

Prueba de grupo. Una eficiente alternativa para estimar prevalencia animal

Group testing. An efficient alternative for estimating animal prevalence

Osva Antonio Montesinos López^a, Abelardo Montesinos López^b, Ignacio Luna Espinoza^c,
Laura Sanely Gaytán Lugo^d, Teodoro Espinosa Solares^e

RESUMEN

La estimación de prevalencia animal o detección de agentes infecciosos, es de vital importancia para una nación o estado para proteger su seguridad alimentaria y garantizar el comercio local y exterior. Sin embargo, realizar estas dos tareas (estimación y detección) eficientemente requiere de significativos recursos materiales y humanos. Por ello, el presente artículo presenta una revisión del proceso de estimación y detección usando la técnica estadística conocida como prueba de grupo (group testing) que puede contribuir al ahorro de recursos cuantiosos para la estimación y detección de enfermedades infecciosas en ciencia animal cuando la prevalencia es menor o igual al 10%. Group testing es una ingeniosa técnica estadística propuesta por Dorfman (1943). Esta técnica consiste en juntar el material de k individuos, mezclarlos perfectamente y en lugar de realizar una prueba de laboratorio para cada individuo, se realiza una sola prueba con la mezcla de los k individuos. De esta forma se logran ahorros significativos. Por ello, se muestra el funcionamiento de algunos métodos para clasificar individuos; se expone el proceso de estimación puntual y por intervalo de prevalencia animal; y se ilustra el cálculo del tamaño de muestra bajo este modelo, con la finalidad de que los investigadores en ciencia animal la utilicen y ahorren recursos significativos.

PALABRAS CLAVE: Prueba de grupo, Prevalencia animal, Tamaño de muestra, Estimación, Detección.

ABSTRACT

An estimate of animal prevalence or the detection of infectious agents is of vital importance to a country or state to protect food safety and to safeguard local and foreign trade. However, significant human and material resources are required to perform these two tasks. This article presents an overview of the estimate and detection processes with the statistical technique known as group testing. In animal science, group testing can help save substantial resources in the estimate and detection of infectious diseases when the prevalence is less than or equal to 10%. Group testing is a sound statistical technique proposed by Dorfman (1943). It consists of mixing the material of k individuals. Instead of a diagnostic test for each individual, only one test is performed with the resulting mixture of the material. This allows significant savings. Furthermore, this paper presents some methods to classify individuals, it describes the process of point and interval estimation of animal prevalence, and it illustrates the calculation of sample size according to this model. The intention is to help researchers in animal science to use this technique and save significant resources.

KEY WORDS: group testing, animal prevalence, sample size, estimate, detection.

El método de prueba de grupo (group testing) fue propuesto por Dorfman⁽¹⁾ para detectar individuos con sífilis durante la segunda guerra mundial. La

During World War II, Dorfman⁽¹⁾ proposed Group testing to detect individuals who suffered from syphilis. This technique consists essentially of

Recibido el 3 de Junio de 2010. Aceptado el 17 de Octubre de 2011.

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

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

c Universidad del Istmo, Campus Ixtepac.

d Facultad de Ingeniería Mecánica y Eléctrica. Universidad de Colima.

e Departamento de Ingeniería Agroindustrial. Universidad Autónoma Chapingo.

esencia de esta técnica consiste en mezclar el material (sangre en el caso de Dorfman) de k individuos y sobre esta mezcla realizar una sola prueba de laboratorio. Si la prueba del grupo (mezcla de k individuos) resulta positiva, entonces se realizan pruebas individuales a los individuos dentro del grupo positivo para identificar a los individuos enfermos. Por otro lado, si la prueba de grupo resulta negativa, se concluye que los k individuos que forman el grupo están libres de dicha enfermedad. Es importante mencionar que la selección de los individuos para formar los grupos se hace de manera aleatoria (Figura 1).

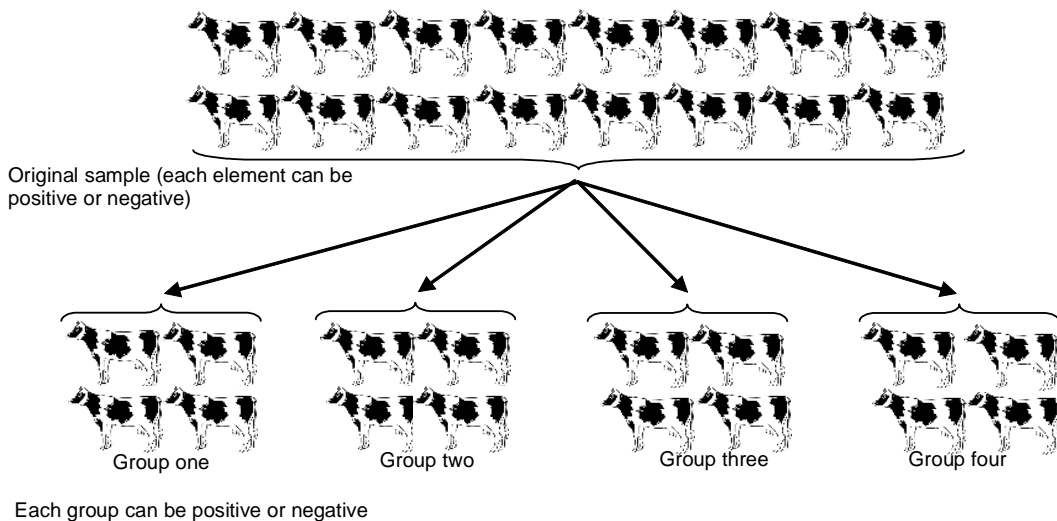
La prueba de grupo se ha utilizado para estimar y detectar diversas enfermedades, como el virus de inmunodeficiencia humana (VIH), hepatitis B y C, y el virus del Nilo (*West Nile virus*)⁽²⁾. También se ha usado para detectar enfermedades en la donación de sangre⁽³⁾; para detectar drogas⁽⁴⁾; para

mixing the material (in the case of the Dorfman research, blood) of k individuals and performing a single diagnostic test on this mixture. If the result for a group is positive for the mixture of k individuals, the people from this group are tested individually to identify those who are sick. If the result is negative, the conclusion is that the k individuals from that group are healthy. It is important to mention that the individuals that form the group are selected randomly (Figure 1).

Group testing has been used to estimate and detect different diseases, such as human immunodeficiency virus, HIV, Hepatitis B and C, and West Nile Virus⁽²⁾. It has also been used to detect diseases before blood donation⁽³⁾, to detect drugs⁽⁴⁾, to estimate human⁽⁵⁾, plant⁽⁶⁾ and animal⁽⁷⁾ disease prevalence and to detect them⁽⁵⁾, to detect transgenic plants^(8,9), to solve problems in information theory⁽¹⁰⁾ and even in science fiction⁽¹¹⁾.

Figura 1. Con la muestra original se realizarían 16 pruebas de laboratorio, una por cada animal. Con la prueba de grupo se realizarían cuatro pruebas, ahorrando 75% de los recursos sin pruebas adicionales. Si un grupo es negativo, los cuatro elementos (animales) no tienen la enfermedad. Si un grupo es positivo, al menos un elemento tiene la enfermedad. La selección de elementos para formar los grupos es aleatoria.

Figure 1. With the original sample 16 animals are tested individually. With the group testing method only 4 diagnostic tests are required. This means 75% savings in resources without additional testing. If a group is negative, the four items (animals) do not suffer from the disease. If a group is positive, at least one animal has the disease. The items that form the groups are selected randomly.



estimar la prevalencia y detección de enfermedades humanas⁽⁵⁾, de plantas⁽⁶⁾ y de animales⁽⁷⁾; para detectar plantas transgénicas^(8,9); para resolver problemas de teoría de la información⁽¹⁰⁾ y hasta de ciencia ficción⁽¹¹⁾.

La gran aceptación de esta técnica obedece principalmente al ahorro significativo de recursos materiales y económicos que produce. En el contexto de enfermedades infecciosas, la prueba de grupo es típicamente usada para: 1) detectar o identificar individuos que tienen cierta enfermedad; y 2) para estimar la prevalencia de enfermedades, es decir, para estimar la proporción p de individuos en la población que tienen una enfermedad específica⁽¹²⁾.

A pesar de la amplia gama de aplicaciones y ventajas por usar la prueba de grupo, ésta es subutilizada en los países en vías de desarrollo. En los países desarrollados, su uso en los laboratorios de salud pública es reciente. Por ejemplo, sólo 12% de los laboratorios de Estados Unidos reportan el empleo de esta técnica⁽¹³⁾. También se reporta que más de 15 % de las bases de datos utilizadas para detectar expresiones genéticas utilizan pruebas de grupo⁽¹⁴⁾. La subutilización de la prueba de grupo obedece básicamente a que la mayor parte de las aplicaciones son altamente técnicas y teóricas, lo cual ha restringido su aplicación directa por directores de laboratorios y técnicos laboratoristas⁽²⁾.

Este trabajo hace una revisión sobre la prueba de grupo, resaltando sus ventajas y desventajas. También se presentan algunos métodos para clasificar individuos, los estimadores apropiados para el proceso de estimación puntual y por intervalo de la prevalencia de enfermedades, la determinación del tamaño de muestra, y las restricciones bajo las cuales la prueba de grupo no produce buenos resultados.

Cuando la prueba de grupo se utiliza para detectar enfermedades y el grupo resulta positivo, entonces

This technique is broadly accepted mainly due to the significant material and financial savings it allows. In the context of contagious diseases, group testing is typically used to 1) detect or identify sick individuals, and to 2) estimate disease prevalence, i.e., to estimate the proportion (p) of individuals in a population that suffer from a specific disease⁽¹²⁾.

Despite the broad range of group testing applications and advantages, it is underused in developing countries. In developed countries, its use in public health laboratories is recent. For instance, only 12% of the laboratories in the USA report using this technique⁽¹³⁾.

Group testing is employed in over 15% of the data bases used to detect genetic expressions⁽¹⁴⁾. Lab directors and technicians have not taken advantages of group testing mainly due that most applications are highly technical and theoretical⁽²⁾.

This paper review the studies on group testing and highlights its advantages and drawbacks. It presents some methods to classify individuals, the appropriate estimators for the point estimate and the interval estimate of disease prevalence, and explains the determination of the sample size. Some constraints, where in certain situations group testing does provide good results, are also presented.

When group testing is used to detect diseases, if the group is found to be positive, then other diagnostic tests are performed on the individuals who belong to this group. This test is on individuals or on sub-groups until all the sick individuals are detected. Therefore, If the group is positive, then it is required retesting to identify the individuals with positive results. However, when the purpose is just to estimate the prevalence there is no retesting on either the individuals or the subpools that belong to the group with positive results.

There are two kinds of algorithms to detect or classify positive individuals: the hierarchical and the matrix arrays. The Dorfman model, the halving

a los individuos que conforman dicho conjunto se les realiza una prueba de laboratorio en forma individual, o en subgrupos hasta identificar a todos los individuos que padecen la enfermedad; es decir, si el grupo resulta positivo, entonces se requieren de pruebas de laboratorio adicionales ("retesting") para identificar a los individuos positivos. Por otro lado, cuando el prop3sito es s3lo la estimaci3n de la prevalencia, si el grupo resulta positivo, entonces ya no se realizan pruebas adicionales a los individuos o subgrupos que conforman al grupo positivo.

Los algoritmos para detectar o clasificar individuos positivos se pueden dividir en jer3rquicos y arreglos matriciales. Dentro de los jer3rquicos se tienen el modelo de Dorfman, el m3todo de Sterrett's, y el m3todo "halving".

El procedimiento de Dorfman realiza una prueba individual a cada elemento dentro de un grupo que result3 positivo. Por lo tanto, si un grupo de tama1o k resulta positivo, entonces se requerir3n $k+1$ pruebas de laboratorio para clasificar a todos los individuos que conforman al grupo positivo; pero si el grupo resulta negativo, entonces se declaran libres de la enfermedad a los k individuos que lo conforman.

El n3mero relativo esperado de pruebas de laboratorio individuales (X) usando el procedimiento de Dorfman, es igual a $E(X) = \frac{E(Y)}{n} = \frac{1}{k} + P$; donde $E(Y)$ es el n3mero esperado de pruebas requeridas usando pruebas por grupo, n (tama1o de muestra) es el n3mero de pruebas requeridas haciendo pruebas individuales, P denota la probabilidad de que un grupo sea positivo y se determina por $P = (1 - Sp)(1 - p)^k + Se(1 - (1 - p)^k)$, Se y Sp son la sensibilidad y especificidad de la prueba de laboratorio, respectivamente, y p es la prevalencia de la enfermedad en la poblaci3n. Por lo tanto, el porcentaje de ahorro esperado realizando pruebas adicionales es igual a $100(1 - E(X))\%$, mientras que sin pruebas adicionales es igual a $100(1 - 1/k)\%$.

process and Sterrett's methods are examples of hierarchical algorithms.

Dorfman's procedure consists of testing each item individually in a positive group. Therefore, if a group of size k is positive, then $k+1$ tests will be required to classify all the individuals who belong to the positive group. But, if the group is negative, then the k -individuals that belong to the group are declared disease free.

The expected relative number of individual diagnostic tests (X) using the Dorfman's method is equal to $E(X) = \frac{E(Y)}{n} = \frac{1}{k} + P$. Where $E(Y)$ is the expected number of required diagnostic tests using group testing, n (the sample size) is the number of diagnostic tests required for individual testing, P denotes the probability a group has to be positive which is determined by $P = (1 - Sp)(1 - p)^k + Se(1 - (1 - p)^k)$. Se and Sp are the diagnostic test sensitivity and specificity, respectively, and p is the disease prevalence in the population. Therefore, the expected percentage of savings with retesting is $100(1 - E(X))\%$, while, without retesting, it is equal to $100(1 - 1/k)\%$. Even when this procedure leads to significant savings, there are some other procedures that are even more efficient⁽¹¹⁾.

Sterrett's method (1957) is more efficient than Dorfman's method, since, it requires fewer diagnostic tests. After a positive group has been found, the method consists of randomly selecting individuals within the positive group and testing each one until a positive one is found. Once the first positive individual has been found, the rest of the individuals from this group are mixed to form a new group and another test is taken. If this new group is found to be negative in the diagnostic test, the classification process ends, and the rest of the individuals in this new group are declared negative. If the new group is found to be positive, the process starts again and individuals are randomly selected and tested in laboratory until a second group is found positive. Once the second

Aunque este procedimiento produce ahorros significativos, existen otros procedimientos que son aún más eficientes⁽¹¹⁾.

El método de Sterrett' (1957) es más eficiente que el de Dorfman porque requiere de menos pruebas de laboratorio. Bajo el supuesto de haber encontrado un grupo positivo, este método consiste en seleccionar al azar a individuos dentro del grupo positivo y a cada individuo realizarle una prueba de laboratorio hasta que se encuentre un positivo. Una vez encontrado el primer individuo positivo, los demás elementos de este grupo se mezclan para formar un nuevo grupo. Si este nuevo grupo resulta negativo en la prueba de laboratorio, el proceso de clasificación termina, y los individuos dentro del nuevo grupo son declarados negativos. Si el nuevo grupo resulta positivo, el proceso comienza nuevamente seleccionando individuos al azar y realizándoles la prueba de laboratorio hasta encontrar otro positivo (el segundo). Una vez que el segundo positivo ha sido encontrado, los demás individuos de este grupo son mezclados nuevamente para formar otro grupo y determinar si este es positivo o negativo. Este proceso se repite hasta que ya no se encuentran grupos o individuos positivos⁽¹¹⁾.

El procedimiento halving es otro método jerárquico y consiste en dividir a la mitad a los grupos que resultan positivos y aplicar una prueba de laboratorio a cada mitad. Si las dos mitades resultan negativas, el proceso de clasificación termina; pero si al menos una resulta positiva, se subdivide a la mitad nuevamente y se aplica otra vez la prueba de laboratorio a cada una de éstas. Por ejemplo, suponiendo que un grupo de tamaño 16 resulta positivo; este grupo se divide en dos subgrupos de tamaño ocho y a cada subgrupo se le aplica una prueba de laboratorio. Cualquier nuevo grupo que resulte positivo se subdivide en grupos de tamaño cuatro. De igual manera, si alguno de estos sub-subgrupos resulta positivo, entonces se divide en grupos de tamaño dos y finalmente a los grupos que resulten positivos se les aplican pruebas individuales. Cuando el grupo

positive individual has been found, the rest of the individuals in this group are mixed again to form another group to determine if this group is positive or negative. This process is repeated until no group or any individual tests positive⁽¹¹⁾.

The halving process is another hierarchical method. It consists of dividing by two the groups that were found positive and testing each half in a laboratory. If both halves are found negative, the classification process ends. But, if one half is positive, it is divided again in two and another test is administered in the laboratory to both halves. For instance, suppose that a group of 16 individuals is positive. This group is divided into two subpools of eight and each subpool is tested in the laboratory. Any new group that tests positive is subdivided into groups of four. In the same manner, if any of these subpools test positive, then it is divided into two groups of two and finally the groups that result to be positive are tested individually. When a group is not divisible between two, groups of different sizes are formed. For example, a group of 11 can be divided into two groups: a group of six and a group of five. The group of five can be subdivided into a group of three and a group of two. Finally, the group of three can be subdivided into a group of two and one individual⁽¹¹⁾.

In the matrix arrays, the squared arrays are the most frequent. Squared arrays consist of accommodating k^2 individuals in a matrix of $k \times k$ order. With the sample, groups of k are formed with items from a same row or a same column. Assuming that there is no false negative, the two groups of k can be analyzed. In this circumstance, all positive individuals will occur at the intersection of a positive row group and a positive column group. When more than a row and a column result to be positive, individual tests are administered to individuals that are at the intersection row-column. In that way, the classification process ends. Sometimes, the sample size is not a multiple of k^2 . For example, suppose a grid of 15x15 for the diagnostic tests and 245 individuals need to be

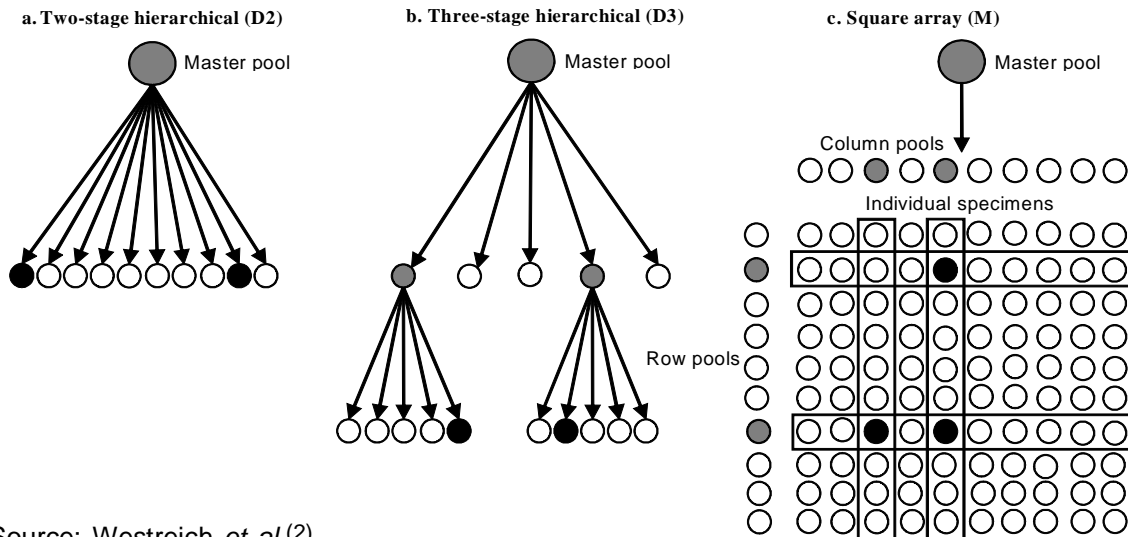
no es divisible entre dos, se forman dos subgrupos de tamaño diferente. Por ejemplo, un grupo de tamaño 11 puede dividirse en dos grupos, uno de tamaño seis y otro de tamaño cinco. Para el grupo de tamaño cinco, la subdivisión puede ser un grupo de tamaño tres y otro de tamaño dos, y finalmente el grupo de tamaño tres puede sub-subdividirse en un grupo de tamaño dos y un grupo con sólo un elemento⁽¹¹⁾.

Dentro de los arreglos matriciales, los arreglos cuadrados son los más comunes. Los arreglos cuadrados consisten en acomodar k^2 individuos en una matriz de orden $k \times k$. Con la muestra se

classified. This grid works well for the first 225 individuals but not for the last 20. These 20 individuals can be treated with row of 15 and the remaining 5 items using Dorfman method. When the objective is only to estimate the prevalence (p), once the status of each group has been determined, positive or negative, retesting is not needed in positive groups. Indeed, information on the status of the original groups is sufficient to estimate the prevalence. No additional test is administered to the positive groups to estimate disease prevalence. The number of positive groups (y) among the g groups of size k is enough. The g groups are formed with the n items that are part of the selected sample.

Figura 2. Diagrama de tres estrategias para formar grupos. Grupos positivos están en gris; individuos positivos están en negro. En D2 (a), el grupo principal positivo es dividido en elementos individuales. Uno o más de estos elementos pueden ser positivos. En D3 (b), el grupo principal positivo es dividido en diversos subgrupos. A cada subgrupo se le realiza la prueba. Los subgrupos positivos son divididos en elementos individuales; y uno o más de estos individuos puede ser positivo. En M (c), si el grupo principal es positivo, a los grupos columna y a los grupos hileras se les aplica la prueba, alguno de los cuales pueden ser positivos. A los individuos de la intersección entre un grupo columna positivo y un grupo hilera positivo, se les realiza la prueba individualmente; algunos o todos ellos pueden ser positivos.

Figure 2. Diagram of the three strategies to form pools. Positive pools are in grey, positive specimens are in black. In D2 (a), the positive master pool is divided into individual items. One or more of these items might be positive. In D3 (b), the positive master pool is divided in subpools. Each pool is tested. The positive subpools are divided into individual items. One or more of these individuals might be positive. In M (c), if the master pool is positive, the column pools and the row pools are tested, some of them might be positive. The individuals at the intersection between a positive column pool and a positive row pool are tested individually. Some or all of them might be positive.



Source: Westreich *et al.*⁽²⁾

forman grupos de tamaño k con los elementos de la misma hilera o la misma columna. Bajo el supuesto de que no existen pruebas falsas negativas, los $2k$ grupos son analizados. En este escenario, todos los individuos positivos ocurrirán en la intersección de un grupo hilera positivo y un grupo columna positivo. Cuando más de una hilera y columna resultan positivas, se realizan pruebas individuales a los individuos que se encuentran en la intersección hilera-columna, completando así el proceso de clasificación. Algunas veces el tamaño de muestra no es múltiplo de k^2 . Por ejemplo, suponiendo que se dispone con una cuadrícula de 15×15 para las pruebas de laboratorio y se necesitan clasificar 245 individuos. Esta cuadrícula hace buen trabajo para los primeros 225 individuos, pero no para los restantes 20. Los 20 individuos pueden tratarse usando una hilera de tamaño 15 y los otros cinco elementos empleando el procedimiento de Dorfman⁽¹¹⁾. En la Figura 2 se muestran gráficamente tres estrategias para la utilización de la prueba de grupo utilizando los algoritmos mencionados.

Es importante mencionar que en los arreglos matriciales también existen arreglos rectangulares, así como otras modificaciones a los procedimientos jerárquicos, los cuales pueden consultarse en Kim y colaboradores⁽¹²⁾.

Cuando el único objetivo es la estimación de la prevalencia (p), una vez que se determina el estado de cada grupo, positivo o negativo, no es necesario realizar retesting en los grupos que resultaron positivos, ya que para realizar la estimación de la prevalencia es suficiente disponer sólo con la información del estado de los grupos originales. Es decir, en la estimación de la prevalencia de una enfermedad no se realizan pruebas adicionales a los grupos que resultaron positivos, sólo es necesario conocer el número de grupos positivos (y) resultantes entre los g grupos de tamaño k . Los g grupos se forman con los n elementos que integran la muestra seleccionada.

The estimator of the proportion, the variance, the confidence intervals and the sample size calculation, according to group testing method, are presented, below. The process is illustrated through disease prevalence estimation.

Without group testing and assuming that the diagnostic test is perfect, the p maximum likelihood estimator (MLE) is $\hat{p} = y/n$, where y is the number of individuals that are positive according to the diagnostic test among the n that form the sample.

When group testing is used and the diagnostic test is not assumed to be perfect, Tu *et al.*⁽¹⁵⁾ suggest that the MLE of the proportion (p) is equal to

$$\hat{p} = 1 - \left(\frac{Se - y/g}{Sp + Se - 1} \right)^{1/k} \dots\dots\dots (Ec. 1)$$

g is the number of groups formed by n individuals from the sample and each has group size k (the group size is the number of individuals that are mixed in each group). y is the number of positive groups, Se is the diagnostic test sensitivity and Sp is the specificity. Both Se and Sp are greater than 0.5, which is the case for most of the diagnostic tests. For the MLE of p to exist is required that $1 - Sp \leq y/g \leq Se$. This guarantees that p is between zero and one. In a context of low prevalence, $1 - Sp \leq y/g$ is necessary. When this condition is not hold and under a given specificity, fewer positive groups than expected will be observed. The result will be a negative value for p . It is worth mentioning that if $Sp = Se = 1$, the MLE of p is reduced to $\hat{p} = 1 - (1 - y/g)^{1/k}$. Moreover, if the group size is equal to $k=1$, the MLE of p in equation (1) is reduced to $\hat{p} = y/n$. On the other hand, if $Se, Sp < 1$ and $k=1$, then $\hat{p} = \frac{y/n + Sp - 1}{Se + Sp - 1}$.

It has been reported that the estimator of p in equation (1) is biased to the right, with a bias approximately equal to $[(k-1)Var(\hat{p})]/[2(1-p)]$ ⁽¹⁵⁾. However, the bias is insignificant when the number of individuals $n = gk$ is large, the size of the group is moderate and the proportion to estimate is less than 10%.

A continuación se presentan los estimadores de la proporción, la varianza, los intervalos de confianza y el cálculo del tamaño de muestra bajo el marco de la prueba de grupo. El proceso se ilustra estimando la prevalencia de una enfermedad.

Sin utilizar la prueba de grupo y considerando que la prueba de laboratorio es perfecta, el estimador de máxima verosimilitud (EMV) de p es igual a: $\hat{p} = y/n$, donde y es el número de individuos que resultaron positivos en la prueba de laboratorio entre los n que conforman la muestra.

Cuando se utiliza la prueba de grupo y se asume que la prueba de laboratorio no es perfecta, de acuerdo con Tu y colaboradores⁽¹⁵⁾ el EMV de la proporción p es igual a:

$$\hat{p} = 1 - \left(\frac{Se - y/g}{Sp + Se - 1} \right)^{1/k} \dots\dots\dots (Ec. 1)$$

donde g es el número de grupos formados con los n individuos de la muestra y cada grupo es de tamaño k (el tamaño de grupo es el número de individuos a mezclar por grupo); y es el número de grupos positivos; Se es la sensibilidad de la prueba de laboratorio y Sp es la especificidad. Se asume que tanto Se como Sp son mayores a 0.5, lo cual aplica en la mayoría de las pruebas de laboratorio. Para que exista el EMV se requiere que $1 - Sp \leq y/g \leq Se$, esto asegura que p esté entre cero y uno. En un contexto de baja prevalencia es necesario que $1 - Sp \leq y/g$. Cuando esta condición no se cumple y bajo una especificidad dada, se observarán menos grupos positivos de lo esperado, dando como resultado un valor negativo para p . Cabe mencionar que si $Sp = Se = 1$, el EMV de p se reduce a $\hat{p} = 1 - (1 - y/g)^{1/k}$. Si además el tamaño del grupo es igual a $k=1$, el EMV de p en la Ecuación (1) se reduce a $\hat{p} = y/n$. Por otro lado,

si $Se, Sp < 1$ y $k=1$, entonces $\hat{p} = \frac{y/n + Sp - 1}{Se + Sp - 1}$.

Se ha documentado que usando prueba de grupo, el estimador de p en la Ecuación (1) es sesgado hacia la derecha, con una magnitud de sesgo

According to Tu et al.⁽¹⁵⁾ the estimated variance of the estimator given in equation (1) is:

$$Var(\hat{p}) = \frac{(Se - P)^{2/k-2} P(1-P)}{gk^2 (Se + Sp - 1)^{2/k}} \dots\dots\dots (Ec. 2)$$

where $P = [1 - (1 - p)^k]Se + (1 - p)^k(1 - Sp)$, is the probability for a group to be positive. Similarly, if $Sp = Se = 1$, then the variance of \hat{p} is reduced to

$$Var(\hat{p}) = \frac{[1 - (1 - p)^k]^2}{gk^2 (1 - p)^{k-2}} = \frac{(1 - P)^{2/k-1} P}{gk^2}, \quad \text{with}$$

$P = 1 - (1 - p)^k$. If P is unknown, an estimate of the variance can be obtained by replacing P with $\hat{P} = y/g$.

Hepworth⁽¹⁶⁾ and Tebbs *et al.*⁽¹⁷⁾ suggest that the corresponding Wald confidence interval (CI) for p equals:

$$\begin{aligned} p_L &= \hat{p} - Z_{1-\alpha/2} \sqrt{V(\hat{p})} \\ p_U &= \hat{p} + Z_{1-\alpha/2} \sqrt{V(\hat{p})} \dots\dots\dots (Ec. 3) \end{aligned}$$

$Var(\hat{p}) = \frac{(Se - y/g)^{2/k-2} P(1-P)}{gk^2 (Se + Sp - 1)^{2/k}}$, $P = [1 - (1 - p)^k]Se + (1 - p)^k(1 - Sp)$ and $Z_{1-\alpha/2}$ is the quantile $1 - \alpha/2$ of the standard normal distribution. \hat{p} is the MLE calculated with equation (1).

The approximate CI given in equation (3) is easy to calculate and allows an analytical solution (a formula) for the size of the sample to be obtained. However, when g and p are small, the normal approximation for the CI is not very good, leading to negative limits for the CI^(16,17). Furthermore, the coverage probability of the CI is often lower than the nominal probability $100(1 - \alpha)\%$.

The Wilson interval is another CI with an analytical solution that has proved to have a coverage that is very close to the nominal probability $100(1 - \alpha)\%$. To construct it, first the confidence interval is set in terms of groups. That is equal to

aproximadamente igual a $[(k-1)Var(\hat{p})]/[2(1-p)]$ (15). Sin embargo, el sesgo es insignificante cuando el número de individuos $n=gk$ es grande, el tamaño del grupo es moderado y la proporción a estimar es menor de 10%.

De acuerdo con Tu y colaboradores(15), la varianza estimada del estimador dado en la Ecuación (1) es:

$$Var(\hat{p}) = \frac{(Se - P)^{2/k-2} P(1-P)}{gk^2 (Se + Sp - 1)^{2/k}} \dots\dots\dots (Ec. 2)$$

donde $P = (1 - (1-p)^k)Se + (1-p)^k(1-Sp)$. P es la probabilidad de que un grupo sea positivo. De igual manera, si $Sp=Se=1$, entonces la varianza de

p se reduce a $Var(\hat{p}) = \frac{[1 - (1-p)^k]}{gk^2 (1-p)^{k-2}} = \frac{(1-P)^{2/k-1} P}{gk^2}$ con

$P = 1 - (1-p)^k$. Si no se conoce P , una estimación de la varianza puede obtenerse sustituyendo P por $\hat{P} = y/g$.

El correspondiente intervalo de confianza (IC) de Wald para p , de acuerdo con Hepworth(16) y Tebbis y colaboradores(17), es igual a:

$$p_L = \hat{p} - Z_{1-\alpha/2} \sqrt{V(\hat{p})}$$

$$p_U = \hat{p} + Z_{1-\alpha/2} \sqrt{V(\hat{p})} \dots\dots\dots (Ec. 3)$$

donde $Var(\hat{p}) = \frac{(Se - y/g)^{2/k-2} P(1-P)}{gk^2 (Se + Sp - 1)^{2/k}}$; $P = (1 - (1-p)^k)Se + (1-p)^k(1-Sp)$; $Z_{1-\alpha/2}$ es el cuantil $1-\alpha/2$ de la distribución normal estándar; y \hat{p} es el EMV calculado con la Ecuación (1).

El IC aproximado dado en la Ecuación (3) es fácil de calcular y permite derivar una solución analítica (fórmula) para el tamaño de muestra. Sin embargo, cuando g y p son pequeños la aproximación normal para el IC no es muy buena, produciendo límites negativos para el IC(16,17). Además, la probabilidad de cobertura del IC es a menudo menor que la probabilidad nominal $100(1-\alpha)\%$.

$$P_L = \frac{y + \frac{Z_{1-\alpha/2}^2}{g}}{g + Z_{1-\alpha/2}^2} - \frac{Z_{1-\alpha/2} g^{1/2}}{g + Z_{1-\alpha/2}^2} \sqrt{\frac{y}{g} \left(1 - \frac{y}{g}\right) + \frac{Z_{1-\alpha/2}^2}{4g}}$$

$$P_U = \frac{y + \frac{Z_{1-\alpha/2}^2}{g}}{g + Z_{1-\alpha/2}^2} + \frac{Z_{1-\alpha/2} g^{1/2}}{g + Z_{1-\alpha/2}^2} \sqrt{\frac{y}{g} \left(1 - \frac{y}{g}\right) + \frac{Z_{1-\alpha/2}^2}{4g}}$$

Then the CI for p is determined as:

$$p_L = 1 - \left(\frac{Se - P_L}{Sp + Se - 1}\right)^{1/k} \quad p_U = 1 - \left(\frac{Se - P_U}{Sp + Se - 1}\right)^{1/k} \dots (Ec. 4)$$

This interval is better than the Wald interval because the p distribution is less symmetrical. It is also possible to construct a CI for p by transforming the Clopper-Pearson(18) interval in terms of p . In the group scale, the lower and higher CI limits are: $P_L = B_{\alpha/2, y, g-y+1}$ and $P_U = B_{1-\alpha/2, y+1, g-y}$, respectively, where $B_{\gamma, a, b}$ is the γ quantile of the $beta(a, b)$ distribution(6). The lower and higher limits for p are obtained by using P_L and P_U in equation (4). The lower limit is equal to zero when $y=0$, while the higher limit equals one when $y=g$. The Clopper-Pearson(18) interval is exact but it does not have an analytical solution. Various researches indicate that the Wilson CI produces better results than the Clopper-Pearson CI, with the advantage that the Wilson CI has an analytical solution.

To obtain the number of pools (sample size) Montesinos-López *et al.*(19) provided an expression that ensures accuracy in the estimate of the proportion (p), assuming sensitivity and specificity of 100%. The expression is given by:

$$s_n = \left\lceil \frac{\sqrt{\frac{(1-P_\gamma)^{2k-1} P_\gamma}{k^2} + \frac{(1-P_\gamma)^{2k-1} P_\gamma + \frac{\omega}{Z_{1-\alpha/2}} Z_\gamma \left[\frac{1}{k} \frac{(1-P_\gamma)^{2k-2}}{\sqrt{(1-P_\gamma)^{2k-1} P_\gamma} \left(1 - \frac{2P_\gamma}{k}\right)} \right] \sqrt{P_\gamma(1-P_\gamma)}}}{\frac{\omega}{Z_{1-\alpha/2}}}} \right\rceil \dots (Ec. 5)$$

where γ represents the desired level of assurance (required probability) for the width of the CI (W) for p not to be greater than the desired value (ω). Z_γ is the γ quantile of the standard normal

Otro IC con soluci3n anal3tica que ha demostrado tener cobertura muy cercana a la nominal $100(1-\alpha)\%$, es el intervalo de Wilson⁽⁶⁾. Para su construcci3n primero se forma el intervalo de confianza en t3rminos de grupos, el cual es igual a:

$$P_L = \frac{y + \frac{Z_{1-\alpha/2}^2}{2}}{g + Z_{1-\alpha/2}^2} - \frac{Z_{1-\alpha/2} g^{1/2}}{g + Z_{1-\alpha/2}^2} \sqrt{\frac{y}{g} \left(1 - \frac{y}{g}\right) + \frac{Z_{1-\alpha/2}^2}{4g}}$$

$$P_U = \frac{y + \frac{Z_{1-\alpha/2}^2}{2}}{g + Z_{1-\alpha/2}^2} + \frac{Z_{1-\alpha/2} g^{1/2}}{g + Z_{1-\alpha/2}^2} \sqrt{\frac{y}{g} \left(1 - \frac{y}{g}\right) + \frac{Z_{1-\alpha/2}^2}{4g}}$$

Posteriormente se determina el IC para p de la siguiente forma:

$$P_L = 1 - \left(\frac{Se - P_L}{Sp + Se - 1} \right)^{1/k}$$

$$P_U = 1 - \left(\frac{Se - P_U}{Sp + Se - 1} \right)^{1/k} \dots\dots\dots (Ec. 4)$$

Este intervalo es superior al de Wald porque la distribuci3n de p es menos sim3trica. Tambi3n se puede construir un IC para p transformando el intervalo de Clopper-Pearson⁽¹⁸⁾ en t3rminos de P . En la escala de grupo, los l3mites inferior y superior del IC son: $P_L = B_{\alpha/2, y, g-y+1}$ y $P_U = B_{1-\alpha/2, y+1, g-y}$, respectivamente, donde $B_{\gamma, a, b}$ es el cuantil γ de la distribuci3n $beta(a, b)$ ⁽⁶⁾. Los l3mites inferior y superior para p se obtienen empleando a P_L y P_U en la Ecuaci3n (4). El l3mite inferior es igual a cero cuando $y=0$, mientras que el superior es igual a uno cuando $y=g$. El intervalo de Clopper-Pearson⁽¹⁸⁾ es exacto pero no tiene soluci3n anal3tica. Diversas investigaciones se3alan que el IC de Wilson produce resultados mejores al IC de Clopper-Pearson, con la ventaja de que el IC de Wilson tiene soluci3n anal3tica.

Para determinar el tama3o de muestra, Montesinos-L3pez y colaboradores⁽¹⁹⁾ proporcionaron una expresi3n que asegura precisi3n en la estimaci3n de la proporci3n p asumiendo sensibilidad y

distribuci3n and $P_g = 1 - (1-p)^k$ is the probability for a group to be positive. Note that if $\gamma = 0.5$, then $Z_{\gamma=0.5} = 0$ (the quantile 50% of a standard normal distribution). In this case, the formula to determine the number of groups, given by equation (5), is reduced to $g = \left(\frac{2Z_{1-\alpha/2}}{\omega k} \right)^2 \frac{1 - (1-p)^k}{(1-p)^{k-2}}$, assuming that the variance of the proportion $V(\hat{p})$ is known and fixed. In practice, $V(\hat{p})$ is unknown; this implies that the desired width of the CI is attained only 50% of the time. But, if $k=1$, equation (5) is reduced to:

$$n = \left(\frac{Z_{1-\alpha/2}}{\omega} \right)^2 \left(\sqrt{p(1-p)} + \sqrt{p(1-p) + \frac{\omega |1-2p| Z_\gamma}{Z_{1-\alpha/2}}} \right)^2 \quad (Ec. 6)$$

This formula (equation 6) is appropriate for the determination of the sample size to estimate the proportion (p) without using group testing (without forming groups). It guarantees a W lower than ω with a probability γ . Equation (6) is the analytical counterpart to the exact sample sizes proposed by Montesinos-L3pez *et al.*⁽¹⁸⁾ in the case of simple random sampling (SRS) and without group testing. In other words, only in $100(1-\gamma)\%$ of the time will W be greater than the desired CI width, ω .

Note also that if $\gamma=0.5$, the equation 6 (without using group testing) is simplified to the formula $n = [4Z_{1-\alpha/2}^2 p(1-p)] / \omega^2$, which is the standard sample size formula to estimate a proportion under SRS; however, the latter formula does not take into account the stochastic nature of the CI width.

The equation (5), proposed by Montesinos-L3pez *et al.*⁽¹⁹⁾, determines the minimal sample size g_m that guarantees that the observed width (W) will be lower than or equal to the desired width (ω) with a probability of at least γ . This is possible because equation (5) considers the stochastic nature of variance $V(\hat{p})$ through the assurance level γ . Because of that $\gamma \geq 0.5$. It is important to mention that the expression (5) was derived by using the Wald interval for group testing and for this reason

especificidad de 100%. La expresión está dada por:

$$g_m = \left[\frac{\sqrt{\frac{(1-P_g)^{2k-1} P_g}{k^2}} + \sqrt{\frac{(1-P_g)^{2k-1} P_g}{k^2} + \frac{\omega}{Z_{1-\alpha/2}} \left| \frac{1 - (1-P_g)^{2k-2}}{\sqrt{(1-P_g)^{2k-1} P_g} \left(1 - \frac{2P_g}{k}\right)} \right| \sqrt{P_g(1-P_g)}}}{\frac{\omega}{Z_{1-\alpha/2}}} \right]^2 \dots \text{(Ec. 5)}$$

donde γ representa el nivel de seguridad deseado (probabilidad requerida) para que la amplitud del IC (W) para p no sea más amplia que el valor deseado (ω); Z_γ es el cuantil γ de la distribución normal estándar; y $P_g = 1 - (1 - p)^k$ es la probabilidad de que un grupo sea positivo. Note que si $\gamma=0.5$, entonces $Z_{\gamma=0.5} = 0$ (el cuantil 50% de una distribución normal estándar); en este caso la fórmula que determina el número de grupos, dada en la Ecuación (5), se reduce a $g = \left(\frac{2Z_{1-\alpha/2}}{\omega k} \right)^2 \frac{1 - (1 - p)^k}{(1 - p)^{k-2}}$, asumiendo que la varianza de la proporción, $V(\hat{p})$, es conocida y fija. En la práctica se desconoce a $V(\hat{p})$; esto implica que la amplitud deseada del IC se alcance sólo 50% de las veces. Por otro lado, si $k=1$, la Ecuación (5) se reduce a:

$$n = \left(\frac{Z_{1-\alpha/2}}{\omega} \right)^2 \left(\sqrt{p(1-p)} + \sqrt{p(1-p) + \frac{\omega |1-2p| Z_\gamma}{Z_{1-\alpha/2}}} \right)^2 \dots \text{(Ec. 6)}$$

Esta fórmula (Ecuación 6) es apropiada para determinar el tamaño de muestra sin usar pruebas de grupo (sin formar grupos), garantizando que W será menor a ω con una probabilidad γ . La Ecuación (6) es la contraparte analítica a los tamaños de muestra exactos propuestos por Montesinos-López y colaboradores⁽¹⁸⁾ en el caso de muestreo aleatorio simple. En otras palabras, solo 100(1- γ)% de las veces W será más grande que la amplitud deseada del IC, ω .

También note que si $\gamma=0.5$, la Ecuación 6 (sin usar pruebas de grupo) se simplifica a la fórmula $n = [4Z_{1-\alpha/2}^2 p(1-p)] / \omega^2$, la cual es el tamaño de muestra estándar para estimar una proporción bajo muestreo aleatorio simple (MAS); aunque en esta

is not exact. The researchers interested in getting accurate sample sizes for group testing, using the Clopper-Pearson interval, may refer to Montesinos-López *et al.*⁽¹⁹⁾ However, these sample sizes do not have an analytical solution.

For the application of this method to be successful, the selection of the pool size (k) must be done very carefully. Because the trait of interest can be diluted and diminish the sensitivity of the diagnostic test and increase the rate of false negatives. Therefore, the right selection of the group size k is crucial to obtain accurate estimates with significant savings of resources. A group that is too large may lead to a dilution of the trait of interest and will increase the rate of false negatives. But, a group that is too small means very few savings.

Many methods have been proposed to determinate the size of the group k . For example, Thompson⁽²⁰⁾ proposes the use of $k = \frac{1.5936 - p}{p}$. This minimizes the mean square error. Chiang and Reeves⁽²⁰⁾ suggest a group size equal to $k = \log(1/2) / \log(1-p)$. This produces positive and negative groups with the same probability⁽²²⁾. Another exclusive form for the Dorfman method⁽¹⁾ in two stages consists in selecting $k = p^{-1/2}$. Hernández-Suárez *et al.*⁽⁸⁾ propose the use of $k = (c/d) - 1$, where c is the concentration of the trait of interest per individual, and d , the detection limit of the diagnostic test, both are assumed known, this method produces very large k values, as the Tompson method does⁽²⁰⁾.

The criteria presented above to select the size of the group are not always the best practical choice. A better alternative consists in gauging the diagnostic tests to ensure accuracy in group testing, as in individual diagnostic tests.

The optimal k is a function of the statistic model and the biochemical nature of the diagnostic test used to detect positive groups. For example, for a prevalence of 1/400, the Thompson method⁽²⁰⁾

última expresión la naturaleza estocástica de la amplitud del IC no es considerada.

La Ecuación (5), propuesta por Montesinos-López y colaboradores⁽¹⁹⁾, determina el tamaño de muestra mínimo, g_m , que garantiza que W será menor o igual a ω con una probabilidad de al menos γ ; esto se logra porque la Ecuación (5) considera la naturaleza estocástica de la varianza, $V(\hat{p})$, vía el nivel de aseguramiento γ . Por esta razón $\gamma \geq 0.5$. Es importante mencionar que la expresión (5) se derivó usando el intervalo de Wald para pruebas de grupo, por lo que estos tamaños de muestra son aproximados. Para investigadores interesados en tamaños de muestra para group testing (prueba de grupo) exactos usando el intervalo de Clopper-Pearson, los puede encontrar en Montesinos-López y colaboradores⁽¹⁹⁾; sin embargo, estos tamaños de muestra no tienen solución analítica.

Para la aplicación exitosa de este método, se debe cuidar la elección del tamaño del grupo (k) para que la sustancia de interés no se diluya y afecte la sensibilidad de las pruebas de laboratorio y, por ende, no se aumente la tasa de falsos negativos. Por lo tanto, la elección apropiada del tamaño de grupo g es vital para obtener estimaciones precisas y con ahorros de recursos significativos. Un grupo demasiado grande puede provocar que se diluya la sustancia de interés y se incremente la tasa de falsos negativos; mientras que un grupo demasiado pequeño produce pocos ahorros.

Para determinar el tamaño del grupo k se han propuesto varios métodos. Por ejemplo, Thompson⁽²⁰⁾ propone utilizar $k = \frac{1.5936 - p}{p}$, el cual minimiza el cuadrado medio del error. Chiang y Reeves⁽²¹⁾ sugieren un tamaño de grupo igual a $k = \log(1/2) / \log(1-p)$, el cual produce grupos positivos y negativos con la misma probabilidad⁽²²⁾. Otra forma exclusiva para el método de Dorfman⁽¹⁾ en dos etapas, consiste en elegir $k = p^{-1/2}$. Hernández-Suárez y colaboradores⁽⁸⁾ proponen usar $k = (c/d) - 1$, donde

recommends a group of 635 while the Chiang and Reeves method⁽²¹⁾ recommends a group of 277. Both are possibly larger than the maximum size diagnostic tests can manage. Most diagnostic tests based on polymerase chain reaction (PCR) can handle a maximum of 100 individuals for acceptable sensitivity and specificity levels⁽²²⁾. Similarly, most diagnostic tests based on antibodies can handle a maximum pool size of 50 units. This shows that the biochemical nature of diagnostic tests imposes the limits of the optimal pool size. It is also important to meticulously evaluate the sensitivity and the specificity of the tests. Different pool sizes might be examined to help choosing the size of the pool that best ensures the detection of positive groups without disregarding the diagnostic test sensitivity and specificity⁽²²⁾.

Suppose a researcher has carried out a study in a region to estimate disease prevalence in animals. The sample the researcher used was $n=10,000$ animals. With SRS three animals were found to be sick. Therefore, the estimated prevalence is $\hat{p} = 3/10,000 = 0.0003$, or 0.03%. Because the researcher did not know group testing, the research was quite expensive. By using group testing, a very accurate estimate is obtained with considerable savings in time, material and financial resources, as shown below.

The first line of the Table 1 shows the prevalence estimate without using group testing ($k=1$). The estimated prevalence of the disease in the population is $\hat{p} = 3/10,000 = 0.0003$. In this case, the percentage of savings is zero percent (0%). Furthermore, by using group testing, following the Dorfman model, with a group size of $k=10$ individuals, and assuming that $y=3$ groups are positive. (Each positive group contains a single sick item, since three animals in this population are assumed to be sick). The number of groups to contemplate is equal to $g=10,000/10=1,000$. Without retesting, there are 90% savings in the number of diagnostic tests. Indeed, instead of 10,000 tests, only 1,000 are necessary. With retesting in positive groups, the expected savings

c , la concentración de la sustancia de interés por individuo, y d , el límite de detección de la prueba de laboratorio, son conocidos. Este método produce valores muy grandes de k , como el método de Thompson⁽²⁰⁾.

Los criterios para seleccionar el tamaño de grupo dados anteriormente, no son siempre la mejor opción en la práctica. Una mejor alternativa consiste en calibrar las pruebas de laboratorio para asegurar precisión en la prueba de grupo, como se realiza con las pruebas de laboratorio individuales.

El tamaño óptimo del grupo k es una función del modelo estadístico y de la naturaleza química de la prueba de laboratorio usada en la detección de grupos positivos. Por ejemplo, si se sospecha que la prevalencia es igual a 1/400, el método de Thompson⁽²⁰⁾ recomienda un tamaño de grupo de 635, mientras el método de Chiang y Reeves⁽²¹⁾

are 89.7%. Since the diagnostic test is perfect ($Se = Sp = 1$), the estimated prevalence with group testing with $k=10$ is equal to:

$$\hat{p} = 1 - \left(\frac{Se - y/g}{Sp + Se - 1} \right)^{1/k} = 1 - \left(\frac{1 - 3/1,000}{1 + 1 - 1} \right)^{1/10} = 0.0003004058 \text{ (0.0300406\%)}$$

The estimate using group testing (0.0300406%) is slightly greater than the estimate done with individual testing (0.03%). However, it is near the estimated value without group testing.

When the group size equals 50 ($k=50$), and assuming there are 3 positive groups ($y = 3$), then 200 groups are formed ($g=10,000/50=200$). This means 98% savings in the number of required diagnostic tests when there is no retesting and 96.5% if there is retesting in the positive groups. In this case, the estimated prevalence is equal to

Cuadro 1. Estimación puntual y por intervalo usando la prueba de grupo. Se asumen diferentes tamaños de grupo k , dos o tres grupos positivos y , y un tamaño de muestra $n=10,000$; g denota el número de grupos, p es la proporción estimada, p_L y p_U son los límites inferior y superior del intervalo de confianza de Wald con 95% de confiabilidad, respectivamente. Se asume sensibilidad y especificidad igual a uno

Table 1. Point and interval estimate using group testing. There are different group sizes of k , two or three positive groups y , and a sample size $n=10,000$. g denotes the number of groups, p is the estimated proportion, p_L and p_U are the lower and higher limits, respectively, of the confidence Wald interval, with 95% reliability. Sensitivity and specificity are assumed to be equal to one.

(n)	k	$g=n/k$	y	p	p_L	p_U	Saving% ¹	Saving% ²
10,000	1	10,000	3	0.0003	0.00006187149	0.0008764745	0.00	00.00
10,000	5	2,000	3	0.0003001802	0.00006189625	0.000877001	80.00	79.85
10,000	10	1,000	3	0.0003004058	0.00006192725	0.0008776605	90.00	89.70
10,000	15	667	3	0.0003004813	0.00006192728	0.0008778821	93.33	92.88
10,000	20	500	3	0.0003008583	0.00006198937	0.000878985	95.00	94.40
10,000	25	400	3	0.0003010853	0.0000620205	0.0008796498	96.00	95.25
10,000	30	334	3	0.0003007087	0.00006192744	0.0008785521	96.66	95.77
10,000	35	286	3	0.0003012379	0.00006202064	0.0008801005	97.14	96.10
10,000	40	250	3	0.000301769	0.00006211413	0.0008816551	97.50	96.31
10,000	45	223	3	0.0003009375	0.00006192771	0.00087923	97.77	96.44
10,000	50	200	3	0.0003022271	0.00006217677	0.000883001	98.00	96.51
10,000	100	100	2	0.0002020067	0.00002434268	0.0007295698	99.00	96.04
10,000	200	50	2	0.0002040891	0.00002446663	0.0007372285	99.50	93.68

1 Percentage of saving without retesting; 2 Percentage of saving with retesting under the Dorfman method.

sugiere un tamaño de grupo de 277. Ambos son probablemente más grandes que el tamaño máximo que pueden manejar las pruebas de laboratorio. La mayoría de las pruebas basadas en PCR (polymerase chain reaction) han mostrado que son capaces de manejar un máximo de 100 individuos por grupo, con niveles de sensibilidad y especificidad aceptables⁽²²⁾. En forma similar, la mayoría de las pruebas basadas en anticuerpos pueden usar un tamaño de grupo máximo de 50 unidades. Esto muestra que la naturaleza bioquímica de las pruebas de laboratorio imponen los límites del tamaño óptimo del grupo. Además, es importante que la sensibilidad y especificidad de las pruebas sea evaluada rigurosamente, considerando varios tamaños de grupo, con la finalidad de elegir el tamaño que asegure la detección de grupos positivos sin hacer un sacrificio importante en la sensibilidad y especificidad de la prueba de laboratorio⁽²²⁾.

Suponiendo que un investigador realizó un estudio en una región para estimar la prevalencia animal de una enfermedad. La muestra que utilizó fue de $n=10,000$ animales, bajo MAS, entre los cuales encontró a tres con dicha enfermedad. Por lo tanto, la prevalencia estimada es de $\hat{p} = 3/10,000$, es decir, 0.03%. Dado que el investigador no conocía la existencia de la prueba de grupo, el estudio fue bastante costoso. A continuación se ilustra como usando la prueba de grupo se obtiene una estimación muy precisa de la prevalencia pero con un considerable ahorro de tiempo y recursos materiales y económicos.

La primera línea del Cuadro 1 muestra la estimación de la prevalencia sin usar prueba de grupo ($k=1$). La prevalencia estimada de la enfermedad en la población es $\hat{p} = 3/10,000 = 0.0003$; en este caso el porcentaje de ahorro es de cero (0%). Por otro lado, usando la prueba de grupo y bajo el modelo Dorfman, con un tamaño de grupo de $k=10$ individuos y asumiendo que $y=3$ grupos son positivos (cada grupo positivo contiene a un solo elemento enfermo dado que se asume que la población tiene exactamente tres individuos

$$\hat{p} = 1 - \left(\frac{Se - y/g}{Sp + Se - 1} \right)^{1/k} = 1 - \left(\frac{1 - 3/200}{1 + 1 - 1} \right)^{1/50} = 0.0003022271 \text{ (0.03022\%)}$$

This shows again that the estimate with group testing is very close to the real prevalence 0.0003 (0.03%). Moreover, all confidence intervals capture the real prevalence (Table 1).

Table 1 shows that with the same number of positive groups (in this case $y = 3$), point estimates and interval estimates of the prevalence are practically identical to the estimate without using group testing. However, testing with larger pool size might lead to the detection of fewer positive groups. For example, if the sizes of the groups are 100 or 200, and only two groups are found to be positive, the estimate of the prevalence (0.0002020067 and 0.0002040891, respectively) is less than the prevalence without using group testing (0.0003). This is due that If the size of the group is very large, a group is likely to have more than two positive individuals.

The results in Table 1 show that even when estimates with group testing are slightly biased, they remain reliable. Furthermore, group testing gives good estimates when the proportion to estimate is less than or equal to 0.1 (10%).

To determinate the sample size, suppose that a researcher is interested in estimating the prevalence (p) of a certain disease. Assume that $p=0.01$, that a CI of 95% is required, and that the group size k equals 25 ($k = 25$). The desired width is equal to $W_x = (p_U - p_L) \leq \omega = 0.007$, and the assurance level is of 99% ($\gamma=0.99$). First, P_g is calculated as:

$$P_g = 1 - (1 - p)^k = 1 - (1 - 0.01)^{25} = 0.2221786$$

$$h(P_g) = \sqrt{\frac{(1 - P_g)^{2/k-1} P_g}{k^2}} = \sqrt{\frac{(1 - 0.2221786)^{2/25-1} (0.2221786)}{25^2}} = 0.0221164,$$

$$h(P_g) = \frac{1}{k} \frac{(1 - P_g)^{2/(k-1)}}{2\sqrt{(1 - P_g)^{2/k-1} P_g}} \left(1 - \frac{2P_g}{k} \right) = \frac{1}{25} \frac{(1 - 0.2221786)^{2/(25-1)}}{2\sqrt{(1 - 0.2221786)^{2/25-1} (0.2221786)}} \left(1 - \frac{2(0.2221786)}{25} \right) = 0.0590768$$

Because a CI of 95% is required, then $Z_{1-0.05/2} = 1.96$.

enfermos), el número de grupos a considerar es igual a $g=10,000/10=1,000$, es decir, sin hacer retesting se tiene un ahorro de 90% en la cantidad de pruebas de laboratorio, ya que en lugar de 10,000 pruebas sólo se realizarán 1,000. Realizando retesting en los grupos positivos, el ahorro esperado es de 89.7%. Por otro lado, dado que la prueba de laboratorio es perfecta ($Se=Sp=1$), la prevalencia estimada con la prueba de grupo con $k=10$ es igual a:

$$\hat{p} = 1 - \left(\frac{Se - y/g}{Sp + Se - 1} \right)^{1/k} = 1 - \left(\frac{1 - 3/1,000}{1 + 1 - 1} \right)^{1/10} = 0.0003004058 \text{ (0.0300406\%).}$$

La estimación usando la prueba de grupo (0.0300406%) es ligeramente mayor a la estimación hecha con pruebas individuales (0.03%), aunque está cercano al valor estimado sin usar pruebas de grupo.

Cuando el tamaño del grupo es $k=50$ y asumiendo el mismo número de grupos positivos $y=3$, se forman $g=10,000/50=200$ grupos, lo cual produce un ahorro de 98% en la cantidad de pruebas de laboratorio requeridas cuando no se realiza retesting y de 96.5% si se realiza retesting en los grupos que resulten positivos. En este caso la prevalencia estimada es igual a:

$$\hat{p} = 1 - \left(\frac{Se - y/g}{Sp + Se - 1} \right)^{1/k} = 1 - \left(\frac{1 - 3/200}{1 + 1 - 1} \right)^{1/50} = 0.0003022271 \text{ (0.030222\%).}$$

Esto muestra nuevamente que la estimación por usar la prueba de grupo es muy cercana a la prevalencia (Cuadro 1). También se observa que mientras se mantenga el mismo número de grupos positivos (en este caso $y=3$) las estimaciones puntual y por intervalo de la prevalencia son prácticamente idénticas a la estimación sin el uso de la prueba de grupo. Sin embargo, al incrementar el tamaño del grupo puede ocasionar que se detecten menos de tres grupos positivos. Por ejemplo, si los tamaños de grupo son de 100 ó 200 y sólo se detectan dos grupos positivos, la estimación de la prevalencia (0.0002020067 y 0.0002040891, respectivamente) será menor que la prevalencia sin usar prueba de grupo (0.0003).

Also, because the assurance level is 99% ($\gamma=0.99$) thus $Z_{0.99}=2.33$, $\omega=0.007$, and $k=25$. Therefore,

$$g_m = \left(\frac{h(P_g) + \sqrt{h(P_g)^2 + \frac{2\omega}{Z_{1-\alpha/2}} Z_\gamma |h(P_g)| \sqrt{P_g(1-P_g)}}}{\omega / Z_{1-\alpha/2}} \right)$$

$$g_m = \left(\frac{0.0221164 + \sqrt{0.0221164^2 + \frac{2(0.007)(2.33)(0.05907768)}{1.96} \sqrt{0.2221786(1-0.2221786)}}}{0.007/1.96} \right)^2 = 200$$

This 200 pools (sample size) guarantee that the observed width of the confidence interval will be less than or equal to the desired width $\omega=0.007$ with a probability of 99%. The double accuracy format was used to calculate this sample size. Indeed, if it had not been used, there would have been a slight overestimate of the sample size. It is worth highlighting that if $\gamma=0.5$, the value of $Z_\gamma=0$ and the number of required groups is equal to 140. But this sample size does not guarantee a good accuracy for estimating the proportion.

Montesinos-López *et al.*⁽¹⁹⁾ provide tables with different values for the parameters and a programme in the software R to determine the sample size easily and rapidly in other contexts. Likewise, if exact sample sizes are desired in using group testing to estimate a proportion, Montesinos-López *et al.*⁽⁹⁾ offer tables with their respective R programme to calculate the sample sizes.

Finally the group testing technique is increasingly used due to the significant savings in time, material and financial resources it allows. However, this technique is useful only to estimate small prevalence, less than or equal to 10%, when the variable response is binary. Moreover, the selection of the pool size needs to be done carefully, because an incorrect selection of the pool size can increase the rate of false negatives. Another concern is that, by using this method, the distribution of the individuals with the characteristic under study is assumed to be homogeneous. Besides these restrictions in its application, and because governments attach great importance to the

Si el tama1o de grupo es muy grande, es muy probable que un grupo tenga m1s de dos individuos positivos.

Los resultados del Cuadro 1 muestran que aunque la prueba de grupo produce estimaciones ligeramente sesgadas, 3stas son muy confiables. Adem1s, hay que tener presente que la prueba de grupo produce buenas estimaciones cuando la proporci3n a estimar es menor o igual a 0.1 (10%).

Ahora se ilustra la determinaci3n del tama1o de muestra. Suponiendo que un investigador est1 interesado en estimar la prevalencia p de cierta enfermedad. Se asume que $p=0.01$; y se requiere un IC de 95%, un tama1o de grupo de $k=25$, una amplitud absoluta igual a $W_x = (p_U - p_L) \leq \omega = 0.007$, y un nivel de aseguramiento de 99 % ($\gamma=0.99$). En primer lugar se calcula P_g :

$$P_g = 1 - (1 - p)^k = 1 - (1 - 0.01)^{25} = 0.2221786$$

$$h(P_g) = \sqrt{\frac{(1 - P_g)^{2/k-1} P_g}{k^2}} = \sqrt{\frac{(1 - 0.2221786)^{(2/25)-1} (0.2221786)}{25^2}} = 0.0221164,$$

$$h'(P_g) = \frac{1}{k} \frac{(1 - P_g)^{2/(k-1)}}{2\sqrt{(1 - P_g)^{2/k-1} P_g}} \left(1 - \frac{2P_g}{k}\right) = \frac{1}{25} \frac{(1 - 0.2221786)^{2(1/25-1)}}{2\sqrt{(1 - 0.2221786)^{(2/25)-1} 0.2221786}} \left(1 - \frac{2(0.2221786)}{25}\right)$$

$$= 0.0590768.$$

Como se requiere un IC de 95%, entonces $Z_{1-0.05/2} = 1.96$. El nivel de aseguramiento es de 99% ($\gamma = 0.99$), as3 $Z_{0.99} = 2.33$,

y $k=25$. Por lo tanto,

$$g_m = \left(\frac{h(P_g) \sqrt{h(P_g)^2 + \frac{2\omega}{Z_{1-\alpha/2}} Z_\gamma |h'(P_g)| \sqrt{P_g(1 - P_g)}}}{\frac{\omega}{Z_{1-\alpha/2}}} \right)^2$$

$$= \left(\frac{0.0221164 \sqrt{(0.0221164)^2 + \frac{2(0.007)}{1.96} (2.33) [0.0590768] \sqrt{0.2221786(1 - 0.2221786)}}}{\frac{0.007}{1.96}} \right)^2 = 200$$

El tama1o de muestra que garantiza con una probabilidad de 99% que la amplitud observada del intervalo de confianza ser1 menor o igual a la

epidemiologic vigilance programs to face the growing threat of epidemics, this technique might be very helpful for the researchers in animal science to detect and estimate the prevalence of specific diseases at low cost. Hence, the risk of plague or other diseases that may cause damage to consumers and producers may diminish. New generalizations are being created to make this technique more efficient. The use of group testing is recommended despite its limitations. It leads to particularly accurate results and really significant savings. Researchers in developing countries will find this technique useful to detect sick animals and estimate the prevalence of certain diseases that threaten some regions.

End of english version

amplitud deseada $\omega=0.007$, es igual a 200. Para el c1culo de este tama1o de muestra se utiliz3 el formato de doble precisi3n, ya que de no usarse ocurrir1 una ligera sobrestimaci3n del tama1o de muestra. Es necesario resaltar que si $\gamma=0.05$, el valor de $Z_\gamma=0$ y el n1mero de grupos requeridos es igual a 140; pero este tama1o de muestra no garantiza buena precisi3n en la estimaci3n de la proporci3n.

Para otros escenarios de estimaci3n puede consultarse Montesinos-L3pez y colaboradores⁽¹⁹⁾, donde se proporcionan tablas con diversos valores de los par1metros y un programa en el paquete R para realizar la determinaci3n del tama1o de muestra de forma f1cil y r1pida. De igual manera, si se desean tama1os de muestra exactos usando pruebas de grupo para estimar una proporci3n, puede consultarse Montesinos-L3pez y colaboradores⁽⁹⁾, donde se proporcionan tablas con su respectivo programa R para calcular los tama1os de muestra.

Finalmente, debido al ahorro significativo de tiempo y recursos materiales y econ3micos al utilizar las pruebas de grupo, el uso de esta t3cnica incrementa

día a día. Sin embargo, hay que tener presente que esta técnica es útil sólo para estimar prevalencias pequeñas, menores o iguales a 10% cuando la variable respuesta es binaria. Además, también hay que ser cuidadoso en la elección del tamaño de grupo para no incrementar la tasa de falsos negativos, y estar conscientes que el método aquí presentado supone que la distribución de individuos con la característica de interés es homogénea. A pesar de estas restricciones para su aplicación, pero debido a que los gobiernos dan gran importancia a los programas de vigilancia epidemiológica por la creciente amenaza de epidemias, esta técnica puede ser de gran ayuda para los investigadores en ciencia animal para detectar y estimar la prevalencia de enfermedades específicas a costos mucho más bajos. Consecuentemente, con ello se podría disminuir el riesgo de plagas o enfermedades que puedan ocasionar daños a los consumidores y productores. Si bien es cierto que se están creando nuevas generalizaciones para hacer esta técnica más eficiente, aún con sus restricciones se recomienda usar esta técnica porque produce resultados muy precisos y ahorros altamente significativos, lo cual puede ayudar a científicos en países en vías de desarrollo para detectar animales enfermos así como para estimar la prevalencia de ciertas enfermedades que amenazan estas regiones.

LITERATURA CITADA

- Dorfman R. The detection of defective members of large populations. *Ann Math Stat* 1943;14(4):436-440.
- Westreich DJ, Hudgens MG, Fiscus SA, Pilcher CD. Optimizing screening for acute human immunodeficiency virus infection with Pooled Nucleic Acid Amplification tests. *J Clin Microbiol* 2008;46(5):1785-1792.
- Dodd R, Notari E, Stramer S. Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross donor population. *Transfusion* 2002;42:975-979.
- Remlinger K, Hughes-Oliver J, Young S, Lam R. Statistical design of pools using optimal coverage and minimal collision. *Technometrics* 2006;48:133-143.
- Verstraeten T, Farah B, Duchateau L, Matu R. Pooling sera to reduce the cost of HIV surveillance: a feasibility study in a rural Kenyan district. *Trop Med Int Health* 1998;3:747-750.
- Tebbs J, Bilder C. Confidence interval procedures for the probability of disease transmission in multiple-vector-transfer designs. *J Agric Biol Environ Stat* 2004; 9(1):79-90.
- Peck C. Going after BVD. *Beef* 2006;42:34-44.
- Hernández-Suárez CM, Montesinos-López OA, McLaren G, Crossa J. Probability models for detecting transgenic plants. *Seed Sci Res* 2008;18:77-89.
- Montesinos-López OA, Montesinos-López A, Crossa J, Eskridge K, Hernández-Suárez CM. Sample size for detecting and estimating the proportion of transgenic plants with narrow confidence intervals. *Seed Sci Res* 2010;20:123-136.
- Wolf J. Born again group testing - multi access communications. *IEEE Transactions on Information Theory* 1985;31(2):185-191.
- Bilder CR. Human or Cylon? Group testing on Battlestar Galáctica. *Chance* 2009;22(3):46-50.
- Kim HY, Hudgens MG, Dreyfuss JM, Westreich DJ, Pilcher CD. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* 2007;63:1152-1163.
- Lindan C, Mathur M, Kumta S, Jerajani H, Gogate A, Schachter J, Moncada J. Utility of pooled urine specimens for detection of Chlamydia trachomatis and Neisseria gonorrhoeae in men attending public sexually transmitted infection clinics in Mumbai, India, by PCR. *J Clin Microbiol* 2005;43(4):1674-1677.
- Kendzioriski C, Irizarry RA, Chen KS, Haag JD, Gould MN. On the utility of pooling biological samples in microarray experiments. *Proc Natl Acad Sci. USA*, 2005;102:4252-4257.
- Tu XM, Litvak E, Pagano M. Studies of AIDS and HIV surveillance, screening tests: can we get more by doing less? *Statistics in Medicine* 1994;13:1905-1919.
- Hepworth G. Exact CIs for proportions estimated by group testing. *Biometrics* 1996;52:1134-1146.
- Tebbs JM, Bilder CR, Moser BK. An empirical Bayes group-testing approach to estimating small proportions. *Communications in statistics: Theory and methods* 2003;32(5):983-995.
- Montesinos-López OA, Montesinos-López A, Santos-Fuentes EE, Valladares-Celis PE, Magaña-Echeverría MA. Tamaños de muestra para estimar prevalencia animal que aseguran cortos intervalos de confianza. *Rev Mex Cienc Pecu* 2011; 2(2):229-245.
- Montesinos-López OA, Montesinos-López A, Crossa J, Eskridge K, Saenz-Casas RA. Optimal sample size for estimating the proportion of transgenic plants using the Dorfman model with a random confidence interval. *Seed Sci Res* 2011; 21:235-245.
- Thompson KH. Estimation of the proportion of vectors in a natural population of insects. *Biometrics* 1962;18:568-578.
- Chiang CL, Reeves WC. Statistical estimation of virus infection rates in mosquito vector populations. *Amer J Hygiene* 1962;75:377-391.
- Katholi CR, Unnasch TR. Important experimental parameters for determining infection rates in arthropod vectors using pool screening approaches. *Am J Trop Med Hyg* 2006;74:779-785.