no sólo de que la magnitud del efecto puede ser evaluado, si no que el efecto en cuestión puede ser fácilmente identificado por un lector. Es

level) does the disease actually occur in the population⁽¹⁾. The implementation of appropriate animal-health monitoring and surveillance systems is required to develop information to address this question. This is why governments of developed and developing countries are giving a great importance to programs of animal-health monitoring and surveillance systems⁽²⁾.

When attempting to estimate animal prevalence or substantiate freedom of populations from infectious agents, sampling methods are of paramount importance⁽³⁾. For this reason, sample size calculation plays an important role in the design of optimal veterinary and agricultural experiments for estimate population proportions including disease prevalence⁽⁴⁾. Because, a sample size too small cannot assure sufficient precision for the estimation of the parameter of interest, while a sample size too large is an unnecessary waste of the limited resources⁽⁵⁾.

With regards to sample size determination, traditionally, statisticians have formulated it in terms of power considerations. This approach is consistent with an emphasis on hypotheses testing for inference and presentation of results in terms of p -values. Recently, there has been a growing interest in the use of confidence intervals (CIs) instead of hypotheses tests for inference-making purposes⁽⁶⁾. In fact, some journals, recognizing possible problems with hypothesis tests, have recently adopted editorial policies or issued editorial statements encouraging the use of CIs in published papers therein, when such intervals are warranted. A major advantage of CIs in the presentation of results is that interval estimates, in common with point estimates, are relatively close to the data, being on the same scale of measurement, whereas the p -value is a probabilistic abstraction⁽⁷⁾. Because, CIs help to assure not only that the magnitude of the effect can be better assessed, but that the effect in question can be readily identified by the reader. Importantly, of course, confidence intervals convey information about magnitude and precision of effect simultaneously, keeping these two aspects of

importante destacar que los intervalos de confianza transmiten información sobre magnitudes y precisión, manteniendo estos dos aspectos de medición cercanamente ligados^(7,8). El usual intervalo de confianza bilateral es simplemente interpretado como un margen de error de una estimación puntual⁽⁷⁾.

Por estas razones una creciente atención se ha dado en el diseño de métodos para calcular tamaños de muestra apropiados para ICs. Este enfoque ha sido denominado *precisión en la estimación de parámetros* (AIPE por sus siglas en inglés), ya que cuando la anchura del IC con $(1-\alpha)$ 100 % de confianza disminuye, la exactitud esperada en la estimación incrementa^(8,9,10). Aunque el enfoque AIPE para la planificación del tamaño de muestra no es nuevo⁽¹¹⁾, ha sido examinado más en las ciencias sociales que en ciencias veterinarias y agrícolas.

Para determinar los tamaños de muestra, es necesario obtener información de algunos parámetros. No obstante, en la práctica, estos parámetros son desconocidos y son usualmente estimados basándose en literatura o estudios piloto. Estas estimaciones son tratadas como parámetros verdaderos. Consecuentemente, no se toma en cuenta la incertidumbre inducida por dichas estimaciones. Como resultado, el tamaño de muestra resultante puede no alcanzar la anchura deseada para estimar un parámetro como fue planeado⁽⁵⁾. Para dar cuenta de tal incertidumbre inducida por usar una estimación del parámetro desconocido, Kelley⁽⁸⁾, y Kupper y Hafner⁽¹²⁾ enfatizan que la naturaleza estocástica de la anchura del IC debe ser considerada para evitar serias subestimaciones de los tamaños de muestra requeridos para lograr la anchura deseada. En particular, muestran este fenómeno de subestimación en forma numérica para estimar el promedio de una muestra que proviene de una distribución normal y para estimar dos medias que provienen de dos muestras normales asumiendo igualdad de varianzas. Wang y Kupper⁽¹³⁾ amplían esta metodología para el caso de muestras aleatorias de dos poblaciones normales con varianzas desiguales.

measurement closely linked^(7,8). The usual two-sided confidence interval is thus simply interpreted as a margin of error about a point estimate⁽⁷⁾.

For these reasons an increasing attention is being given to design-stage methods for calculating appropriate sample sizes for CIs. This approach to sample size estimation has been termed accuracy in parameter estimation (AIPE), because when the width of the $(1-\alpha)$ 100% CI decreases, the expected accuracy of the estimate increases^(8,9,10). Although the AIPE approach to sample size planning is not new⁽¹¹⁾, it has been examined more in social science than in veterinary and agricultural sciences.

To perform sample size calculation, it is necessary to obtain information regarding some study parameters. In practice, however, those parameters are unknown and are usually estimated from literature or some pilot studies. Those estimates are then treated as true parameters. Consequently, it fails to account for the uncertainty induced by such estimations. As a result, the resulting sample size may not achieve the desired width for estimating a parameter as was planned⁽⁵⁾. To account for such an uncertainty induced by using an estimate of the unknown parameter, Kelley⁽⁸⁾, and Kupper and Hafner⁽¹²⁾ emphasize that the stochastic nature of CI width will be taken into account to avoid serious underestimation of sample sizes required to achieve the desired width. In particular, they show numerically this underestimation phenomenon when using either one normal random sample or equal-sized random samples from two normal populations with common variance to make statistical inferences about population means. Wang and Kupper⁽¹²⁾ extend this methodology to the case of unequal-sized random samples from two normal populations with unequal variances.

The objective of this research is to propose an exact method for sample size (n) determination for estimating the animal prevalence (binomial proportion; p) that assure a narrow CI, that is, assure accuracy in the estimation of p . The method requires three steps and was constructed treating

El objetivo de esta investigación es proponer un método exacto para la determinación de tamaños de muestra (n) para la estimación de la prevalencia animal (proporción binomial; p) que asegura un estrecho IC, esto es, asegurar precisión en la estimación de p . El método requiere tres pasos y fue construido tratando la amplitud del IC de manera aleatoria. El primer paso del método propuesto permite a los investigadores planear los tamaños de muestra para que la anchura relativa completa del IC para p sea lo suficientemente pequeña. Una modificación permite un grado de seguridad deseado (γ) que está incorporado al método, para que el IC obtenido sea lo suficientemente pequeño con alguna probabilidad específica; es decir, el segundo paso toma en cuenta la naturaleza estocástica de la amplitud del IC. El tercer paso proporciona el número de conglomerados requeridos cuando el esquema de muestreo es por conglomerados en dos etapas.

Al planearse el tamaño de la muestra, para obtener una amplitud suficientemente estrecha del IC calculado con el enfoque AIPE, se necesita usar un proceso iterativo⁽⁸⁾. Sea $p_L(X) = p_L$ el límite inferior de confianza de p y $p_U(X) = p_U$ el límite superior de confianza de p ; para un nivel de confiabilidad especificado, donde X representa la matriz de datos observados sobre los cuales el intervalo de confianza (IC) está basado. La amplitud completa del IC con el método de Clopper-Pearson^(7,14) es $w = p_U - p_L$, donde $P_L = B_{\alpha/2, y, n-y+1}$ y $P_U = B_{1-\alpha/2, y+1, n-y}$ (límites inferior y superior para basados en una distribución Beta(a, b)). No obstante, la amplitud relativa completa del IC (ARCIC) es igual a

$$w_r = (p_U - p_L) / p \dots\dots\dots (1)$$

Se usará re como la amplitud relativa deseada del IC (error relativo), especificada previamente por el investigador. También, para planear el tamaño de la muestra con los objetivos de AIPE, debe ser conocido (o estimado). Para obtener (w_r), es

CI widths as random. The first step of the proposed method allows researchers to plan sample size so that the relative mean width of the CI for p is sufficiently narrow. A modification allows a desired degree of certainty (γ) that is incorporated into the method, so that the obtained CI will be sufficiently narrow with some specific probability, that is, the second step take into account the stochastic nature of CI width. The third step provides the number of clusters required when the sampling scheme is under cluster sampling in two stages.

When planning sample size, under the AIPE approach in order to obtain a width of the computed CI sufficiently narrow, we need to use an iterative process⁽⁸⁾. Let $p_L(X) = p_L$ be the random lower confidence limit for p , and $p_U(X) = p_U$ be the random upper confidence limit for p at the specified confidence level, where X represents the observed data matrix on which the confidence interval (CI) is based. The entire width of the obtained CI under the Clopper-Pearson method^(7,14) is $w = p_U - p_L$, where $p_L = B_{\alpha/2, y, n-y+1}$ and $p_U = B_{1-\alpha/2, y+1, n-y}$ (the lower and upper confidence limits of p). Therefore, the entire relative width of the CI (CIRW) will be defined as

$$w_r = (p_U - p_L) / p \dots\dots\dots (1)$$

Let re be the desired relative width of the CI (relative error), specified a priori by the researcher. Also, in order to plan sample size under the goals of AIPE, p must be known (or estimated). To obtain (w_r), it is necessary to know the sample size (n) as well as the number of success (y). Nevertheless, since the sample size (n) is unknown, it is not possible to know how many individuals obtained from the sampling procedure will be positives (m). Thus, w_r cannot be determined exactly. However, according to Vollset⁽¹⁴⁾ and Newcombe⁽⁷⁾, it is possible to determine the exact mean CIRW value by averaging the widths of the intervals associated with the number of successes $0 \leq y \leq n$, given n , and p using the expression

necesario conocer el tamaño de la muestra (n) así como también el número de eventos (y). Sin embargo, dado que el tamaño de la muestra (n) es desconocido, no es posible conocer cuántos individuos resultarán positivos (y) tras el muestreo. Por lo tanto, w_r no puede ser determinada exactamente. No obstante, de acuerdo con Vollset⁽¹⁴⁾ y Newcombe⁽⁷⁾, es posible determinar el valor exacto de la ARCIC promediando las amplitudes de los intervalos asociados con el número de eventos de entre $0 \leq y \leq n$, dados n , y y p usando la expresión

$$\bar{w}_r(p, n) = \sum_{y=0}^n w_r \binom{n}{y} [p]^y [1-p]^{n-y} \dots\dots\dots (2)$$

Donde w_r es la ARCIC, que se calcula con la ecuación (1), para $Y = y$ individuos positivos, dados la proporción (p) y el tamaño de muestra (n).

Para determinar el tamaño de muestra óptimo que asegura cortos intervalos de confianza, se requieren dos pasos. El primero, calcula el tamaño de muestra inicial (o preliminar). Este paso consiste en empezar con algún tamaño de muestra mínimo, digamos n_0 , para que podamos encontrar el valor de n que satisfaga

$$\bar{w}_r(p, n) = \sum_{y=0}^n w_r \binom{n}{y} [p]^y [1-p]^{n-y} \leq re \dots\dots (3)$$

Para simplificar, escribiremos $\bar{w}_r(p, n)$ como \bar{w}_r , tomando en cuenta que la amplitud relativa necesariamente dependerá de $1-\alpha$, p y n . Si la ARCIC es más ancha que el valor deseado (re), el tamaño de muestra debe de ser aumentado en uno, y la ARCIC debe de ser determinada nuevamente. Este proceso iterativo de incrementar el tamaño de muestra y recalculando la ARCIC deberá de continuar hasta que $\bar{w}_{ri} \leq re$ (donde i representa la i -ésima iteración). Este método encuentra el tamaño de muestra requerido para conseguir que la ARCIC sea lo suficientemente reducida. Aún así, no garantiza que para cualquier IC la anchura observada será lo suficientemente pequeña porque la ARCIC,

$$\bar{w}_r(p, n) = \sum_{y=0}^n w_r \binom{n}{y} [p]^y [1-p]^{n-y} \dots\dots\dots (2)$$

Where w_r is the CIRW, and is computed using equation (1), for $Y = y$, given p and n .

To determine the optimal sample size that ensures short confidence intervals, two steps are required. The first calculates the initial sample size (or preliminary). The second consists of starting at some minimal sample size, say n_0 , so that we can find the value of n that satisfies

$$\bar{w}_r(p, n) = \sum_{y=0}^n w_r \binom{n}{y} [p]^y [1-p]^{n-y} \leq re \dots\dots (3)$$

For simplicity, we will write $\bar{w}_r(p, n)$ as \bar{w}_r , realizing that the relative width will necessarily depend on $1-\alpha$, p and n . If the CIRW is wider than the desired value (re), the sample size should be increased by one, and the CIRW should be determined again. This iterative process of increasing sample size and recalculating the CIRW should continue until $\bar{w}_{ri} \leq re$ (where i represents the particular iteration number). This method finds the required sample size for achieving a mean CIRW that is sufficiently narrow. However, it does not guarantee that for any particular CI the observed width will be sufficiently narrow because the mean CIRW, \bar{w}_r , will be a random variable ($\hat{\bar{w}}_r$) that will fluctuate from sample to sample if we don't know the exact value of p . To demonstrate this, we need to calculate the probability of obtaining CIRW's less than the the specified value (re), which can be computed as

$$P(\hat{\bar{w}}_r \leq re) = \sum_{y=0}^n I(w_r, y, p) \binom{n}{y} [p]^y [1-p]^{n-y} \dots\dots (4)$$

where $I(w_r, y, p)$ is an indicator function showing whether or not the actual CIRW calculated with equation (1) is no greater than re , and $\hat{\bar{w}}_r$ is considered a random variable because we don't know the exact value of p .

We show a numerical example for sample size determination under the AIPE approach in Table 1,

\widehat{w}_r , es una variable aleatoria (\widehat{w}_r) que oscila de muestra a muestra si no sabemos el valor exacto de p . Para demostrar esto, necesitamos calcular la probabilidad de obtener ARCICs menores que el valor especificado (re), que puede ser calculado con

$$P(\widehat{w}_r \leq re) = \sum_{y=0}^n I(w_r, y, p) \binom{n}{y} [p]^y [1-p]^{n-y} \dots \quad (4)$$

Donde $I(w_r, y, p)$ es una variable indicadora que muestra si la ARCIC actual calculada con la ecuaci3n (1) no es m3s grande que re , y \widehat{w}_r es considerada una variable aleatoria puesto que no sabemos el valor exacto de p .

En el Cuadro 1, se ilustra la determinaci3n del tama1o de la muestra con el enfoque AIPE para un IC de 95% y para varios valores de la proporci3n

with a CI of 95 %, with various values of the proportion in the population (p). Under the AIPE approach the required preliminary sample size is calculated with equation (3) and the probability $P(\widehat{w}_r \leq re)$ that the mean CIRW \widehat{w} is less than the specified value ($re = 0.3$) is calculated with equation (4). This is done with the objective of showing that the initial sample size (n_p , second column) computed using equation (3) has a probability of around 0.50 (50 %) that the mean CIRW $\widehat{w}_r \leq re = 0.3$ (third column). For example, when $p = 0.1$, the preliminary (initial) sample size (n_p) is 1,600 and the probability of obtaining a CIRW $\leq re = 0.3$ is 0.4878. With $p = 0.25$, $n_p = 536$, we can only be 48.34 % certain that the CIRW will be $\leq re = 0.3$. When the sample size is increased by 100 (n_{m100} , fourth column),

Cuadro 1. Tama1o de muestra preliminar (n_p) y tres incrementos, cada uno con su correspondiente probabilidad de que el intervalo de confianza esperado para la anchura relativa completa (ARCIC) \widehat{w} es menor que el valor especificado ($re = 0.3$) ($P(\widehat{w}_r < re)$). IC de 95%. La proporci3n en la poblaci3n es p , $re = 0.3$ es la ARCIC deseada (error relativo), n_p es el tama1o de muestra preliminar calculado usando la ecuaci3n 3. $P(\widehat{w}_r < re)$ es la probabilidad de que la ARCIC es menor que el valor especificado ($re = 0.3$) calculado usando la ecuaci3n 4; n_T es el tama1o de muestra tradicional con la aproximaci3n normal ($n_T = [4(Z^2 * q)/(re^2 * p)]$)

Table 1. Initial sample size (n_p) and three increments, each with its corresponding probability that the expected confidence interval relative width (CIRW) \widehat{w} is less than the specified value ($re = 0.3$) ($P(\widehat{w}_r < re)$). For a 95% CI. The proportion in the population is p , $re = 0.3$ is the desired CIRW, and n_p is the preliminary sample size computed using Eq. 3. $P(\widehat{w}_r < re)$ is the probability that the CIRW is less than the specified value ($re = 0.3$) calculated using Eq.4. n_T is the traditional sample size with normal approximation ($n_T = [4(Z^2 * q)/(re^2 * p)]$)

p	n_p	$P(\widehat{w}_r < re)$	n_{m100}	$P(\widehat{w}_r < re)$	n_{m300}	$P(\widehat{w}_r < re)$	n_{m800}	$P(\widehat{w}_r < re)$	n_T
0.05	3374	0.4984	3474	0.6775	3674	0.9097	4174	0.9999	3245
0.1	1600	0.4878	1700	0.8439	1900	0.9992	2400	1.00	1537
0.15	1009	0.5094	1109	0.9607	1309	1.00	1809	1.00	968
0.2	713	0.5000	813	0.9982	1013	1.00	1513	1.00	683
0.25	536	0.4834	636	0.9999	836	1.00	1336	1.00	513
0.3	417	0.4773	517	1.00	717	1.00	1217	1.00	399
0.35	333	0.5000	433	1.00	633	1.00	1133	1.00	318
0.4	270	0.4769	370	1.00	570	1.00	1070	1.00	257
0.45	220	0.4211	320	1.00	520	1.00	1020	1.00	209
0.5	181	0.5522	281	1.00	481	1.00	981	1.00	171

(p) en la población y se comparan con los del método tradicional ($n_T = \{4Z^2q/re^2p\}$). Con el enfoque AIPE el tamaño de muestra requerido preliminar se calcula con la ecuación (3) y la probabilidad de $P(\hat{w}_r \leq re)$ que la ARCIC \bar{w} es menor del valor especificado ($re = 0.3$) con la ecuación (4). Esto es hecho con el objetivo de mostrar que el tamaño de muestra preliminar (n_p , segunda columna) calculado con la ecuación (3) tiene una probabilidad de alrededor de 0.50 (50 %) de que la ARCIC sea menor o igual al error relativo especificado $re = 0.3$ (tercera columna). Por ejemplo, cuando $p=0.1$, el tamaño de muestra preliminar (n_p) es 1,600 y la probabilidad de obtener una $ARCIC \leq re = 0.3$ es 0.4878. Con $p=0.25$, el tamaño de muestra preliminar es 536 y sólo podemos estar el 48.34 % seguros de que la ARCIC será menor o igual a 0.3. No obstante, cuando los tamaños de muestra aumentan en 100 (n_{m100} , cuarta columna), la probabilidad de que se cumpla la amplitud deseada del IC incrementa. Por ejemplo, cuando $p=0.1$ y hay 1,700 unidades en la muestra, $P(\hat{w}_r \leq 0.3) = 0.8439$, y para un tamaño de muestra de 1,900 unidades, $P(\hat{w}_r \leq 0.3) = 0.9992$, es decir, incrementa significativamente la probabilidad de que se cumpla la ARCIC.

Los resultados en el Cuadro 1 muestran que para lograr $P(\hat{w}_r \leq re = 0.3)$ con una alta probabilidad de éxito, necesitamos aumentar significativamente el tamaño de muestra. Por lo tanto, con el enfoque AIPE se puede observar que el tamaño de muestra preliminar (n_p) es insuficiente (probabilidad de alrededor del 50 %) para garantizar que se cumpla la amplitud deseada (re) de la ARCIC, por esto es de vital importancia agregar cierto nivel de aseguramiento. Por otro lado, comparando los tamaños de muestra preliminar (n_p) con los tradicionales que asumen normalidad se puede ver que los tamaños de muestra tradicionales producen tamaños de muestra considerablemente más pequeños que los exactos sin aseguramiento con el enfoque AIPE. Esto sugiere que se debe abandonar el uso de los tamaños de muestra tradicionales ($n_T = \{4Z^2q/re^2p\}$ cuando se usa la amplitud relativa completa del IC o $n_T = \{Z^2q/re^2p\}$ cuando se usa

$P(\hat{w}_r \leq re = 0.3)$, increases. For example, when $p=0.1$ and there are 1,700 units in the sample, $P(\hat{w}_r \leq 0.3) = 0.8439$, and for a sample size of 1,900 units, $P(\hat{w}_r \leq 0.3) = 0.9992$.

The results in Table 1 show that to achieve $P(\hat{w}_r \leq re = 0.3)$ with a high probability of success, we require an increasing sample size. Therefore, under the AIPE approach we can see that the preliminary sample size is not enough (probability of about 50 %) to ensure that the CIRW will be satisfied (re), so it is vital to add some level of assurance. On the other hand, comparing the preliminary sample sizes (n_p) with the traditional sample size assuming normality we can see that the traditional sample sizes are smaller than the exact approach without assurance ($n_T = \{4Z^2q/re^2p\}$) under AIPE. This suggests that we must abandon the use of traditional sample sizes ($n_T = \{Z^2q/re^2p\}$ when using the full relative width of the CI or when using the half relative width of the CI) because they produce a serious underestimation and the probability that the CIRW is less than the specified a priori relative error is less than 50 %.

Therefore, it requires a second step to ensure a sample size that ensures that the CIRW is not wider than the desired width at a certain level of probability. That is, it is required a modification of the previous sample size procedure so that a probabilistic statement can be incorporated into the sample size planning procedure, in order to guarantee a sufficiently narrow CI. This procedure allows us to use the initial sample size (n_p) and compute a modified sample size. In any given sample, the estimator will not be equal to p , even if the true value of p is known. Approximately half of the time, the sample p will be larger (or smaller) than p . For this reason, the preliminary sample size (n_p) produced by equation (3) only assure approximately a 50 % chance that the interval will be narrower than required (Table 1), because p will vary from sample to sample.

To increase the likelihood that the obtained CIRW will be no wider than required, we propose a

la mitad de la amplitud relativa del IC) porque producen una subestimaci3n muy grande y la probabilidad de que la ARCIC sea menor que el error relativo especificado *a priori* es menor al 50 %.

Por lo tanto, se requiere de un segundo paso para garantizar un tama1o de muestra que asegure una ARCIC no m1s ancha que la amplitud deseada con un cierto nivel de probabilidad. Es decir, se requiere de una modificaci3n al procedimiento previo para la determinaci3n del tama1o de muestra con el objeto de garantizar un IC suficientemente estrecho. Para ello, se incorpora una afirmaci3n (declaraci3n) probabilística dentro del procedimiento de planificaci3n del tama1o de muestra. Por lo tanto, se utilizar1 el tama1o de muestra inicial (n_p) para poder calcular el tama1o de muestra modificado. En cualquier muestra dada, el valor de calculado no ser1 igual a p , a1n si se conoce el valor verdadero de p . Aproximadamente la mitad de las veces la estimaci3n de p ser1 m1s grande (o m1s peque1a) que p . Por esta raz3n, el tama1o de muestra preliminar (n_p) producido por la ecuaci3n (3) solamente asegura una probabilidad de 50 % de que el intervalo sea m1s angosto que el requerido (Cuadro 1), porque p variar1 de muestra a muestra.

Para incrementar la probabilidad de que la ARCIC no sea m1s amplia de lo requerido, proponemos una modificaci3n al tama1o de muestra inicial calculado en el paso 1 (Ecuaci3n 3). Esto producir1 un tama1o de muestra modificado (n_m) de tal manera que el investigador puede estar γ por ciento seguro de que la ARCIC (\bar{w}_r) observada ser1 menor o igual al requerido (re). La ecuaci3n (3) brinda el tama1o de muestra preliminar y para obtener n_m (tama1o de muestra modificado) necesitamos resolver

$$P(\hat{w}_r \leq re) = \sum_{y=0}^n I(w_r, y, p) \binom{n}{y} [p]^y [1-p]^{n-y} \geq \gamma \dots (5)$$

El procedimiento involucra incrementar el tama1o de muestra preliminar (n_p) por una unidad y recalculer la ecuaci3n (5) cada vez hasta lograr el grado deseado de certeza (γ). Esto producir1 el

modification of the initial sample size calculated in step 1 (Equation 3). This will produce a modified sample size (n_m) such that the researcher can be γ percent confident that the observed \bar{w}_r will be less than or equal to the required sample size. Using equation (3) gives the preliminary sample size and, to obtain n_m (modified sample size), we need to solve

$$P(\hat{w}_r \leq re) = \sum_{y=0}^n I(w_r, y, p) \binom{n}{y} [p]^y [1-p]^{n-y} \geq \gamma \dots (5)$$

The procedure involves increasing the initial sample size (n_p) by one unit and recalculating equation (5) each time until the desired degree of certainty (γ) is achieved. This will produce the modified sample size (n_m) that assures, with a probability of at least γ , that the \bar{w}_r will be no wider than re . In other words, n_m ensures that the researcher will have approximately 100 γ % certainty that a computed CI will be of the required width or less. For example, if the researcher requires 90% confidence that the obtained \bar{w}_r will be no larger than the desired width (re), $1-\gamma$ would be defined as 0.10, and there would be only a 10 % chance that the width of the CI, around \hat{p} , would be larger than specified (re)⁽⁹⁾.

As for choosing a 100 γ percent CI for estimation as compared with choosing the 100(1- α) percent CI for hypothesis testing, important distinctions can be pointed out. The most obvious difference in the present context is that (1- γ) represents the probability of obtaining a CIRW that is larger than the specified required width of re , whereas α is the probability of rejecting a null hypothesis that is true. Also, α is thought as one of two essentially constant values, 0.05 (5 %) or 0.01 (1 %), while γ ($\gamma \geq 0.5$ (50%)) is chosen by the researcher to achieve some desired degree of certainty that the precision of the estimated parameter will be achieved. Thus CIs formed in the context of hypothesis testing have different goals than CIs formed when the researcher wishes to obtain a precise estimate of the parameter of interest^(8,9).

The above two steps produce a sample size (n_m) that ensures narrow confidence intervals but under

tamaño de muestra modificado (n_m) que asegure con una probabilidad de por lo menos γ , que la ARCIC (\bar{w}_r) no será más ancha que re . En otras palabras, n_m asegura que el investigador tendrá aproximadamente 100 γ % de certeza de que el IC calculado será de la amplitud requerida o menos. Por ejemplo, si el investigador requiere un nivel de aseguramiento de 90 % de que la ARCIC (\bar{w}_r) no sea más grande que la anchura deseada (re), $1-\gamma$ sería definido como 0.10, y habría solamente un 10 % de probabilidad de que la amplitud del IC, alrededor de \hat{p} , sea más grande que el valor especificado (re)⁽⁹⁾.

En cuanto a escoger un nivel de aseguramiento de 100 por ciento para la estimación de la proporción y a escoger un IC del 100(1- α) por ciento para pruebas de hipótesis, se pueden señalar importantes diferencias. La diferencia más obvia es que (1- γ) representa la probabilidad de obtener una ARCIC observada (\bar{w}_r) que es más grande que la amplitud requerida, re , mientras que α es la probabilidad de rechazar la hipótesis nula dado que es verdadera (Error tipo I). También, α comúnmente toma los valores de 0.05 (5%) ó 0.01 (1%), mientras que γ ($\gamma \geq 0.5(50\%)$) es escogido por el investigador para lograr algún grado de certeza deseado de que la precisión del parámetro deseado se logre. Esto es, el IC formado en el contexto de pruebas de hipótesis tiene diferentes metas al IC formado cuando el investigador desea obtener un cálculo preciso del parámetro de interés^(8,9).

Hay que tener presente que los dos pasos anteriores producen un tamaño de muestra (n_m) que asegura cortos intervalos de confianza pero con muestreo aleatorio simple. Sin embargo, en el área de veterinaria debido a la limitada disponibilidad de marcos de muestreo confiables y altos costos de transporte, es usualmente imposible e impráctico seleccionar una muestra aleatoria simple de individuos de una población⁽¹⁵⁾. Una solución bastante usada en la mayoría de las encuestas es tomar una muestra de conglomerados (grupo de individuos) donde el hato o la comunidad es la unidad de muestreo y seleccionar aleatoriamente

simple random sampling (SRS). However, in the veterinary area due to the limited availability of reliable sampling frames and high transport costs, it is usually impossible and impractical to select a simple random sample of individuals in a population⁽¹⁵⁾. The solution most utilized in surveys is to take a sample of clusters (groups of individuals) where the herd or the community is the sampling unit and randomly it is selected a number of animals per herd or community (cluster sampling in two stages). However, it must be remembered that for a given number of individuals, the precision of a cluster sample is lower than that obtained through simple random sampling, due to the intra cluster correlation (rate of homogeneity). That is, 20 animals of the same herd provide less information than 20 animals from different herds^(15,16). This reduction in precision should be considered when calculating the sample size in cluster sampling and multi stages procedures⁽¹⁵⁾. A measure of the standard error increased in the prevalence due to the sampling procedure used is given by the square root of the design effect, $Deff = 1 + (b-1)ro$ where b is the average number of individuals measured per cluster and ro is the intra cluster correlation coefficient of the disease under study. In a random sampling of clusters ro is equivalent to the intra class correlation⁽¹⁵⁾.

In cluster sampling the decision that usually has to be considered is how many clusters of size b will be needed to obtain the desired accuracy, since b is usually limited by the size of the herd and other practical considerations. If the design effect is known, the number of clusters can be calculated

$$c = \frac{Deff * n_m}{b} \dots\dots\dots(6)$$

Where: n_m is the sample size preliminary or modified under simple random sampling (provided by the two steps above using equations (3) and (5), $Deff = 1 + (b-1)ro$ is the design effect, but you need to know ro and b . Approximate values can be obtained from a pilot sample or results of previous surveys, whereas ro vary from region to region and from poll to poll^(15,16).

un número de animales de cada hato o comunidad (muestreo por conglomerados en dos etapas). Sin embargo, hay que tener presente que para un número dado de individuos, la precisión de una muestra por conglomerados es menor a la obtenida por medio de muestreo aleatorio simple, debido a la correlación intraconglomerados (tasa de homogeneidad); es decir, 20 animales de un mismo hato proporcionan menos información que 20 animales de diferentes hatos^(15,16). Esta reducción en la precisión debe ser considerada al calcular el tamaño de muestra en un muestreo por conglomerados o etapas múltiples⁽¹⁵⁾. Una medida del incremento del error estándar de la prevalencia debido al procedimiento de muestreo utilizado está dado por la raíz cuadrada del efecto de diseño, $Deff = 1 + (b-1)ro$ donde: b es el número promedio de individuos medidos por conglomerado y ro es el coeficiente de correlación intraconglomerados de la enfermedad en estudio. En un muestreo aleatorio simple de conglomerados ro es equivalente a la correlación intraclase⁽¹⁵⁾.

En muestreo por conglomerados la decisión que usualmente tiene que ser considerada es, cuántos conglomerados de tamaño b serán necesarios para obtener la precisión deseada, ya que b suele estar limitado por el tamaño del hato y otras consideraciones prácticas. Si el efecto de diseño es conocido, el número de conglomerados puede calcularse con

$$c = \frac{Deff * n_m}{b} \dots\dots\dots (6)$$

Donde: n_m es el tamaño de muestra modificado o preliminar bajo muestreo aleatorio simple (proporcionado por los dos pasos anteriores usando las ecuaciones (3) y (5), $Deff = 1 + (b-1)ro$ es el efecto de diseño y para calcularlo se necesita conocer ro y b . Valores aproximados de ro pueden obtenerse de una muestra piloto o de resultados de encuestas previas, considerando que ro variará de región a región y de encuesta a encuesta^(15,16).

Aunque el apéndice que se presenta proporciona información sobre la implementación de los métodos

Although the Appendix provides information on implementing the methods proposed and discusses how to obtain sufficiently narrow CIs for any combination of p , re , γ , α , ro , b , using the R package⁽¹⁷⁾, tables of selected conditions are provided. The tables are not meant to include all potentially interesting conditions; rather, they are intended to provide researchers with: 1) a convenient way to plan sample size when the situation of interest is approximated by the scenarios covered by the tables; and 2) a way to illustrate the relation between p , re , γ , α , ro , b , and the necessary sample size (n). The values in Tables 2 are based on numbers that should be useful for detecting and estimating p with small values of p .

Sample sizes are shown for p values of 0.05 to 0.5 by 0.05, re values of 0.05, 0.1, 0.2, 0.3, subtables with the modified sample size (n_m) and γ values of 0.90 and 0.99, each for CI coverage of 95 % and 99 % (Table 2). Each condition is crossed with all other conditions in a factorial manner; thus there are a total of 240 situations for planning an appropriate sample size.

Suppose that a researcher is interested in estimating the animal prevalence (p) of avian influenza in the south of México. To determinate the sample size we need an estimate of p . After doing a literature review of other studies for estimating this disease, it is hypothesized that $p=0.1$, a CI of 95 %, $ro=0.196$ (the intra cluster correlation), $b=7$ (average number of individuals measured per cluster) and it is assumed that the desired final CIRW is $w_r = [(p_U - p_L) / p] \leq re = 0.1$. The application of the method leads to a preliminary sample size of $n_p = 14026$. This sample size is contained in the first sub-table (n_p without γ , $\gamma=0.5$) of Table 2 where $p=0.1$, and $re=0.1$. Realizing that $n_p = 14026$ will lead to a sufficiently narrow CI only about 50 % of the time, the researcher incorporates an assurance of $\gamma=0.99$, which implies that the width of the 95 % CI will be larger than required (i.e., 0.1) no more than 1 % of the time. From the third sub-table of Table 2 (n_m with $\gamma=0.99$), it can be seen that the modified sample

TAMAÑOS DE MUESTRA PARA ESTIMAR PREVALENCIA ANIMAL

Cuadro 2. Tamaño de muestra requerido (n_p preliminar y n_m modificado) para 95% y 99% de confianza. La proporción en la población es p , γ es el nivel de aseguramiento deseado para que la ARCIC para p no sea más grande que re ; re es la ARCIC deseada (error relativo), n_m es el tamaño de muestra modificado cuando γ es especificado

Table 2. Required sample size (n) for 95% and 99% CI. The proportion in the population is p , γ is the desired degree of certainty of achieving a CI for p that is no wider than re ; re is the desired relative CIRW, n_m is the required modified sample size when γ is specified

IC=95%												
n_p without γ ($\gamma \approx 0.5$)					n_m with $\gamma = 0.90$				n_m with $\gamma = 0.99$			
p	0.05	0.1	0.2	0.3	0.05	0.1	0.2	0.3	0.05	0.1	0.2	0.3
0.05	117577	29592	7496	3374	119370	30483	7937	3662	120820	31194	8280	3889
0.1	55714	14026	3554	1600	56509	14420	3747	1729	57154	14736	3899	1826
0.15	35093	8838	2240	1009	35558	9069	2353	1082	35932	9250	2442	1138
0.2	24782	6243	1583	713	25080	6391	1656	760	25321	6507	1711	796
0.25	18596	4687	1189	536	18795	4784	1238	567	18952	4862	1274	589
0.3	14472	3649	927	417	14604	3714	959	439	14708	3764	981	453
0.35	11526	2908	739	333	11610	2949	760	346	11677	2980	772	354
0.4	9316	2352	598	270	9366	2376	610	277	9402	2393	616	280
0.45	7598	1919	489	220	7619	1930	493	223	7634	1935	494	223
0.5	6223	1573	401	181	6224	1574	402	182	6224	1574	402	182

IC=99%												
n_p without γ ($\gamma \approx 0.5$)					n_m with $\gamma = 0.90$				n_m with $\gamma = 0.99$			
p	0.05	0.1	0.2	0.3	0.05	0.1	0.2	0.3	0.05	0.1	0.2	0.3
0.05	202500	50824	12805	5735	204867	52002	13380	6120	206774	52938	13840	6416
0.1	95939	24083	6068	2717	96991	24606	6325	2886	97834	25021	6525	3020
0.15	60419	15169	3823	1712	61029	15473	3974	1811	61524	15716	4090	1888
0.2	42659	10712	2700	1209	43053	10907	2796	1273	43369	11063	2870	1320
0.25	32003	8038	2026	907	32266	8168	2091	949	32474	8269	2140	980
0.3	24899	6255	1577	706	25074	6341	1619	734	25212	6408	1651	753
0.35	19825	4982	1256	562	19937	5037	1284	580	20025	5079	1302	591
0.4	16019	4027	1016	455	16084	4058	1032	465	16135	4081	1041	470
0.45	13059	3284	829	371	13088	3298	835	375	13108	3306	838	375
0.5	10691	2689	679	304	10692	2690	680	305	10692	2690	680	305

propuestos, y discute como obtener ICs suficientemente angostos para cualquier combinación de p , re , γ , α , ro , b , usando el paquete R⁽¹⁷⁾, también se proporcionan cuadros para algunas condiciones. Los cuadros no incluyen todas las condiciones potencialmente interesantes; en lugar de esto se pretende que proporcionen al investigador:

size procedure yields a necessary sample size of $n_m = 14736$. Using a sample size of 14,736 will thus provide 99 % assurance that the obtained CI for p will be no wider than 0.1 units, but under SRS. However, because the selection is required under cluster sampling the $Deff = 1 + (b-1)ro = 1 + (7-1)0.196 = 2.34$. Therefore, the number of

una forma conveniente de planear el tamaño de muestra cuando la situación de interés es aproximada por los escenarios cubiertos por las tablas; y también una forma de ilustrar la relación entre p , re , γ , α , ro , b , y el tamaño de muestra (n) y número de conglomerados, C , requeridos. Los valores en el Cuadro 2 se basan en condiciones que deberían ser útiles para detectar y calcular la proporción de interés para algunos valores de p .

Los tamaños de muestra se presentan para valores p de 0.05 a 0.5 con incrementos de 0.05, para valores re de 0.05, 0.1, 0.2, 0.3, para valores de γ de 0.5 0.90 y 0.99, para los cuales se presentan o sub-cuadros (tamaño de muestra preliminar, n_p , sin aseguramiento ($\gamma=0.5$) y tamaños de muestra modificado, n_m , para valores de de 0.90 y 0.99, cada uno para coberturas del IC de 95 y 99 %) (Cuadro 2). Cada condición se cruza con todas las otras condiciones de una manera factorial, resultando un total de 240 situaciones para planear un tamaño de muestra.

Suponiendo que un investigador está interesado en calcular la prevalencia animal (p) de la influenza aviar en el sur de México. Para determinar el tamaño de la muestra se necesita una estimación de p . Después de hacer una revisión de literatura de otros estudios, para la estimación de la prevalencia de esta enfermedad, se hipotetiza que $p=0.1$, el IC de 95%, $ro=0.196$ (coeficiente de correlación intraconglomerados), $b=7$ (promedio de individuos a extraer por conglomerado) y se asume que la ARCIC final deseada es $w_r = [(p_U - p_L) / p] \leq re = 0.1$. La aplicación del método conduce a una muestra preliminar de $n_p = 14,026$. Este tamaño de muestra está contenido en el primer sub-cuadro (n_p sin γ ; $\gamma=0.5$) del Cuadro 2 donde $p=0.1$, y $re=0.1$. Tomando en cuenta que $n_p = 14,026$ conducirá a un IC relativamente estrecho solamente el 50 % de las ocasiones, el investigador incorpora un nivel de aseguramiento de $\gamma=0.99$, lo cual implica que la amplitud del IC relativo de 95 % será más grande que lo requerido ($re=0.1$) no más del 1 % por ciento de las veces. Del tercer sub-cuadro del Cuadro 2 (n_m con $\gamma=0.99$), se puede

clusters (ranches) required from the population is equal to $C = [(Deff * n_m) / b] = [(2.34 * 14736) / 7] = 4927$, that is, 4,927 clusters are required to assure a 99 % probability that the resulting CI width is no wider than 0.1 units. Similarly, in the lower part Table 2 shows the required sample size for a CI of 99 %, but these are under SRS. If we want to use cluster sampling we need to multiply these by the design effect and divide between the average numbers of individuals per cluster (b).

To be sure that the reader understand very well the proposed procedure to get the required sample size we will present another example. Suppose a researcher it is interested in estimating p for adventitious present of unwanted transgenic plants (AP) maize in the region of Oaxaca, Mexico, where Quist and Chapela^(6,7) reported finding AP maize. First of all, it is necessary to corroborate this information and estimate p . With this information and after doing a literature review of other studies, it is hypothesized that $p=0.05$, $b=20$, $ro=0.35$. We want a CI of 99 % and we assume that the final CIRW is $w_r = [(p_U - p_L) / p] \leq re = 0.2$. The application of the method leads to a required (preliminary) sample size of $n_p = 12,805$ (under SRS). This sample size is contained in the second sub-table (n_p without γ ; $\gamma=0.5$;) of Table 2, where $p=0.05$ and $re=0.2$.

Realizing that $n_p = 12,805$ will lead to a sufficiently narrow CI only about half of the, the researcher incorporates an assurance of $\gamma=0.99$, which implies that the width of the 99 % CI will be larger than required (i.e., 0.2) no more than 1 % of the time. From the second sub-table of Table 2 (n_m with $\gamma=0.99$), it can be seen that the modified sample size procedure yields a sample size $n_m = 13,840$ (under SRS). Using a sample size of 13,840 will thus provide 99 % assurance that the obtained CI for p will be no wider than 0.2 units if the test is perfect.

However, since it is impractical to use SRS, the selection will be under cluster sampling. Therefore, the $Deff = 1 + (b-1)ro = 1 + (20-1)0.35 = 7.65$. This

apreciar que el tamaño de muestra modificado requerido es de $n_m = 14736$, pero este tamaño de muestra es con muestreo aleatorio simple (MAS). Sin embargo, dado que se requiere la selección con muestreo por conglomerados el $Deff = 1 + (b-1)ro = 1 + (7-1)0.196 = 2.34$. Por lo tanto, el número de conglomerados (que pueden ser ranchos) a seleccionar de la población es igual a $C = [(Deff * n_m) / b] = [(2.34 * 14736) / 7] = 4927$, es decir, se requieren 4,927 conglomerados para garantizar con una probabilidad de 99 % de que el IC relativo obtenido para será no más ancho que 0.1 unidades. De manera similar, en la parte inferior del Cuadro 2 se muestra el tamaño de muestra requerido para un IC de 99 %, pero estos son con MAS, y si se desean con muestreo por conglomerados hay que multiplicar estos por el efecto de diseño y dividirse entre el promedio de individuos por conglomerado (b).

Para mejorar la comprensión del procedimiento propuesto, se presenta otro ejemplo. Suponiendo que un investigador está interesado en estimar la proporción (p) de plantas de maíz transgénico (maíz AP) en la región de Oaxaca, México, donde Quist y Chapela^(18,19) reportaron encontrar maíz AP. Después de una revisión de literatura, se hipotetiza que $p=0.05$, $b=20$, $ro=0.35$, se desea un IC de 99 % y se asume que la ARIC es $w_r = [(p_U - p_L) / p] \leq re = 0.2$. La aplicación del método propuesto conduce a un tamaño de muestra preliminar de $n_p = 12,805$ (bajo MAS). Este tamaño de muestra está contenido en el segundo sub-cuadro (n_p sin γ ; $\gamma=0.5$) del Cuadro 2, donde $p=0.05$ y $re=0.2$.

Tomando en cuenta que $n_p = 12,805$ conducirá a un IC suficientemente angosto, solamente la mitad de las veces el investigador incorpora un nivel de aseguramiento de $\gamma=0.99$, lo cual implica que la anchura del IC relativo de 99 % será más amplio que lo requerido (i.e., 0.2) no más que el 1 % de las veces. Del segundo sub-cuadro del Cuadro 2 (n_m con $\gamma=0.99$), se puede ver que el tamaño de muestra modificado es $n_m = 13,840$ (también con MAS) que proporcionará un 99 % de seguridad de que el IC obtenido por p no será más ancho que

implies that $C = [(Deff * n_m) / b] = [7.65 * 13840] / 20 = 5294$ clusters are required (which may be ranchos) to ensure a 99 % probability that the final CI is no wider than 0.2 units.

With Table 2 you can obtain the required sample size that ensures narrow confidence intervals under SRS. However, this Table (Table 2) only included some conditions (only some values of p) for getting sample sizes. For this reason, we propose an R program (Appendix) where the researcher can get other scenarios. If the selection scheme is under cluster the provided sample sizes (under SRS) must be multiplied by the design effect and divided between b (average number of individuals per cluster) for getting the required number of clusters.

To be confident that the reader can get the required number of clusters, Table 3 presents the intra cluster correlation (ro) and design effects ($Deff$) for 14 serovars of *Leptospira* for Yucatan Mexico reported by Segura-Correa *et al*⁽¹⁵⁾. Then for each serovar it was calculated the sample size assuming normality using the traditional expression $n_T = [4(Z^2 * q) / (re^2 * p)]$ with $p=0.2$, $re=0.3$ and a CI of 95 % ($Z=1.96$). Subsequently, using the proposed method (AIPE approach) it was calculated for each serovar the preliminary sample sizes (n_p sin γ ; $\gamma=0.5$), modified sample sizes with assurance levels of 90 % ($\gamma=0.90$) and 99% ($\gamma=0.99$) (assuming a proportion of 0.2, a relative error of 0.3 and a CI of 95 %). Here it should be stressed again that the traditional sample size ($n_T = [4(Z^2 * q) / (re^2 * p)]$) is less than the preliminary and modified, which clearly shows that the expression for calculating the traditional sample size produces a serious underestimation. Finally, for each of these four sample sizes it is reported the number of required clusters of size 7 that were calculated with the expression $C = [(Deff * n_i) / b]$ for $i = T, p, m1$ and $m2$. It is also noted that the number of clusters under the traditional approach (C_T) underestimates the number of clusters required.

In conclusion, the method is useful for planning sample size. Although the method proposed in this

Cuadro 3. Número de conglomerados requeridos para cada serovariedad. Se asume para todas las serovariedades la proporción $p = 0.2$, el error relativo $re = 0.5$ y nivel de confianza de 95%, ro es el coeficiente de correlación intraclase. $Deff = 1 + (b-1)ro$ es el efecto de diseño y b es el número promedio de individuos medidos por conglomerado ($b=7$); n_T es el tamaño de muestra tradicional con la aproximación normal $n_T = [4(Z^2 * q)/(re^2 * p)]$, n_p es el tamaño de muestra preliminar calculado con la ecuación (3) y sin aseguramiento ($\gamma \approx 0.5$), n_{m1} y n_{m2} son los tamaños de muestra modificados calculados con el procedimiento propuesto de dos pasos (calculados con las ecuaciones 3 y 5) con niveles de aseguramiento de 90% ($\gamma = 0.9$) y 99% ($\gamma = 0.99$). C_T, C_p, C_{m1}, C_{m2} representan el número de conglomerados requeridos con el método tradicional (CT) y con el método propuesto de dos pasos sin aseguramiento (C_p) y con 90 y 99% de aseguramiento (C_{m1}, C_{m2}). Los ro y $Deff$ para las 14 serovariedades de *Leptospira* en Yucatán, México son reportados por Segura-Correa *et al.*(15)

Table 3. Number of clusters required for each serovar. It is assumed for all serotypes the proportion $p = 0.2$, the relative error $re = 0.5$ and a confidence level of 95%. ro is the intra cluster correlation. $Deff = 1 + (b-1)ro$ is the design effect and b is the average number of individuals measured per cluster ($b = 7$); n_T is the traditional sample size with normal approximation $n_T = [4(Z^2 * q)/(re^2 * p)]$, n_p is the preliminary sample size calculated with the equation (3) and without assurance ($\gamma \approx 0.5$), n_{m1} and n_{m2} are the modified sample sizes calculated with the two-step procedure proposed (calculated with equations 3 and 5) with assurance levels of 90 % ($\gamma = 0.9$) and 99 % ($\gamma = 0.99$). C_T, C_p, C_{m1}, C_{m2} represent the number of clusters required under the traditional method (CT) and the proposed two-step method without assurance (CP) and 90 and 99 % assurance (, Cm1, Cm2). The ro and $Deff$ for the 14 serovars of *Leptospira* in Yucatan, Mexico are reported by Segura-Correa *et al.* (2002)

Serovar	ro	$Deff$	n_T	n_p	n_{m1}	n_{m2}	C_T	C_p	C_{m1}	C_{m2}
Javanica	0.5192	4.58	683	713	760	796	447	467	498	521
Wolffi	0.317	3.18	683	713	760	796	311	324	346	362
Hardjo	0.2809	2.94	683	713	760	796	287	300	320	335
Tarassovi	0.2803	2.93	683	713	760	796	286	299	319	334
Hebdomadis	0.2689	2.85	683	713	760	796	279	291	310	325
Panama	0.2679	2.84	683	713	760	796	278	290	309	323
Bratislava	0.2633	2.81	683	713	760	796	275	287	306	320
Botavia	0.2085	2.44	683	713	760	796	239	249	265	278
Canicola	0.1961	2.35	683	713	760	796	230	240	256	268
Pomona	0.196	2.34	683	713	760	796	229	239	255	267
Grippothyphosa	0.157	2.08	683	713	760	796	203	212	226	237
Sherman	0.1235	1.85	683	713	760	796	181	189	201	211
Icterohamorrhagiae	0.1029	1.71	683	713	760	796	167	175	186	195
Pyrogen	0.0985	1.68	683	713	760	796	164	172	183	192

0.2 unidades. Sin embargo, dado que es impráctico usar MAS, la selección se realizará con muestreo por conglomerados. Por lo tanto, el $Deff = 1 + (b-1)ro = 1 + (20-1)0.35 = 7.65$. Esto implica que se requieren $C = [(Deff * n_m)/b] = [7.65 * 13840]/20 = 5294$ conglomerados (que pueden ser predios) para garantizar con una probabilidad de 99 % de que el

paper guarantees precision in the estimation of p , it cannot be solved in an analytical way; that is why tables of the required sample size are provided for a variety of scenarios that may help researchers planning an appropriate sample that guarantee a narrow CI in the estimation of a binomial proportion. However, it is not possible to cover all

IC relativo obtenido para p será no más ancho que 0.2 unidades.

Con el Cuadro 2 se puede obtener el tamaño de muestra requerido que asegura cortos intervalos de confianza con MAS, pero, este cuadro incluye solamente algunas condiciones (solamente algunos valores de p) para la obtención de tamaños de muestra. Por esta razón, se propone un programa en R (Apéndice), donde el investigador puede obtener tamaños de muestra para otros escenarios, y si el esquema de selección es por conglomerados, los tamaños de muestra proporcionados se tienen que multiplicar por el efecto de diseño y dividirse entre b (promedio de individuos por conglomerado) para obtener el número de conglomerados requeridos.

Para afianzar como obtener el número de conglomerados requeridos. En el Cuadro 3 se presentan los coeficientes de correlación intraconglomerados (ro) y efectos de diseño ($Deff$) para 14 serovariedades de *Leptospira* para Yucatán México reportados por Segura-Correa *et al*⁽¹⁵⁾. Los efectos de diseño se obtienen con la expresión: $Deff = 1 + (b-1)ro$, asumiendo un valor de $b=7$ (número promedio de animales medidos por rancho). En seguida para cada serovariedad se calcula el tamaño de muestra tradicional con la expresión $n_T = [4(Z^2 * q)/(re^2 * p)]$ asumiendo $p=0.2$, $re=0.3$ y un IC de 95% ($Z=1.96$). Posteriormente, con el método propuesto (enfoque AIPE) se calculan para cada serovariedad los tamaños de muestra preliminar (n_p sin γ ; $\gamma=0.5$), tamaños de muestra modificados con niveles de aseguramiento de 90 % ($\gamma=0.90$) y 99 % ($\gamma=0.99$) de igual manera asumiendo una proporción de 0.2, un error relativo de 0.3 y un IC de 95 %. Aquí se puede resaltar nuevamente que el tamaño de muestra tradicional (n_T) es menor que el preliminar y modificado, lo que muestra claramente que la expresión ($n_T = [4(Z^2 * q)/(re^2 * p)]$) para el cálculo del tamaño de muestra tradicional produce una seria subestimación. Finalmente, para cada uno de estos cuatro tamaños de muestra se reporta el número de conglomerados requeridos de tamaño 7

possible combinations of ($p, re, \gamma, \alpha, ro, b$); this is why appropriate programs in R package⁽¹⁷⁾ can be applied to allow the user to quickly and simply plan the sample size according to her/his requirements or needs (Appendix). Finally, it is important to clarify that the methods presented here assume a perfect test (sensitivity=specificity=1) and that the sample selection is under simple random sampling, which must be taken into account when designing a study. Also, if you want to use cluster sampling the user must know or estimate the value of the intra cluster correlation (ro) from a pilot study and the value of b (average number of individuals measured per cluster).

End of english version

que se calculan con la expresión $C = [(Deff * n_i) / b]$, para $i=T, p, m1$ y $m2$. También se observa que el número de conglomerados con el enfoque tradicional (C_T) subestima el número de conglomerados requeridos.

Se concluye que el método propuesto es útil para planear tamaños de muestra, cuando se desea precisión en la estimación de la proporción, y aunque garantiza precisión en la estimación de p , no puede ser resuelto en forma analítica; es por esto que se proporcionan cuadros con tamaños de muestra para algunos escenarios, que podrían ayudar a los investigadores a planear su tamaño de muestra que les garantice un reducido IC en la estimación de una proporción binomial. No obstante, no es posible cubrir todas las combinaciones posibles de ($p, re, \gamma, \alpha, ro, b$); es por esto que se realizó un programa en el paquete R⁽¹⁷⁾ para que el usuario de una forma rápida y sencilla pueda calcular el tamaño de muestra que requiere de acuerdo a sus necesidades (Apéndice). Es importante aclarar que el método presentado supone una prueba perfecta (sensibilidad=especificidad=1) y que si se desea usar muestreo por conglomerados el usuario debe de conocer o estimar con una muestra piloto el valor del coeficiente de correlación intraconglomerados (ro)

y además especificar el valor de b (promedio de individuos a medir por conglomerado).

LITERATURA CITADA

1. Salman MD. Animal disease surveillance and survey systems: methods and applications, Iowa USA: Iowa State Press; 2003.
2. Montesinos-López OA, Hernández-Suárez CM, Sáez RA, Luna-Espinoza I. Modelo estocástico para determinar la mejor estrategia de cuarentena en granjas ganaderas. *Tec Pecu Méx* 2009;47(3):271-284.
3. Hernández-Suárez CM, Montesinos-López OA, McLaren G, Crossa J. Probability models for detecting transgenic plants. *Seed Sci Res* 2008;18:77-89.
4. Fosgate GT. A cluster-adjusted sample size algorithm for proportions was developed using a beta-binomial model. *J Clin Epidemiol* 2007;60:250-255.
5. Wang H, Chow SC, Chen M. A bayesian approach on sample size calculation for comparing means. *J Biopharm Stat* 2005;15:799-807.
6. Pan Z, Kupper L. Sample size determination for multiple comparison studies treating confidence interval width as random. *Stat Med* 1999;18:1475-1488.
7. Newcombe RG. Two-sided CIs for the single proportion: comparison of seven methods. *Stat Med* 1998;17:857-872.
8. Kelley K, Maxwell SE, Rausch JR. Obtaining power or obtaining precision: Delineating methods of sample size planning. *Eval Health Prof* 2003;26:258-287.
9. Kelley K, Maxwell SE. Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychol Methods* 2003;8:305-321.
10. Kelley K, Rausch JR. Sample size planning for the standardized mean difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychol Methods* 2006;11:363-385.
11. Mace AE. Sample size determination. New York: Reinhold; 1964.
12. Kupper LL, Hafner KB. How appropriate are popular sample size formulas? *Am Stat* 1989;43:101-105.
13. Wang Y, Kupper LL. Optimal sample sizes for estimating the difference in means between two normal populations treating confidence interval length as a random variable. *Commun Stat* 1997;26:727-741.
14. Vollset SE. CIs for a binomial proportion. *Stat Med* 1993;12:809-824.
15. Segura-Correa JC, Segura-Correa VM, Solís-Calderón JJ. Efecto de diseño para 14 serovariedades de *Leptospira* en Yucatán, México. *Rev Biomed* 2002;13:33-36.
16. Montesinos-López OA, Luna-Espinoza I, Hernández-Suárez CM, Tinoco-Zermeño MA. Muestreo Estadístico. Tamaño de muestra y estimación de parámetros. Colima, México: Editorial Universidad de Colima; 2009.
17. R Development Core Team. *R: A language and environment for statistical computing* [Computer software and manual], R Foundation for Statistical Computing 2007. Retrieved from www.r-project.org.
18. Quist D, Chapela IH. Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. *Nature* 2001;414:541-543.
19. Quist D, Chapela IH. Quist and Chapela reply. *Nature* 2002;416:602.

APÉNDICE

Para calcular el tamaño de muestra apropiado hemos desarrollado el paquete SamBin en R. El paquete SamBin R esta disponible en la siguiente dirección de internet http://www.carloscruz.org.mx/portal/wp-content/SamBin_1.0.zip. Puesto que SamBin es un paquete opcional, debe de ser cargado con cada nueva sesión en R en donde las rutinas serán usadas. El Paquete R es cargado con el comando `library()` el cual es ilustrado con SamBin como sigue:

```
R> library (SamBin)
```

Usando el ejemplo del Cuadro 2, donde $p = 0.05$ y $re = 0.2$ para un IC de 95%, la función `ss.bin.aipe()` puede ser usada para determinar el tamaño de muestra que asegura que la amplitud relativa completa del IC es suficientemente reducida y se obtiene con:

```
R> ss.bin.aipe(p=0.05, conf.level=0.95, re=0.2),
```

Donde p es la prevalencia poblacional (i.e., p), *conf.level* es el nivel de confianza (i.e., $1 - \alpha$), y re es la ARCIC deseada. La implementación de esta función proporciona un tamaño de muestra de 12,805 (como se reporta en el texto y en el Cuadro 2).

El nivel de aseguramiento deseado puede ser incorporado en la función anterior especificando el nivel de certeza requerido, el cual es un argumento opcional en la función `ss.bin.aipe()`. La función deberá ser especificada como:

```
R> ss.bin.aipe(p=0.05, conf.level=0.95, re=0.2, assurance=0.99).
```

Esto regresa el tamaño de muestra requerido, 13,840 (como se mostró en el texto y en el Cuadro 2), que provee un 99 % de seguridad de que la ARCIC obtenida para p no será mayor que 0.2 unidades. Estos tamaños de muestra son con MAS y si el investigador lo requiere en términos de conglomerados los debe de ajustar para obtenerlo en términos de conglomerados con: $C = [(Deff * n_m) / b]$ donde n_m es el tamaño de muestra con AIPE (método propuesto y con el paquete R explicado anterior). Además, para calcular el *Deff* el investigador debe de proporcionar el promedio de individuos a medir por conglomerados (b) y el coeficiente de correlación intraconglomerados (ro) $Deff = 1 + (b-1)ro$.

APPENDIX

To calculate the appropriate sample size we develop the package SamBin in R. The SamBin R package is available at the following Internet address http://www.carloscruz.org.mx/portal/wp-content/SamBin_1.0.zip. Because SamBin is an optional package, it must be loaded with each new R session where its routines will be used. Packages in R are loaded with the `library()` command, which is illustrated with SamBin as follows:

```
R> library (SamBin)
```

Using the example in Table 2, where $p = 0.05$ and $re = 0.2$ for a 95 % CI, the `ss.bin.aipe()` function can be used so that the relative width is sufficiently narrow and given as:

```
R> ss.bin.aipe(p=0.05, conf.level=0.95, re=0.2),
```

Where p is the population prevalence (i.e., p), *conf.level* is the confidence level (i.e., $1 - \alpha$), and re is the desired CIRW. Implementation of this function yields 12,805 as the necessary sample size (as reported in the text and in Table 2).

The desired degree of certainty can be used in the function by specifying the certainty, which is an optional argument in the `ss.bin.aipe()` function. The function would thus be specified as:

```
R> ss.bin.aipe(p=0.05, conf.level=0.95, re=0.2, assurance=0.99).
```

This returns the required sample size, 13,840 (as reported in the text and in Table 2), that provides 99 % assurance that the obtained CIRW for p will be no wider than 0.1 units. These sample sizes are under simple random sampling and if the researcher requires it, in terms of clusters it must be adjusted with: where n is the sample size under AIPE (proposed method and the R package explained above). Moreover, to calculate the *Deff* the investigator must provide the average number of individuals to measure cluster (b) and the intra cluster correlation coefficient (ro) $Deff = 1 + (b-1)ro$.