

Estudios experimentales: una perspectiva veterinaria

Saavedra, P.

RESUMEN: Se analiza la metodología de los estudios experimentales, especialmente en las investigaciones más frecuentes en el campo de las ciencias veterinarias, a saber: los ensayos clínicos en medicina animal, y los estudios experimentales en producción animal y bromatología. Se discuten algunos aspectos acerca de los modelos de datos para tales estudios y se presentan finalmente algunas cuestiones metodológicas relacionadas con un ensayo clínico realizado en la Facultad de Veterinaria de Las Palmas.

Palabras clave: Diseño de estudios experimentales; Ensayos clínicos; Tamaño muestral.

ABSTRACT: Methodology in experimental studies is analysed, mainly in the most usual surveys in the veterinary sciences subject, namely; the clinical trials in animal medicine and the experimental studies in animal production and food science. Several aspects of the data models for a such studies are examined, and finally, any methodology questions about a trial clinic carried out in the Faculty of Veterinary of Las Palmas are showed.

Key words: Design of experimental studies; Clinical trials; Sample size.

Introducción

El objetivo esencial de cualquier estudio experimental consiste habitualmente en evaluar la relación de causalidad entre una *intervención* sobre los elementos de una población y una serie de posibles *efectos*. En el campo de las ciencias veterinarias, el estudio puede consistir en evaluar la respuesta inmunológica producida por una cierta vacuna (medicina animal), el rendimiento de una explotación agropecuaria atribuible a un protocolo de manejo de un ganado (producción animal), o la acción de un conservante sobre el crecimiento de la flora bacteriana en alimentos (bromatología). Los elementos u objetos de la población analizada (animales o alimentos) generalmente se comportan como *cajas negras*, lo que significa que no es posible por lo general observar plenamente el proceso que desencadena la intervención evaluada en el objeto o unidad experimental. Esto supone que cuando se observa una cierta respuesta con posterioridad a una intervención (remisión de la enfermedad o aumento en la producción de leche), no siempre es posible atribuirla a tal intervención. Así por ejemplo, si un sujeto remite de una infec-

ción con posterioridad a la administración de un antibiótico, no podemos en general atribuirle a éste el efecto de la remisión, dado que ésta podría deberse total o parcialmente a la respuesta inmunológica del sujeto. ¿Cómo puede entonces evaluarse el efecto de intervenciones, sobre todo en sujetos vivos, de tal forma que pueda establecerse una relación de causalidad verosímil entre intervención y efecto? De esto se ocupa precisamente la metodología de los estudios experimentales.

Un tipo especial de estudio experimental es el ensayo clínico. Su objetivo es evaluar intervenciones terapéuticas sobre sujetos humanos o animales. Cuando la intervención consiste en la administración de un medicamento, el objetivo del ensayo es propiamente establecer la farmacodinamia de ese medicamento, esto es; determinar si existe interacción entre el medicamento y los componentes específicos del organismo llamados *receptores*. En España, la realización de ensayos clínicos con medicamentos aplicados a sujetos humanos está regulada por el RD 223/2004. En este trabajo haremos una breve revisión de tres aspectos importantes de los protocolos de los

estudios experimentales, a saber: el diseño, tamaño muestral y la modelación del conjunto de datos.

Diseño de estudios experimentales

La respuesta observada en un objeto tras una intervención experimental podría atribuirse a causas diferentes a la propia intervención. Por tanto, para evaluar su efecto real deben utilizarse *controles*, esto es: objetos similares a los experimentales pero que en el transcurso del estudio no reciben ningún tratamiento o bien reciben un placebo. De esta forma, si las unidades experimentales muestran conjuntamente una respuesta que difiere de la mostrada por los controles, tal diferencia puede razonablemente atribuirse al efecto de la intervención experimental.

En general, en los estudios experimentales se comparan simultáneamente varias intervenciones. Dependiendo de la naturaleza del estudio, hay esencialmente dos estrategias de comparación. La primera consiste en que cada objeto del estudio (unidad experimental a partir de ahora) recibe una única intervención o tratamiento en el curso del estudio. La segunda estrategia permite que una unidad experimental pueda recibir a lo largo del estudio y en sucesivas fases más de un tratamiento. El diseño de los estudios que utilizan la primera estrategia reciben el nombre de *grupos paralelos* y mientras que los que utilizan la segunda se denominan *estudios cruzados*.

En los diseños de grupos paralelos, los objetos que satisfacen las condiciones de inclusión en el estudio son aleatoriamente asignados a uno y sólo uno de los tratamientos evaluados. La estrategia de asignar los tratamientos al azar recibe el

nombre de aleatorización y su propósito es que los grupos formados sean *comparables*, esto es; que los llamados factores pronósticos se distribuyan de la misma forma en todos los grupos experimentales. De esta forma, las eventuales diferencias que se produzcan entre las respuestas observadas tras la aplicación de los tratamientos podrían atribuirse razonablemente a los efectos de éstos.

Ilustramos estas ideas mediante el estudio 4S, el cual se diseñó para evaluar la efectividad de un fármaco (*simvastatina*) en la reducción de la morbi-mortalidad en pacientes con elevadas cifras de LDL y que habían previamente sufrido un episodio cardiovascular (IAM y/o angina de pecho). El diseño del estudio fue de dos grupos paralelos (fármaco experimental frente a placebo) con asignación aleatoria de tratamientos, contándose finalmente con 2221 pacientes evaluables en el brazo experimental frente a 2223 en el control. La distribución de los factores pronósticos más importantes, previa a las intervenciones, se muestra en la siguiente tabla.

De los 2223 pacientes intervenidos en el grupo placebo, el número de muertes atribuibles a enfermedad coronaria en un seguimiento de seis años fue de 180 (8,5%) mientras que en el grupo experimental fue de 111 (5%). Ello supone una reducción en la mortalidad del 41,2% (IC-95% = 26,2% a 53,1%). Razonablemente, tal mortalidad sólo puede atribuirse a la intervención realizada en el grupo experimental, puesto que cualquier otro factor pronóstico se ha comportado de la misma forma en ambos grupos de estudio.

Supóngase ahora que los pacientes se hubiesen asignado a los tratamientos mediante un criterio como el que sigue: *dado que la simvastatina podría tener efectos secundarios graves no previstos, ésta se asignará preferentemente a los pacientes más jóvenes, dado que éstos resistirían mejor cualquier efecto adverso que pudiera producirse*. En tal caso, una superior respuesta en el grupo experimental podría ser atribuible total o parcialmente a los sujetos experimentales.

Tabla 1. Resumen de la distribución de diversos factores pronósticos en los brazos de experimentales correspondientes al estudio 4S, previa a las intervenciones.

	Placebo (n=2223)	Simvastatina (n=2221)
Edad media (Hombres)	58.1	58.2
Edad media (Mujeres)	60.5	60.5
Sólo angina de pecho	21%	21%
Sólo infarto de miocardio (IM)	62%	63%
Angina e IM	17%	16%
HTA (Hipertensión arterial)	26%	26%
Fumadores	27%	24%
Colesterol total (mmol/L)	6.75	6.74
LDL (mmol/L)	4.87	4.87

Podría afirmarse que hay un imperativo ético en el uso de criterios como éste, pero desde luego, invalidan claramente el estudio.

En los estudios de grupos paralelos es preferible que las distintas intervenciones se hagan con una cierta simultaneidad (*concurrancia*). Esto no significa que todos los tratamientos se apliquen simultáneamente, sino que los periodos de aplicación de tratamientos sean similares para todos los grupos experimentales. Esta condición es indispensable para mantener la comparabilidad de los grupos experimentales. En medicina humana hay situaciones en las que se tiene un alto grado de seguridad acerca de la superioridad de un nuevo tratamiento experimental *E* frente a un control *C*, y que sus efectos adversos son irrelevantes. Esto ocurre frecuentemente en la evaluación de nuevos antibióticos. Por motivos éticos, los investigadores deben ofertar a todos los pacientes el tratamiento *E* y realizar la comparación con datos históricos (respuestas observadas en el pasado al tratamiento estándar *C*). Esta falta de concurrencia en la aplicación de ambos tratamientos puede producir dudas acerca de las eventuales diferencias que se observen. Por ejemplo, podría ocurrir que *E* fuera realmente superior a *C*, pero esta superioridad no se pone de manifiesto en el curso del estudio debido a que *C* se empleó en el pasado frente a un germen que era menos resistente que el que ahora se trata con *E*.

Es recomendable y a veces necesario que la evaluación de los tratamientos se realice sin que el evaluador conozca el tratamiento aplicado. Esto es particularmente importante cuando la

evaluación de la respuesta pudiera estar afectada por la subjetividad del observador. En tal caso el estudio debe ser *enmascarado*; esto es: determinar la respuesta desconociéndose el tratamiento aplicado. En ensayos clínicos con pacientes humanos es frecuente que el sujeto del estudio tampoco conozca el tratamiento que ha recibido en orden a evitar que la sugestión inducida en el paciente altere la respuesta (efecto placebo). Cuando ni el paciente ni el evaluador conocen el tratamiento asignado se dice que el estudio se realiza a *doble ciego*.

En el análisis de datos correspondientes a un estudio de grupos paralelos, hay que contemplar dos fuentes principales de valoración, a saber: la atribuible al tratamiento y la específica de la unidad muestral. Esto significa que si se comparan las respuestas observadas en unidades experimentales que han recibido diferentes tratamientos, parte de la diferencia que se observe se deberá al tratamiento, pero otra parte a las propias unidades muestrales. De hecho, unidades muestrales que reciben el mismo tratamiento, darán por lo general lugar a respuestas distintas. Cuando la variabilidad atribuible a las unidades experimentales sea muy grande en comparación con la correspondiente a los tratamientos, se requerirán por lo general grandes tamaños muestrales para poner de manifiesto las presuntas diferencias entre tratamientos, lo que en muchos casos dificulta la realización del estudio. Este problema puede solventarse en aquellas situaciones en las que cada unidad muestral pueda

recibir más de un tratamiento en fases sucesivas. Naturalmente esto no es posible cuando el primer tratamiento produzca un efecto irreversible. Así por ejemplo, si a un sujeto con infección bacteriana se le trata con un antibiótico y la infección remite, sería absurdo aplicarle en una segunda fase otro antibiótico para comparar su efecto con el primero. Hay no obstante situaciones en las que esto es posible. Así por ejemplo, el tratamiento con insulina en sujetos diabéticos tiene como objetivo regular las concentraciones de glucosa en sangre y desde luego, su uso no modifica prácticamente el estatus de la enfermedad. Si se deseara comparar el efecto sobre la glucemia de una insulina *A* con otra *B*, cabría la posibilidad de tratar durante un periodo con *A* y durante un siguiente periodo, con *B*. En tal caso, el sujeto del estudio es control de sí mismo y de esta forma podría eliminarse el efecto de variabilidad individual. Surge entonces la siguiente pregunta: ¿cuál es el orden en que deben aplicarse los tratamientos? Sin duda, la forma más equilibrada de realizar el estudio sería que una parte de los pacientes (generalmente la mitad) recibieran la secuencia *A-B* y la otra parte, la secuencia *B-A*. Naturalmente, las unidades muestrales deberían asignarse a las secuencias también de forma aleatorizada. Este tipo de diseño, recibe el nombre de *diseño cruzado*, y sus ventajas son obvias cuando sea posible su aplicación. Estos diseños pueden a menudo resumirse en lo que se llama un cuadrado latino. Así por ejemplo, para comparar cuatro tratamientos *A, B, C* y *D*, de tal forma que todos los pacientes reciban todos los tratamientos en periodos consecutivos, habría que formar inicialmente 4! (24°) secuencias (todas las posibles ordenaciones de los cuatro elementos). Ello supondría manejar en el curso del ensayo 16 grupos experimentales, lo que desde luego es poco práctico por razones diversas. Dado que en realidad lo que se pretende es que en cada fase se utilicen todos los tratamientos, los 24 grupos podrían reducirse a cuatro

de acuerdo con el siguiente esquema (cuadrado latino de orden 4):

Tabla 2. Cuadro latino para un diseño cruzado con cuatro tratamientos.

Grupo	Fase			
	I	II	III	IV
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

Este tipo de diseño puede mejorar notablemente la calidad del estudio. Hay que advertir no obstante que podría ocurrir que el efecto de un tratamiento aplicado en una fase perdurara total o parcialmente en la siguiente (*efecto carry-over*). En tal caso, es conveniente que los modelos de análisis de datos recojan esta eventualidad.

Consideraciones sobre el tamaño muestral

Una vez decidido el diseño del estudio experimental (grupos paralelos o cruzados) surge el problema de determinar el tamaño de la muestra; esto es, el número de unidades muestrales que han de asignarse a cada grupo experimental, el cual se corresponderá con un tratamiento (grupos paralelos) o secuencia de tratamientos (cruzado). En este punto es esencial definir el *objetivo principal de valoración* del estudio, pues frecuentemente el tamaño muestral se determina en orden a que el conjunto de datos tenga *suficiente potencia* para detectar los efectos que sobre la variable principal de valoración puedan producir los diferentes tratamientos. Más concretamente, el procedimiento estadístico de comparación (*test de hipótesis*) tiene asociado dos parámetros que reciben el nombre de *significación* y *potencia*. Para entender el concepto de significación, supóngase que se van a comparar dos intervenciones que realmente producen el mismo efecto sobre la variable principal de valoración. Podría ocurrir que por el azar

de las muestras seleccionadas, el test de hipótesis concluyera erróneamente que las intervenciones evaluadas producen efectos diferentes. En tal caso se dice que se ha producido un error *alpha* y la significación es precisamente la probabilidad de que tal error se produzca. Por el contrario, supóngase que hay una diferencia determinada entre ambas intervenciones (por ejemplo, una de ellas tiene asociada una tasa de respuestas favorables del 60% y la otra del 65%). En ese caso, la potencia es la probabilidad de que tal diferencia sea detectada por el test de hipótesis.

Es muy frecuente que el test de hipótesis utilizado tenga una significación del 5%. Esto significa que si no hubiera diferencia real de efectos entre las intervenciones comparadas, habría no obstante una probabilidad del 5% de que se produzca un error *alpha*. Tal margen de error parece asumible. Pero esto supone que debe establecerse con claridad un único criterio principal de valoración antes de la realización del estudio. Considérese un estudio diseñado en orden a evaluar si dos tratamientos producen alguna diferencia, a priori no especificada. Una vez inducidas las respuestas a los tratamientos podrían compararse múltiples variables entre los grupos experimentales. Supóngase que se examinan 100 variables entre las que realmente no hay ninguna diferencia relevante. Sin embargo, si la significación de los test de comparación es la habitual del 5%, cada vez que se realiza una comparación hay una probabilidad del 5% de que se produzca un error *alpha*. Esto supone que en las 100 comparaciones cabe esperar que alrededor de cinco variables mostrarían diferencias significativas entre los grupos de estudio, cuando en realidad, tales diferencias no existen. Obviamente esta metodología de estudio es inaceptable, pues a través de ella podríamos fácilmente establecer efectos de intervenciones realmente inexistentes. A este modo de proceder se le suele llamar *salir de cacería*.

¿Significa esto que no deben tenerse en consideración hallazgos *post-hoc*? (no especificados a priori). Tales hallazgos, aunque no sean concluyentes, deben ser informados. Cuando se desee valorar alguno específico, debe planificarse un nuevo estudio experimental que lo considere como objetivo principal de valoración.

Modelos de análisis de datos

En líneas generales, un estudio experimental aleatorizado, con suficiente tamaño muestral y eventualmente enmascarado de tal forma que permita suponer la comparabilidad de los grupos experimentales no requiere generalmente el uso de modelos extraordinariamente sofisticados como ocurre en los estudios observacionales en los que a menudo hay presencia de factores de confusión. La comparación de los efectos principales de varios tratamientos sobre una variable continua, correspondiente a un estudio de grupos paralelos puede realizarse a través de un simple modelo de análisis de la varianza a una vía. La modelación se realiza sobre el conjunto de datos obtenido de la realización del estudio. Para el caso que nos ocupa tiene la forma:

$$(D1) \{y_{ij} : i = 1, \dots, p; j = 1, \dots, m\}$$

donde y_{ij} presenta la respuesta de la j -ésima unidad muestral perteneciente al i -ésimo grupo de tratamiento, p es el número de tratamientos y m el tamaño muestral dentro de cada grupo experimental. Representamos por μ_i el valor esperado de y_{ij} (en la práctica representa la media de todas las posibles respuestas de la variable de valoración a ese tratamiento). Este parámetro consiste en definitiva en el efecto principal del i -ésimo tratamiento. Por tanto, podemos suponer que el conjunto de datos D1 obedece al modelo:

$$(M1) y_{ij} = \mu_i + e_{ij} : i = 1, \dots, p; j = 1, \dots, m$$

representando ahora e_{ij} la variabilidad específica de la correspondiente unidad muestral. El modelo (M1) recibe el nombre de modelo de análisis de la varianza con un factor de variación.

El parámetro $\theta = (\mu_1 + \dots + \mu_p)/p$ representa obviamente la media aritmética de todos los efectos de los tratamientos. Si hacemos $\alpha_i = \mu_i - \theta$, es fácil comprobar que el modelo M1 puede expresarse como:

$$(M2) y_{ij} = \theta + \alpha_i + e_{ij} : i = 1, \dots, p; j = 1, \dots, m$$

donde ahora la media $(\alpha_1 + \dots + \alpha_p)/p = 0$. Puede ahora interpretarse (M2) de la siguiente forma: el parámetro θ es el efecto medio global de todos los tratamientos, α_i representa la desviación producida por el i -ésimo tratamiento sobre la respuesta y finalmente, e_{ij} corresponde a la variabilidad específica de la unidad muestral.

junto de datos obtenido podría expresarse ahora por:

$$(D2) \{y_{ijk} : i = 1, 2, 3; j = 1, 2; k = 1, \dots, m\}$$

donde y_{ijk} representa la respuesta de la k -ésima unidad muestral que ha recibido la dieta i -ésima dieta y j -ésimo antibiótico. El modelo de efectos principales para estos datos tiene la forma:

$$(M3) y_{ijk} = \theta + \delta_i + \alpha_j + e_{ijk} : i = 1, 2, 3; j = 1, 2; k = 1, \dots, m$$

adonde por simple analogía con (M1), la media de los efectos de dieta y la media de los efectos de antibiótico son nulas. De las consideraciones realizadas anteriormente acerca de los parámetros que representan los efectos de los tratamientos, puede considerarse $\delta_1 = \alpha_1 = 0$ y $\alpha_2 = \alpha$. En ese caso, los efectos de los tratamientos pueden expresarse mediante la siguiente matriz:

Tabla 3. Resumen del modelo de análisis de la varianza con dos factores de variación.

Antibiótico	Dieta	
	1	2
1	θ	$\theta + \alpha$
2	$\theta + \delta_2$	$\theta + \delta_2 + \alpha$
3	$\theta + \delta_3$	$\theta + \delta_3 + \alpha$

Esta representación es útil para los diseños factoriales, donde cada tratamiento corresponde a una combinación de factores. Un ejemplo simple de diseño factorial en producción animal es el siguiente: supóngase que se desea evaluar sobre una cierta variable (producción de leche en un periodo fijo) el efecto de tres clases de dieta y dos tipos de antibiótico. Cada tratamiento consistiría ahora en una combinación dieta-antibiótico. Considérese que a cada tratamiento se asignan m unidades muestrales (animales). La asignación de las unidades muestrales a los tratamientos debe ser aleatoria, y de esta forma, este diseño factorial es un caso especial de diseño de grupos paralelos. El con-

La idea del modelo es que cada unidad experimental tendrá una respuesta igual a la suma de efectos expresada en su celda-tratamiento, más su especificidad que se recoge en el término de variabilidad atribuible a la unidad experimental e_{ijk} . Podría ocurrir que una dieta interactuara con un antibiótico, esto es; que cuando se utilizan conjuntamente una determinada dieta y antibiótico, se produzca un efecto añadido sobre la respuesta. Por ejemplo, si la interacción se produce entre la dieta 3 y antibiótico 2, el efecto final sería $\theta + \delta_3 + \alpha + (\delta\alpha)_{3,2}$. En general, el modelo (M3) con interacciones se expresaría por:

$$(M4) y_{ijk} = \theta + \delta_i + \alpha_j + e_{ijk} : i = 1, 2, 3; j = 1, 2; k = 1, \dots, m$$

La respuesta de cada unidad experimental a su tratamiento puede consistir en una secuencia de datos obtenida a lo largo de un periodo específico (datos longitudinales). Ilustramos este tipo de análisis con un ensayo clínico llevado a cabo en la Facultad de Veterinaria de Las Palmas, en el cual se compararon los efectos de dos vacunas A y B y un control C sobre la respuesta inmunológica (inmunoglobulina IGG). El diseño del ensayo fue de grupos paralelos, en el cual 40 animales fueron asignados aleatoriamente a recibir la vacuna A y otros 40 la vacuna B. Los valores IGG fueron determinados en los días post-vacunación 0, 25, 45, 100, 150, 180 y 210. Nótese que en este caso las diferencias de efectos pueden variar a lo largo del periodo de observación. En este caso la modelación es algo más compleja. Podemos representar el conjunto de datos por:

$$(D3)y_{ijk} : i=A,B,C; j=1,\dots,7;k=1,\dots,40$$

donde naturalmente el primer índice se refiere al tipo de vacuna, el segundo al periodo de observación y el tercero a la unidad experimental. Una modelación que recoge bien este proceso es:

$$(M5)y_{ijk} = \theta + \tau_i + \delta_j + (\tau\delta)_{ij} + \gamma_{k(i)} + \epsilon_{ijk}; i=A,B,C; j=1,\dots,7;k=1,\dots,40$$

donde θ es el efecto medio global, τ_i el efecto basal de la vacuna, $(\tau\delta)_{ij}$ la interacción vacuna-periodo, y finalmente, $\gamma_{k(i)}$ es el efecto de la unidad experimental, el cual está anidado en el tratamiento. Téngase en cuenta que ahora aparece un efecto propio de unidad experimental (animal) que puede ser evaluado por el hecho de haber realizado sucesivas

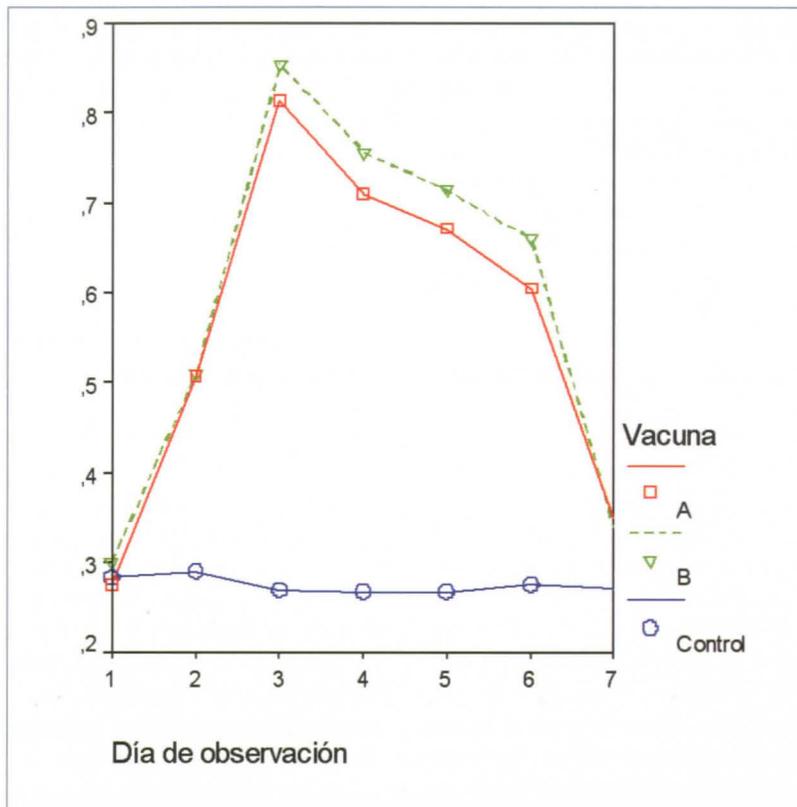


Figura 1. Medidas estimadas para la IGG correspondientes al modelo (M5).

Tabla 4. Contrastes sobre los efectos del modelo (M5).

Fuente de variación	IGG (Mm Mcoll)	
	Suma de cuadrados	F-test (p-valor)
Vacuna	25.442	< 0.001
Periodo	15.616	< 0.001
Interacción	10.544	< 0.001
Animal	0.856	
Residual	4.421	
Total	56.879	

observaciones intra-animal. La siguiente tabla resume los efectos que son estadísticamente significativos correspondientes al modelo (M5).

Nótese que existe efecto de vacuna, de periodo e interacción vacuna-periodo. Esto puede observarse en el siguiente gráfico.

Es obvio que el control no tiene ningún efecto sobre las respuestas IGG. La vacuna B muestra entre el tercer y sexto día de observación una respuesta superior a la A. Existe por tanto efecto de vacuna, obviamente de día de observación y finalmente, interacción vacuna-periodo.