

## Estadísticas para estimación del riesgo (Statistics of risk estimation)

Badii, M.H., A. Guille & J.L. Abreu  
UANL, Cd. Universitaria, San Nicolás, N.L., México

**Resumen.** En el presente trabajo se presentan y discuten algunos métodos estadísticos estándares aplicadas a la ciencia de la salud, particularmente en epidemiología. Se definen los conceptos de Odds, Odds ratio, riesgos tanto relativo, absoluto como atribuible. Se realizan el cálculo de los arriba mencionados y el nivel de la reducción de los riesgos relativos y absolutos para las vacunas Pfizer, Astra Zeneca y Sputnik utilizados contra el Covid-19. Se discuten la relevancia de proporcionar datos no solamente sobre el riesgo relativo, sino también sobre el riesgo absoluto por su precisa importancia epidemiológica. También, para los casos de las tres vacunas, se calculan el número necesario de personas que deben estar vacunadas para asegurar proteger un individuo. Se estiman, en función de los datos disponible, la necesidad de contar con cierto porcentaje de población vacunada para obtener la inmunidad del rebaño.

**Palabras clave.** Inmunidad del rebaño, odds ratio, riesgo absoluto, riesgo relativo

**Abstract.** Standard statistical methods applied in health Sciences, particularly in epidemiology, are presented and discussed. Concepts of Odds, Odds ratio, relative, absolute and attributable risks are defined. Their estimation and the level of reduction of relative and absolute risks with Pfizer, Astra Zeneca and Sputnik vaccines against Covid-19 are calculated. The importance of providing data not only on relative risks but also on absolute risks due to its epidemiological relevance are discussed. The necessary number of persons to be treated in order to protect one individual for all three above vaccines are also calculated. Based on available data, the percentage of people that need be vaccinated in order to obtain herd immunity is estimated.

**Keywords.** Absolute risk, herd immunity, Odds ratio, relative risk

### Introducción

En los estudios en el campo de la salud, frecuentemente, el interés radica en estimar la magnitud del nivel de la interrelación entre dos o más variables. Es decir, en comparar dos variables, por ejemplo el impacto de un producto innovador o estándar versus el efecto de un producto experimental, a pesar que los dos productos puedan tener impactos positivos, todavía nos interesa comparar el grado del efecto de un en comparación con el otro. Para esto debemos estar familiarizada con algunos indicadores. Por ejemplo si estamos usando la escala nominal u ordinal, se emplean varios indicadores, como proporciones, porcentajes, relaciones y tasas. La estadística y los métodos estadísticos han sido empleados de forma amplia en las ciencias naturales y sociales entre otros campos de conocimiento (Badii et al, 2007a,b,c; Badii & Castillo, 2008; Badii et al, 2008; Badii & Castillo, 2009, Badii & Guillen, 2010; Badii et al, 2010; Badii et al, 2011; Badii et al, 2014).

### Conceptos y definiciones

A continuación se presentan algunos conceptos y definiciones esenciales para la mejor comprensión del contexto del presente trabajo.

### **Proporciones**

Se trata de número de observaciones de un rasgo determinado “x” (personas con Covid) dividida entre el número total de observaciones “x + y” (total de personas ajo el estudio), por tanto:

$$\text{Proporción} = P = x / (x + y)$$

La “P” siempre se define como una parte dividida entre el total. La proporción es útil para datos numéricos tanto en escala nominal como la escala ordinal. Por ejemplo si de total de 834 personas bajo el estudio 172 tuvieron el rasgo de interés o el Covid, entonces la proporción es:  $P = 172 / 834 = 0.206$ .

### **Porcentaje**

El porcentaje es igual a la proporción multiplicada por 100. Para el ejemplo arriba, el porcentaje o % es igual a  $0.206 * 100 = 20.6\%$ .

### **Relaciones**

La relación es igual al número de las observación en un grupo bajo del estudio con un rasgo de interés (x) dividida entre el número de las observaciones en el grupo sin el rasgo de interés (y). En otras palabras, la relación =  $x / y$ . En el ejemplo arriba sobre la proporción, de un total de 834, el 172 tuvieron el rasgo de interés, es decir  $834 - 172 = 662$  no tuvieron el rasgo de interés. Por tanto, para este ejemplo, la relación =  $172 / 662 = 0.259$ .

### **Tasas**

La tasa en realidad es similar a la proporción, excepto en el caso de la tasa que se calcula para un periodo determinado, se utiliza un multiplicador o base (por ejemplo, 1000, 10,000, 100,000, o 1,000,000). Por tanto la tasa o T se calcula de forma siguiente

$$T = [x / (x + y)] * \text{multiplicador}$$

Por ejemplo en el caso de la proporción de arriba (0.206), suponiendo de un estudio de un año de duración, se multiplica el valor de 0.206 por 10,000 y por tanto la tasa sería  $0.206 * 10,000 = 2,060$  pacientes por año.

## **Técnicas estadísticas**

Existen técnicas adecuadas para medir la relación y también la magnitud de la relación entre las variables. Por ejemplo el modelo de  $J_i^2$  de Pearson (1904), el modelo de McNemar (1947).

## **Modelo $J_i^2$ de Pearson**

### **Descripción**

Se utiliza para 1, 2 o más muestras independientes con varias variables en escala Nominal. Cuando el tamaño de muestra es mayor que 30 reduce la probabilidad (evita cometer) error tipo I y II. Sin embargo, toma en cuenta las 2 siguientes condiciones:

**A.-** Cuando maneja Tabla de Contingencia de 2X2, el  $n_{opt}$  es  $\leq 30$ , y hay 0

observaciones a algunas celdas, se usa la prueba Exacta de Fisher & Yates.

**B.-** Con múltiples grupos (muestras) pero con frecuencias menores que 5, se usa

Prueba  $(J_i)^2$  de Proporciones.

Estas 2 opciones limitan error Tipo I e incrementan Potencia & eficacia de las muestras pequeñas. Donde eficacia es reducir la probabilidad de rechazar un “Ho” cierto y aumentar la probabilidad de rechazar un “Ho” falso (en otras palabras, reducir el nivel de alfa e incrementar la potencia estadística).

**Ecuación:**  $X^2 = \sum [(O - E)^2 / E]$

Donde,

$X^2$  = Valor calculado del modelo,  $\sum$  = Signo de sumatorio, O = Frecuencia observada, E = Frecuencia esperada o teórica,  $gl = K - 1$ , donde, K = número de muestras o grupos.

### **Ejemplo 1**

Hay 120 pacientes con una enfermedad (1 grupo o muestra), un médico aplica cuatro tipos de terapia (drogas) y encontró datos diferentes sobre mortalidad (36 muertos) de diferentes grupos. Se analizó los 36 muertos para ver diferencia entre los tratamientos (Tabla 1).

Tabla 1. Número de muertos bajo tratamiento.					
<b>Frecuencias</b>	<b>Droga 1</b>	<b>Droga 2</b>	<b>Droga 3</b>	<b>Droga 4</b>	<b>Total</b>
<b>Observado (O)</b>	12	7	8	9	36
<b>Esperado (E)</b>	9	9	9	9	36

### **Hipótesis**

**Ho:** Las diferencias en la mortalidad por tratamientos son un efecto aleatorio

**Ha:** Las diferencias en la mortalidad por tratamientos son un efecto aleatorio y de hecho la droga uno ocasionó más mortalidad que otras drogas

**Nivel de significancia.**  $\alpha = 0.05$ ., **Zona de rechazo.** Para todo valor de  $p \leq 0.05$ , Ho se rechaza.

**Procedimiento.** Primero, hay que calcular los valores esperados. En este caso para una situación aleatorio del efecto de droga se maneja  $36/4 = 9$  y por tanto: para cuantificar el valor calculado del modelo de  $X^2 = \sum [(O - E)^2/E]$ , se procede con  $X^2 = [(12 - 9)^2 + (7 - 9)^2 + (8 - 9)^2 + (9 - 9)^2] / 9 = 1.55$

El gl es igual a  $k - 1$ , es decir,  $4 - 1 = 3$ , y el valor tabulado para 3 gl es igual a 7.815

**Resultado.** Valor calculado (1.55) es menor que valor crítico de 7.815, por tanto Ho se Acepta.

**Interpretación.** Todos los tratamientos (drogas) tienen el mismo efecto, y por tanto el hecho de que la droga uno aparentemente está causando mayor mortalidad se debe a un efecto aleatorio, es decir, no hay diferencia entre las 4 drogas en término de mortalidad.

## Modelo McNemar

### Descripción

Se utiliza este modelo para dos muestras dependientes (relacionados) y solamente se aplica cuando existen 2 momentos de medición: uno antes y otro después, (estudios longitudinales). Según McNemar, se utiliza esta prueba para cuando los datos son dicotómicos, es decir, la variable en la escala nominal tiene dos posibles valores. El arreglo de los datos para el modelo de McNemar se observa en la Tabla 2.

Tabla 2.- El arreglo de los datos para el modelo McNemar.			
		Después (Experimental)	
		Si	No
Antes (Control)	Si	A	B
	No	C	D

**Ecuación:**  $Z = (B - C) / \sqrt{(B+C)}$

**A** = Número de pares donde ambos sujetos desempeñan el papel, **D** = Número de pares donde Ningún sujeto desempeñan el papel, **B** = Número de pares donde Sujeto Control (antes) desempeñan el papel, **C** = Número de pares donde Sujeto Experimental (después) desempeñan el papel, **N** = Gran total = **A+B+C+D**, **A+B** = Total de sujeto de Control (antes) que desempeñan el papel, **A+C** = Total de sujeto de Experimental (Después) que desempeñan el papel.

### Hipótesis

**Ho:** Proporción 1 = Proporción 2

**Ha:** Proporción 1  $\neq$  Proporción 2

Debido a que **Ho:** [Proporción 1 = Proporción 2] o [Proporción 1 – Proporción 2 = 0],

Entonces: [(A+B)/N = (A+C)/N] o [(A+B)/N – (A+C)/N = 0], por ende: **Ho:**  $(B-C)/N = 0$

Según McNemar, La Prueba adecuada para la hipótesis Ho es:  $Z = (B - C) / \sqrt{(B+C)}$ .

### Toma de decisión

Si el valor calculado de Z se ubica fuera del rango  $\pm 1.96$  ( $\alpha=0.05$ ), Ho se rechaza y viceversa.

### **Ejemplo 3**

En un estudio, 216 víctimas de la enfermedad fueron emparejadas (en término de la edad y el género) con el mismo número de individuos, es decir 216 individuos libre de enfermedad. Luego se preguntó a los miembros de cada par si habían sido diagnosticados anteriormente con exposición (donde la exposición significa estar en espacios cerrados). Los resultados se encuentran en la Tabla 3.

Tabla 3.- Datos de exposición y enfermedad.			
Exposición	enfermedad		Total
	Si	No	
Si	69	36	107
No	147	178	325
Total	216	216	GT = 432

Entre los 216 con enfermedad, 69 habían sido diagnosticado con exposición y 147 no. Entre los 216 que no tuvieron la enfermedad solo 36 eran expuestos. Tenemos interés en saber acerca de las proporciones de expuestos en 2 grupos (con y sin enfermedad). En la tabla arriba con el formato de una tabla de contingencia de 2 x 2, uno desea aplicar el modelo de  $J_i^2$ . Sin embargo, hay que notar que hay 432 observaciones pero solo 216 pares. Cada par emparejado nos proporciona dos respuestas: una respuesta para el individuo que sufrió la enfermedad y uno para el individuo que no sufrió la enfermedad. Debido a que la prueba de  $J_i^2$  ignora la naturaleza de datos emparejados, no se puede utilizar esta prueba en esta situación. Nosotros debemos tomar en cuenta la noción de emparejamiento en el análisis de los datos. Si observamos de nuevo a los datos crudos (Tabla 4) del estudio pero ahora clasificando los datos de la siguiente manera:

Tabla 4.- datos con y sin enfermedad.			
Con enfermedad	Sin enfermedad		Total
	Si Exposición	No Exposición	

<b>Si Exposición</b>	15 (celda I o <i>a</i> )	58 (celda II o <i>b</i> )	73
<b>No Exposición</b>	24 (celda III o <i>c</i> )	119 (celda IV o <i>d</i> )	143
<b>Total</b>	39	177	GT=216

De las 73 personas que tuvieron la enfermedad y fueron diagnosticados como expuestos, 15 fueron emparejados con individuos expuestos y 58 con individuos libre de exposición. De los 143 individuos que no tuvieron la enfermedad, 24 fueron emparejados con los individuos expuestos y 119 no. Ahora cada entrada en la tabla corresponda a una combinación de respuestas para un par en lugar de un solo individuo. Ahora se puede conducir un análisis adecuado de datos emparejados. Sin embargo, primero debemos cambiar el enunciado de la hipótesis para reflejar el formato nuevo de los datos. Es decir, en lugar de probar si la proporción de los expuestos dentro de los individuos con la enfermedad es igual a la proporción de los expuestos en los individuos sin la enfermedad, debemos probar lo siguiente:

**H<sub>0</sub>:** No hay asociación entre exposición y ocurrencia de la enfermedad (la asociación es al azar)

**H<sub>1</sub>:** Hay asociación entre exposición y ocurrencia de la enfermedad (la asociación no es al azar)

### Selección de pares adecuados de datos

Los pares de respuestas en donde o dos expuestos (celda I) o dos no-expuestos (celda IV) están emparejados, proporcionan ninguna información para probar la hipótesis nula acerca de la diferencia en el estatus de exposición. Por tanto, descartamos estos pares y enfocamos solamente en los datos discordantes, es decir, pares de respuestas en donde un individuo con exposición se empareja con un individuo sin exposición (celda II y celda III).

La celda II representa el número de las pares víctimas de la enfermedad que eran expuestos, y la celda III representa el número de las pares víctimas de la enfermedad que no eran expuestos.

Si la hipótesis nula es cierto, entonces celda II y celda III deben ser iguales. Por tanto si diferencia entre estas dos celdas es grande rechazaríamos de hipótesis nula de no asociación. Para realizar la prueba de McNemar se puede utilizar tanto el  $X^2$  o la Z:

**1.- Modelo de  $X^2$**

**1.-  $X^2 = (|Celda II - celda III| - 1)^2 / (celda II + celda III)$**

Esta razón se aproxima a la distribución de Ji con un grado de libertad. El término “-1” en el numerador es para la corrección de la continuidad, ya que estamos utilizando conteos discretos para estimar la distribución de  $X^2$ .

Por tanto:  $X^2 = (|58 - 24| - 1)^2 / (58 + 24) = (|34 - 1|^2) / (82) = (1,089) / (82) = 13.28$

**Resultado.** Valor tabulado con  $gl = 1$  es igual a 3.841.

El valor calculado (13.28) es mayor que valor tabulado (3.841), por tanto  $H_0$  se rechaza.

**Conclusión.** Las víctimas de la enfermedad tienen mayor probabilidad de sufrir exposición versus individuos no-victimas de ña enfermedad, una vez emparejada por la edad y el género.

**2.- Modelo de Z**

**2.-  $Z = (B - C) / \sqrt{(B+C)}$**

La Tabla 5 demuestra el templete para el cálculo del valor de Z.

Tabla 5.- Templete para el cálculo de valor de Z		
Con enfermedad	Sin enfermedad	
	Si Exposición	No Exposición
Si Exposición	Celda I = a	Celda II = b
No Exposición	Celda III = c	Celda IV = d

Por tanto:  $Z = (58 - 24) / \sqrt{(58+24)} = 34 / \sqrt{82} = 34 / 9.05 = 3.756$

$Z_{calculado} = 3.756$ ,  $Z_{crítico} = \pm 1.96$  (desviación normal estándar),  $\alpha = 0.05$  (nivel de significancia), En la prueba de McNemar los grados de libertad siempre son 1.

**Resultado.** En razón a que el valor  $Z$  calculado (3.756) no se ubica dentro del rango  $\pm 1.96$  ( $\alpha=0.05$ ), se rechaza la hipótesis nula de no asociación entre exposición y la ocurrencia de la enfermedad.

**Conclusión.** Existe una asociación entre la ocurrencia de la enfermedad y el sufrir por exposición.

### **Odds ratio (OR)**

El modelo de  $\chi^2$  permite determinar la existencia de una asociación entre dos variables independientes en la escala nominal, además el modelo de McNemar hace lo mismo para variables emparejados dicotómicos, sin embargo, ninguno de estos dos modelos proporciona una medida de la magnitud de la asociación. Una medida que proporciona esta respuesta es el *odds ratio* o *relative Odds*.

La odds de ocurrencia de algo es la probabilidad de su ocurrencia dividido entre la probabilidad de su no ocurrencia:  $p / (1 - p)$ . En una muestra la odds de tener un resultado o efecto o evento (enfermedad) con el factor de riesgo es:  $a / b$ . Por ejemplo según los datos de la Tabla 6,  $Odds=a/b=716/4,984=0.144$ . Para los sujetos sin el factor del riesgo,  $Odds=c/d$ . Por tanto,  $Odds=458/5,490=0.083$ . Hay que resaltar que los *Odds* no se expresen en porcentaje.

El *Odds Ratio (OR)* es el Odds de tener la enfermedad cuando se tiene el factor del riesgo dividida entre el Odds de tener la enfermedad sin tener el factor del riesgo. De manera alterna, se puede definir el *OR* como los *odds* a favor de la exposición de los individuos enfermos dividido entre *odds* a favor de la exposición de los individuos no enfermos. Por tanto, se puede estimar el *OR* tanto para los estudios con cohorte como los estudios con caso-control.  $OR = (a/b) / (c/d) = 0.144 / 0.083 = 1.73$ . Por tanto, aquellos con el factor del riesgo tienen 1.73 veces el *Odds* de tener la enfermedad que aquellos sin el factor del riesgo. Hay que resaltar que se puede calcular el *Odds Ratio* para las muestras naturales, de tipo cohorte y también caso-control.

También se debe mencionar que el *OR* no es lo mismo que el riesgo relativo (*RR*). Recordar que  $OR = (a/b) / (c/d) = ad/bc$ . Mientras que el riesgo relativo o  $RR = [a/(a + c)]/[b/(b + d)] = a(b + d)/b(a + c)$ . Cuando los valores de *a* y *b* son muy pequeñas en relación a los valores de *c* y *d*, es decir, cuando la enfermedad es de rara ocurrencia, el valor de *RR* se aproxima al valor de *OR*, es decir,  $RR = a(b + d)/b(a + c) \approx ad/bc = OR$ .

El valor de 1.73 es un valor puntual y como tal no tiene sentido y no presente su significancia y por tanto, hay que presentar cualquier dato puntual con su intervalo de confianza, es decir un rango de valores alrededor de este valor donde indica con cierto nivel de probabilidad la ocurrencia de valor puntual dentro de este rango. Consecuentemente, se calcula el intervalo de confianza (IC) al 95% alrededor de *OR* de forma siguiente. Debido a que la distribución de la probabilidad de *OR* no es simétrica y esta sesgada hacia la derecha se debe calcular el intervalo de confianza en la escala de logaritmo natural. Es decir,

$$\ln(OR) \pm 1.96 EE [\ln(OR)],$$

Donde,

Ln = Logaritmo natural, y EE [ln(OR)] = Error Estándar,

$$\ln(OR) = \ln(1.73) = 0.548,$$

$$95\% \text{ IC} = (\ln(OR)/W) \text{ hasta } (\ln(OR)*W),$$

$$W = \exp^{(1.96*EE)},$$

Donde,

*exp* es la base del logaritmo natural, y *EE* es el error estándar.

$$EE = \sqrt{[(b + c) / (b * c)]} = \sqrt{[(4,984+458) / (4,984*458)]} = \sqrt{0.002383} = 0.0488.$$

$$W = \exp^{(1.96*EE)} =$$

$$W = \exp^{(1.96*0.0488)} = 1.10$$

$$95\% \text{ IC} = (\ln(OR)/W) \text{ hasta } (\ln(OR)*W).$$

$$95\% \text{ IC} = (0.548/1.10) \text{ hasta } (0.548*1.10) = 0.49 \text{ hasta } 0.60$$

Asumiendo que  $\ln(OR)$  tiene una distribución normal, se calcula el IC al 95% de forma siguiente.

$$95\% \text{ IC} = (OR / W) \text{ hasta } (OR * W).$$

$$W = \exp^{(1.96*EE)},$$

Donde,

$exp$  es la base del logaritmo natural, y  $EE$  es el error estándar.

$$EE = \sqrt{[1/a + 1/b + 1/c + 1/d]}.$$

$$EE = \sqrt{(1/716 + 1/4,984 + 458 / 5,484 = \sqrt{0.00396} = 0.0629}.$$

$$W = exp^{(1.96*0.0629)} = 1.13.$$

95% IC = (1.73/1.13) hasta (1.73\*1.13), es decir, 1.53 hasta 1.95.

En ambos casos los valores de IC al 95% no incluyen el valor de 1. Un valor igual a la unidad indicaría que el valor de  $OR$  es idéntico para los sujetos expuestos y los no expuestos.

### Estimación del riesgo

Un factor del riesgo (la causa o la exposición) puede incrementar la probabilidad de la ocurrencia de un resultado (efecto o enfermedad), por ejemplo, estar en lugares cerradas sin considerar las medidas recomendadas sanitarias en contra de la enfermedad (la Causa) aumenta la probabilidad de adquirir dicha enfermedad (el efecto).

El riesgo relativo ( $RR$ ) es igual al riesgo con exposición dividido entre el riesgo sin exposición. El ejemplo imaginario siguiente ilustra la menor de la estimación del riesgo. Vamos a asumir que tenemos una muestra del tamaño  $n$  con los datos arreglados en una tabla de contingencia de  $2 \times 2$ . La siguiente tabla (Tabla 6) demuestra la relación de los datos entre la enfermedad (el efecto) y el hábito de estar en lugares cerradas (la causa) sin medidas sanitarias (exposición) o no estar en dichos lugares (no exposición).

Tabla 6. Datos de enfermedad y exposición en una muestra de 11,648.			
Exposición	Enfermedad		Total
	Si	No	
Si	716 (a)	4,984 (b)	5,700 (a + b)
No	458 (c)	5,490 (d)	5,948 (c + d)
Total	1,174 (a + c)	10,474 (b + d)	GT=11,648(a + b + c + d)

De un total de 11,648 sujetos en el estudio, 5,700 fueron expuestos y el resto (5,948) no fue expuesto. El riesgo o la probabilidad de adquirir la enfermedad se estima como la proporción de los sujetos en cualquier grupo tanto expuesto como no expuesto que adquiere la

enfermedad. Por tanto aquí el riesgo ( $R$ ) en el grupo expuesto es:  $R=a/(a+b) = 716/5,700=0.126$  (12.6%) y en grupo no expuesto:  $R = c/(c+d)=458/5,948=0.077$  (7.7%).

*Riesgo relativo (RR)* es una medida de la fuerza de relación y es igual a la probabilidad de ocurrencia de la enfermedad en un grupo expuesto dividido entre la probabilidad de ocurrencia de la enfermedad en el grupo no-expuesto.  $RR = p(\text{enfermedad} \mid \text{expuesto}) / p(\text{enfermedad} \mid \text{no-expuesto})$ . Dicho de otra forma, el riesgo relativo o cociente del riesgo ( $RR$ ) es igual a la tasa del riesgo con el factor del riesgo (exposición) dividido entre la tasa del riesgo cuando sin el factor del riesgo (exposición). En el ejemplo arriba, el riesgo relativo ( $RR$ ) se calcula de forma siguiente.  $RR=[a/(a+b)]/[c/(c+d)] = 0.126/0.077=1.64$ . Por tanto los sujetos expuestos tienen 64% más probabilidad de adquirir la enfermedad en comparación con los no expuestos.

Un valor de  $RR$  igual a 1 indica que la probabilidad de desarrollar la enfermedad es igual para el grupo expuesto y el grupo no expuesto, un valor mayor y menor de uno indica mayor y menor probabilidad de desarrollar enfermedad, respectivamente, para el grupo expuesto en comparación con el grupo no expuesto.

El riesgo relativo solamente se calcula para las muestras naturales y también para los estudios con cohortes, pero no para los estudios de caso-control. La razón es que se estiman los riesgos para sujetos tanto expuestos y no expuestos, y en el caso de los estudios de caso-control, estos sujetos se seleccionan de forma arbitraria.

El riesgo histórico en individuos no-expuestos se denomina el *riesgo absoluto (RA)*. Otro indicador útil es el *riesgo atribuible* lo cual indica la cantidad adicional de enfermedad en la población debido a la exposición. El riesgo atribuible considera la carga de la enfermedad en la población y qué cantidad de enfermedad puede prevenirse si se elimina la exposición.  $\text{Riesgo atribuible} = (\text{incidencia en el grupo expuesto}) / (\text{incidencia en el grupo no-expuesto})$ . El riesgo atribuible es proporcional al grado de asociación entre el desarrollo de la enfermedad y la exposición. El riesgo atribuible, también indica la proporción de riesgo adicional en individuos expuestos y puede expresarse de forma siguiente,  $\text{Riesgo atribuible} = \{[(\text{Incidencia en el grupo expuesto}) - (\text{Incidencia en el grupo no-expuesto})]/(\text{Incidencia en el grupo expuesto})\}$ .

El valor de 1.64 para  $RR$  es un valor puntual y en la estadística hay que calcular un rango de error alrededor de este valor puntual. A este rango se denomina el intervalo de confianza al 95% (IC al 95%) y se calcula de manera siguiente.

IC al 95% =  $(RR/W)$  hasta  $(RR*W)$ ,

Donde,

$W = \exp(1.96*EE)$ ,

$\exp$  es la base de logaritmo natural,

$EE$  es el error estándar, lo cual se calcula como.  $EE = \sqrt{[b/(a*(a+b))] + [d/(c*(c+d))]} = \sqrt{[4984/(716(5,700))] + [5,490/(458(5,490))]} = 0.057$ .

Ahora,

$W = \exp(1.96*EE) = \exp(1.96*0.0567) = 1.12$ .

Por tanto,

IC al 95% =  $(RR/W)$  hasta  $(RR*W)$ , o  $(1.64/1.12)$  hasta  $(1.64*1.12)$  resultando en 1.46 hasta 1.84.

En este ejemplo, IC al 95% no incluye el valor de 1. Un valor igual a la unidad indicaría que el valor de  $RR$  es idéntico para los sujetos expuestos y los no expuestos.

### **Reducción del riesgo**

Reducción del riesgo en caso del riesgo relativo ( $RR$ ) se denomina  $RRR$  y se calcula como: (Incidencia en grupo tratado) / (Incidencia en grupo placebo), y la reducción del riesgo en caso del riesgo absoluto ( $RA$ ) se denomina  $RRA$  y se calcula como: (Incidencia en grupo placebo) - (Incidencia en grupo tratado). El valor de  $RRR$  puede ser muy grande pero en caso de  $RRA$  muy pequeña cuando se aplica a la población. Este ocurre especialmente con los tratamientos muy eficaces pero con bajo nivel de prevalencia de la enfermedad.

Basado en los datos de Logunov et al. 2021; Noordzi et al. 2017; Polak et al. 2020; Voysey et al. 2021 y Zhang et al. 2021; podemos calcular los valores de  $RRR$  y  $RRA$  para los 3 vacunas de Pfizer, Sputnik & Astra Zeneca.

**\*Ejemplo 1.** Caso de la vacuna de Pfizer contra el Covid-19.\* De un total de 43,448 voluntarios, 21,720 recibieron la vacuna y el resto (21,728) recibieron el placebo. De los vacunados 8

dieron síntomas positivas para el Covid-19 y de los que recibieron el placebo 162 dieron positivo. Por tanto, la proporción de los casos positivos en el grupo protegido con vacuna es  $8/21,720 = 0.0004$  y el valor correspondiente para el grupo no protegido (placebo) es  $162/21,728 = 0.0075$ .

La reducción en el riesgo relativo (*RRR*), lo que la compañía reclama como la eficacia es  $RRR = 0.0004/0.0075 = 0.0533 * 100 = 5.33\%$ . Por tanto la eficacia es  $100 - 5.33 = 94.67\%$ . En otras palabras, una eficacia de 95%. Esto indica que 95 de cada 100 personas vacunadas están protegidas contra el Covid-19.

La reducción en el riesgo absoluto (*RRA*) es  $RRA = 0.0075 - 0.0004 = 0.0071 * 100 = 0.71\%$ , es decir, la reducción del riesgo absoluto en la población es 0.71% o inferior al 1%. El Número Necesario a Tratar (*NNT*) para proteger una persona es:  $NNT = (1 / RRA) * 100 = (1/0.71) * 100 = 141$ . En otras palabras, se requiere vacunar mínimo 141 persona para evitar un acontecimiento, es decir, para proteger una persona. Ahora bien, aunque el valor de *NNT* no considera el costo y los efectos secundarios del uso de tratamiento, sin embargo, podemos formar criterios acerca de esta medida. Por ejemplo, si el tratamiento es seguro y barato, se recomendaría su uso aun con un valor alto de *NNT*.

Poner este concepto de *NNT* en perspectiva a nivel mundial con una población de cerca de 7.8 billones, entonces al vacunar al 100% de la población mundial tendremos  $7.8 \times 10^9 / 141 = 55319148$  personas protegidas, es decir menos de 1% (en realidad 0.7%) de la población global será protegida si vacunamos toda la población del mundo. En otras palabras, matemáticamente, es imposible proteger a totalidad de la población mundial contra el Covid a pesar de vacunar toda la población.

**\*Ejemplo 2.** Caso de la vacuna de Sputnik contra el Covid-19.\* De un total de 19,866 voluntarios, 14,964 recibieron la vacuna y 4,902 recibieron el placebo. El número de los casos positivos para los vacunados y para no vacunados fue de 16 y 26, respectivamente.

Por tanto, la proporción de los casos positivos en el grupo protegido con vacuna es  $16/14,964 = 0.001069$  y el valor correspondiente para el grupo no protegido (placebo) es igual a  $26/4,902 = 0.012647$ .

La reducción en el riesgo relativo (*RRR*), lo que la compañía reclama como la eficacia es  $RRR = 0.001069 / 0.012647 = 0.0845 * 100 = 8.45\%$ . Por tanto, la eficacia es  $100 - 8.45 = 91.6\%$ .

En otras palabras, una eficacia de 92%. Esto indica que 92 de cada 100 personas vacunadas están protegidas contra el Covid-19.

La reducción en el riesgo absoluto (*RRA*) es  $RRA = 0.012647 - 0.001069 = 0.0011578 * 100 = 1.2\%$ , es decir, la reducción del riesgo absoluto en la población es inferior al 2%. Aquí el valor de  $NNT = 100/1.2 = 83.3$ .

**\*Ejemplo 3.** Caso de la vacuna de Astra Zeneca contra el Covid-19.\* De un total de 11,636 voluntarios, 5,807 recibieron la vacuna y 5,829 recibieron el placebo. El número de los casos positivos para los vacunados y para no vacunados fue de 30 y 101, respectivamente.

Por tanto, la proporción de los casos positivos en el grupo protegido con vacuna es  $30/5,807 = 0.005166$  y el valor correspondiente para el grupo no protegido (placebo) es igual a  $101/5,829 = 0.017327$ .

La reducción en el riesgo relativo (*RRR*), lo que la compañía reclama como la eficacia es igual a  $0.005166 / 0.017327 = 0.2981 * 100 = 29.8\%$ . Por tanto, la eficacia es  $100 - 29.8 = 70.2\%$ . En otras palabras, una eficacia de 70%. Esto indica que 70 de cada 100 personas vacunadas están protegidas contra el Covid-19.

La reducción en el riesgo absoluto (*RRA*) es  $RRA = 0.017327 - 0.005166 = 0.012161 * 100 = 1.2\%$ , es decir, la reducción del riesgo absoluto en la población es inferior al 2%. Aquí el valor de  $NNT = 100/1.2 = 83.3$ .

Cabe resaltar que las compañías no proporcionan los datos de la *RRA* para que uno pueda tener una idea más precisa sobre la reducción del riesgo al usar la vacuna contra la enfermedad. Esto significa que hay que presentar tanto los valores de *RRR* como los de *RRA*, de lo contrario, al presentar solamente la reducción del riesgo relativo, tiende a sobreestimar el efecto. Uno no puede interpretar un riesgo sin tomar en cuenta el otro y además el riesgo absoluto da una mejor representación e información más relevante para el paciente en el sentido de que dada esta información real uno puede tomar una decisión más objetiva y racional sobre la aplicación de la vacuna (Noordzi et al. 2017). Este es una práctica común en las investigaciones en el campo de la salud, en donde no se reportan los valores de *ARR*. De hecho las normas de FDA claramente indican que en estos tipos de investigación se debe reportar y hacer público los datos de ambos *RRR* y *ARR*. El valor de *RRR*, llamada la *eficacia de la vacuna*, es relevante para la investigación, y no proporciona información importante

para la toma de decisión en parte de las autoridades de salud pública, en término de utilizar el tratamiento, mientras que el valor de  $ARR$ , sí es relevante para tomas de decisión en la salud pública. El valor de  $RRR$  es importante en término de resumir y comparar los resultados de los ensayos clínicos, sin embargo, no es relevante para los trabajadores clínicos, los médicos, las enfermeras, las autoridades de la salud pública, los políticos y la totalidad de la población.

Entonces la respuesta a la pregunta que 91.2 (caso Sputnik) o 70 (caso Astra Zeneka) o 95 (caso Pfizer) de 100 personas vacunadas van a estar protegidas del Covid-19 es correcto en término del  $RR$ . Sin embargo, en término absoluto y en la población, menos de 2 de cada 100 personas, están protegida en caso de cada una de las tres vacunas.

### **Capacidad de transmisión**

$R_0$  denominada el número reproductivo se usa para medir la capacidad de distribución una enfermedad.  $R_0$  es el promedio número de personas al cual una persona infectada transmite la infección. Por ejemplo, si una enfermedad tiene un  $R_0$  de 6, indica que una persona que tiene la enfermedad la transmitirá a un promedio de otras 6 personas.  $R_0$  proporciona una estimación de la posible expansión de la transmisión. El valor de la  $R_0$  para COVID-19 es una mediana de 2.2 a 2.7, Este significa que una persona con COVID-19 puede potencialmente transmitir el coronavirus a entre 2 o 3 personas. Por otro lado,  $R_0 = e^{rt}$ , donde,  $r$  = la tasa Instantánea natural del crecimiento,  $t$  = el tiempo,  $e$  = base de logaritmo natural. También se puede calcular la  $R_0$  de forma siguiente,  $R_0 = N_{t+1} / N_t$ , donde,  $N_t$  = el número de personas infectadas en el tiempo “ $t$ ”, y  $N_{t+1}$  = el número de personas infectadas en el tiempo “ $t+1$ ”.

Existen tres posibilidades de transmisión potencial o disminución de una enfermedad, dependiendo de su valor  $R_0$ .

Si  $R_0$  es menor que 1, cada infección existente causa menos de una nueva infección. En este caso, la enfermedad disminuirá y eventualmente desaparecerá.

Si  $R_0$  es igual a 1, cada infección existente provoca una nueva infección. La enfermedad se mantendrá viva y estable, pero no habrá un brote ni una epidemia.

Si  $R_0$  es más de 1, cada infección existente causa más de una nueva infección. La enfermedad se transmitirá entre personas y puede haber un brote o una epidemia.

Ahora en término de  $r$  o la tasa Instantánea natural del crecimiento, los valores de  $r$  menor de cero, igual a cero y mayor de cero, indican reducción, estado estacionario o incremento de la enfermedad, respectivamente.

El tiempo de duplicación ( $T_2$ ) de Covid-19 actualmente varia de 2.6 a 5.6, esto significa que el valor de “ $r$ ” es igual a 0.266 (para  $T_2 = 2.6$ ) a 0.123 (para  $T_2 = 5.6$ ), ya que  $T_2 = \ln 2/r$ , donde  $\ln 2 = 0.693$ .

### **Inmunidad del rebaño**

La inmunidad del rebaño ( $IR$ ) se obtiene cuando suficiente personas, en término de porcentaje, han adquirido inmunidad de forma natural o inducida, por ejemplo, por medio de la vacunación. La ecuación de  $IR$ :  $IR = 1 - (1/R_0)$ , donde,  $R_0$  = el número reproductivo. Por ejemplo, si el valor de  $R_0$  para el Covid es igual a 2.6, entonces  $IR = 1 - (1/2.6) = 1 - (0.38) = 0.62 * 100 = 62\%$ , o en caso de  $R_0 = 5.6$ ,  $IR = 1 - (1/5.6) = 1 - (0.178) = 0.822 * 100 = 82.2\%$ . Esto significa que se debe vacunar entre el 62% o el 82% de la población (dependiendo a la fuente de datos sobre el valor de  $R_0$ ) para obtener la inmunidad del rebaño.

### **Bases de los ensayos clínicos**

Todos los ensayos clínicos como el caso de las vacunas, deben tomar en cuanto los principios de ICH, es decir, Conferencia Internacional de Armonización (ICH). ICH-E6 Buenas Prácticas Clínicas (BPC). A continuación se mencionan de forma breve estos principios (<https://www.nejm.org/doi/full/10.1056/NEJMoa2034577>).

### **Principios de ICH E6**

Los ensayos clínicos son una parte fundamental de la investigación clínica que respalda el desarrollo de nuevos medicamentos o usos de medicamentos existentes. Los ensayos clínicos bien diseñados y realizados ayudan a responder preguntas clave en el cuidado de la salud y

el desarrollo de medicamentos. Sus resultados son esenciales para las decisiones de salud basadas en evidencia. Los ensayos con un diseño inadecuado y / o ensayos realizados de manera deficiente pueden poner en riesgo la seguridad de los participantes y producir pruebas inadecuadas o poco fiables. Desperdician recursos, esfuerzos y tiempo de investigadores y participantes.

Los principios de GCP están diseñados para ser flexibles y aplicables a una amplia gama de ensayos clínicos. Esta guía, junto con ICH E8, fomenta la consideración y la planificación cuidadosas para abordar aspectos específicos y potencialmente únicos de un ensayo clínico individual. Esto incluye la evaluación de las características del ensayo, como los elementos de diseño, el producto en investigación que se evalúa, la afección médica que se aborda, las características de los participantes, el entorno en el que se realiza el ensayo clínico y el tipo de datos que se recopilan. Se necesita una consideración cuidadosa de los factores relevantes para asegurar la calidad del ensayo para cada ensayo clínico.

Los principios están destinados a respaldar enfoques mejorados y más eficientes para el diseño y la realización de ensayos. Los diseños de ensayos clínicos que llevan el ensayo a los participantes y sus comunidades pueden mejorar la representatividad de la población participante y permitir una participación más amplia. El proceso de incorporación de la calidad en el diseño del ensayo puede ser apoyado por la participación de los directamente involucrados. Los diseños y procesos de prueba deben evaluarse para minimizar la complejidad y la carga innecesarias.

Los siguientes principios generales proporcionan un marco flexible para la realización de ensayos clínicos. Están estructurados para brindar orientación durante todo el ciclo de vida del ensayo clínico. Estos principios son aplicables a los ensayos que involucran a participantes humanos, es decir, pacientes o voluntarios sanos. Los principios son interdependientes y deben considerarse en su totalidad para asegurar la realización ética del ensayo y resultados confiables.

**1.-** Los ensayos clínicos deben realizarse de acuerdo con los principios éticos que tienen su origen en la Declaración de Helsinki y que son consistentes con las buenas prácticas clínicas (BPC) y los requisitos regulatorios aplicables.

**2.-** Los ensayos clínicos deben diseñarse y realizarse de manera que garanticen los derechos, la seguridad y el bienestar de los participantes.

- 3.- El consentimiento informado es una característica integral de la conducta ética de un ensayo. La participación en un ensayo clínico debe ser voluntaria y debe basarse en un proceso de consentimiento que garantice que los participantes estén bien informados.
- 4.- Los ensayos clínicos deben estar sujetos a una revisión objetiva por parte de una junta de revisión institucional (IRB) / comité de ética independiente (IEC).
- 5.- Los ensayos clínicos deben ser científicamente sólidos para su propósito previsto y estar basados en conocimientos y enfoques científicos sólidos y actuales.
- 6.- Los ensayos clínicos deben ser diseñados y realizados por personas calificadas.
- 7.- La calidad debe integrarse en el diseño científico y operativo y en la realización de los ensayos clínicos.
- 8.- Los procesos, medidas y enfoques de los ensayos clínicos deben ser proporcionales a los riesgos para los participantes y a la confiabilidad de los resultados del ensayo.
- 9.- Los ensayos clínicos deben describirse en un protocolo claro, conciso y operativamente factible.
- 10.- Los ensayos clínicos deben generar resultados fiables.
- 10.5.- Los ensayos clínicos deben incorporar procesos eficientes y bien controlados para la gestión de la información a través de una gestión adecuada de la integridad de los datos, la trazabilidad y la protección de la información personal, permitiendo así el reporte, interpretación y verificación precisos de la información relacionada con el ensayo clínico.
- 11.- Los roles, tareas y responsabilidades en los ensayos clínicos deben ser claros y documentados adecuadamente.
- 12.- Los productos en investigación utilizados en un ensayo clínico deben fabricarse de acuerdo con las normas de Buenas Prácticas de Fabricación (GMP) aplicables y deben almacenarse, enviarse y manipularse de acuerdo con las especificaciones del producto y el protocolo del ensayo.

## **Conclusión**

Se puede calcular los riesgos relativos y absolutos, utilizando los modelos estándares de epidemiología. Las reclamaciones de las empresas productoras de las vacunas con respecto al porcentaje de protección proporcionado por cada vacuna son correctas, sin embargo,

solamente en cuanto se trata del riesgo relativo. Es imperativo tener datos sobre los riesgos absolutos por su relevancia epidemiológica para la totalidad de la población. Hay que resaltar que reducción del riesgo relativo es 95%, 92% y 70% para las vacunas de Pfizer, Sputnik y Astra Zeneca, respectivamente. Sin embargo la reducción del riesgo absoluto es de 0.71%, 1.2% y 1.2% para las vacunas de Pfizer, Sputnik y Astra Zeneca, respectivamente. Además para las vacunas de Pfizer, Sputnik y Astra Zeneca, se requiere vacunar a 141, 83 y 83 individuos, respectivamente, para poder dar protección a un individuo. La inmunidad de rebaño se obtiene cuando entre 62% a 82% (dependiente de los valores de  $R_0$ ) adquieren inmunidad por cualquier forma.

## Referencias

- Armonización de la PCI para una mejor salud. 19 de abril de 2021. 7pp.  
<https://www.nejm.org/doi/full/10.1056/NEJMoa2034577>.
- Badii, M.H., J. Castillo, J. Landeros & K. Cortez. 2007a. Papel de la estadística en la investigación científica. *InnOvaciones de Negocios*. 4(1): 107-145.
- Badii, M.H., J. Castillos, R. Foroughbakhch & K. Cortez. 2007b. Probability and scientific research. *Daena*, 2(2): 358-369.
- Badii, M.H., J. Castillo, M. Rodríguez, A. Wong & P. Villalpando. 2007c. Diseños experimentales e investigación científica. *InnOvaciones de Negocios*, 4(2): 283-330.
- Badii, M.H. & J. Castillo. 2008. Una tabla estadística para los K-valores. *CULCYT*, 5(29): 24-29.
- Badii, M.H., J. Castillo & A. Wong. 2008. Uso de análisis de covarianza (ANCOVA) en investigación científica. *InnOvaciones de Negocios*, 5(1): 25-38.
- Badii, M.H. & J. Castillo. 2009. Distribuciones probabilísticas de uso común. *Daena*, 4(1): 149-178.
- Badii, M.H. & A. Guillen. 2010. Decisiones estadísticas: bases teóricas: *Daena international J. Good Conscience*. 5(1): 185-207.
- Badii, M.H., A. Guillen & L.A. Araiza. 2010. Estimaciones estadísticas: Un acercamiento analítico. *Daena international J. Good Conscience*. 5(1): 237-255.

- Badii, M.H., A. Guillen, E. Cerna & J. Landeros. 2011. Dispersión espacial: El requisito esencial para el muestreo. *Daena International J. of Good Conscience*. 6(1): 40-71.
- Badii, M.H., A. Guillen & J.L. Abreu. 2014. Tamaño Óptimo de Muestra en Ciencias Sociales y Naturales. *Daena* 9(2):41-51.
- Logunov et al. 2021. *The lancet*. Doi: [https://doi.org/10.1016/50140-6737\(21\)00234-8](https://doi.org/10.1016/50140-6737(21)00234-8).
- McNemar, Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 12: 153 – 157.
- Noordzi et al. 2017. *Nephrol Dial Transpl* <http://doi.org/10.1093/ndt/gfw456>.
- Pearson, K. 1904. *Mathematical contributions to the theory of evolution*. XIII. On the theory of contingency and its relation to association and normal correlation. *Draper's Co. Res. Mem. Biometric Ser.* 1. 35.
- Polak et al. 2020. *The New Englan J Medicine*. Doi: 10.1056/NEJMoa2034577.
- Voysey et al. 2021. *The lancet*. Doi: [https://doi.org/10.1016/50140-6737\(20\)32661-81](https://doi.org/10.1016/50140-6737(20)32661-81).
- Zhang et al. 2021. *The lancet*. Doi: [https://doi.org/10.1016/51473-3099\(20\)30843-84](https://doi.org/10.1016/51473-3099(20)30843-84).