



Pérez, Natalia

Kovalevski, Leandro

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística

ESTUDIO DE POTENCIA EN DISEÑOS “CLUSTER CROSS-OVER” CON RESPUESTA BINARIA.

Introducción

En distintas áreas de investigación, usualmente se persigue el objetivo de evaluar el efecto que produce una determinada intervención (tratamiento, terapia, método de enseñanza, etc.) sobre alguna variable respuesta. Los ensayos clínicos, estudios prospectivos en humanos comúnmente utilizados al momento de probar la eficacia de algún tratamiento, son una herramienta posible para evaluar los efectos de las distintas intervenciones que se asignan al azar a los diferentes individuos.

En los ensayos clínicos los individuos son aleatorizados a recibir o no la intervención para lograr formar grupos comparables con respecto a factores de riesgo conocidos y desconocidos. La aleatorización elimina el sesgo del investigador en la asignación de los participantes a la intervención, y garantiza que las pruebas estadísticas tengan niveles de significación válidos. La aleatorización de los tratamientos generalmente se realiza a nivel individual aunque puede realizarse en forma grupal. Estos grupos comúnmente se denominan “clusters” o conglomerados, en este trabajo se utilizarán indistintamente los términos cluster, conglomerados o grupos.

La utilización de diseños con aleatorización por grupos es ventajosa cuando la asignación aleatoria sobre individuos no es viable por cuestiones logísticas, económicas, éticas o bien para minimizar la “contaminación” por cruce de información entre sujetos asignados a distintos tratamientos. Y a menudo se utiliza cuando los individuos se encuentran en estrecha proximidad, por ejemplo, en la misma comunidad, institución o práctica médica.

Una de las principales desventajas de un diseño los conglomerados es que los participantes dentro de cualquier grupo tienden a ser más parecidos y a responder de una manera similar. Por lo tanto, al analizar los datos ya no se pueden suponer independientes unos de otros. Por ejemplo, en el caso de que la aleatorización sea por escuelas, los alumnos pueden compartir vivencias, hábitos nutricionales, costumbres, por lo que los resultados serán más similares entre sí que en las otras escuelas. Es ahí donde es importante involucrar el concepto de correlación intra-grupo o de varianza entre los conglomerados para describir el grado en que dos miembros de un grupo presentan más semejanzas que dos de distintos grupos.

Dentro de los diseños aleatorizados por conglomerados se pueden mencionar los diseños en **paralelo**, donde los clusters son aleatorizados mitad al tratamiento A y la otra mitad al tratamiento B; los diseños **cross-over o cruzado**, donde cada grupo recibe las dos intervenciones en dos períodos



diferentes, en un orden aleatorio, por lo que la duración del estudio siempre será mayor que la de un diseño en paralelo; y **los diseños stepped wedge o secuencial** donde los grupos comienzan el estudio en un tratamiento (generalmente el control) y en distintas etapas se aleatorizan secuencialmente para recibir el otro tratamiento (que se supone o puede suponer de mayor efectividad).

En este trabajo centraremos nuestro estudio sobre los **diseños cross-over (cruzado)**, los cuales son utilizados cada vez con mayor frecuencia (Reich N. G. y otros 2012) y por lo tanto se torna importante avanzar en su estudio. Un paso crítico en la fase de planificación y desarrollo de los ensayos clínicos es el cálculo del tamaño de la muestra. Para muchos diseños y metodologías de análisis se dispone o se puede deducir una manera analítica para determinar el tamaño de muestra. Sin embargo, esa deducción no siempre es simple cuando los ensayos clínicos tienen diseños o son analizados con modelos estadísticos complejos como es el caso de los **estudios aleatorizados por cluster de diseño cross-over**.

Los métodos de prueba para los diseños por conglomerado son relativamente menos desarrollados que para los tradicionales ensayos aleatorizados individuales. La característica principal de éstos es que requieren una metodología adecuada que considere la correlación entre las observaciones dentro de los conglomerados. Existen algunos estudios que desarrollan en profundidad estos diseños como el de Hughes (2003) donde se deduce una fórmula general para determinar el tamaño de muestra de manera analítica, práctica y útil para los distintos diseños por conglomerados presentados (paralelo, cross-over, stepped wedge) cuando se resume la información a nivel de cluster. Sin embargo cuando se prefiere analizar datos a nivel individual con resultados binarios son menos los avances realizados, y no se encuentran muchos trabajos que profundicen sobre la determinación del tamaño muestral y el cálculo de potencias asociadas a las pruebas de hipótesis.

Para estudios de diseños complejos, como el cluster cross-over con respuesta binaria, no existen fórmulas sencillas para calcular la potencia que reflejen adecuadamente la variabilidad de los datos. En estos casos, los métodos de simulación son una alternativa válida para la estimación de la potencia. Además, estos métodos tienen añadida la ventaja de poder ser aplicados en distintas situaciones o escenarios tanto simples como complejos.

Los métodos de estimación mediante simulación generan numerosos conjuntos de datos aleatorios, donde cada uno representa una versión hipotética del estudio para ser llevado a cabo. Los conjuntos de datos son generados asumiendo que una hipótesis alternativa específica es verdadera (generalmente que existe efecto tratamiento). En cada uno de esos conjuntos de datos se aplica la metodología de análisis pensada y se registra si se detecta ese efecto simulado y la estimación (empírica) de la potencia resulta de la proporción de veces que se detecta ese efecto específico en los conjuntos de datos generados.



Metodología

Un diseño en cluster cross-over en donde se comparan dos tratamientos en dos períodos se puede especificar (Parianti y Kuss. (2006)) con: el número de clusters en el ensayo (C) y el número total de sujetos en el estudio (N). Cada cluster ($i = 1, \dots, C$) representa una unidad de asignación al inicio del ensayo clínico. Al comienzo, cada cluster es asignado al grupo tratamiento o control. Luego de un cierto período de tiempo o de cierta cantidad de períodos, cada cluster es reasignado al grupo contrario. Cuando hay solo dos períodos ($J = 2$) cada cluster recibe ambos tratamientos y lo que se aleatoriza es el orden en que reciben los tratamientos. El tamaño de muestra total es la suma

de la cantidad de individuos por cluster en cada período
$$N = \sum_j \sum_i n_{ij}$$
 si $n_{ij} = n, \forall ij$, entonces $N = 2.C.n$. Se podría generalizar para más de dos períodos.

En estos estudios es fundamental que exista un período de lavado "washout" entre la aplicación del tratamiento y control o viceversa, dentro de cada cluster, para que la ejecución de la segunda intervención no se vea afectada por la anterior. También es importante que la condición (por ejemplo, la enfermedad que se quiere tratar) no cambie a través del tiempo.

En los diseños en cluster, es conveniente que al analizar la variable respuesta sobre unidades pertenecientes a un mismo cluster, se tenga en cuenta que las observaciones están correlacionadas entre sí. Esta correlación se denomina correlación intracluster (ICC) y se simboliza con ρ_c .

La variancia total de la variable binaria se define como la suma de variancia entre cluster (σ_c^2) y variancia dentro del cluster (σ_a^2), $\sigma_y^2 = \sigma_c^2 + \sigma_a^2$. Y la correlación intracluster $\rho_c = \frac{\sigma_c^2}{\sigma_y^2}$, siendo $\sigma_c^2 = \frac{\sum_i (\pi_i - \pi)^2}{C}$, $\sigma_a^2 = \pi(1 - \pi)$ donde π es la probabilidad de éxito poblacional y π_i es la probabilidad de éxito en el cluster i . El ICC refleja la proporción de la variancia total explicada por la variancia entre cluster (Hayes, Moulton, 2009).

Metodología propuesta por Hughes (2003)

Este análisis propone resumir la respuesta de interés a nivel de conglomerado y ajustar el siguiente modelo para describir y_{ij} , la respuesta resumen en el conglomerado i en el tiempo j :

$$y_{ij} = \mu + \alpha_i + \gamma_j + \beta x_{ij} + \varepsilon_{ij}$$

donde y_{ij} es el promedio (o la proporción en el caso de respuesta binaria) de la respuesta en el cluster i para el período j , μ representa a la media general de la respuesta de la población,



$\alpha_i \sim N(0, \sigma_c^2)$ es un efecto aleatorio del conglomerado ($i = 1, C$), γ_j es un efecto arbitrario del tiempo ($j = 1, J$) y β es el efecto del tratamiento, el parámetro de interés. Los $\epsilon_{ij} \sim N(0, \sigma_\alpha^2)$.

Los valores x_{ij} corresponden a una variable indicadora del tratamiento tal que

$$x_{ij} = \begin{cases} 0 & \text{si el conglomerado } i, \text{ en el tiempo } j, \text{ recibe el control} \\ 1 & \text{si el conglomerado } i, \text{ en el tiempo } j, \text{ recibe el tratamiento} \end{cases}$$

y σ_c^2 representa la variabilidad "entre-conglomerados".

Para probar el efecto tratamiento se plantea la siguiente hipótesis

$$H_0: \beta = 0$$

$$H_1: \beta = \beta_\alpha \text{ con } \beta_\alpha \neq 0$$

Y la estadística de prueba:

$$z = \frac{\hat{\beta} - \beta}{\sqrt{\text{Var}(\hat{\beta})}} \sim N(0, 1)$$

Este análisis hizo foco en el cálculo de la potencia para el test del efecto tratamiento ($H_0: \beta = 0$), partiendo de la variancia del efecto tratamiento bajo el modelo propuesto (Hughes 2003).

$$V(\hat{\beta}) = \frac{C \sigma_y^2 (\sigma_y^2 + J \sigma_c^2)}{(CV - W) \sigma_y^2 + (V^2 + CJV - JW - CU) \sigma_c^2}$$

donde:

C : cantidad de clusters ,

J : número de periodos ,

$$V = \sum_{ij} X_{ij} = C$$

$$W = \sum_j \left(\sum_i X_{ij} \right)^2 = 2 \cdot \left(\frac{C}{2} \right)^2$$

$$U = \sum_i \left(\sum_j X_{ij} \right)^2 = C$$

llegando a la potencia estimada:



$$Potencia = \Phi^{-1}\left(\sqrt{\frac{\beta_a^2}{Var(\hat{\beta})}} + Z_{\frac{\alpha}{2}}\right) + \Phi^{-1}\left(\sqrt{\frac{\beta_a^2}{Var(\hat{\beta})}} - Z_{\frac{\alpha}{2}}\right) \quad (2)$$

donde Φ es la función de distribución normal, Z_p es el percentil p de la distribución normal y β_a efecto tratamiento bajo H_1 .

Metodología de evaluación del efecto a nivel individual

En un diseño de cluster cross-over se analizan C clusters, en J periodos de estudio, con N_{ij} participantes en el i -ésimo cluster y el j -ésimo período. Se define Y_{ijk} como la variable aleatoria binaria representando la presencia o ausencia del resultado de interés para el k-ésimo individuo en el i -ésimo cluster durante el j -ésimo período del estudio.

Algunos estudios (Milanés V. 2010, Reich N. G. 2012, Parianti J. 2006) que compararon distintos métodos estadísticos de análisis en un ensayo clínico de diseño cruzado con aleatorización por cluster cuando la variable respuesta es binaria proponen utilizar el modelo de regresión logística con efectos aleatorios. Por lo tanto, para simular los datos se plantea un modelo lineal generalizado mixto de la siguiente forma:

$$\text{logit}(\pi_{ijk}) = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \alpha_i + \beta X_{ij} \quad (1)$$

donde β es el efecto tratamiento fijo, α_i es el efecto de cluster aleatorio siguiendo una distribución normal $N(0, \sigma_e^2)$ y X_{ij} es una variable indicadora para determinar si el cluster i durante el período j es asignado al tratamiento [$X_{ij} = 1$] o al control [$X_{ij} = 0$]. Vale destacar que por simplicidad se supone que no existe efecto período.

Para probar el efecto tratamiento se plantea la siguiente hipótesis

$$H_0: \beta = 0$$

$$H_1: \beta = \beta_a \text{ con } \beta_a \neq 0$$

Y la estadística de prueba es el test de razón de verosimilitud, que consiste en comparar las log-verosimilitudes maximizadas de dos modelos, el modelo reducido ($\beta = 0$) con el modelo completo (β_a).

$$G = -2(\hat{l}_{red} - \hat{l}_{comp}) \text{ que bajo } H_0 \text{ se distribuye } \chi_{gl}^2$$

Donde \hat{l}_{red} es la log-verosimilitud del modelo reducido, \hat{l}_{comp} es la log-verosimilitud del modelo completo y gl es la diferencia de parámetros de los dos modelos, que en este caso es igual a 1.



Estudio por simulación

Para estimar mediante un estudio de simulación la potencia en distintos escenarios de un diseño en cluster cross-over cuando la respuesta es binaria como una alternativa para el cálculo del tamaño de muestra, se simularon distintos conjuntos de datos para comparar diferentes valores de los parámetros y evaluar el comportamiento de la potencia.

Los datos fueron generados bajo un diseño de cluster cross-over para comparar dos tratamientos en dos períodos suponiendo igual cantidad de individuos por cluster en cada período para todos los cluster.

La respuesta binaria se simuló bajo la estructura del modelo de regresión logística:

$$\text{logit}(\pi_{ijk}) = \alpha_i + \beta X_{ij}$$

Donde, α_i es el efecto aleatorio del i -ésimo cluster que permite que las observaciones dentro de un mismo cluster estén correlacionadas $\alpha_i \sim N(0, \sigma_c^2)$, β es el efecto tratamiento y X_{ij} es el indicador de tratamiento dado por la combinación período-cluster que toma el valor 1 en caso de haber aplicado el tratamiento y el valor 0 para el control.

Para estudiar cómo varía la potencia se generaron distintos escenarios variando el efecto tratamiento (β) y la variabilidad entre cluster (σ_c^2) para distinta cantidad de clusters (C) y distinta cantidad de individuos por período-cluster (n_{ij}).

Para el efecto tratamiento (β) se eligieron arbitrariamente los valores 0; 0,25; 0,50; 0,75 y 1. Los valores de efecto tratamiento se combinaron con diferentes valores de la variabilidad entre los cluster $\sigma_c^2 = 0, 1, 2 \text{ y } 3$, cuando σ_c^2 toma el valor 0 se está considerando que las observaciones de un mismo cluster son independientes. Las distintas cantidades de cluster (C) elegidas fueron 5, 10, 15, 20 y 25, y distintas cantidades de observaciones por período-cluster (n_{ij}) fueron 5, 10, 15, 20, 25, 30, 35, 40, 45 y 50. Generándose en total 1000 escenarios diferentes y para cada escenario se generaron 1000 iteraciones.

Para generar la respuesta binaria, en cada escenario se llevan a cabo los siguientes pasos:

- I. Se genera la estructura de los datos según el diseño cluster cross-over donde se restringió solo a dos períodos y dos tratamientos.
- II. Efecto aleatorio de los clusters se genera a través de una distribución normal

$$\alpha_i \sim \text{Normal}(0, \sigma_c^2)$$



- III. Se determinan los valores del $\text{logit}(\pi_{ijk}) = \alpha_i + \beta X_{ij}$ y luego, por ende, los de la probabilidad $\pi_{ijk} = \frac{\exp(\alpha_i + \beta X_{ij})}{1 + \exp(\alpha_i + \beta X_{ij})}$.
- IV. Se generan valores al azar de una distribución uniforme definida en el intervalo $[0,1]$ ($U(0,1)$) y se comparan con los valores π_{ijk} obtenidos en el paso anterior.
- V. Si el valor de π_{ijk} es mayor que la variable $U(0,1)$ generada entonces se asigna a esa observación el valor 1 para la respuesta binaria. En caso contrario, se le asigna el valor 0. De esta forma los valores altos de la probabilidad generan valores de la respuesta igual a 1 con alta probabilidad y valores bajos de probabilidad generan valores de la respuesta igual a 1 con baja probabilidad.

Para el cálculo de la potencia que es el resultado de interés, en cada iteración se ajusta el modelo (1) a los datos simulados y se crea una variable indicadora que toma el valor 1 cuando se rechaza la hipótesis nula de no efecto tratamiento $H_0: \beta = 0$. La potencia para cada escenario se estima a través de la proporción de rechazos (en el test de razón de verosimilitud) en las 1000 iteraciones.

El valor simulado de $\beta = 0$ significa que no hay diferencia entre los grupos, es decir que la proporción de eventos simulada en los grupos tratamiento y control sería la misma, y por lo tanto las potencias estimadas serán estimaciones de la probabilidad de error de tipo I (rechazar H_0 cuando es cierta).

La proporción de eventos de interés esperada para la respuesta binaria del grupo control se fijó en todos los escenarios en 0,5 (modelo sin intercepto) y sólo se varió la proporción de eventos esperados en el grupo tratado. Se podrían calcular las potencias estimadas para nuevos escenarios donde la proporción de eventos esperados en el grupo control sea distinta de 0,5 siguiendo la metodología propuesta en este trabajo. Otra de las consideraciones en la simulación de los datos fue suponer que no existe interacción entre el periodo y el cluster.

Dado que en una etapa inicial se realizaron ensayos con 1000, 2000 y 5000 iteraciones para diversas situaciones, y los resultados de potencia alcanzados fueron muy similares, se halló conveniente la reducción de tiempo mediante la utilización de 1000 iteraciones en cada escenario.

Para el análisis estadístico de los conjuntos de datos simulados se utiliza el software SAS 9.2, la programación en lenguaje IML para la generación de datos, y PROC GLIMMIX para el ajuste de los modelos.

Resultados

En esta sección se presentará una breve descripción de la respuesta binaria simulada en algunos escenarios, luego el análisis de la potencia estimada en todos los escenarios planteados y



finalmente la potencia calculada a partir de la metodología de Hughes para los diseños en cluster cross-over.

En la *Tabla 1* se resume la proporción de eventos de la variable respuesta binaria generada en la simulación de los datos. Se presenta, a modo de ejemplo, una de las 1000 simulaciones para el escenario con 10 clusters, con un efecto tratamiento de 0,5, con 30 individuos por cluster-periodo y con las 4 opciones propuestas para la variabilidad entre clusters: 0, 1, 2 y 3. En esta tabla se puede apreciar claramente el efecto tratamiento, siendo la proporción de eventos (en promedio) mayor en el grupo tratado para los cuatro escenarios con distinta variabilidad entre clusters. En el caso en que los datos se simularon sin variabilidad entre clusters la diferencia promedio de la proporción de eventos entre los tratados y el control es 21% mientras que en los tres escenarios que se simuló variabilidad entre clusters la diferencia fue del 10%.

También podemos observar que cuando los datos se simularon sin variabilidad entre cluster la diferencia entre tratados y no tratados varía de un 13% a favor de los controles (cluster 9) a un 50% a favor de los tratados (cluster 5). Mientras cuando la variabilidad entre cluster fue igual a 1, la diferencia entre tratados y no tratados varía de un 10% a favor de los controles (cluster 10) a un 34% a favor de los tratados (cluster 8). Cuando la variabilidad simulada fue 2, la diferencia entre tratados y no tratados varía de un 3% a favor de los controles (cluster 7) a un 23% a favor de los tratados (cluster 1). Y por último cuando la variabilidad simulada fue 3, la diferencia entre tratados y no tratados varía de un 6% a favor de los controles (cluster 3) a un 24% a favor de los tratados (cluster 1). En ese último escenario con variabilidad simulada igual a 3 se observa que hay clusters con proporciones de eventos muy altas en tratados y no tratados como los clusters 2 y 4, y también hay clusters con proporciones de eventos bajas en los dos grupos como los clusters 1 y 6. Esta situación es contraria al escenario sin variabilidad entre clusters donde no hay porcentajes tan extremos y justamente está reflejando la variabilidad entre los clusters.



Tabla 1: Porcentaje de eventos simulados para el escenario con 10 clusters, con un efecto tratamiento de 0,5 y con 30 individuos por cluster-periodo, en los distintos clusters para las distintas opciones de la variabilidad entre clusters: 0, 1, 2 y 3

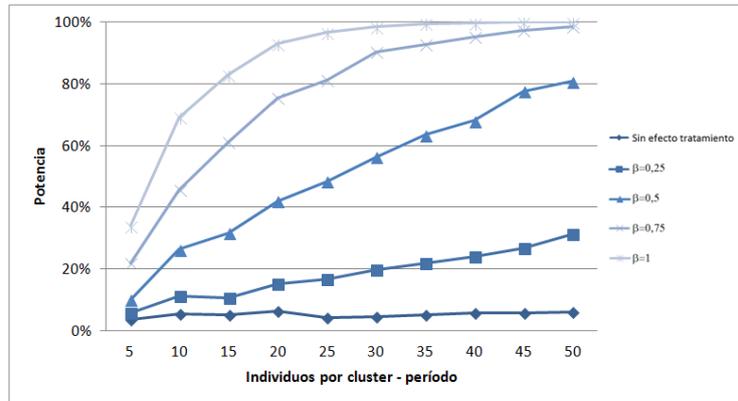
Cluster	Variabilidad entre cluster							
	0		1		2		3	
	Control	Tratamiento	Control	Tratamiento	Control	Tratamiento	Control	Tratamiento
1	40%	60%	80%	83%	67%	90%	3%	27%
2	43%	87%	30%	63%	90%	90%	90%	93%
3	37%	47%	67%	70%	47%	63%	93%	87%
4	53%	57%	63%	80%	70%	80%	97%	97%
5	20%	70%	37%	33%	37%	40%	40%	60%
6	37%	60%	27%	40%	50%	57%	23%	27%
7	53%	67%	77%	77%	93%	90%	37%	60%
8	40%	63%	13%	47%	40%	60%	63%	80%
9	60%	47%	50%	53%	70%	87%	63%	83%
10	40%	73%	70%	60%	10%	10%	90%	87%
Total	42%	63%	51%	61%	57%	67%	60%	70%

Para analizar la potencia en distintas situaciones, se presenta la *Tabla 2* donde quedan resumidos los 1000 escenarios planteados: los cinco valores elegidos para el efecto tratamiento (β): 0; 0,25; 0,50; 0,75 y 1, combinados con los diferentes valores de la variabilidad de los cluster (σ_c^2): 0, 1, 2 y 3, combinados a su vez con las distintas cantidades de cluster (C): 5, 10, 15, 20 y 25, y con las 10 distintas cantidades de observaciones por periodo-cluster (n_{ij}): 5, 10, 15, 20, 25, 30, 35, 40, 45 y 50. A partir de la *Tabla 2* se observa que cuando hay 15 clusters; un efecto tratamiento de 0,5, no habiendo variabilidad entre clusters y con 25 individuos por cluster-período, la potencia estimada es 93%.

En la *Tabla 2* también podemos apreciar que, como era de esperar, la cantidad de individuos por cluster aumenta, la potencia estimada crece. Como ejemplo, en el caso de 5 clusters que no tienen variabilidad (*Gráfico 1*) donde para los cinco valores elegidos para el efecto tratamiento (β): 0; 0,25; 0,50; 0,75 y 1, se ve claramente el crecimiento de la potencia (de 10% a 81% en el caso de un efecto tratamiento de 0,5) a medida que aumenta la cantidad de individuos por cluster. Este comportamiento persiste cuando hacemos el análisis para las cantidades de cluster (C) propuestas (5,10,15,20 y 25) y para los diferentes valores de la variabilidad de los cluster (σ_c^2): 0, 1, 2 y 3. En el *Anexo 1* se pueden ver los gráficos correspondientes para cada escenario en estudio.



Gráfico 1: Potencia estimada para cada uno de los valores de efecto tratamiento, en las diferentes cantidades de individuos por cluster, en un escenario con 5 clusters sin variabilidad entre clusters

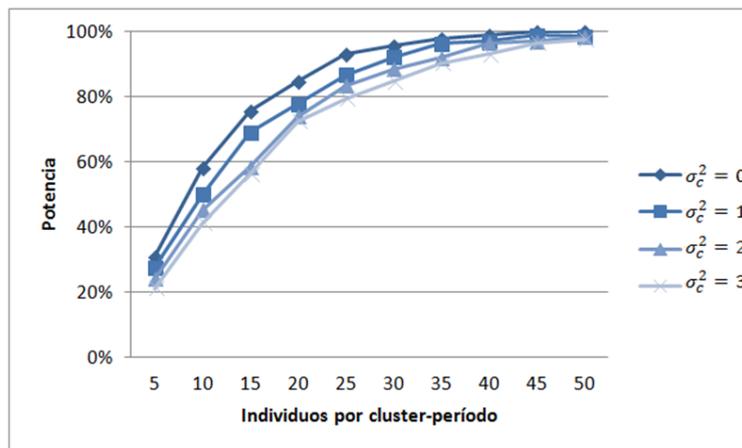


También en la *Tabla 2* podemos observar cómo la potencia estimada crece rápidamente para valores altos del efecto tratamiento (β) y cantidad de individuos por cluster (n). En el caso de 5 clusters, un efecto de 0,25 ($\beta = 0,25$) y cuando las observaciones son independientes ($\sigma_c^2 = 0$), la potencia estimada para la mayor cantidad de individuos por periodo-cluster propuesta ($n = 50$) no supero el 31%. Pero a medida que aumenta la diferencia entre los grupos, la potencia crece rápidamente para valores altos de individuos por período cluster (n). Por ejemplo cuando ($\beta = 0,75$) y tenemos 30 individuos por periodo cluster ($n = 30$) se alcanza una potencia del 91% y luego en ($\beta = 1$) con 20 individuos por periodo cluster ($n = 20$) ya se llega a una potencia del 93%. Cuando la diferencia entre los grupos tratados es grande ($\beta = 1$) para más de 35 individuos por periodo cluster ($n \geq 35$) la potencia toma valores del 100%.

Al analizar la variabilidad entre clusters, se destaca que a mayor variabilidad (σ_c^2 crece), es decir, cuando los individuos dentro de los clusters son más parecidos, la potencia disminuye. Aunque esa disminución no es tan pronunciada, ya que cuando el efecto y la cantidad de individuos por periodo-cluster son altos (cuando $\beta \geq 0,75$ y $n > 35$) se alcanzan potencias del 100%. Para cuantificar esta disminución presentamos el *Gráfico 2* con el ejemplo de 15 clusters, un efecto tratamiento de 0,5 ($\beta = 0,5$) y los individuos por cluster varían (están en el eje horizontal del gráfico). Cuando se considera que no hay variabilidad entre los clusters se tiene una potencia estimada de 31%, en el mismo escenario pero cuando la variabilidad entre los clusters es 1 ($\sigma_c^2 = 1$) tenemos una potencia estimada del 28%, cuando la variabilidad entre los clusters es 2 ($\sigma_c^2 = 2$) tenemos una potencia estimada del 24% y cuando la variabilidad entre los clusters es 3 ($\sigma_c^2 = 3$) la potencia estimada disminuyo al 21%.



Gráfico 2: Potencia estimada para los diferentes valores de variabilidad de los cluster, en las diferentes cantidades de individuos por cluster, en un escenario con 15 clusters y con un efecto tratamiento de 0,5



Las potencias estimadas de la *tabla 2* resultan de gran utilidad porque nos permiten determinar el tamaño de muestra necesario para un efecto tratamiento dado. Supongamos que queremos llevar a cabo un estudio para probar un efecto tratamiento que se supone igual a 0,50, es decir, que se busca comparar una proporción de eventos del 50% en el grupo control contra una

proporción de eventos del 62% $\left(\frac{\exp(0,50)}{1 + \exp(0,50)} \right)$ en el grupo tratado; si se busca una potencia del 90% y solo se pueden reclutar individuos de 15 cluster, se necesitarán por lo menos 25 individuos por cluster cuando no haya variabilidad entre cluster, 30 individuos por cluster cuando la variabilidad es igual a 1 y 35 individuos por cluster si la variabilidad es 2 o 3. De la misma manera las potencias estimadas en la *tabla 2* pueden ser utilizadas para el cálculo del tamaño de muestra en escenarios con diferentes efectos tratamientos, cantidades de cluster y variabilidad entre cluster.

Cabe destacar que cuando no hay diferencia entre los grupos ($\beta = 0$), es decir cuando la proporción de eventos simulada en los grupos tratamiento y control es la misma, las potencias estimadas son en realidad estimaciones de la probabilidad de error de tipo I (rechazar H_0 cuando es cierta). En la *Tabla 2* se tienen valores para esas potencias cercanos al 5% ya que se trabajó con una probabilidad de error de 0,05.

Potencias a través de la metodología propuesta por Hughes (2003)

Hughes (2003) propone resumir la respuesta de interés a nivel de conglomerado y sugiere una fórmula para estimar el tamaño de muestra válida para distintos diseños, que al aplicarla en un diseño cluster cross-over el efecto no tiene en cuenta la variabilidad entre clusters



En la *Tabla 3* podemos observar la potencia estimada mediante la metodología propuesta por Hughes para algunos de los escenarios propuestos en este trabajo, los mismos están constituidos por las siguientes combinaciones: los cinco valores elegidos para el efecto tratamiento (β) 0; 0,25; 0,50; 0,75 y 1, las distintas cantidades de cluster (C): 5, 10, 15, 20 y 25, y las 10 distintas cantidades de observaciones por periodo-cluster (n_{ij}): 5, 10, 15, 20, 25, 30, 35, 40, 45 y 50. No fueron considerados los diferentes valores de la variabilidad de los cluster dado que Hughes propone una estimación de potencia que no es sensible a la variabilidad del efecto aleatorio de los clusters.

Tabla 3: Potencia estimada mediante la metodología propuesta por Hughes (en porcentajes) para los escenarios formados por las combinaciones de cluster (C): 5, 10, 15, 20 y 25, individuos por cluster (n_{ij}): 5, 10, 15, 20, 25, 30, 35, 40, 45 y 50 y efecto tratamiento (β) 0; 0,25; 0,50; 0,75 y 1.

Número de clusters	Efecto tratamiento	Cantidad de individuos por cluster-periodo									
		5	10	15	20	25	30	35	40	45	50
5	0	5	5	5	5	5	5	5	5	5	5
	0,25	7	9	11	14	16	18	20	23	25	27
	0,5	14	24	33	42	50	57	64	70	75	79
	0,75	28	49	66	78	86	92	95	97	98	99
	1	45	74	89	96	98	99	100	100	100	100
10	0	5	5	5	5	5	5	5	5	5	5
	0,25	9	14	18	23	27	32	36	40	44	48
	0,5	24	42	57	70	79	86	90	94	96	97
	0,75	49	78	92	97	99	100	100	100	100	100
	1	74	96	99	100	100	100	100	100	100	100
15	0	5	5	5	5	5	5	5	5	5	5
	0,25	11	18	25	32	38	44	50	55	60	65
	0,5	33	57	75	86	92	96	98	99	100	100
	0,75	66	92	98	100	100	100	100	100	100	100
	1	89	99	100	100	100	100	100	100	100	100
20	0	5	5	5	5	5	5	5	5	5	5
	0,25	14	23	32	40	48	55	62	68	73	77
	0,5	42	70	86	94	97	99	100	100	100	100
	0,75	78	97	100	100	100	100	100	100	100	100
	1	96	100	100	100	100	100	100	100	100	100
25	0	5	5	5	5	5	5	5	5	5	5
	0,25	16	27	38	48	57	65	72	77	82	86
	0,5	50	79	92	97	99	100	100	100	100	100
	0,75	86	99	100	100	100	100	100	100	100	100
	1	98	100	100	100	100	100	100	100	100	100

Si se comparan las estimaciones obtenidas por Hughes (*Tabla 3*) con las calculadas mediante el método de simulación propuesto en este trabajo en los escenarios determinados por las combinaciones de cantidades de cluster (C): 5, 10, 15, 20 y 25; cantidades de observaciones por periodo-cluster (n_{ij}): 5, 10, 15, 20, 25, 30, 35, 40, 45 y 50; y los cinco valores elegidos para el efecto tratamiento (β) 0; 0,25; 0,50; 0,75 y 1 podemos ver que cuando no existe variabilidad entre los cluster las estimaciones obtenidas con ambos métodos son similares (*Tabla 4*). Mientras que cuando la variabilidad entre los cluster aumenta (σ_c^2): 1, 2 y 3 la diferencia de las estimaciones por ambos métodos crece, mostrando una sobreestimación de las potencias por parte del método de Hughes (*Tabla 5*). Esto se debe a que el método propuesto por Hughes no considera la variabilidad entre clusters en el cálculo de la estimación de la potencia.



Tabla 5: Diferencia (en porcentajes) entre la potencia estimada mediante la metodología propuesta por Hughes y la de simulación propuesta en este trabajo para los escenarios formados por las combinaciones de cluster (C): 5, 10, 15, 20 y 25, individuos por cluster (n_{ij}): 5, 10, 15, 20, 25, 30, 35, 40, 45 y 50, efecto tratamiento (β): 0; 0,25; 0,50; 0,75 y 1 y cuando la variabilidad entre los cluster es (σ_c^2): 1, 2 y 3

Cluster	β/n	Variabilidad entre cluster=1										Variabilidad entre cluster=2										Variabilidad entre cluster=3																			
		5	10	15	20	25	30	35	40	45	50	5	10	15	20	25	30	35	40	45	50	5	10	15	20	25	30	35	40	45	50										
5	0	0	2	1	0	0	0	1	0	0	-1	0	0	0	1	1	0	0	0	0	0	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0			
	0,25	3	0	1	2	0	1	5	5	3	3	2	4	2	3	3	5	3	3	5	2	4	1	3	3	5	5	7	4	7	6	4	7	6	4	7	6				
	0,5	3	6	8	6	8	9	11	11	7	8	7	6	10	9	10	13	14	15	14	16	7	7	9	14	16	18	18	16	20	18	7	7	9	14	16	18	18	16	20	18
	0,75	10	12	12	12	11	9	7	5	4	3	12	14	21	17	17	14	12	11	7	6	13	22	26	26	25	16	16	14	11	9	13	22	26	26	25	16	16	14	11	9
	1	14	17	13	8	5	3	1	1	0	0	18	20	21	12	10	6	4	2	1	1	23	29	24	18	13	8	5	3	2	1	23	29	24	18	13	8	5	3	2	1
10	0	0	1	2	0	-1	0	1	0	0	0	0	2	0	1	-1	0	0	0	0	1	0	1	1	-1	-1	1	0	1	1	1	0	1	1	1	0	1	1			
	0,25	0	-1	3	2	0	5	6	0	1	6	3	1	4	5	3	7	7	9	11	12	2	4	5	5	6	8	10	10	12	12	2	4	5	5	6	8	10	10	12	12
	0,5	5	9	10	9	9	7	8	6	4	3	7	10	15	13	12	14	10	9	7	6	8	11	14	19	21	17	16	13	11	9	8	11	14	19	21	17	16	13	11	9
	0,75	14	13	7	6	3	1	1	0	0	0	17	18	17	10	6	4	1	1	0	0	23	24	19	12	8	5	3	2	1	1	23	24	19	12	8	5	3	2	1	1
	1	17	9	2	1	0	0	0	0	0	0	21	13	4	2	1	0	0	0	0	0	24	20	8	3	0	0	0	0	0	0	24	20	8	3	0	0	0	0	0	0
15	0	0	0	0	0	2	-1	-1	1	0	0	0	1	-1	-1	0	1	0	-1	1	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
	0,25	2	2	3	3	10	3	3	4	6	5	1	4	5	8	8	8	10	11	10	15	2	6	8	8	14	13	10	16	13	15	2	6	8	8	14	13	10	16	13	15
	0,5	5	7	6	8	6	4	2	2	1	1	9	12	16	12	9	7	6	3	3	1	12	16	18	13	13	11	8	6	3	2	12	16	18	13	13	11	8	6	3	2
	0,75	11	10	3	2	0	0	0	0	0	0	19	15	6	4	1	0	0	0	0	0	23	19	11	6	1	0	0	0	0	0	23	19	11	6	1	0	0	0	0	0
	1	14	2	0	0	0	0	0	0	0	0	20	4	1	0	0	0	0	0	0	0	23	7	1	0	0	0	0	0	0	0	23	7	1	0	0	0	0	0	0	0
20	0	1	1	0	0	0	0	0	0	1	1	2	0	0	-1	1	1	1	1	1	0	0	-1	1	0	-2	1	-1	-1	-1	0	0	-1	1	0	-2	1	-1	-1	-1	
	0,25	2	4	5	4	4	7	7	4	7	6	2	5	4	9	11	8	11	12	11	13	1	6	7	10	16	13	17	15	15	16	1	6	7	10	16	13	17	15	15	16
	0,5	7	9	7	3	4	1	1	0	0	0	9	14	12	9	6	2	2	2	1	0	14	20	18	12	10	6	3	2	1	1	14	20	18	12	10	6	3	2	1	1
	0,75	13	7	1	0	0	0	0	0	0	0	17	8	3	1	0	0	0	0	0	0	23	13	5	2	0	0	0	0	0	0	23	13	5	2	0	0	0	0	0	0
	1	8	1	0	0	0	0	0	0	0	0	13	2	0	0	0	0	0	0	0	0	16	2	0	0	0	0	0	0	0	0	16	2	0	0	0	0	0	0	0	0
25	0	0	0	1	1	0	1	0	1	1	-1	0	1	-1	0	0	-1	0	0	-1	-1	-1	0	1	0	0	0	0	1	-1	0	0	1	0	0	0	1	-1	0		
	0,25	2	3	6	7	2	5	4	5	5	3	2	6	10	7	10	10	10	12	11	9	2	9	12	13	13	15	16	15	13	15	2	9	12	13	13	15	16	15	13	15
	0,5	7	5	6	5	2	0	0	0	0	0	14	13	8	5	3	2	1	0	0	0	10	19	17	10	7	3	1	0	0	0	10	19	17	10	7	3	1	0	0	0
	0,75	10	2	0	0	0	0	0	0	0	0	15	5	1	0	0	0	0	0	0	0	20	8	2	0	0	0	0	0	0	0	20	8	2	0	0	0	0	0	0	0
	1	5	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	11	1	0	0	0	0	0	0	0	0	11	1	0	0	0	0	0	0	0	0

Consideraciones finales

Los estudios con diseño de aleatorización por conglomerados resultan ser muy ventajosos en aquellos casos donde se deben realizar investigaciones sólo a través de grupos. Éstos llegan a ser muchas veces más económicos y adecuados, reduciendo la "contaminación" de los resultados que podría producirse si los participantes fueran asignados al azar individualmente a cada tratamiento.

En los últimos años, el número de estudios que utilizaron un diseño aleatorizado de cluster ha crecido, y también la discusión sobre los estándares de diseño y análisis para este tipo de estudio.

A través de un estudio de simulación se logró evaluar las modificaciones de la potencia estimada empíricamente como la proporción de rechazos en las pruebas de hipótesis sobre el efecto tratamiento en un modelo de regresión logística para distintos tamaño muestrales propuestos.

En el presente trabajo se usaron valores arbitrarios para cantidad de clusters, cantidad de individuos por cluster, efecto (proporción para el grupo tratado), proporción para el grupo control y variabilidad del efecto aleatorio de los clusters pero se podrían utilizar otros valores para estimar la potencia dependiendo del escenario que sea de interés evaluar. Las potencias estimadas en este



trabajo resultan de gran utilidad para el cálculo del tamaño de muestra en esos escenarios simulados.

La estimación por simulación es sensible al aumento de la variancia mientras que la fórmula propuesta por Hughes no lo toma en cuenta. Si no existe variabilidad entre los cluster las estimaciones de potencia son muy similares.

Se observó que para los valores propuestos la potencia estimada crece considerablemente a medida que se incrementa la cantidad de clusters, la cantidad de individuos por cluster y la diferencia entre los grupos sometidos a investigación (efecto tratamiento), y que levemente decrece cuando es mayor la variabilidad entre-cluster, es decir cuánto más parecidos son los individuos dentro de los grupos (cluster).

En los escenarios analizados siempre se consideró el mismo tamaño de individuos por cluster en cada periodo. Se deja propuesto para próximas investigaciones evaluar la potencia estimada cuando hay desbalance en el tamaño de cluster- periodo.

La proporción de eventos de interés esperada para la respuesta binaria del grupo control se fijó en todos los escenarios en 0,5 (modelo sin intercepto) y sólo se varió la proporción de eventos esperados en el grupo tratado. Se podrían calcular las potencias estimadas para nuevos escenarios donde la proporción de eventos esperados en el grupo control sea distinta de 0,5 siguiendo la metodología propuesta en este trabajo. Otra de las consideraciones previas en la simulación de los datos fue suponer que no existe interacción entre el periodo y el cluster.

Bibliografía

Austin P. C. (2010). A Comparison of the Statistical Power of Different Methods for the Analysis of Repeated Cross-Sectional Cluster Randomization Trials with Binary Outcomes. The International Journal of Biostatistics: Vol. 6 : Iss. 1, Article 11. (www.bepress.com/ijb/vol6/iss1/11).

Boggio G. S. (2007). Evaluación de criterios de bondad de ajuste para modelos logit marginales. Tesis de maestría, Universidad Nacional de Rosario, Rosario, Argentina.

Chen J., Stock S., Deng C. Talecris Biotherapeutics Inc., Research Triangle Park, NC 27709 Department of Biostatistics, Harvard University, 02115. (2008). Sample Size Estimation Through Simulation of a Random Coefficient Model by Using SAS.

Diez M. B. (2013). Estudio de las pruebas de hipótesis para el efecto tratamiento, en regresiones P-Spline, cuando los datos provienen de diseños por conglomerados. Tesina de grado, Universidad Nacional de Rosario, Rosario, Argentina.

Feng Z., Diehr P., Peterson P., McLerran D. (2001). Selected statistical issues in group randomized trials. Annu. Rev. Public Health 2001. 22:167–87.



Giraudeau B., Ravaud P., Donner A. (2008). Sample size calculation for cluster randomized cross-over trials. *Statist. Med.* 2008; 27:5578–5585 Published online 21 July 2008 in Wiley InterScience (www.interscience.wiley.com).

Hayes R. J., Moulton L. H.. Cluster randomized trials.

Holford N. H. G., Kimko H. C., Monteleone J. P. R., Peck C. C. (2000). Simulation of clinical trials. *Annu. Rev. Pharmacol. Toxicol.* 2000. 40:209–34.

Hughes J. et al. (2003). Design of the HIV Prevention Trials Network (HPTN) Protocol 054: A cluster randomized crossover trial to evaluate combined access to Nevirapine in developing countries.

Hussey M. A., Hughes J. P. b. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials* 28 (2007) 182–191 (www.sciencedirect.com).

Isern G. (2012). Diseños "Stepped wedge". Estudio de la potencia del test de hipótesis sobre el efecto tratamiento. Tesina de grado, Universidad Nacional de Rosario, Rosario, Argentina.

King G., Tomz M., Wittenberg J. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*, Vol. 44, No. 2, April 2000, Pp. 341–355.

Lee J., Schell M. J., Roetzheim R. (2009). Analysis of Group Randomized Trials with Multiple Binary Endpoints and Small Number of Groups. *PLoS ONE* 4(10): e7265. doi:10.1371/journal.pone.0007265.

Milanés V. (2010). Ensayos clínicos de diseño cruzado con aleatorización por cluster. Comparación de métodos de análisis. Tesina de grado, Universidad Nacional de Rosario, Rosario, Argentina.

Parienti J. J., Kuss O. (2006). Cluster-crossover design: A method for limiting clusters level effect in community-intervention studies.

Peters T. J., Richards S. H., Bankhead C. R., Ades A. E., Sterne JAC. (2003). Comparison of methods for analysing cluster randomized trials: an example involving a factorial design. *International Journal of Epidemiology* 2003;32:840–846 (ije.oxfordjournals.org).

Reich N. G., Myers J. A., Obeng D., Milstone A. M., Perl T. M. (2012). Empirical Power and Sample Size Calculations for Cluster-Randomized and Cluster-Randomized Crossover Studies. *PLoS ONE* 7(4): e35564. doi:10.1371/journal.pone.0035564.

Turner R. M., White I. R., Croudace T. for the PIP Study Group. (2006). Analysis of cluster randomized cross-over trial data: A comparison of methods. *Statist. Med.* 2007; 26:274–289. Published online 15 March 2006 in Wiley InterScience (www.interscience.wiley.com).

Ukoumunne O. C., Thompson S. G. (2001). Analysis of cluster randomized trials with repeated cross-sectional binary measurements. *Statist. Med.* 2001; 20:417-433.