

Métodos matemáticos y modelos estadísticos aplicados a la estimación de la curva de dosis-respuesta en un estudio de exposición oral de cerdos gnotobióticos a Norovirus humano

Tesis presentada para optar al título de Magister de la Universidad de Buenos Aires,
Área Biometría y Mejoramiento

Viviana Parreño

Bioquímica - Universidad de Morón- 1995

Doctora, área Microbiología e Inmunología Veterinaria

(Universidad de Buenos Aires), 2002

Lugar de trabajo: INCUINTA, Centro de investigaciones en Ciencias Ceterinarias y Agronómicas, CICV y A. INTA, Castelar, Bs. As., Argentina



Escuela para Graduados Ing. Agr. Alberto Soriano
Facultad de Agronomía – Universidad de Buenos Aires

2021

COMITÉ CONSEJERO

Director de tesis

Sebastián Munilla

Ing. Agr. (Universidad de Buenos Aires, Argentina), 2004
Doctor de la Universidad de Buenos Aires, Área Ciencias Agropecuarias
(2011)

Co-Director:

Lijuan Yuan

Ph.D. (Inmunología y virología), 2000.
Food Animal Health Research Program, Department of Veterinary
Preventive Medicine, Ohio Agriculture Research and Development Center
(OARDC), The Ohio State University.
Magister (Inmunología y Virología), 1991.
Capital Institute of Pediatrics, Beijing, China
Bioquímica y Farmacia, 1985.
Beijing University, China
Experta en modelo de cerdo gnotobiótico
de infección y diarrea por norovirus humano
Politécnico de Virginia, EE. UU.

Jurados

Dra. Viviana Mbayed
Master Adriana Perez
Dr. Sergio Bramardi

Dedicatoria

*A mi querida Laura Marangunich,
mi maestra de estadística,
ejemplo de dedicación y compromiso,
por su amor por la docencia de esta disciplina en todos los niveles
gracias por tu apoyo y tu amistad.*

Agradecimientos

A mi director de Tesis, Sebastián Munilla, por ofrecer su ayuda para dirigirme, por su acompañamiento permanente en la elaboración de este trabajo, por su buena onda, su paciencia y voluntad infinitas. Por su rigurosidad académica. Gracias de todo corazón!

A la Dra. Lijuan Yuan, co-directora de esta tesis, por invitarme a realizar el análisis estadístico de este ensayo dosis-respuesta y orientarme en todo lo referido al modelo animal. Por nuestra amistad de tantos años.

A la Escuela de Graduados y a todos los docentes y compañeros con los que compartí esta maestría desde 2014 a 2017.

Al INTA, por permitirme siempre acceder a una educación de calidad.

A los Dres Christian Ritz, Philip Schmidt y Mónica Emelko por su asesoramiento estadístico.

Una frase celebre

Remember that “All models are wrong, but some are useful”

George Box

Presentaciones a congresos y publicaciones derivadas

Presentación oral

Evaluation of different dose-response models for the calculation of 50% infectious dose of human norovirus Cin-1 in gnotobiotic pigs. Ashwin Ramesh,[§] Viviana Parreño,[§] Shaohua Lei, Weiming Zhong, Xi Jiang 3 and Lijuan Yuan. 7ma Conferencia Internacional de Calicivirus. 16 de Octubre de 2019, Sidney, Australia 2019.

Estimation of the median infective and median diarrhea dose in a gnotobiotic pig model of Human Norovirus based on the área under the curve of virus shedding and diarrhea score. Viviana Parreño, Ashwin Ramesh, Lijuan Yuan and Sebastián Munilla. 2nd conference of Woman in Bioinformatic and data science LA. 22-24 de Septiembre de 2021.

Publicación

Evaluation of the 50% Infectious Dose of Human Norovirus Cin-2 in Gnotobiotic Pigs: A Comparison of Classical and Contemporary Methods for Endpoint Estimation. Ashwin K. Ramesh 1,† , **Viviana Parreño** 2,†, Philip J. Schmidt 3, Shaohua Lei 1,‡ , Weiming Zhong 4, Xi Jiang 4,5, Monica B. Emelko 3 and Lijuan Yuan 1,* *Viruses* 2020, 12, 955; doi:10.3390/v12090955 † comparten ambos autores la 1er autoría del paper.

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en ésta u otra institución.

Índice

ABREVIATURAS	9
RESUMEN	13
ABSTRACT	15
1. INTRODUCCIÓN	17
1.1. Modelos de dosis respuesta aplicados al estudio de exposición oral a patógenos entéricos	18
1.2. Nuestro ejemplo biológico: modelo de cerdo gnotobiótico para reproducir la infección y diarrea por Norovirus Humano	18
1.3. Bases biológicas de los modelos de relación dosis respuesta	20
<i>a. Exposición -> Característica particulada del inóculo y efectos de agregación.....</i>	<i>21</i>
<i>b. Infección -> Modelos de relación dosis-infección</i>	<i>22</i>
<i>c. Enfermedad: Diarrea por Norovirus -> Modelos de relación dosis enfermedad</i>	<i>23</i>
1.4. Modelos aplicados a los estudios dosis-respuesta y cálculo de la dosis efectiva 50%	23
1.5. Criterios utilizados en la selección de modelos estadísticos.....	23
2. OBJETIVOS	25
3. MATERIALES Y MÉTODOS.....	27
3.1. Datos y diseño experimental	28
3.2. Variables registradas	28
3.3. Análisis exploratorio	29
3.4. Agregación viral: análisis de regresión del tamaño de los acúmulos virales en función del número de copias genómicas en el inóculo	29
3.5. Métodos clásicos aplicados al cálculo de las dosis infecciosa y diarrea 50%.....	30
<i>a. Método de Reed-Muench</i>	<i>30</i>
<i>b. Método de Dragstedt-Beherens</i>	<i>31</i>
<i>c. Método no paramétrico de Spearman-Kärber.....</i>	<i>32</i>
3.3: Modelos lineales generalizados: regresión logística	35
3.4: Métodos de regresión no lineal.....	37
a. Regresión log-logística de 2 parámetros	37
b. Modelos mecanicistas de dosis-respuesta: Exponencial y Beta-Poisson	38
c. Modelo de regresión log-logística de 3 parámetros aplicado sobre variables continuas	40
3.5: Estimación de parámetros y bondad de ajuste de los modelos empleados.....	41
4. RESULTADOS.....	44
4.1. Exploración de los datos y naturaleza de las variables bajo estudio	45
<i>a. Decodificación de las variables.....</i>	<i>46</i>
<i>b. Estadística descriptiva de las variables asociadas a la infección.....</i>	<i>47</i>
<i>c. Magnitud de la excreción viral.....</i>	<i>49</i>

d.	<i>Estadística descriptiva de las variables asociadas al desarrollo de la enfermedad</i>	53
e.	<i>Magnitud de la severidad del cuadro clínico</i>	55
4.2.	Comparación de medias para las variables asociadas a la infección y enfermedad causadas por el NoV humano en el ensayo con cerdos gnotobióticos	57
4.3.	Análisis del grado de agregación del inóculo	64
4.4.	Cálculo $DI_{50\%}$ y $DD_{50\%}$ por métodos clásicos	68
4.5.	Cálculo de las $DI_{50\%}$ y $DD_{50\%}$ utilizando modelos lineales generalizados	70
a.	<i>Regresión logística</i>	70
b.	<i>Modelo de Log-log (herramienta online del NCBI)</i>	74
4.6.	Cálculo de las $DI_{50\%}$ y $DD_{50\%}$ utilizando modelos de regresión no lineal	74
a.	<i>Regresión log-logística de dos parámetros 2PL.L.</i>	74
b.	<i>Modelos mecanicistas: Exponencial y Beta Poisson</i>	77
4.6.	Cálculo $DI_{50\%}$ y $DD_{50\%}$ utilizando variables continuas	87
4.7.	Resumen de resultados $DI_{50\%}$ and $DD_{50\%}$	90
4.8.	Aplicación práctica: dosis óptima de desafío del modelo cerdo gnotobiótico para NoV humano	93
5.	DISCUSIÓN	96
6.	CONCLUSIONES	102
7.	ANEXO	104
1.	<i>Resultados 4.4. Cálculo $DI_{50\%}$ y $DD_{50\%}$ métodos clásicos. Ejemplo de desarrollo del cálculo por R&M</i>	105
2.	<i>Resultado 4.5. Intervalos de confianza para las curvas de dosis-respuesta de los modelos mecanicistas por el método de Cadenas de Markov Montecarlo, MCM</i>	107
8.	BIBLIOGRAFÍA	108

Indice de Figuras y tablas

Figura 1. Modelo cerdo gnotobiotico de infección y diarrea por Norovirus humano, Dra. Yuan. En este modelo los lechones inoculados con una dosis adecuada de virus proveniente de materia fecal humana de paciente infectado desarrolla diarrea leve, viremia transciente y excreción viral en heces, Fuente:(Cheetham et al. 2007; Souza et al. 2007).	19
Figura 2. Principales etapas en el proceso de una enfermedad infecciosa transmitida por los alimentos.	20
Figura 3. Modelo estadístico-matemático ajustado a datos de un ensayo dosis respuesta.	21
Figura 4. Diseño Experimental.	28
Figura 5. Método de Reed y Muench 1938. (Figura Modificada de (Siev 2018)).	30
Figura 6. Ilustración del Método de Dragstedt-Behrens (Figura extraída de Siev, 2018).	32
Figura 7. Conceptualización del método de Spearman Karber, extraído de (Siev 2018).	33
Figura 8. Esquemización geométrica del método de extrapolación de Spearman Karber.	33
Figura 9. Box & Whisker del momento de inicio de la excreción viral (ONSETS) en cada tratamiento.	48
Figura 10. Box & Whisker de la duración de la excreción viral (SDUR) en cada tratamiento.	49
Figura 11. Log10 del área bajo la curva promedio de excreción viral (logAUC) y promedio del área bajo la curva de los score fecal (AUCD) en cada grupo de cerdos gnotobióticos inoculado con dosis crecientes de NoV humano.	52
Figura 12. Box & Whisker del momento de inicio de la excreción viral (SDUR) en cada tratamiento.	53
Figura 13. Box & Whisker de la duración de la enfermedad (DDUR) en cada tratamiento.	54
Figura 14. Box & Whisker ANOVA. A. Período de incubación de la infección (momento de inicio de la excreción viral en días post inoculación). B. Duración de la excreción viral. C. Magnitud de la excreción viral en términos de área bajo la curva de cantidad de virus excretado en los 7 días que duró el ensayo. Barras con diferente letra difieren significativamente (ver resumen del análisis estadístico en la Tabla R10. Los scripts y salidas de R e infostat están disponibles a solicitud)	61
Figura 15. Análisis de regresión lineal del inicio y duración de la diarrea en función de la dosis de NoV inoculada en los cerdos gnotobióticos. (detalle de Scripts y salidas disponibles a solicitud).	62
Figura 16. Box & Whisker. A. Período de incubación de la enfermedad (momento de inicio de la diarrea en días post inoculación). B. Duración de la diarrea en días. C. Severidad de la diarrea en términos de área bajo la curva de score fecal durante los 7 días que duró el ensayo. Barras con diferente letra difieren significativamente (ver resumen de analisis estadisco en Tabla R10, scripts y salidas de Infostat y R disponibles a solicitud).	63
Figura 17. Se indican los tres cerdos, dos del grupo F (inóculo: 3200 copias de genoma viral) y uno del grupo G (inóculo: 800) que excretan títulos de virus muy elevados en sus heces.	64
Figura 18. Fotografía de microscopia electrónica del inóculo de NoV Cin-2. a) Dosis: 8×10^4 , el recuento de partículas virales indicó la presencia de agregados con un tamaño de alrededor de 175 partículas; b) Dosis: 2×10^5 , agregados con alrededor de 390 partíc	64
Figura 19. Scatter plot del tamaño de acúmulos virales en función del número de copias de genoma viral en cada dosis de inóculo. Los puntos grises indican cada dato y los puntos rojos la media en cada dosis.	67
Figura 20. Análisis de Regresión de Poisson	67
Figura 21. Distribución datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano y un octavo grupo control negativo que recibe solución salina.	68
Figura 22. Ajuste de curva de modelo de regresión logística a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano. Línea roja: respuesta media; líneas grises: IC95%; puntos: datos observados proporción por grupo de dosis).	72
Figura 23. Curva de de dosis-respuesta ajustando el modelo de regresión 2PL.L. a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con 7 dosis crecientes de NoV humano y un octavo grupo control negativo que recibe solución salina en lugar de virus. Los círculos representan el valor predicho promedio. El sombreado y las barras los IC95%.	75
Figura 24. Histograma de las estimaciones del parámetro r del modelo Exponencial, con 10.000 pseudo-réplicas.	81
Figura 25. Estimación bootstrap de la incertidumbre de los parámetros μ y $N50$ del modelo Beta-Poisson al ajustarse a los datos de (a) infección y (b) enfermedad. Elipses de confianza para los percentiles 90, 95 y 99 (10.000 pseudo réplicas), scripts de Weir, 2017	82

Figura 26. Curvas de probabilidad estimada e intervalos de confianza del modelo Exponencial ajustado a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano. Los intervalos de confianza fueron calculados por bootstrap implementado en el script de Weir et al. (2017).	83
Figura 27. Curvas de probabilidad estimada e intervalos de confianza del modelo Beta-Poisson ajustado a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano. Los intervalos de confianza fueron calculados por el método de bootstrap implementado en el script de Weir et al. (2017).	84
Figura 28. Ajuste del modelo de regresión log-logística de tres parámetros a los datos del área bajo la curva de excreción viral (a) y de los scores fecales AUCD. Los intervalos de confianza 95% fueron calculados por el método delta del paquete drc. Ritz, 2015.	89
Figura 29. Comparación de los diferentes modelos para los datos de infección.	92
Figura 30. Comparación de los diferentes modelos para los datos de enfermedad.	92

Tabla M&M 1. Datos simples para calcular el título en término de dosis infecciosas 50% de una suspensión viral	34
Tabla M&M 2. Principales modelos lineales generalizados	36

Tabla R 1. Datos del ensayo de la infección y la enfermedad causada por NoV humano en cerdos gnotobióticos experimentalmente desafiados con dosis crecientes de virus	45
Tabla R 2. Estadística descriptiva. Período de incubación de la infección: ONSETS .	47
Tabla R 3. Estadística descriptiva. Duración de la excreción viral: SDUR .	48
Tabla R 4. Excreción viral determinada por qRT PCR (número de copias de genoma viral).	49
Tabla R 5. Estadística descriptiva. Logaritmo del área bajo la curva de excreción viral: LogAUCS .	51
Tabla R 6. Estadística descriptiva. Período de incubación de la enfermedad ONSETD .	53
Tabla R 7. Estadística descriptiva. Duración de la enfermedad (diarrea) DDUR .	54
Tabla R 8. Datos de score fecal de cada cerdo individual, 1 a 7 días.	55
Tabla R 9. Estadística descriptiva. Severidad del cuadro clínico: AUCD .	56
Tabla R 10. Comparación de medias entre grupos para las variables asociadas a la infección y diarrea causadas por la inoculación de NoV humano en los cerdos gnotobióticos.	60
Tabla R 11. Resultados del tamaño de los acúmulos virales en cada dosis.	65
Tabla R 12. DI50% y DD50% por métodos clásicos.	69
Tabla R 13. Regresión logística.	70
Tabla R 14. Estimaciones por regresión logística: diferentes métodos.	71
Tabla R 15. Probabilidad de infección y enfermedad estimado a partir del modelo de regresión logística.	73
Tabla R 16. Resultado del ajuste de los datos de infección y diarrea mediante regresión log-logística 2P L.L.	76
Tabla R 17. Parámetros y bondad de ajuste del modelo Exponencial para estimar la DI50% y la DD50% de NoV humano en cerdos gnotobióticos.	78
Tabla R 18. Parámetros y bondad de ajuste del modelo mecanicista Beta-Poisson para estimar DI50% y la DD50% de NoV humano en cerdos gnotobióticos.	79
Tabla R 19. Probabilidad de infección a partir de los modelo Exponencial y Beta-Poisson aproximado.	85
Tabla R 20. Probabilidad de diarrea a partir de los modelos Exponencial y Beta-Poisson aproximado.	85
Tabla R 21. Prueba de comparación entre modelos mecanicistas.	86
Tabla R 22. Resultado del ajuste de la regresión log-logística de tres parámetros a los datos de excreción viral (Log AUCS) y de score fecal de diarrea (AUCD).	88
Tabla R 23. Predichos e intervalos de confianza y predicción del 95% para el AUCS de excreción viral (infección) y el AUCD de severidad de la diarrea (enfermedad).	89
Tabla R 24. Resumen de resultados DI50% and DD50%	90
Tabla R 25. Comparación de infección y enfermedad en por NoV en cerdos gnotobióticos versus voluntarios humanos.	93
Tabla R 26. Comparación de la dosis desafío de cerdos gnotobióticos de dos cepas de NoV humano.	94

Abreviaturas

AIC	Criterio de información de Akaike
AUC	Área bajo la curva
CV	coeficiente de variación
DB	Dragstedt-Beherens
DI _{50%}	Dosis infecciosa 50%
DD _{50%}	Dosis diarrea 50%
D.E.	desvío estándar de la muestra
DP	distancia proporcional
drc	Paquete Dosis respuesta de R
Cin-1 Cin-2	Cepas de Norovirus Cincinati 1 y 2
ELISA	Enzyme linked immunosorbent assay
GLM	Modelos lineales generalizados
HBGA	Carbohidratos del sistema de antígenos histosanguíneos
IC	Interval de confianza
qPCR	Reacción en cadena de la polimerasa en tiempo real cuantitativa
LMM	Modelo lineal mixto
LRT	Tests de cocientes de verosimilitud
MCM	Cadenas de Markov Monte Carlo
MLE	Estimación de Máxima verosimilitud
NoV	Norovirus
PID	Días post inoculación
RT	Retrotranscripción a ADN a

RM	Reed y Muench
SE	Error estándar
SK	Spearman Karber
skrmdb	Paquete de R Spearman Karber, Reed y Muench Dragstedt-Beherens
2P L.L. 3P L.L 4P L.L.	Regresión Log-logística de 2, 3 y 4 parámetros

Resumen

Los modelos animales son utilizados en estudios pre-clínicos para reproducir las infecciones por patógenos humanos. En el proceso de estandarización de estos modelos resulta indispensable conocer cómo se producen la infección y enfermedad a medida que aumenta la dosis del patógeno. Los modelos dosis-respuesta son funciones matemáticas que permiten describir la relación entre la dosis de un patógeno y la severidad de la infección y enfermedad que genera en un hospedador dado. Los métodos aplicados a la estimación de los parámetros de estas curvas se han desarrollado desde hace varios decenios. Inicialmente se empleaban métodos de interpolación matemática, como el método de Reed-Muench y el método de Dragstedt-Beherens, que permiten estimar la dosis de una suspensión viral dada que causa infección o enfermedad en la mitad de los individuos desafiados y denominada dosis efectiva 50%. Si bien estos métodos siguen utilizándose en la actualidad, carecen de base estadística. Más tarde, se desarrolló el método no paramétrico de Spearman-Kärber, que es el que se emplea actualmente a nivel de las agencias regulatorias. Con el advenimiento de nuevas herramientas estadísticas y la capacidad computacional surgieron los modelos lineales generalizados o GLM (por ejemplo, el modelo de la regresión logística). Estos modelos permiten realizar inferencia sobre estadísticos importantes como la dosis efectiva 50% teniendo en cuenta su variabilidad e incertidumbre. Recientemente, y partir de un conocimiento más cabal de la biología detrás de una infección y el desarrollo posterior de una enfermedad, se han desarrollado modelos con base mecanicista. Estos últimos modelos resultan muy valiosos para el estudio de análisis de riesgo asociado a la contaminación de agua y alimentos con agentes virales, como, por ejemplo, Norovirus humano. El presente trabajo de tesis de maestría se propuso emplear los métodos de interpolación matemática clásicos de Reed-Muench y Dragstedt-Beherens, el método no paramétrico de Spearman-Kärber, un modelo lineal generalizado, la regresión logística, y los modelos de regresión no lineal log-logístico y mecanicistas Exponencial y Beta-Poisson para estimar la dosis de infección ($DI_{50\%}$) y enfermedad ($DD_{50\%}$) 50% en un ensayo de dosis-respuesta en cerdos gnotobióticos desafiados con una cepa de Norovirus humano. Todos estos modelos utilizan la condición infectado/no infectado o enfermo/sano o bien la proporción de individuos afectados dentro del grupo que recibió la misma dosis como variable respuesta. Como alternativa novedosa, la tesis exploró también la estimación de la $DI_{50\%}$ y $DD_{50\%}$ ajustando una regresión log-logística sobre el área de la curva de excreción viral y score fecal, dos variables continuas que reflejan la severidad de la infección y del cuadro clínico de diarrea. Los parámetros de los diferentes modelos estadísticos se estimaron por máxima verosimilitud y su incertidumbre se evaluó mediante la estimación de intervalos de confianza aplicando diferentes métodos. La bondad de ajuste y performance relativa de cada modelo se evaluó mediante el criterio de información de Akaike. Los resultados obtenidos de aplicar los diferentes modelos a los datos de proporciones de animales infectados y enfermos indicaron que la $DI_{50\%}$ se encontró alrededor de las 2500 copias de genoma viral y la $DD_{50\%}$ alrededor de las 21000 copias. Independientemente del método empleado, las estimaciones obtenidas arrojaron valores similares. Los modelos mecanicistas Exponencial y Beta-Poisson fueron los que mejor ajustaron los datos de infección. Por su parte, para los datos de enfermedad fue el modelo Beta-Poisson el que arrojó el mejor ajuste. Llamativamente, el ajuste del modelo de regresión log-logístico sobre las variables continuas de área bajo la curva de excreción viral y severidad del cuadro clínico arrojó valores para la $DI_{50\%}$ y la $DD_{50\%}$ que se encuentran en el mismo orden que los calculados en base a los datos discretos, lo que sugiere que este enfoque también podría comenzar a considerarse para estimar las dosis infecciosas y enfermedad 50% para un patógeno viral. Los métodos contemporáneos ofrecen mayor precisión

y exactitud que los métodos clásicos para estimar las dosis efectivas 50%. A su vez, permiten el análisis de datos generados utilizando diluciones no equidistantes y con un número desbalanceado de unidades experimentales entre dosis. La dosis de desafío de la cepa de NoV humano Genogrupo II.2, Cin-2, evaluada en este estudio quedó establecida en 2×10^5 copias de genoma viral, la cual logra reproducir en cerdos gnotobióticos los tiempos de incubación, duración y severidad de la infección y diarrea por NoV que ocurren en voluntarios humanos experimentalmente desafiados con cepas del genogrupo I y II. Estimaciones precisas de los parámetros de infección y enfermedad bajo análisis estadísticos robustos contribuyen a la correcta estandarización del cerdo gnotobiótico como un modelo animal adecuado para estudios pre-clínicos de pruebas de vacunas y otras estrategias antivirales contra Norovirus humanos.

Palabras clave: dosis infecciosa mediana; dosis diarrea mediana, Reed–Muench; Spearman–Karber; Exponencial, beta-Poisson aproximado; regresión logística, 2P.L., 3P.L., AUC, Norovirus, cerdo gnotobiótico.

Abstract

In the process of standardizing animal models in pre-clinical studies, which reproduce infections by human pathogens, it is essential to know how infection and disease occur as the dose of the pathogen increases so that the animal model represents as precisely as possible what happens in humans. Dose-response models are mathematical functions that allow describing the relationship between the dose of a pathogen and the severity of the infection and disease it generates in a given host. The methods applied to the estimation of the parameters of these curves have been developed for several decades. Initially, mathematical interpolation methods were used, such as the Reed-Muench and the Dragstedt-Beherens methods, which allow estimating the 50% effective dose of a given viral suspension, which represents the viral dose that causes infection or disease in the half of the individuals challenged. Although these methods are still in use today, they lack statistical foundation. Later, the nonparametric Spearman-Kärber method was developed, which is the one currently used at the level of regulatory agencies. Recently, with the advent of new statistical tools and computational capacity, generalized lineal models, GLM, emerged including the parametric models such as logistic regression, which possess a sigmoidal shape and allowed a better fit and inference to the dose response data. From a more thorough understanding of the biology behind an infection and the subsequent development of a disease, within GLMs the mechanistically based models like the Exponential and the Beta Poisson regressions have been developed. Under the fulfillment of certain biological assumptions, mechanistic models allow modeling the dose-response relationship with great precision and making statistical inferences considering, at the same time, its variability and uncertainty. These last models are very valuable for the study of risk analysis associated with the contamination of water and food with viral agents, such as, for example, human Norovirus. In the same line of study of dose response trials, sigmoid regressions can also be applied on both discrete binary variables and continuous variables. The latter approach is widely used in pharmacology and toxicology but has not been applied in virology for the 50% dose calculation. The present Master's thesis work proposes to use the classic mathematical interpolation methods of Reed-Muench and Dragstedt-Beheren, the non-parametric method of Spearman-Kärber, logistic regression and the Exponential and Beta-Poisson mechanistic models to estimate the dose of infection ($ID_{50\%}$) and disease 50% ($DD_{50\%}$) from discontinuous, dichotomous or binary variables such as the proportion of infected and diseased animals obtained in a dose-response test in gnotobiotic pigs challenged with a strain of human norovirus. The thesis also explores the dose response effect and the estimation of the $ID_{50\%}$ and $DD_{50\%}$ using sigmoid models as the two parameter log-logistic distribution from the binary variable mentioned above and also a three parameter log-logistic distribution applied on the continuous variables corresponding to the area of the viral excretion curve and fecal score, which reflect the severity of the infection and the clinical picture of diarrhea. The parameters of the different statistical models were estimated by the maximum likelihood method and their uncertainty was evaluated by the estimation of confidence intervals using different calculations (Delta method, Bootstrap simulation method, etc). The goodness of fit and relative performance of each model was evaluated using the Akaike information criterion (AIC). The results obtained from applying the different models to the data of proportions of infected and diseased animals indicate that the $ID_{50\%}$ is around 2500 genome of viral genomes, and the $DD_{50\%}$ around 21000 copies. All the results obtained from these dichotomous variables yielded similar values regardless of their methodological basis. Of the mechanistic methods, both the Exponential distribution and the Poisson beta were optimal to fit the infection data. However, for the disease

data, the Poisson beta distribution was the one that yielded the best fit. Interestingly, the adjustment of a 3P L.L. model on the continuous variables AUC of virus shedding and AUC of diarrhea score yielded $\log DI_{50\%} = 3.85$. and $\log DD_{50\%}=4.46$ which are in the same order as those calculated based on the binary data of proportions of infected and diseased animals, which suggests that this approach could also begin to be considered to estimate the infectious doses and disease 50% for a viral pathogen. Contemporary methods offer greater precision and accuracy in estimating 50% effective doses than classical methods. In turn, they allow the analysis of data generated using non-equidistant dilutions and with an unbalanced number of experimental units between doses. The challenge dose of the human NoV Genogroup II.2, Cin-2 strain, evaluated in this study was established at 2×10^5 copies of viral genome, which manages to reproduce in gnotobiotic pigs the incubation times, duration and severity of infection and diarrhea by NoV that occur in human volunteers experimentally challenged with NoV group I and II. Precise estimates of the infection and disease parameters under robust statistical analysis contribute to the correct standardization of the gnotobiotic pig as a suitable animal model for pre-clinical studies of human NoV.

Keywords: median infectious dose; Reed–Muench; Spearman–Karber; approximate beta-Poisson; dose–response, germ-free pig model, Norovirus, beta Poisson, Exponential, Logistic regression, 2P L.L., 3P L.L., AUC

1. INTRODUCCIÓN

1.1. Modelos de dosis respuesta aplicados al estudio de exposición oral a patógenos entéricos

En el contexto de los análisis de riesgo asociados al consumo de agua o alimentos contaminados, los estudios de dosis-respuesta estiman la probabilidad de respuesta en términos, por ejemplo, de infección, enfermedad y muerte frente a la exposición a una dosis conocida de un enteropatógeno dado. Los modelos dosis-respuesta son funciones matemáticas que permiten describir la relación entre la dosis del patógeno y la respuesta en términos de severidad de infección y enfermedad que se genera en un hospedador dado. Para un punto final particular (del inglés, *endpoint*), un patógeno específico, una ruta de exposición y un hospedador (humano o modelo animal) hay una relación dosis-respuesta única y, consecuentemente, un modelo de dosis-respuesta que mejor ajusta y describe a este fenómeno (Haas, Rose, and Gerba 2014).

Los modelos dosis-respuesta se han desarrollado desde hace varios decenios y se aplican fundamentalmente en el campo de la toxicología y la farmacología. Pero también se han utilizado en estudios de riesgo asociados a la microbiología del agua y los alimentos. En particular, se han utilizado modelos estadísticos para describir la relación dosis-respuesta a la exposición a agua y alimentos contaminados con microorganismos, incluyendo a agentes virales como Norovirus (NoV) humanos. Estos modelos proporcionan información útil, teniendo en cuenta al mismo tiempo la variabilidad y la incertidumbre del fenómeno (Chaudhry et al. 2017; Haas, Rose, and Gerba 2014).

1.2. Nuestro ejemplo biológico: modelo de cerdo gnotobiótico para reproducir la infección y diarrea por Norovirus Humano

Los NoV humanos son los agentes etiológicos más comunes asociados a gastroenteritis agudas epidémicas y esporádicas que afectan a personas de todas las edades. Los NoV son responsables de alrededor de 125 millones de casos anuales de diarrea alimentaria a nivel mundial (Payne et al. 2013). La infección por NoV se caracteriza por diarrea aguda, náuseas y vómitos, síntomas que pueden desencadenar un cuadro de deshidratación que requiere hospitalización. Por este motivo, generan un importante impacto económico en términos de salud pública. Estadísticas recientes estiman que la enfermedad causa un gasto de alrededor de \$4,2 mil millones de dólares al sistema de salud y más de \$60 mil millones a nivel social global por año (Iturriza-Gómara and Lopman 2014; Bartsch et al. 2016). Actualmente no se cuenta con vacunas u otros tipos de estrategias de prevención o tratamiento, pero varias vacunas a *virus like particles* se encuentran actualmente en ensayos clínicos y dan resultados prometedores de inmunogenicidad y eficacia protectora (Atmar and Estes 2012; Atmar, Ramani, and Estes 2018; Bernstein et al. 2015).

Los NoVs humanos solo logran multiplicarse en grandes cantidades en el intestino humano ($\sim 10^{11}$ partículas virales por gramo de heces), por lo que las pruebas de estrategias preventivas o terapéuticas, como vacunas, drogas antivirales o estrategias de inmunidad pasiva, se evalúan en ensayos de desafío oral en voluntarios. Las personas enroladas en estos ensayos reciben por vía oral un filtrado de heces de pacientes infectados que contienen una cantidad de virus capaz de causar infección y diarrea (Frenck et al. 2012; Bernstein et al. 2015). Este es el *gold standard* y la dosis de virus para reproducir el cuadro clínico natural de infección y

enfermedad se la denomina *dosis desafío*. Antes de realizar pruebas con humanos, sin embargo, es necesario llevar a cabo ensayos pre-clínicos en modelos animales.

La comunidad científica ha realizado enormes esfuerzos para lograr tanto la replicación de NoV *in vitro* así como para desarrollar modelos animales adecuados para reproducir la infección y diarrea por NoV. A la fecha, se ha desarrollado un modelo de NoV murino (Taube et al. 2013) y un modelo chimpancé (Bok et al. 2011). Sin embargo, el único modelo animal que reproduce la infección y enfermedad de forma aceptable es el cerdo gnotobiótico (animal libre de gérmenes mantenido en aisladores estériles), experimentalmente inoculado con muestras de materia fecal de pacientes infectados con el virus, dado que el cerdo posee una fisiología e inmunidad similar a la de los seres humanos y en las condiciones de aislamiento y esterilidad se puede controlar muy bien el curso de la infección y diarrea (Yang and Yuan 2014; Bui et al. 2013).

El cerdo posee en la mucosa intestinal carbohidratos de histogrupos sanguíneos HBGA similares a los de los humanos (mucina gástrica de cerdo), que funcionan como los receptores celulares que utilizan los NoV humanos para ingresar a la célula y comenzar su ciclo de replicación. El fenotipo de HBGA en humanos puede ser secretor ó no secretor, este último es naturalmente resistente a la infección con NoV. El fenotipo de HBGA de los cerdos da como resultado en todos los casos individuos secretores susceptibles a la infección. Hasta el momento, los cerdos han sido infectados exitosamente con dos variantes virales (GII.4/HS66 y GII.4/2006b), arrojando niveles comparables de infectividad al obtenido por las cepas HuNoV GI.1/Norwalk y GII.4/2003 Cin-1 en ensayos en humanos (Souza et al. 2008, 2007; Todd and Tripp 2019) (**Figura 1**).

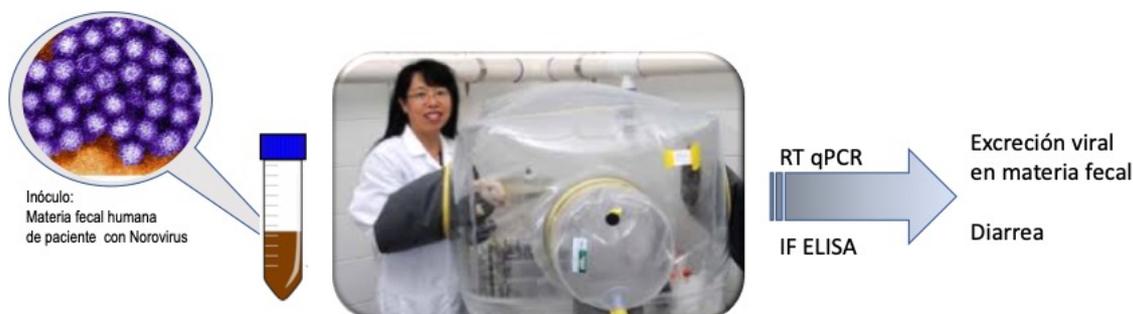


Figura 1. Modelo cerdo gnotobiótico de infección y diarrea por Norovirus humano, Dra. Yuan. En este modelo los lechones inoculados con una dosis adecuada de virus proveniente de materia fecal humana de paciente infectado desarrolla diarrea leve, viremia transiente y excreción viral en heces, Fuente:(Cheetham et al. 2007; Souza et al. 2007).

Para garantizar que los ensayos de desafío sean repetibles y reproducibles, se debe administrar una *dosis desafío* viral fija que reproduzca la infección y enfermedad con una duración e intensidad similar a la infección natural. Para establecer dicha dosis de desafío se realizan ensayos dosis-respuesta donde se estiman la dosis infecciosa 50% ($DI_{50\%}$) y/o la dosis diarrea 50% ($DD_{50\%}$), que representan a aquella dosis de virus de desafío que genera la respuesta esperada en el 50% de las unidades experimentales del ensayo. La dosis de desafío final del ensayo se expresa como un múltiplo de estos valores de referencia (Teunis et al. 2008; Philip J. Schmidt 2015; Bui et al. 2013).

1.3. Bases biológicas de los modelos de relación dosis respuesta

La base biológica de los modelos que se utilizan para describir la relación dosis-respuesta deriva de las etapas críticas que ocurren en el proceso que tiene lugar cuando un hospedador se expone a un patógeno presente en un vehículo contaminado, que puede ser agua, un alimento o fomites. En la **Figura 2** se exponen las principales etapas del proceso infeccioso. En cada una de estas etapas intervienen numerosos factores biológicos. El patógeno debe superar múltiples barreras de defensa del hospedador para lograr la infección y, por lo tanto, cada patógeno tiene cierta probabilidad de infectar al hospedador (Haas, Rose, and Gerba 2014).

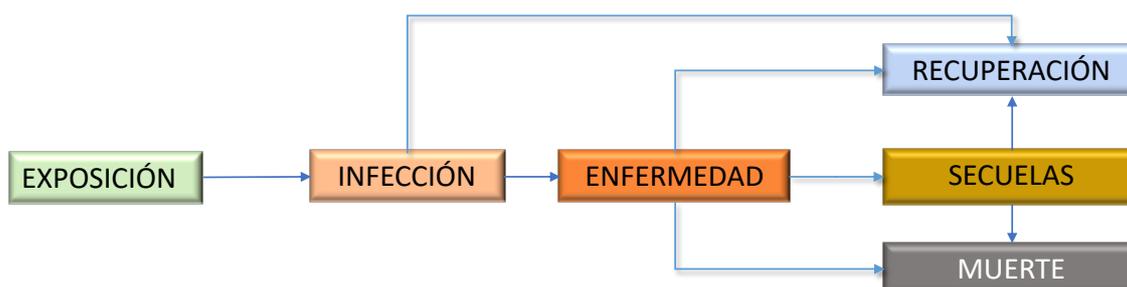


Figura 2. Principales etapas en el proceso de una enfermedad infecciosa transmitida por los alimentos.

En el caso de la infección por NoV en humanos estos pasos se cumplen en la mayoría de los casos hacia la recuperación, aunque existen pacientes de riesgo que pueden desarrollar como secuela una infección crónica y/o muerte. El modelo cerdo gnotobiótico se estandariza para reproducir una infección aguda por NoV humanos, que cumple los pasos hacia la recuperación dentro de los siete días post exposición (Bui et al. 2013; Yang and Yuan 2014).

Un estudio dosis-respuesta describe la probabilidad de observar una respuesta específica (por ejemplo, número de individuos infectados ó enfermos del total de expuestos) en función de la dosis de un patógeno concreto sobre una población de hospedadores determinada. Esta función se basa en datos empíricos y normalmente se expresa en forma de relación o función matemática. El uso de modelos estadístico-matemáticos es necesario porque resulta útil en el análisis de riesgo frente a la exposición de alimentos o agua contaminada, la cual se suele producir en concentraciones bajas del virus o en circunstancias excepcionales. Dado que no se pueden hacer ensayos que abarquen todo el rango de dosis necesaria, los modelos matemáticos permiten extrapolar que ocurrirá a dosis muy elevadas o a la exposición de dosis muy bajas.

Los datos de respuesta de un grupo de individuos expuestos a diferentes dosis del patógeno (a menudo expresada en escala logarítmica) muestran con frecuencia una forma sigmoidea (Figura 3), que se puede ajustar por medio de un gran número de funciones matemáticas. Sin embargo, cuando se realiza una extrapolación fuera de la región de los datos observados, cada una de estas funciones puede pronosticar resultados diferentes (Ryan et al. 2014; Chaudhry et al. 2017). Por consiguiente, es necesario seleccionar una de entre las numerosas funciones que representan la relación dosis-respuesta observada, considerando cuidadosamente los aspectos biológicos de la interacción patógeno-hospedador-matriz. El modelo matemático también nos permite establecer el rango de dosis que produce una respuesta lineal y estimar la pendiente de dicha recta (Figura 3). Asimismo, dado que los patógenos como NoV presentes en los alimentos y el agua normalmente no se encuentran dispersos al azar, sino que aparecen formando grupos o aglomeraciones, los modelos

matemáticos actuales permiten considerar estos efectos, así como la existencia de respuesta inmune frente a exposiciones previas al patógeno, al momento del ajuste a los datos empíricos de los ensayos de dosis-respuesta. Por último, el tamaño de los grupos experimentales es limitado y se necesitan modelos estadístico-matemáticos, incluso en experimentos bien controlados, para distinguir las variaciones al azar de los verdaderos efectos biológicos.

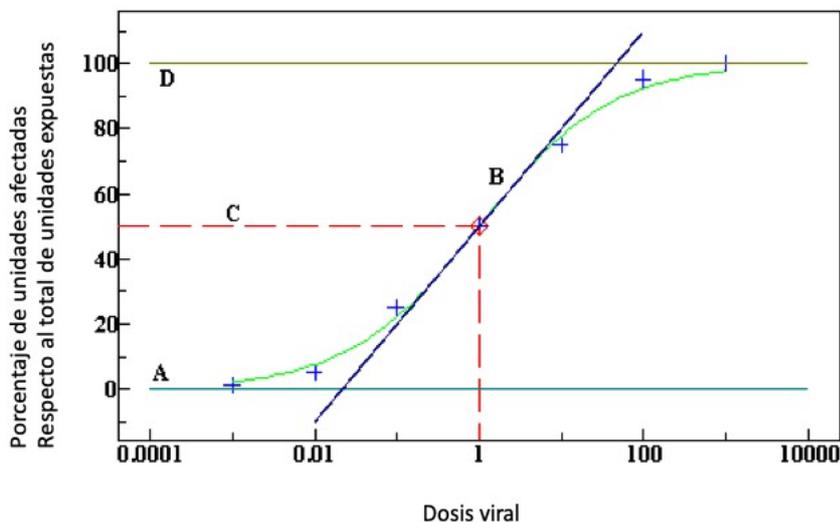


Figura 3. Modelo estadístico-matemático ajustado a datos de un ensayo dosis respuesta.

Eje y: porcentaje de individuos infectados/enfermos del total de expuestos en función de la dosis viral (Eje x). A: Límite inferior dosis que no logran una respuesta. B: región lineal de respuesta proporcional a la dosis. D: Plateau superior dosis a partir de las cuales el 100% de los expuestos se infecta/enferma. Punto rojo indica la dosis viral correspondiente a la dosis efectiva 50%.

a. Exposición -> Característica particulada del inóculo y efectos de agregación

En general, los modelos de relación dosis-respuesta para patógenos virales deben tener en cuenta su carácter discreto (particulado) y, en consecuencia, deben basarse en el concepto de infección por uno o más "supervivientes" de una dosis inicial. La dosis es la cantidad de virus infeccioso presente en el volumen de inóculo administrado. La dosis viral se puede medir por efecto citopático sobre células susceptibles, por ELISA para evaluar masa antigénica, por recuento de partículas virales al microscopio electrónico o por métodos moleculares. Es importante aclarar que el NoV humano no se replica en cultivos celulares convencionales y, por lo tanto, la cantidad de virus presente en el inóculo administrado a cada grupo experimental y la cantidad de virus excretada por cada individuo expuesto a una dosis dada se expresa como el número de copias de genoma viral determinado por retro transcripción seguida por la reacción en cadena de polimerasa en tiempo real (RT-PCR *real time*) cuantitativa (qRT-PCR). Luego, la dosis promedio (el número promedio de virus infeccioso ingerido por un grupo determinado de individuos) se obtiene multiplicando la concentración viral del inóculo por el volumen ingerido o administrado del mismo. **El número de virus infecciosos o dosis ingerida por un individuo en particular es una variable aleatoria centrada en este promedio y caracterizada por una distribución de probabilidad.** Cuando el virus está presente en una suspensión de materia fecal bien homogenizada, en donde los acúmulos virales que constituyen las "unidades infecciosas virales" se distribuirán más o menos al azar, se suele utilizar la **distribución Poisson** para

caracterizar a esta variable (Philip J. Schmidt, Emelko, and Reilly 2010; Philip J. Schmidt and Emelko 2010; P. J. Schmidt, Emelko, and Thompson 2013).

En dos estudios de dosis respuesta de infección por NoV realizados en humanos los autores sugirieron la presencia de un efecto de agregación (Messner, Berger, and Nappier 2014; Teunis et al. 2008) debido a una alta concentración de proteínas en la muestra de la cepa viral utilizada como inóculo. Estos agregados no pudieron disociarse aun por sonicación y tampoco por el efecto de dilución. En una fotografía obtenida por microscopía electrónica se observaron acúmulos virales de al menos 18 partículas. Para evaluar como el efecto de agregación modificaba la curva dosis respuesta los autores prepararon una muestra desagregada fresca por pasaje en humanos y se realizaron los ensayos dosis-respuesta con ambos inóculos (Philip J. Schmidt 2015). El análisis de los datos mediante un modelo de dosis-respuesta Beta-Poisson Exacto (a.k.a. hypergeometric) permitió estudiar este efecto así como el tamaño de los agregados (Nelder and Wedderburn 1972), considerando en la ecuación del modelo que el número de agregados se encontraba distribuido en el inóculo siguiendo una distribución de Poisson y que el tamaño de los agregados variaba según una distribución log-series.

b. Infección -> Modelos de relación dosis-infección

Cada unidad viral infecciosa de la dosis ingerida tiene que superar una serie de barreras hasta alcanzar el lugar específico de adsorción al tejido susceptible, penetrar en la célula y dar inicio al ciclo de replicación viral intracelular. La relación entre el número real de partículas virales infecciosas supervivientes (dosis efectiva) y la probabilidad de infección del hospedador es un concepto fundamental en la derivación de modelos matemáticos utilizados en la descripción de la relación dosis-respuesta. La infección se define casi siempre como una situación en la cual el agente viral, tras la ingestión y la superación de todas las barreras, se multiplica activamente en su sitio específico de replicación. La infección se suele medir como la proporción de individuos con presencia o ausencia de infección respecto del total de expuestos. La infección se puede evidenciar por los mismos métodos que se enumeraron para el inóculo: midiendo la cantidad partículas virales infecciosas excretadas, su masa antigénica o su genoma viral, en materia fecal. El uso de variables de respuesta continua (por ejemplo, el área bajo la curva de excreción viral en un período de tiempo dado) puede ser útil para el perfeccionamiento ulterior de los modelos matemáticos aplicados al estudio de la relación dosis-respuesta.

Los NoV humanos son altamente contagiosos. Unos pocos estudios han evaluado el riesgo de infección y han calculado que la dosis infecciosa 50% (DI_{50}) de la cepa de referencia Norwalk en humanos sería de alrededor de 2.800 copias de genoma viral (equivalente a ~ 14-18 unidades víricas infecciosas) para infectar a un humano adulto voluntario sano (Frenck et al. 2012). Cuando el hospedador recibe una dosis baja de NoV, se infecta y excreta el virus en heces en recuentos bajos y por un plazo de tiempo limitado, sin presentar síntomas, es decir, sufriendo una infección asintomática. Al recibir dosis mayores del patógeno la infección será cada vez más severa, con una excreción elevada de virus y enfermedad sintomática de severidad creciente (WHO, 2016). En el modelo cerdo gnotobiótico (**Figura 1**) se cuenta con un único estudio que estimó la DI_{50} de la cepa de NoV GII.4/2006b, realizado por Bui y col, 2013. El mismo involucró lechones de dos grupos etarios, 4-5 días de edad ($DI_{50} \leq 2.74 \times 10^3$) y 33-34 días de edad ($DI_{50} \leq 6.43 \times 10^4$). El cálculo se realizó con el método clásico de Reed y Muench y no se empleó ningún método basado en modelos estadísticos (Bui et al. 2013). En el presente trabajo de tesis se analizaron los datos de un nuevo ensayo dosis respuesta en animales de 33-34 días.

c. Enfermedad: Diarrea por Norovirus -> Modelos de relación dosis enfermedad

La severidad de la enfermedad asociada a la infección causada por un virus entérico es un proceso que se mide como una variable categórica denominada *score fecal* que combina variables cuantitativas, como el número de deposiciones al día y la temperatura corporal, y variables cualitativas como la consistencia, color y olor de las heces o la presencia o ausencia de sangre o mucosidad, etc. En los estudios de dosis-respuesta y en los estudios de evaluación de riesgos la enfermedad, la respuesta se evalúa como una variable dicotómica: presencia o ausencia de enfermedad. En general, en el caso de la infección por NoV, que genera una diarrea acuosa no enterocolítica ni hemorrágica, se utiliza un *score fecal* que va de 1 a 4. Las materias fecales con un score de 2 o más son consideradas diarrea. Los individuos expuestos dosis infecciosas de entre 2.800 y 64.000 partículas del virus contraen la enfermedad, que se caracteriza por síntomas de diarrea y vómitos. Luego de un promedio de dos a tres días el cuadro clínico se autolimita (Teunis et al. 2008). El modelo de infección de cerdos gnotobióticos (**Figura 1**) se debe estandarizar muy bien para reproducir el cuadro clínico que ocurre en los humanos (Bui et al. 2013).

1.4. Modelos aplicados a los estudios dosis-respuesta y cálculo de la dosis efectiva 50%

En virología, los métodos comúnmente utilizados para estimar la dosis infecciosa 50% ($DI_{50\%}$) y la dosis diarrea 50% ($DD_{50\%}$) son los métodos de interpolación de Reed-Muench (Reed and Muench 1938) y Dragstedt-Beherens (Dragstedt and Lang 1928) y el método no paramétrico de Spearman Karber (Atkinson 1961; Ramakrishnan 2016; Kä r ber 1931; Spearman 1908). Estos métodos “clásicos”, que proporcionan una estimación puntual de la dosis que infecta al 50% de los sujetos experimentales, fueron desarrollados antes de la disponibilidad generalizada de la computación. Actualmente, también se emplean algunos modelos estadísticos de la familia de los modelos lineales generalizados (GLM). Un ejemplo es la regresión logística, que permiten la estimación de la dosis infecciosa para cualquier proporción de los individuos desafiados y el análisis apropiado de los datos aún cuando las dosis no están igualmente espaciadas (en escala logarítmica) o el número de sujetos por grupo es variable. Sin embargo, y más allá de estas ventajas, en última instancia la regresión logística solo proporciona un ajuste empírico a los datos de dosis-respuesta. La regresión log-logística de dos parámetros (2P L.L.), en cambio, también ajusta una curva sigmoidea y simétrica alrededor de su punto de inflexión, pero la distribución está indexada por un parámetro que representa la dosis efectiva 50%. Finalmente, los modelos mecanicistas (e.g. Exponenciales y Beta-Poisson) comparten las ventajas de la regresión logística y log-logística, pero al tener en cuenta aspectos propios de la biología de la infección, sus parámetros tienen una interpretación biológica de utilidad y, en consecuencia, la inferencia estadística cobra otro sentido.

1.5. Criterios utilizados en la selección de modelos estadísticos

A la hora de seleccionar los modelos estadísticos más apropiados para ajustar datos de un ensayo de dosis-respuesta se pueden aplicar distintos criterios. Según la Organización Mundial de la salud (WHO, 2016), para que un modelo sea aceptable debe satisfacer el criterio estadístico de la **bondad de ajuste** y, a la vez, ser **flexible**. Un modelo flexible es aquel que permite describir el rango completo de respuesta con precisión, incluso en dosis bajas. La adición

de parámetros suele mejorar la flexibilidad de un modelo y su bondad de ajuste, pero la utilización de un modelo con demasiados parámetros puede dar lugar a una mayor incertidumbre de las estimaciones, especialmente para las dosis extrapoladas. Por otro lado, un modelo basado en postulados sólidos y pocos parámetros podría ser demasiado restrictivo y dar lugar a una infravaloración de la incertidumbre en las estimaciones de riesgos. En general, la literatura recomienda una primera selección de modelos de dosis-respuesta en base a una serie de postulados biológicamente admisibles para, luego, realizar un análisis estadístico con los modelos que se consideren más verosímiles (World Health Organization 2016; P. J. Schmidt and Emelko 2011).

En la **selección de modelos** también es importante tener en cuenta que dos (o más) formas funcionales distintas (basadas en postulados diferentes) pueden describir igualmente bien los mismos datos de la relación dosis-respuesta. Esto puede dar lugar a curvas ajustadas muy diferentes si los datos contienen escasa información o bien prácticamente a las mismas curvas si los datos contienen información abundante. Sin embargo, incluso en el último caso la extrapolación del modelo puede ser muy diferente (Ryan et al. 2014).

En general en un estudio dosis respuesta, una vez seleccionado el modelo que se utilizará para caracterizar dicha relación, se ajusta a los datos para estimar los parámetros de sus ecuaciones. Los datos de un ensayo de dosis-respuesta muestran la relación entre la **mediana** de la concentración de la dosis de patógeno administrada (por ejemplo, copias del genoma de NoV) y la proporción de unidades experimentales (pacientes, animal experimental; e.g. cerdo gnotobiótico) que experimentan un efecto determinado (infección, enfermedad), el denominado *end point* ó punto final. Con los parámetros estimados, luego, se evalúa la bondad de ajuste del modelo. El método preferible es el de **máxima verosimilitud (MLE)** (WHO 2016; Haas, Rose, and Gerba 2014). Este método se basa en asignar una distribución de probabilidad a la respuesta condicional en la dosis. Las distribuciones más frecuentes son la Binomial, Exponencial y Beta-Poisson. Luego, se obtienen los valores numéricos de los parámetros de los modelos derivados de estas distribuciones que maximizan la probabilidad de observar los datos. Esto se realiza para cada modelo y se selecciona, entre todos los modelos ajustados, aquel que minimiza la *deviance* y el índice de información de Akaike, AIC (Spiegelhalter et al. 2002; Akaike 1981, 1974).

Para caracterizar la incertidumbre de los parámetros estimados se calculan intervalos de confianza. En algunos casos estos se derivan del mismo método de máxima verosimilitud y, en otros, se obtienen mediante técnicas de re-muestreo como Bootstrap (Literatur et al. 2003; Weir et al. 2017) o cadenas de Markov Montecarlo, MCM (Haas, Rose, and Gerba 2014). Por ejemplo, para el caso de datos binarios se pueden generar réplicas mediante un muestreo aleatorio a partir de una distribución binomial en cada dosis, con un número de ensayos igual al número de sujetos expuestos y una probabilidad expresada por la fracción del cociente entre el número de sujetos infectados y el de expuestos. Luego se ajusta el modelo a cada uno de estos conjuntos de datos replicados y se obtiene una muestra aleatoria de estimaciones de parámetros, una para cada réplica. Éstas se pueden utilizar posteriormente en la creación de un intervalo de confianza para la curva de dosis-respuesta o bien para evaluar la incertidumbre con una dosis determinada. Por su parte, los métodos MCM son métodos adaptativos de muestreo por rechazo y constituyen un instrumento poderoso y eficaz cuando es necesario analizar modelos con muchos parámetros en el contexto de conjuntos de datos reducidos, con apenas unos pocos grupos de dosis y escasos individuos expuestos en cada grupo experimental (Philip J. Schmidt et al. 2013; Philip J. Schmidt 2015).

2. OBJETIVOS

En este trabajo de tesis de maestría se utilizaron datos obtenidos de un ensayo de dosis-respuesta en cerdos gnotobióticos experimentalmente desafiados por vía oral con dosis crecientes de una cepa de NoV humano. Como resultado del ensayo, se contaba con variables de recuento que indicaban, para cada grupo experimental, la proporción de cerdos con excreción viral y con diarrea (variables binaria o dicotómicas: infectado/no infectado y enfermo/sano, respectivamente). A partir de estos datos se emplearon diferentes abordajes (incluyendo métodos de interpolación matemática clásicos y diferentes modelos estadísticos) con el objetivo de calcular la dosis 50% de infección y enfermedad (diarrea).

El **objetivo general** de la tesis fue estimar la dosis infecciosa 50% ($DI_{50\%}$) y la dosis de diarrea 50% ($DD_{50\%}$) a partir de datos generados en un ensayo preclínico de reproducción de la infección y diarrea causadas por Norovirus GII.4, la variante prevalente en humanos a nivel mundial, en cerdos gnotobióticos utilizando diferentes enfoques matemáticos y estadísticos. Los objetivos específicos fueron:

1. Realizar un análisis exploratorio de los datos.
2. Comparar las medias asociadas a la infección y diarrea entre los grupos de cerdos que recibieron diferentes dosis virales.
3. Investigar el grado de agregación del inóculo mediante un análisis de regresión de Poisson entre el tamaño de los acúmulos virales y el número de copias de genoma por dosis.
4. Estimar la dosis infecciosa 50% ($DI_{50\%}$) y dosis de diarrea 50% ($DD_{50\%}$) a partir de los métodos clásicos (Reed y Muench, Dragstedt-Behrens y Spearman Karber).
5. Estimar las $DI_{50\%}$ y $DD_{50\%}$ ajustando modelos lineales generalizados de regresión logística y el modelo no lineal de regresión log-logística de dos parámetros sobre los datos de proporción de cerdos afectados.
6. Estimar las $DI_{50\%}$ y $DD_{50\%}$ ajustando los modelos mecanicistas Exponencial y Beta-Poisson sobre los mismos datos.
7. Estimar las $DI_{50\%}$ y $DD_{50\%}$ ajustando un modelo de regresión log-logística de tres parámetros a las variables continuas área bajo la curva de excreción viral y de severidad de diarrea.
8. Aplicar los resultados para establecer la dosis adecuada para desafiar los animales y compararla con la dosis de desafío utilizada en voluntarios humanos.

Para evaluar el grado de incertidumbre en la estimación de las $DI_{50\%}$ y $DD_{50\%}$ se construyeron intervalos de confianza para las curvas de dosis-respuesta aplicando diferentes métodos. Los diferentes modelos lineales generalizados, mecanicistas y no mecanicistas, se compararon en términos de su bondad de ajuste mediante el criterio de información de Akaike (AIC).

3. Materiales y Métodos

3.1. Datos y diseño experimental

Se partió de una base de datos obtenida a partir de un ensayo dosis-respuesta realizado en cerdos gnotobióticos que recibieron dosis crecientes de NoV humano (materia fecal de voluntario humano, cepa Cin-2). Brevemente, un total de 32 cerdos gnotobióticos de 33-34 días de edad, alojados en aisladores estériles, fueron asignados aleatoria e independientemente del género y del peso al nacer a los ocho grupos experimentales. Los animales enrolados en los primeros 7 grupos fueron desafiados por vía oral con un volumen de inóculo de 5 ml con volúmenes crecientes de materia fecal con NoV tal que recibieron las dosis virales detalladas en la **Figura 4** medidas por la técnica de RT-qPCR y expresada como número de copias de genoma viral. Todos los cerdos se controlaron para registrar diarrea y excreción viral durante siete días, momento en el cual se realizó la eutanasia. Este ensayo tuvo lugar en el laboratorio de la Dra. Ljuan Yuan, Virginia Tech University, Blacksburg, Virginia, EE.UU., entre julio de 2017 y marzo de 2018. La cepa viral se denomina Cin-2 y corresponde a la muestra de materia fecal # 103041 recolectada de un voluntario humano desafiado con la cepa Cin-1 (Dra. Yuan, comunicación personal).

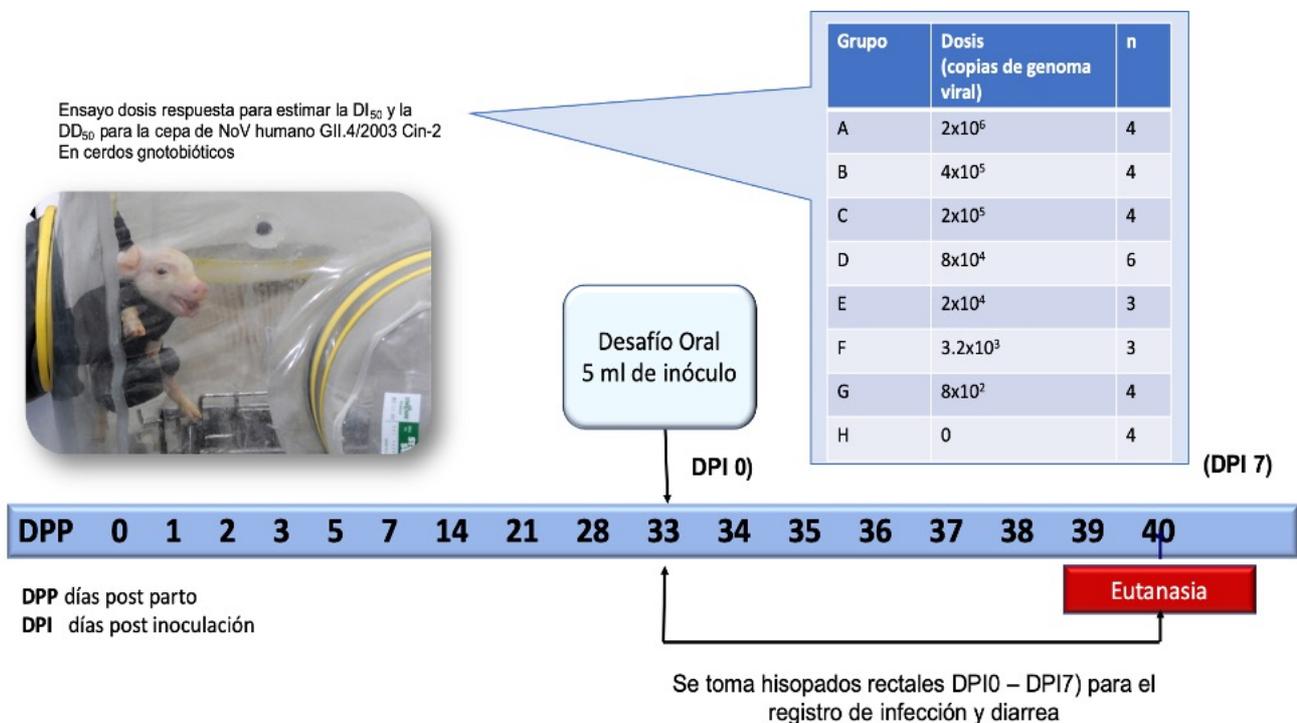


Figura 4. Diseño Experimental.

3.2. Variables registradas

Durante el experimento se registró el porcentaje de animales infectados (con excreción viral), el momento del inicio de excreción (número días postinoculación), la excreción promedio para cada grupo experimental, la duración de la excreción (días) por cerdo. También se registró la cantidad de virus excretado, expresada como el número de copias de genomas de NoV humano por RT-qPCR en muestras de frotis rectales recolectadas después de la inoculación con las diferentes dosis de la cepa de NoV Cin-2. Con los datos obtenidos se calculó el área bajo la curva de excreción viral durante los 7 días de duración del ensayo para cada animal y el promedio por grupo, utilizando el comando AUC del paquete DescTools de R (<https://cran.r-project.org/web/packages/DescTools/index.html>) y la opción “spline”.

Con respecto al desarrollo de la enfermedad, se registró el porcentaje de animales enfermos (con diarrea) en cada grupo. Se registró el día promedio de inicio y el promedio de días de duración de la diarrea. Se estableció una medida de la severidad de la diarrea en función de la consistencia fecal. Las puntuaciones de consistencia fecal se determinaron sobre la base de estudios previos realizados por el grupo de la Dra. Yuan (Lei et al. 2016b, 2016a): 0 sólida; 1 pastosa 2 semilíquidos; 3 líquido. Los cerdos con una puntuación de consistencia fecal de 2 o más (≥ 2) se consideraron diarreicos. Finalmente, se calculó el área bajo la curva de diarrea durante los 7 días de duración del ensayo para cada animal y el promedio por grupo utilizando el comando AUC del paquete DescTools de R bajo la condición "spline".

3.3. Análisis exploratorio

Inicialmente la base de datos se exploró para investigar la naturaleza de cada variable registrada. Las variables relacionadas con la infección (excreción viral) se analizaron mediante un modelo lineal mixto anidado (LMM), considerando el grupo o tratamiento (cerdos que recibieron diferentes dosis de NoV humano) como un factor fijo y los cerdos gnotobióticos anidados dentro de los grupos como una unidad experimental independiente. Se utilizó una estructura *varIdent* de matriz de varianza-covarianza para modelar la heterogeneidad de la varianza. Las variables asociadas al desarrollo de la enfermedad que cumplieron con los supuestos de normalidad y homocedasticidad, por su parte, se analizaron mediante ANOVA de una vía. En todos los casos, se realizaron comparaciones múltiples post-ANOVA utilizando el método de Tukey para las comparaciones entre las siete dosis diferentes. Para las variables momento de inicio y duración de la excreción viral y diarrea se realizó un análisis de regresión lineal. En todos los casos, la significancia estadística se consideró con un $p < 0,05$ para todas las comparaciones. Los análisis se realizaron con el programa Infostat® (<https://www.infostat.com.ar>) conectado al software R (R Core Team) y en R Studio.

3.4. Agregación viral: análisis de regresión del tamaño de los acúmulos virales en función del número de copias genómicas en el inóculo

Dado que era necesario corroborar que el inóculo administrado a los animales fuera una suspensión viral homogéneamente dispersa para que se ajuste a los supuestos de los modelos mecanicistas a utilizar, antes de comenzar con los cálculos del $DI_{50\%}$ y $DD_{50\%}$ se decidió investigar el potencial efecto de agregación que se produce en muestras congeladas por tiempos prolongados. Para ello se cuantificó el tamaño de los acúmulos virales (número de viriones por acúmulo) en cada dilución de virus correspondiente a las dosis probadas en cerdos. Se prepararon diluciones frescas partir de diferentes volúmenes del stock descongelado y disuelto en 5 ml de medio. Con cada suspensión viral se prepararon grillas de microscopía electrónica que fueron teñidas por tinción negativa y observadas al microscopio electrónico de barrido. Se contó el número de viriones en los acúmulos virales presente en 10 cuadrículas de cada grilla / dosis. Luego, se estudió la relación entre el tamaño de los acúmulos (expresado en número de viriones) y el título del virus en cada dosis (expresado en número de copias del genoma). Los datos fueron analizados por una regresión de Poisson con un componente lineal y cuadrático y se seleccionó el modelo de mejor ajuste y de menor índice de información de Akaike, con el paquete glm de R.

3.5. Métodos clásicos aplicados al cálculo de las dosis infecciosa y diarrea 50%

Los métodos utilizados para estimar la $DI_{50\%}$ y la $DI_{50\%}$ fueron, en primer lugar, los métodos de interpolación de Reed-Muench (Reed and Muench 1938) y Dragstedt-Beherens (Dragastedt and Lang 1928) y el método no paramétrico de Spearman-Kärber (Kärber 1931; Spearman 1908).

a. Método de Reed-Muench

El método Reed-Muench, descrito por primera vez en 1938, es un método que ha sido ampliamente utilizado para el cálculo de dosis infecciosas virales 50% y emplea dos ecuaciones. La primera ecuación involucra el cálculo previo de las sumas acumuladas de individuos afectados y no afectados y los porcentajes asociados a cada una de esas sumas. A modo de ejemplo, en la **Figura 5** ilustra los pasos necesarios para calcular el porcentaje acumulado de individuos afectados.

(a) Respuestas observadas (Pos, Neg), (b) respuestas hipotéticas (Cum pos, Cum Neg)
 Respuestas hipotéticas ajustadas (AdjCumPos, AdjCumNeg)

Pos	Neg	CumPos	CumNeg	AdjCumPos	AdjCumNeg
0	10	0	28	0.0	27.6
3	7	3	18	2.9	17.9
4	6	7	11	6.8	11.2
6	3	13	5	13.2	5.4
7	2	20	2	20.7	2.1
10	0	30	0	30.4	0.0

(b) respuestas hipotéticas

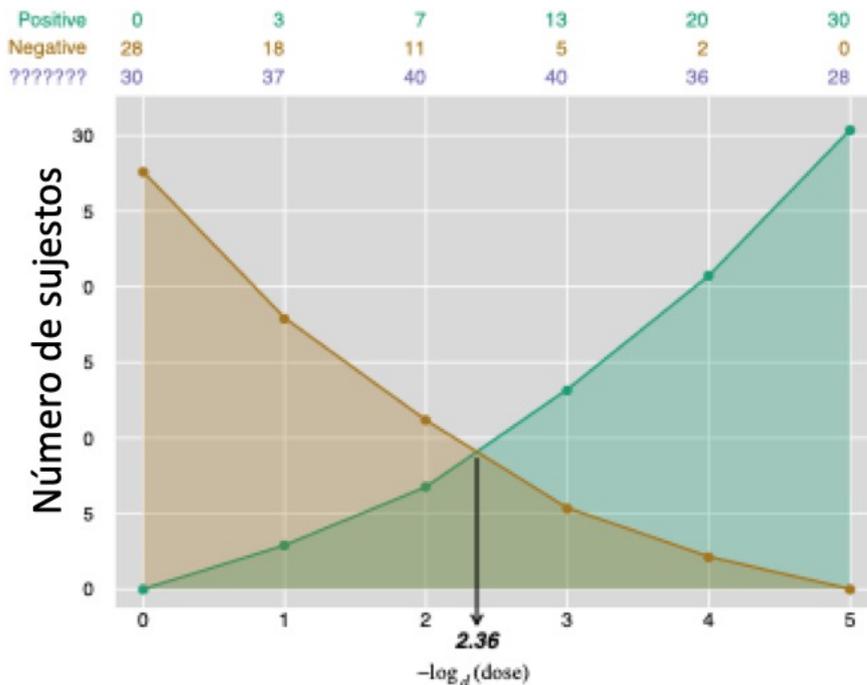


Figura 5. Método de Reed y Muench 1938. (Figura Modificada de (Siev 2018)).

Luego, con dichos porcentajes se calcula la distancia proporcional (DP) entre las diluciones por encima y por debajo del punto final del 50%. En el caso ilustrativo, el punto final son los individuos que

manifiestan la infección a través de la excreción del virus. La siguiente ecuación describe el cálculo de la distancia proporcional:

$$(1) \quad DP = \frac{\%_A - 50}{\%_A - \%_B},$$

donde

$\%_A$ = el porcentaje acumulado de individuos afectados en el grupo inmediato por encima del 50%;

$\%_B$ = el porcentaje acumulado de individuos afectados en el grupo inmediato por debajo del 50%.

A continuación, se calcula la dosis 50% aplicando la siguiente ecuación:

$$(2) \quad \text{Log}(\text{punto final}_{50\%}) = \text{Log}(\text{dilución por encima del 50\%}) - DP \times \text{Log}(\text{factor de dilución})$$

El factor de dilución utilizado en esta ecuación es la proporción entre los dos títulos de inóculo por encima y por debajo de la respuesta del 50%.

b. Método de Dragstedt-Beherens

El método Dragstedt-Beherens (Dragstedt and Lang 1928) es similar al método Reed-Muench, pero en lugar de trabajar con las sumas acumulativas directamente, utiliza la frecuencia de las sumas acumuladas de individuos afectados en cada dosis. La $DI_{50\%}$ se estima interpolando la línea que conecta las fracciones hipotéticas de las dosis por encima y debajo del 50%. Al igual que el método anterior, el método Dragstedt-Beherens se basa en estimar el porcentaje acumulado de individuos afectados y no afectados e interpolar el valor 50% (**Figura 6**). Los valores obtenidos son en realidad estimaciones de medias no de medianas y, si bien resultan prácticos, carecen de base estadística y pueden derivar en errores (Siev 2018). Ambos métodos requieren, además, contar con dosis espaciadas equidistantemente en escala logarítmica, con cada grupo conteniendo el mismo número de unidades experimentales (por ejemplo, pocillos de células, cerdos, ratones, humanos, etc.) para una estimación precisa de la relación dosis-respuesta. Por su relevancia histórica, ambos métodos todavía se utilizan ampliamente en la investigación científica moderna como el principal modo de calcular los títulos virales y las dosis de desafío de modelos animales y voluntarios humanos (Lei et al. 2016a; Bui et al. 2013; Ramakrishnan 2016). El cálculo de $DI_{50\%}$ y $DD_{50\%}$ mediante los dos métodos clásicos mencionados fueron realizados utilizando el paquete **skrmdb de R** (Siev 2018).

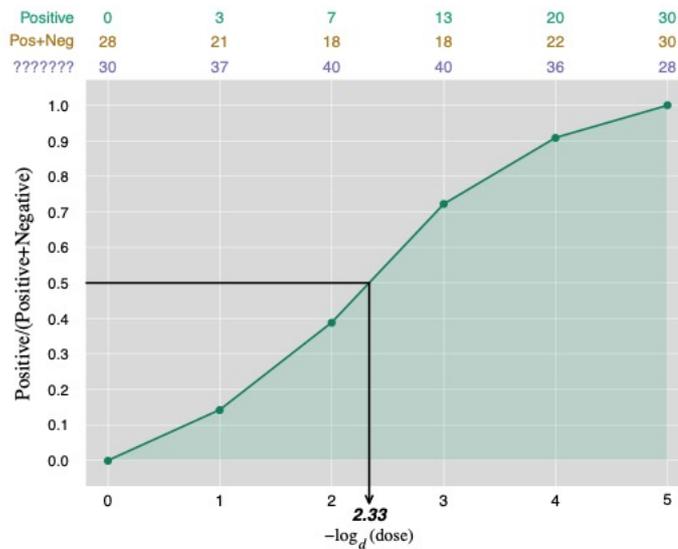


Figura 6. Ilustración del Método de Dragstedt-Behrens (Figura extraída de Siev, 2018).

c. Método no paramétrico de Spearman-Kärber

Una alternativa a los métodos previamente descritos es el método no paramétrico de **Spearman-Kärber** (Atkinson 1961; Spearman 1908; Kä r ber 1931), muy utilizado para el cálculo de dosis virales por las agencias regulatorias (Siev, 2018). Este método da una estimación de la de $DI_{50\%}$ o la $DD_{50\%}$ a partir de su función de probabilidad de masa (PMF) empírica que en el caso de distribuciones discretas describe la probabilidad específica en un punto. Los datos observados dan una estimación empírica de la función de distribución de probabilidad acumulada (CDF) del porcentaje de casos en función de la dosis (**Figura 7**). La función de probabilidad de masa (PMF) empírica se deriva, luego, de la CDF mediante diferenciación. El estimador es $(\sum x \cdot f(x))$, el habitual para la media de una distribución discreta, excepto que aquí $f(x)$ es la PMF empírica obtenida a partir de los datos observados. Este estimador depende de contar con una dosis de respuesta nula y una de respuesta máxima máxima, lo cual puede no ocurrir en un experimento en particular. Si la CDF no cubre todo el dominio de x , una práctica común es extenderlo asumiendo que la siguiente dosis más baja produciría una respuesta cero y la siguiente dosis más alta produciría una respuesta completa. Si bien la distribución de tolerancia no es discreta, está discretizada en virtud del intervalo de censura inherente a los experimentos de este tipo (Siev, 2018) (**Figura 8**).

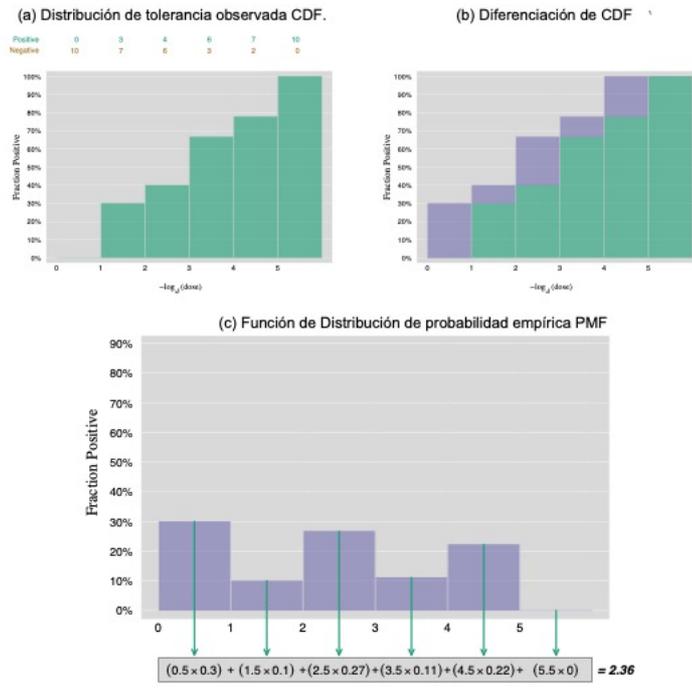


Figura 7. Conceptualización del método de Spearman Karber, extraído de (Siev 2018).

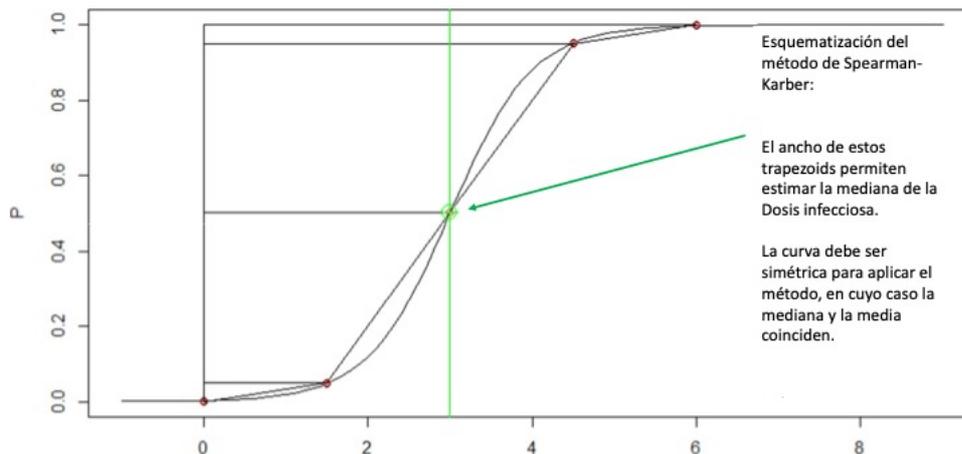


Figura 8. Esquematzación geométrica del método de extrapolación de Spearman Karber.

El método no paramétrico de **Spearman-Karber** utiliza la fórmula descrita en la siguiente ecuación (Ramakrishnan 2016):

$$(3) \quad \text{Log(dilución de punto final 50\%)} = -\left(x_0 - \frac{d}{2} + d \sum_i^n \frac{r_i}{n_i}\right),$$

donde

x_0 = es el logaritmo base 10 de la inversa de la máxima dilución a la cual todos los individuos están afectados;

d = \log_{10} del factor de dilución;

n_i = número de animales incluidos en cada dilución;

r_i = número de animales positivos en cada dilución (del total, n_i).

La sumatoria comienza en la dilución x_0 .

A modo de ejemplo, se ilustra el uso de la ecuación (3) con los datos de la Tabla M&M 1 para estimar el título de una suspensión viral:

Tabla M&M 1. Datos simples para calcular el título en término de dosis infecciosas 50% de una suspensión viral

Grupo	Log (dosis)	cerdos con excreción viral (infectados)	cerdos inoculados	% afectados
1	-1	10	10	100
2	-2	10	10	100
3	-3	10	10	100
4	-4	10	10	100
5	-5	10	10	100
6	-6	6	10	60
7	-7	1	10	10

En este caso, $x_0 = 5$; $d = 1$ y

$$\text{Log}(\text{dilución de punto final } 50\%) = -\left(5 - \frac{1}{2} + 1 \frac{17}{10}\right) = -6.2$$

La dilución correspondiente al 50% = $10^{-6.2}$ y el título viral = $10^{6.2}$ DI₅₀/ml.

Para calcular las DI_{50%} y DD_{50%} mediante el método de Spearman Karber se utilizó el paquete **skrmdb de R** y, como verificación, también se utilizó una herramienta disponible *on line*: *NCBI Statistical Computational Biology Infectious Dose or Dilution (ID50) Server*—https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html/id50/id50.cgi. Esta herramienta está basada en VACMAN, un programa computacional que calcula estadísticas para datos de infectividad *in vitro* e *in vivo* (Spouge 1992).

En las simulaciones realizadas por el programa online, el resultado estimado de DI_{50%} o DD_{50%} al aplicar el método de Spearman-Kärber depende sólo de las replicas utilizadas en cada dilución o dosis. El error estándar obtenido es usualmente muy bajo por lo que se lo considera alejado de la realidad y por eso es un dato que no se informa por defecto. Asimismo, como es un método no paramétrico tampoco se informa el p-valor. Los intervalos de confianza se derivan de las estimaciones del error medio y estándar con aproximación normal, como se indica en (Spiegel 1975).

3.6: Modelos lineales generalizados: regresión logística

John Nelder y Robert Wedderburn formularon en 1972 los modelos lineales generalizados, GLM, como una forma de unificar modelos estadísticos como la regresión lineal, la regresión logística y la regresión de Poisson. Los GLM permiten una generalización flexible del modelo lineal, relacionando las variables predictoras con la variable de respuesta a través de una **función de enlace**. Como consecuencia, permiten que la magnitud de la varianza de cada medición sea una función de su valor predicho. Los GLM son especialmente adecuados para variables respuesta discretas, ya sean nominales u ordinales, que no cumplen los supuestos de linealidad y normalidad. Más generalmente, son adecuados para cualquier variable respuesta con una distribución de probabilidad arbitraria, en la medida que pueda especificarse una función de dicha variable respuesta (la función de enlace) que varíe linealmente con los predictores (Lopez-Gonzalez and Ruiz-Soler 2011). Ejemplos de este tipo de variables son las de clasificación binaria, como infectado/no infectado, variables discretas como la proporción de unidades experimentales afectadas sobre las expuestas y scores acumulados, entre otras.

En la **Tabla M&M 2** se presentan y clasifican los principales modelos lineales generalizados. Como se puede ver, el modelo lineal representa un caso particular del modelo lineal generalizado, en el cual la función de enlace es la función identidad y cuyo componente aleatorio es la distribución Normal. Una diferencia importante entre el modelo lineal y un modelo lineal generalizado es que el primero debe cumplir con el supuesto de homocedasticidad, mientras que para los GLMs no existe ese concepto, ya que salvo para el caso de la normal, la varianza está relacionada matemáticamente con la media. Por otro lado, los parámetros de un GLM se estiman por el método de máxima verosimilitud en lugar del método de mínimos cuadrados estándar de un modelo lineal. La validación de los GLMs, como se explicará más adelante, se realiza empleando tests de cocientes de verosimilitud (*likelihood ratio test*, LRT o Test de Wald) para evaluar su significancia (Lopez-Gonzalez and Ruiz-Soler 2011).

Tabla M&M 2. Principales modelos lineales generalizados

Naturaleza de la variable respuesta	Componente		Función de enlace	Tipo de análisis	Modelo
	Sistemático	Aleatorio			
Numérica continua	Numérico	Normal	Identidad	Regresión Lineal	Lineal ML
	Catagórico	Normal	Identidad	ANOVA	
	Mixto	Normal	Identidad	ANCOVA	
Catagórica binaria	Mixto	Binomial (1)	Logit	Regresión logística	Lineal Generalizado GLM
No agrupada					
Agrupada (frecuencias)	Catagórico	Bernoulli Binomial (n)	Probit	Regresión Probit	
			Logit	Análisis Logit	
Catagórica politómica	Mixto	Multinomial	Logit generalizado	Regresión logística Multinomial	
No agrupada					
Agrupada (frecuencias)	Catagórico	Multinomial	Logit generalizado	Análisis multinomial	
Recuento	Mixto	Poisson	Logarítmica	Regresión de Poisson	
Frecuencias	Catagórico	Poisson	Logarítmica	Análisis loglineal	

Modificado de Lopez-Gonzalez and Ruiz-Soler 2011.

En este trabajo se utilizó el modelo de regresión logística para ajustar los datos del ensayo de dosis-respuesta en cerdos gnotobióticos desafiados con el NoV humano. El modelo de regresión logística, como todos los GLMs, tiene tres componentes básicos (Lee and Nelder 1996):

- i) **Componente aleatorio:** identifica a la variable respuesta y su distribución de probabilidad. En el caso de los datos obtenidos en el ensayo de dosis-respuesta, asume típicamente un carácter binario, infectado/no infectado o enfermo/sano, y, por lo tanto, se modela como una distribución binomial.
- ii) **Componente sistemático:** identifica a las variables explicativas, independientes o predictoras; en datos de ensayos de dosis-respuesta, la dosis opera como variable predictora.
- iii) **una función de enlace (*link function*, en inglés):** es una función que expresa el valor esperado de la variable respuesta como una combinación lineal de las variables predictoras. En el caso de una regresión

logística, por ejemplo, se emplea típicamente la función logit, definida como el logaritmo natural del *odd* cociente de proporciones, como función de enlace:

$$(4) \quad \text{logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \text{dosis} \times \beta,$$

donde $\pi = P(\text{infección/enfermedad} \mid \text{dosis})$. Este modelo ajusta a los datos una curva sigmoidea simétrica a partir de la estimación de los parámetros α y β . La regresión logística es un caso especial de un modelo lineal generalizado (GLM) que se encuentra en un punto medio entre los métodos clásicos descritos en la sección anterior y los modelos GLM dosis-respuesta mecanicistas que explicaremos más adelante. En el contexto de este trabajo, la regresión logística es considerado un modelo no mecanicista, dado que sus parámetros no tienen una interpretación biológica. Para su ajuste se utilizó el comando `glm` del paquete `stats` de R (Chambers, Hastie, and Pregibon 1990) y, complementariamente, la herramienta online de NCBI bajo la opción *curve fit*. Los errores estándares de la estimación se calcularon con el método delta de linealización aproximada según la ecuación 2.3.15, página 14 de (Hammersley and Handscomb 1975).

Alternativamente, empleando también la herramienta online de NCBI (opción *model fit*), se ajustó un modelo GLM con la siguiente función de enlace de un único parámetro (denominada por los autores *Poisson* o *Modelo complementario Log-Log*):

$$(5) \quad \ln[-\ln(1-\pi)] = \alpha + \text{dosis}$$

Mayores detalles de su derivación pueden consultarse en el sitio web del NCBI: https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html/index/presentations.html.

3.7: Métodos de regresión no lineal

Los métodos de regresión no lineal constituyen una forma de análisis de regresión en la cual las observaciones son modeladas a través de una combinación no lineal de parámetros y variables predictoras (Ritz, 2015). En esta tesis se ajustó un modelo de regresión log-logística de dos parámetros y dos modelos mecanicistas, el Exponencial y el Beta-Poisson, a los datos de infección y enfermedad obtenidos en el ensayo con cerdos gnotobióticos. En esta sección se describen estos modelos.

a. Regresión log-logística de 2 parámetros

El modelo de regresión log-logística de dos parámetros (2PL.L.), al igual que la regresión logística, ajusta una curva sigmoidea a los datos del ensayo de dosis-respuesta con variables respuesta infectado/no infectado o enfermo/no enfermo. Este modelo también asume que la respuesta se ajusta a una distribución binomial, pero a diferencia de esta última, contempla las bases biológicas del fenómeno en estudio, dado que modela la relación causa-efecto entre la dosis y la respuesta a través de sus dos parámetros: el coeficiente e , que corresponde al punto de inflexión y representa la dosis efectiva 50%, y el coeficiente b , que representa la pendiente de la curva en dicho punto de inflexión indicando la velocidad de variación de la variable respuesta con la dosis viral administrada. En el caso particular de 2PL.L., los límites inferior y superior se fijan en 0 y 1, ya que la curva ajusta probabilidades de ocurrencia de una respuesta biológica que va de ningún individuo afectado a todos los individuos afectados. La ecuación del modelo de regresión log-logística de dos parámetros es la siguiente:

$$(6) \quad \pi = \frac{1}{1 + (b(\ln D - \ln e))^2}$$

donde $\pi = P(\text{infección/enfermedad} \mid D)$, D es la dosis, y b y e son los parámetros del modelo que representan la pendiente y el punto de inflexión de la curva, respectivamente.

Para ajustar este modelo se utilizó en paquete `drc` de R (Ritz and Streibig 2016). En este paquete, la estimación de los parámetros se realiza por máxima verosimilitud. Por otro lado, y a diferencia de la mayoría de los otros programas para el análisis estadístico de datos de dosis-respuesta, `drc` admite dosis = 0 sin la necesidad de forzar algún valor (como 0.1 ó 0.01) para ajustar la respuesta a ese nivel (Ritz and Streibig 2016; Ritz et al. 2015).

b. Modelos mecanicistas de dosis-respuesta: Exponencial y Beta-Poisson

La estimación de la $DI_{50\%}$ y $DD_{50\%}$ también se realizó ajustando los modelos mecanicistas Exponencial y Beta-Poisson aproximado, que ya han sido utilizados para calcular la dosis de NoV para desafíos en humanos (Teunis et al. 2008; Grytdal et al. 2016; Philip J. Schmidt 2015). En esta sección se describe en detalle la base matemática de estos modelos mecanicistas, basados en la hipótesis del *single-hit*. Para ello se seguirá de cerca el documento elaborado por WHO (2016, Recuadro 1). Derivaciones detalladas pueden consultarse en el libro de Haas et al. (2014, capítulo 8).

Los modelos de *single-hit* (un solo ataque) se derivan a partir del supuesto de que el hospedador ingiere exactamente una sola unidad infecciosa del patógeno. La probabilidad de que este patógeno, en nuestro caso, una unidad infecciosa del NoV, sobreviva a todos los obstáculos y barreras defensivas del hospedador y logre entrar a una célula susceptible para iniciar su replicación tiene un valor distinto de cero y se define igual a p_m . En consecuencia, la probabilidad de que el hospedador no se infecte es $1 - p_m$. Si ingiere una segunda unidad infecciosa y suponemos una acción independiente, la probabilidad de que el huésped no se infecte será $(1 - p_m)^2$. Para n patógenos, la probabilidad de no infectarse es $(1 - p_m)^n$ y la probabilidad de infección se puede expresar como:

$$(7) \quad p_{inf}(n|p_m) = 1 - (1 - p_m)^n$$

A partir de esta ecuación básica se deriva la amplia familia de modelos matemáticos *single-hit* que permiten describir la relación dosis-respuesta. Estos modelos estadísticos permiten estimar no sólo la dosis infecciosa 50% sino el comportamiento a lo largo de todo un rango de datos. Esto resulta especialmente útil para la estimación de riesgo de entrar en contacto con dosis bajas de estos patógenos entéricos (similar a las que contaminan aguas o alimentos). Asimismo, estos modelos son flexibles en cuanto que permiten la estimación aun cuando las dosis probadas no están equidistantes entre sí (en escala logarítmica) o más aún cuando no se cuenta con un número balanceado de individuos entre los grupos.

Los modelos mecanicistas de dosis-respuesta basados en la ecuación (7) utilizados con mayor frecuencia son el **Exponencial** y los **Beta-Poisson**, que se basan en diferentes postulados sobre la distribución de los patógenos en el inóculo. Estos modelos son ampliamente reconocidos como modelos adecuados para describir relaciones dosis-respuesta para diversas infecciones virales en humanos y otros modelos animales. Suponen, como ya hemos visto, que una partícula de virus es suficiente para causar una infección y que los virus están desagregados y homogéneamente distribuidos en la suspensión, de modo que el número exacto de virus administrados en una dosis se distribuye siguiendo una distribución Poisson. El cumplimiento de estos supuestos permite la descripción mecanicista de los efectos de la dosis y la susceptibilidad del hospedador sobre la dosis-respuesta. Cuando no se cumplen (por ejemplo, cuando las partículas virales se agregan), estos modelos aún pueden ser útiles para ajustar o describir una relación dosis-respuesta observada, pero esto excluye la interpretación mecanicista de sus parámetros.

En la Ecuación (8) se presenta el modelo mecanicista de dosis-respuesta Exponencial, que se obtiene asumiendo que el parámetro p_m es una constante r :

$$(8) \quad p_{inf}(D|r) = 1 - e^{-D \times r},$$

donde D es la dosis ingerida media. El modelo Exponencial ha sido utilizado para el estudio de relaciones dosis-respuesta de la infección de NoV en voluntarios humanos (ver revisión de Schmidt, 2015) y fue uno de los modelos utilizados para analizar los datos de la presente tesis de maestría.

Si, en cambio, se considera que la probabilidad de supervivencia del patógeno varía de hospedador a hospedador siguiendo una distribución Beta, con parámetros α y β , en ese caso la probabilidad una infección puede describirse según:

$$(9) \quad p_{inf}(D|\alpha, \beta) = 1 - {}_1F_1(\alpha; \alpha + \beta; -D),$$

La ecuación (9) representa el modelo Exacto de Beta Poisson, donde ${}_1F_1$ representa la función hipergeométrica confluyente de Kummer, que corresponde a la función característica de la distribución Beta. Se cumple que cuando $\alpha \ll \beta$ y $\beta \gg 1$, ${}_1F_1$ es aproximadamente igual a la fórmula Beta-Poisson exacta y, en consecuencia:

$$(10) \quad p_{inf}(D|\alpha, \beta) \approx 1 - \left(1 + \frac{D}{\beta}\right)^{-\alpha}$$

El modelo exacto de dosis-respuesta de Beta-Poisson es una expansión del modelo exponencial que permite un mayor nivel de realismo al modelar la relación dosis-respuesta. En su libro, Haas et al. (2014) presentan una forma aproximada del modelo Beta-Poisson que se expresa como:

$$(11) \quad p_{inf}(D|\alpha, N_{50}) = 1 - \left[1 + \frac{D \times \left(2^{\frac{1}{\alpha}} - 1\right)}{N_{50}}\right]^{-\alpha}$$

Este modelo expresa el parámetro β como una función de α y una nueva variable N_{50} , que representa una estimación de la dosis de infección/enfermedad 50% con unidades de recuento (es decir, dosis en la que se infectan o enferman la mitad de las unidades experimentales). En particular,

$$(12) \quad \beta = \frac{N_{50}}{2^{\alpha} - 1}$$

es válido si se cumple con el criterio $\alpha \ll \beta$ y $\beta \gg 1$. Esta forma aproximada del modelo Beta-Poisson conserva la linealidad a dosis bajas, una característica fundamental de estos modelos de dosis-respuesta para

enfermedades infecciosas (Haas, Rose, and Gerba 2014; Weir et al. 2017). Este fue el otro modelo mecanicista que se ajustó a los datos de dosis-respuesta en la presente tesis.

Para cerrar esta sección, se repasarán los supuestos a los que están sujetos todos los modelos mecanicistas descriptos:

- a. El número de unidades infecciosas (acúmulos virales) en la suspensión viral se distribuye aleatoriamente siguiendo una distribución de Poisson. Esta suposición ayuda a explicar el hecho de que existe la posibilidad de recibir uno o varios acúmulos. Es decir, que en un mismo grupo de individuos que reciben la misma dosis promedio, algunos individuos resultan altamente infectados y otros no.
- b. El tamaño de los agregados varía de acuerdo con una distribución de serie logarítmica, pero no sería un problema a la hora de ajustar este tipo de modelos.
- c. El procedimiento de dilución no modifica la condición de agregación de la muestra.
- d. Una unidad infecciosa es capaz de producir una infección si llega a un sitio apropiado. Una vez que los virus entran en el hospedador, se desagregan e infectan de forma independiente. En otras palabras, los virus tienen una probabilidad independiente e idéntica de sobrevivir para alcanzar e infectar en un sitio apropiado en el caso del modelo Exponencial.
- e. Probabilidades de supervivencia del patógeno no constantes están dadas por una distribución Beta en el caso del modelo Beta-Poisson.

El ajuste de los modelos mecanicistas Exponencial y Beta-Poisson se realizó mediante el *script* publicado por Weir y col. en 2017 que permite comparar ambos métodos.

c. Modelo de regresión log-logística de 3 parámetros aplicado sobre variables continuas

En los métodos que hemos presentado hasta ahora la variable respuesta era la proporción de individuos dentro del grupo que había recibido la misma dosis que manifestaba la infección o enfermedad. Pero la respuesta también puede medirse a través de una variable continua. Por lo general, la forma de una curva dosis-respuesta de una variable respuesta continua es también una curva sigmoidea, donde la respuesta esperada (y) en los individuos está en torno a una media para un valor dado de dosis (x) y sujeta a variación muestral (Ritz et al. 2015). Esto requiere la especificación de un modelo estadístico que describa la variación aleatoria. En general, se postulan modelos que caracterizan el valor de esperado de y en términos de una función f que depende de la dosis x y un conjunto de parámetros, β :

$$(13) \quad E(y) = f(x, \beta)$$

Los valores de los parámetros del modelo β se estiman a partir de los datos para obtener la función de mejor ajuste. Existen diferentes modelos en función de la distribución de probabilidad que se asuma para y (Ritz, 2015; Streibig, 1988).

En este trabajo, se utilizaron el área bajo la curva de excreción viral y el área bajo la curva de severidad de diarrea (Log AUCS y AUCD) como variables respuesta continuas. Ambas variables toman valores positivos con un límite superior definido por la respuesta causada por un inóculo tal de virus que logra la máxima severidad de infección y replicación, y un límite inferior determinado por dosis virales tan bajas que no logran

infectar y/o enfermar a los individuos desafiados y donde se entiende que existe la limitante biológica de que un inóculo con 0 partículas virales (control negativo de solución salina) da una respuesta nula.

El modelo más comúnmente utilizado para las curvas dosis-respuesta sigmoideas de variables continuas es el modelo log-logístico (L.L.), en sus variantes con tres, cuatro o cinco parámetros. En la ecuación 14 se detalla la fórmula del modelo log-logística de cuatro parámetros (4P L.L.).

$$(14) \quad f[x, (b, c, d, e)] = c + \frac{d - c}{[1 + e^{b(\ln(Dosis) - \ln(e))}]}$$

Donde,

b: indica la pendiente en el punto de inflexión de la curva dosis-respuesta y refleja la velocidad de aumento de la respuesta en función de la dosis. Los parámetros c y d: corresponden a los límites inferior y superior, respectivamente, y el parámetro e: representa el punto de inflexión de la curva y, biológicamente, la dosis efectiva 50%.

La función log-logística de 4 parámetros es una función simétrica alrededor del parámetro e, punto de inflexión de la curva, que a su vez representa la mediana de la respuesta observada entre los límites superior e inferior. Este punto se suele denominar dosis efectiva 50%. Asimismo, cuando el parámetro c = 0, la ecuación (13) de 4P L.L. se transforma en el modelo de tres parámetros 3P L.L. (ecuación 15), donde el límite inferior es cero. Esta restricción tiene sentido para describir procesos biológicos como crecimientos de cultivos, cinéticas enzimáticas o, como aquí, para medir la respuesta de una infección viral.

$$(15) \quad f[x, (b, d, e)] = \frac{d}{1 + e^{b(\ln(Dosis) - \ln(e))}}$$

Para el cálculo de la $DI_{50\%}$ y $DD_{50\%}$ de Norovirus humano en el modelo cerdo gnotobiótico se partió de las variables continuas de $\log_{10}AUC$ de excreción viral y de AUC del *score fecal*. Los datos fueron modelados utilizando la distribución log-logística de 3 parámetros (ec. 15), utilizando el paquete drc desarrollado por Ritz y colaboradores. Este paquete ofrece una interface amigable para especificar los supuestos de los modelos a ajustar (Ritz et al. 2015; Ritz and Strebig 2016) y reporta las dosis efectivas a diferentes puntos de la curva (DE10, DE50, DE90, etc.), los intervalos de confianza de los parámetros estimados y la respuesta media predicha y los intervalos de confianza y predicción de nuevas observaciones, que calcula, por defecto, con el método "delta" (Vaart 2013) CRAN, R core, <https://cran.r-project.org/web/packages/drc/drc.pdf>.

3.8: Estimación de parámetros y bondad de ajuste de los modelos empleados

Para calcular la dosis infecciosa 50% ($DI_{50\%}$) y la dosis diarrea 50% (DD_{50}) a partir de los diferentes modelos empleados (las regresiones logística y log-logística y los modelos mecanicistas Exponencial y Beta-Poisson) se deben estimar, en primer lugar, los parámetros de los diferentes modelos. La estimación se realizó por el método de **máxima verosimilitud (MLE)** (WHO 2016; Haas, Rose, and Gerba 2014) mediante los *scripts* de R de los paquetes ya comentados.

En general, cualquiera de estos modelos puede caracterizarse como una función que predice la proporción de individuos de cada grupo que responde a la dosis y que está indexada por un conjunto de parámetros. Siguiendo la notación de Haas et al. (2014), se define en forma genérica

$$(16) \quad \pi_i = p_{inf}(Dosis_i|\theta)$$

como la probabilidad de infección (o desarrollo de la enfermedad) del grupo que recibió la i -ésima dosis. En (10), θ representa el conjunto de parámetros que indexan esa función de probabilidad. Por ejemplo, $\theta = r$ para el modelo Exponencial y $\theta = \{\alpha, N_{50}\}$ para el modelo Beta-Poisson aproximado. Por su parte,

$$(17) \quad \pi_i^0 = \frac{k_i}{n_i}$$

representa la proporción observada de individuos infectados (o enfermos) en el i -ésimo grupo, definida como el cociente entre el número de individuos afectados (k_i) sobre el total del grupo (n_i).

La función de verosimilitud es una función de los parámetros condicional a los valores observados de infección (o enfermedad) en cada grupo y, en este caso, puede escribirse como:

$$(18) \quad L(\theta|k_i, n_i) = \prod_i (\pi_i)^{k_i} (1 - \pi_i)^{n_i - k_i}$$

El método de MLE permite hallar los valores de los parámetros θ que maximicen esta función de verosimilitud. Las estimaciones de la $DI_{50\%}$ y la $DD_{50\%}$ se derivan luego de la función de probabilidad de infección (o enfermedad) estimada con estos parámetros.

Para caracterizar la incertidumbre de los parámetros estimados (r , α y N_{50}) para los modelos mecanicistas Exponencial y Beta-Poisson aproximado se empleó el método de remuestreo o Bootstrap, tal como describe Weir (2017), y un algoritmo cadenas de Markov Montecarlo (MCM) disponible en el paquete de OpenBUGS de R (Versión 3.2.3) (ver P. J. Schmidt, Emelko, and Thompson 2013). Por su parte, para los parámetros b , d , e de los modelos de regresión log-logística de 2 y 3 parámetros se utilizó el método delta (Ritz, 2015).

Para evaluar la bondad de ajuste de los diferentes modelos se utilizó el método del cociente de verosimilitudes. En particular, se obtuvo la *deviance* de cada modelo, definida por

$$(19) \quad Deviance \equiv Y = -2(\ln L_1 - \ln L_2),$$

donde L_1 y L_2 representan las funciones de verosimilitud de los modelos completo o de ajuste perfecto y el modelo real o restringido, respectivamente. La *deviance* es la diferencia del logaritmo natural de las funciones de verosimilitud de estos dos modelos jerarquizados. La distribución de este estadístico aproxima asintóticamente a una distribución *Chi-cuadrado*, con grados de libertad igual al número de grupos de dosis ($k = \sum_i k_i$) menos el número de parámetros (p) del modelo. Luego, el ajuste se considera no satisfactorio si $Y > \chi^2(k - p, \alpha)$. La prueba sólo es válida para modelos que están jerarquizados, es decir, que el modelo más general se puede convertir en el menos general mediante una manipulación de los parámetros y que se aplican al mismo conjunto de datos (Bates et al. 2015).

Finalmente, para comparar la performance relativa de los diferentes modelos evaluados se computó el criterio de información de Akaike (AIC). El AIC es igual a la *deviance* de cada modelo, penalizada por la abundancia de parámetros para equilibrar la bondad del ajuste con la parsimonia del modelo, y representa una medida no ambigua del grado de soporte empírico que recibe un modelo (Akaike 1981, 1974):

$$(20) \quad AIC = 2p + Y$$

Si la bondad de ajuste de los modelos a evaluar resulta significativa, se elige el modelo con el menor AIC.

4. Resultados

4.1. Exploración de los datos y naturaleza de las variables bajo estudio

Es esta sección se presenta la estadística descriptiva de los datos obtenidos del ensayo pre-clínico de dosis-respuesta en cerdos gnotobióticos que recibieron dosis crecientes de NoV humano. A partir de estos datos se construyeron las variables que se utilizaron en este estudio. En la **Tabla R1** se presentan los datos completos del ensayo. Cada registro de la tabla corresponde un cerdo incluido en uno de los ocho grupos experimentales en relación a la infección y la enfermedad causada por NoV humano.

Tabla R 1. Datos del ensayo de la infección y la enfermedad causada por NoV humano en cerdos gnotobióticos experimentalmente desafiados con dosis crecientes de virus

TRAT	DOSIS	Log DOSIS	Gn pig	ONSETS	SDUR	AUCS	Log AUCS	ONSETD	DDUR	AUCD
A	2000000	6.30	1	3	4	9149	3.96	1	6	9.50
A	2000000	6.30	2	2	2	16538	4.22	1	5	10.00
A	2000000	6.30	3	1	3	6832	3.83	3	5	8.25
A	2000000	6.30	4	2	1	5503	3.74	1	4	8.50
B	400000	5.60	1	3	1	3485	3.54	4	2	8.50
B	400000	5.60	2	6	1	4643	3.67	3	3	9.75
B	400000	5.60	3	1	2	10796	4.03	7	1	6.25
B	400000	5.60	4	6	1	2002	3.30	8	0	7.25
C	200000	5.30	1	1	7	158122	5.20	3	2	5.25
C	200000	5.30	2	1	6	171773	5.23	4	3	7.50
C	200000	5.30	3	1	7	137353	5.14	2	5	12.00
C	200000	5.30	4	2	5	39849	4.60	2	6	12.50
D	80000	4.90	1	3	2	4318	3.64	2	5	10.00
D	80000	4.90	2	1	3	20251	4.31	2	2	7.50
D	80000	4.90	3	1	4	17544	4.24	1	7	11.75
D	80000	4.90	4	2	4	32610	4.51	2	6	9.50
D	80000	4.90	5	1	2	3288	3.52	6	1	3.50
D	80000	4.90	6	1	2	5014	3.70	6	2	2.50
E	20000	4.30	1	1	1	28	1.44	8	0	11
E	20000	4.30	2	8	0	0	0.00	8	0	10
E	20000	4.30	3	1	2	225	2.35	8	0	6
F	3200	3.51	1	3	1	4241	3.63	8	0	3.50
F	3200	3.51	2	8	0	0	0.00	6	0	3.00
F	3200	3.51	3	1	1	3761	3.58	1	1	3.00
G	800	2.90	1	8	0	0	0.00	8	0	3.25

G	800	2.90	2	8	0	0	0.00	8	0	3.50
G	800	2.90	3	3	2	11889	4.08	8	0	2.75
G	800	2.90	4	8	0	0	0	8	0	1.50
H	1	0	1	NA	0	0	0	NA	0	0
H	1	0	2	NA	0	0	0	NA	0	0
H	1	0	3	NA	0	0	0	NA	0	0
H	1	0	4	NA	0	0	0	NA	0	0

Nota: los animales que no desarrollaron diarrea y excreción viral reciben en valor arbitrario de 8 para los cálculos del inicio. El grupo H que no recibe virus de desafío, dosis cero, recibe el valor 1 copia de genoma para que el logaritmo base 10 sea cero.

a. Decodificación de las variables

TRAT: variable categórica correspondiente a los ocho tratamientos, simbolizados de A a H. El tratamiento H, corresponde al grupo control negativo que recibió solución salina en lugar de virus y por lo tanto no excreta virus ni desarrolla diarrea durante el ensayo.

DOSIS: indica la dosis de NoV Cin-2 recibida por cada grupo experimental de cerdos gnotobióticos expresada en número de copias de genoma viral. Los animales del grupo H que no reciben virus, se colocó el valor 1 copia de genoma viral para que el log₁₀ de la dosis sea cero.

LogDOSIS: logaritmo base 10 de la dosis.

Gn pig: identificación de cada cerdo enrolado dentro de cada tratamiento o dosis.

ONSETS: período de incubación de la infección medido como el número de días postinoculación en el que se comienza a detectar la excreción viral. Los animales que no desarrollan excreción viral durante los 7 días que dura el ensayo, reciben un valor arbitrario de 8 para los cálculos.

SDUR: período de duración de la excreción viral en días postinicio de la excreción viral, ONSETS.

AUCS: Área bajo la curva de excreción viral de cada animal desde el día 1 a 7 postinoculación. En el caso de los animales que presentaron 0 copias de genoma, se colocó un AUCS de 1 para que el log₁₀ resulte 0 para los cálculos.

LogAUCS: log₁₀AUCS.

ONSETD: período de incubación de la diarrea. Corresponde al día postinoculación en el que se comienza a detectar la presencia de diarrea. Los animales que no desarrollan diarrea durante los 7 días que dura el ensayo, reciben un valor arbitrario de 8 para los cálculos.

DDUR: período de duración de la diarrea en días postinicio de la diarrea, ONSETD.

AUCD: Área bajo la curva del score fecal de cada animal desde el día 1 a 7 post inoculación.

b. Estadística descriptiva de las variables asociadas a la infección

En la **Tabla R2** se presentan estadísticos descriptivos del período de incubación de la excreción viral (ONSETS). En la **Figura 9** se ilustra a través de gráficos de caja y bigote (Box & Whisker) para cada dosis. Se observa, en general, que tanto la media como la mediana y el desvío estándar de los días de incubación aumenta cuanto menor es la dosis. El tratamiento B mostró una excepción a esta tendencia.

Tabla R 2. Estadística descriptiva. Período de incubación de la infección: **ONSETS**.

TRAT	DOSIS	n	Media	D.E.	CV	Mín	Máx	Mediana
A	2000000	4	2	0.82	40.82	1	3	2
B	400000	4	4	2.45	61.24	1	6	4.5
C	200000	4	1.25	0.5	40	1	2	1
D	80000	6	1.50	0.84	55.78	1	3	1
E	20000	3	3.33	4.04	121.24	1	8	1
F	3200	3	4	3.61	90.14	1	8	3
G	800	4	6.75	2.5	37.04	3	8	8

Referencias: n=número de cerdos en el grupo de dosis;

D.E. = desvío estándar de la muestra; CV =coeficiente de variación

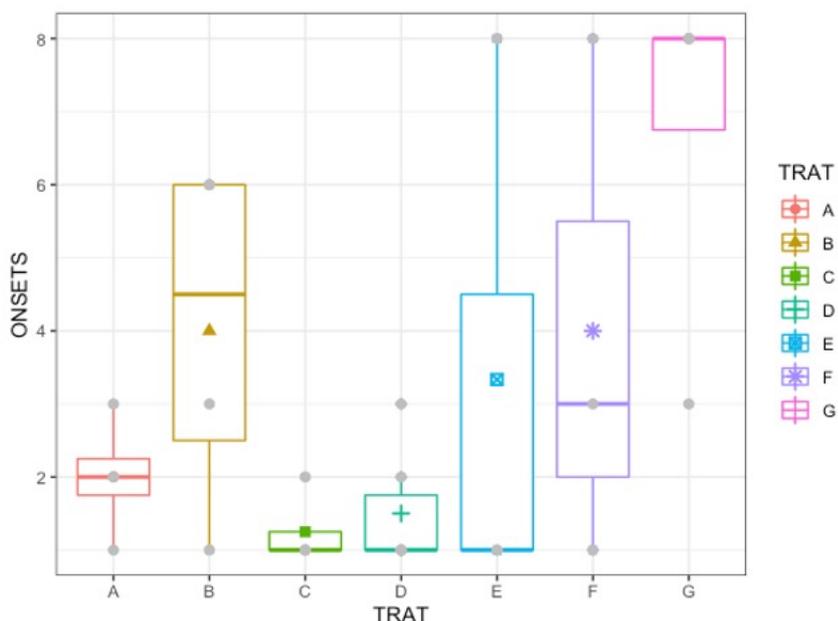


Figura 9. Box & Whisker del momento de inicio de la excreción viral (ONSETS) en cada tratamiento.

En la **Tabla R3** se presentan estadísticos descriptivos del período de duración de la excreción viral. La **Figura 10** ilustra la distribución de la duración en cada dosis. Se observa una reducción del promedio de duración de la excreción con la reducción de la dosis de desafío. Se destaca, en particular, la reducida duración del grupo B en relación a la elevada duración de grupo C. A partir de la dosis E se registran animales sin excreción viral.

Tabla R 3. Estadística descriptiva. Duración de la excreción viral: SDUR.

TRAT	DOSIS	n	Media	D.E.	CV	Mín	Máx	Mediana
A	2000000	4	2.39	1.29	51.64	1	4	2.5
B	400000	4	1.25	0.5	40	1	2	1
C	200000	4	6.25	0.96	15.32	5	7	6.5
D	80000	6	2.83	0.98	34.7	2	4	2.5
E	20000	3	1	1	100	0	2	1
F	3200	3	1.33	1.53	114.56	0	3	1
G	800	4	0.5	1	200	0	2	0

Referencias: n=número de cerdos en el grupo de dosis;

D.E. = desvío estándar de la muestra;

CV =coeficiente de variación

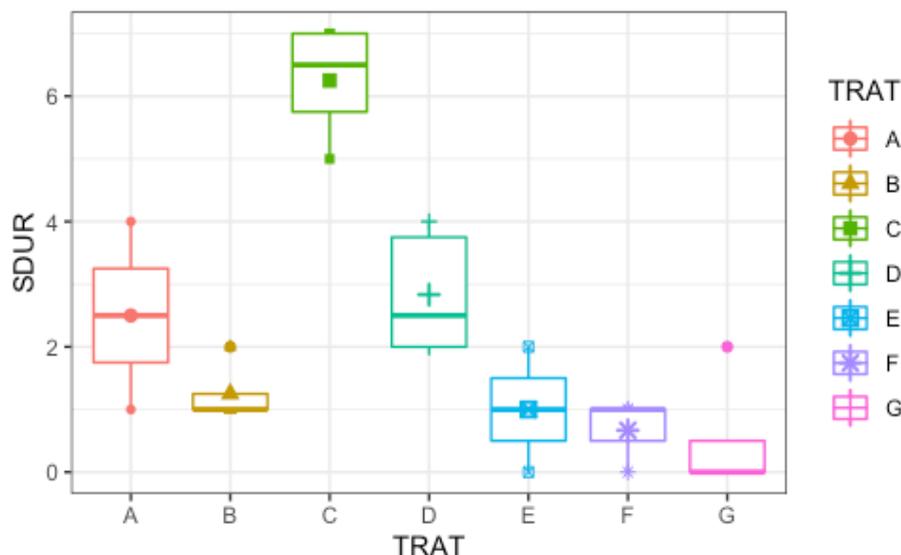


Figura 10. Box & Whisker de la duración de la excreción viral (SDUR) en cada tratamiento.

c. Magnitud de la excreción viral

Tal como fuera descripto en Materiales y métodos, para realizar el seguimiento de la evolución de la infección se tomó una muestra de hisopado rectal de cada cerdo durante los siete días que duró el ensayo, a partir de la cual se determinó el número de copias de genoma viral excretado por qRT-PCR. Los datos individuales de cada cerdo dentro de cada dosis se detallan en la **Tabla R4**. A partir de estos datos se calculó el área bajo la curva de excreción viral y su logaritmo (variables AUCS y logAUCS en la Tabla R1) para cada animal. Como se puede observar, la excreción viral de cada individuo fue variable: algunos cerdos presentaron excreciones sostenidas mientras que otros presentaron excreciones intermitentes o puntuales. En el grupo C se alcanza el objetivo de que todos los animales se infecten y excreten virus por un período de 6 a 7 días. Luego, claramente se observa un descenso de la magnitud de la excreción con la reducción de la dosis viral administrada.

Tabla R 4. Excreción viral determinada por qRT PCR (número de copias de genoma viral).

TRAT A: Dosis		2 x 10 ⁶ de NoV humano Cin-2 GII.4/2003						
Cerdo ID	PID1	PID2	PID3	PID4	PID5	PID6	PID7	AUCS
1. Gp5-5-17	0	0	2701	1842	2750	1856	0	9149
2. Gp5-6-17	0	1064	0	0	0	15474	0	16538
3. Gp5-7-17	2297	0	4018	0	0	1666	0	6832
4. Gp5-8-17	0	5503	0	0	0	0	0	5503
TRAT B: Dosis		4 x 10 ⁵						
Cerdo ID	PID1	PID2	PID3	PID4	PID5	PID6	PID7	AUCS

1.Gp5-9-17	0	0	3485	0	0	0	0	3485
2.Gp5-10-17	0	0	0	0	0	4643	0	4643
3.Gp5-11-17	5274	8159	0	0	0	0	0	10796
4.Gp5-12-17	0	0	0	0	0	2002	0	2002

TRAT C: Dosis 2×10^5

Cerdo ID	PID1	PID2	PID3	PID4	PID5	PID6	PID7	AUCS
1.Gp6-9-17	3618	7248	32670	39862	49992	13571	25940	158122
2. Gp6-10-17	10127	2691	45654	21033	38310	59022	0	171773
3. Gp6-11-17	2895	1379	45435	23934	48063	8961	16264	137353
4. Gp7-8-17	0	2549	0	20329	5226	9548	4395	39849

TRAT D: Dosis 8×10^4

Cerdo ID	PID1	PID2	PID3	PID4	PID5	PID6	PID7	AUCS
1.Gp5-13-17	0	0	3541	0	0	0	1554	4318
2.Gp5-14-17	5002	0	17008	0	0	0	1484	20251
3.Gp5-15-17	1370	1702	13923	0	0	0	2468	17544
4.Gp5-16-17	0	1064	24047	0	4489	3010	0	32610
5.Gp3-11-18	4116	0	1230	0	0	0	0	3288
6.Gp3-12-18	7226	1400	0	0	0	0	0	5014

TRAT E: Dosis 2×10^4

Cerdo ID	PID1	PID2	PID3	PID4	PID5	PID6	PID7	AUCS
1.Gp5-4-18	55	0	0	0	0	0	0	28
2.Gp5-5-18	0	0	0	0	0	0	0	0
3. Gp5-8-18	38	206	0	0	0	0	0	225

TRAT F: Dosis 3.2×10^3

Cerdo ID	PID1	PID2	PID3	PID4	PID5	PID6	PID7	AUCS
1.Gp7-5-17	0	0	4241	0	0	0	0	4241
2.Gp7-6-17	0	0	0	0	0	0	0	0

3.Gp7-7-17	7522	0	24278	0	91886	0	0	3761
TRAT G: Dosis		8×10^2						
Cerdo ID	PID1	PID2	PID3	PID4	PID5	PID6	PID7	AUCS
1.Gp1-5-18	0	0	0	0	0	0	0	0
2.Gp1-6-18	0	0	0	0	0	0	0	0
3.Gp1-10-18	0	0	9012	0	2877	0	0	11889
4.Gp1-11-18	0	0	0	0	0	0	0	0

En la **Tabla R5** se presenta la estadística descriptiva de la variable logAUCS por tratamiento. En la **Figura 11**, panel de la derecha, se ilustra el área bajo la curva promedio de cada grupo. Como se puede observar el grupo C es el que presenta la mayor excreción promedio y mediana. Luego comienza a descender registrándose animales no infectados en los últimos tres grupos.

Tabla R 5. Estadística descriptiva. Logaritmo del área bajo la curva de excreción viral: LogAUCS.

TRAT	DOSIS	n	Media	D.E.	CV	Mín	Máx	Mediana
A	2000000	4	3.94	0.21	5.26	3.74	4.22	3.9
B	400000	4	3.64	0.31	8.39	3.3	4.03	3.6
C	200000	4	5.04	0.3	5.91	4.6	5.23	5.17
D	80000	6	4.22	0.48	11.48	3.64	4.9	4.28
E	20000	3	1.36	1.22	89.46	0	2.35	1.74
F	3200	3	2.9	2.62	90.14	0	5.08	3.63
G	800	4	1.02	2.04	200	0	4.08	0

Referencias: n=número de cerdos en el grupo de dosis;

D.E. = desvío estándar de la muestra;

CV =coeficiente de variación

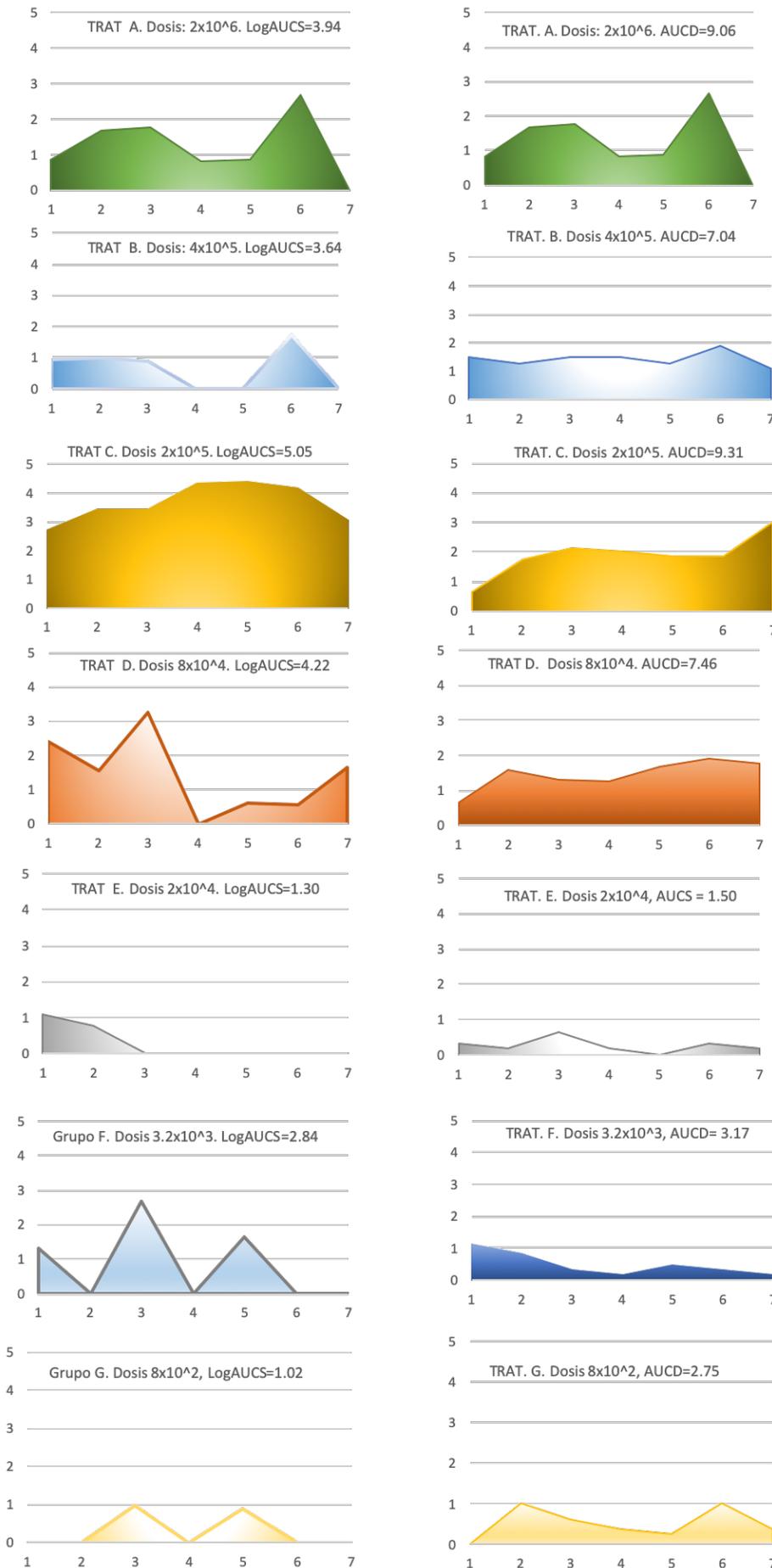


Figura 11. Log10 del área bajo la curva de excreción viral (logAUC) y promedio del área bajo la curva de los score fecal (AUCD) en cada grupo de cerdos gnotobióticos inoculado con dosis crecientes de NoV humano.

d. Estadística descriptiva de las variables asociadas al desarrollo de la enfermedad

En las **Tablas R6 y R7** se detallan los estadísticos descriptivos de las variables inicio y duración de la diarrea, ONSETD y DDUR, respectivamente. En las Figuras 12 y 13 se ilustra la distribución de ambas variables en cada tratamiento. Se observa que el período de incubación promedio de los síntomas clínicos es muy variable entre los grupos. Los animales con números medios, medianos, máximos y mínimos de 8 representan individuos que no desarrollaron la enfermedad.

Tabla R 6. Estadística descriptiva. Período de incubación de la enfermedad ONSETD.

TRAT	DOSIS	n	Media	D.E.	CV	Mín	Máx	Mediana
A	2E+06	4	1.5	1	66.67	1	3	1
B	400000	4	5.5	2.38	43.28	3	8	5.5
C	200000	4	2.75	0.96	34.82	2	4	2.5
D	80000	6	3.17	2.23	70.38	1	6	2
E	20000	3	2.33	0.58	24.74	2	3	2
F	3200	3	3	2.65	88.19	1	6	2
G	800	4	8	0	0	8	8	8

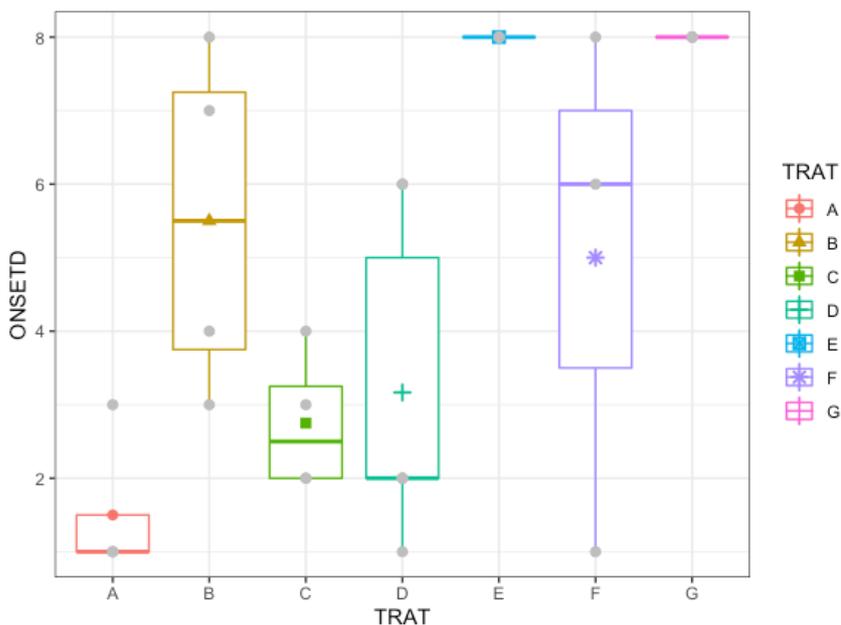


Figura 12. Box & Whisker del momento de inicio de la excreción viral (SDUR) en cada tratamiento.

Por su parte, la duración promedio de la enfermedad se reduce con el descenso de la dosis llegando a solo un animal afectado en el grupo F y ningún animal afectado en el grupo G, cumpliendo con abarcar el rango completo de dosis respuesta deseado.

Tabla R 7. Estadística descriptiva. Duración de la enfermedad (diarrea) DDUR.

TRAT	DOSIS	n	Media	D.E.	Var(n)	CV	Mín	Máy	Mediana
A	2E+06	4	5	0.82	0.5	16.33	4	6	5
B	400000	4	1.5	1.29	1.25	86.07	0	3	1.5
C	200000	4	4	1.83	2.5	45.64	2	6	4
D	80000	6	3.83	2.48	5.14	64.78	1	7	3.5
E	20000	3	3.33	2.08	2.89	62.45	1	5	4
F	3200	3	1	0	0	0	1	1	1
G	800	4	0	0	0	sd	0	0	0

Referencias: n=número de cerdos en el grupo de dosis;

D.E. = desvío estándar de la muestra;

CV =coeficiente de variación

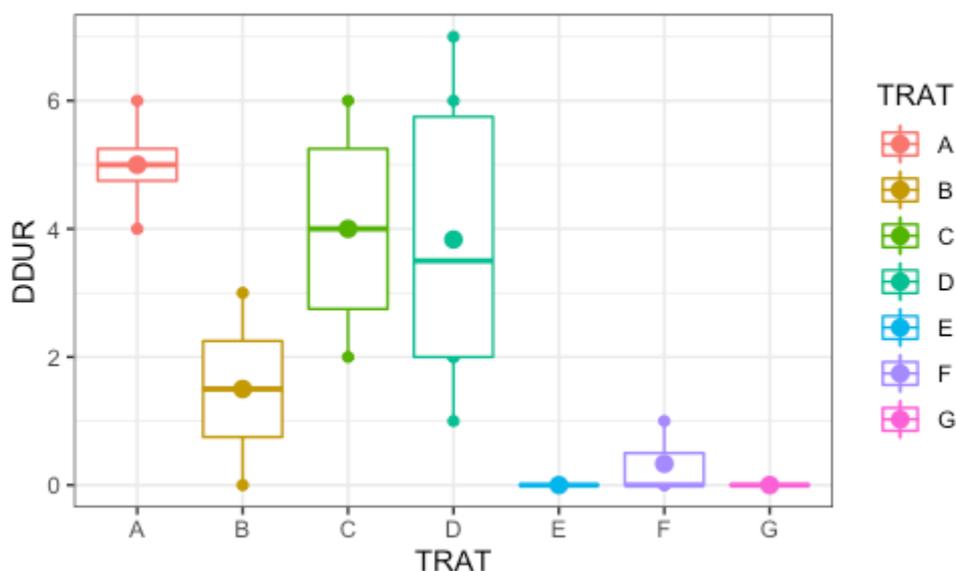


Figura 13. Box & Whisker de la duración de la enfermedad (DDUR) en cada tratamiento.

e. Magnitud de la severidad del cuadro clínico

La severidad de la enfermedad, diarrea, se calculó de la misma manera que lo descripto para AUCS, pero en este caso se computaron los scores fecales de cada cerdo, diariamente, durante los siete días del ensayo (Tabla R8). Se calculó el área bajo la curva de dicho score para cada animal (Tabla R8 y Tabla R1 resumen) y luego el promedio por grupo (Figura 11, panel izquierdo, y Tabla R8) y sus estadísticos descriptivos (Tabla R9).

Tabla R 8. Datos de score fecal de cada cerdo individual, 1 a 7 días.

TRAT A 2.000.000 HuNoV GII.4/2003, PPD33													
Cerdo #	Cerdo ID	PID0	PID1	PID2	PID3	PID4	PID5	PID6	PID7	duración	Score acumulado	Inicio	AUCD
1	Gp5-5-17	1	2.0	0.5	2.0	2.0	2.0	2.0	2.5	6	13	1	9.50
2	Gp5-6-17	1.5	2.0	1.5	2.0	2.5	1.0	2.0	2.0	5	13	1	10.00
3	Gp5-7-17	1	0.0	0.5	2.5	2.0	2.0	2.5	2.0	5	12	3	8.25
4	Gp5-8-17	1.5	2.0	1.0	1.5	2.0	1.0	2.0	2.0	4	12	1	8.50
TRAT B 400.000 HuNoV GII.4/2003, PPD34													
Cerdo #	Cerdo ID	PID0	PID1	PID2	PID3	PID4	PID5	PID6	PID7	duración	Score acumulado	Inicio	AUCD
1	Gp5-9-17	1	1.5	1.0	1.5	2.0	1.5	2.0	0.5	2	10	4	8.5
2	Gp5-10-17	1	1.5	1.5	2.0	2.0	1.5	2.5	1.5	3	13	3	9.75
3	Gp5-11-17	0	1.5	1.0	1.0	1.0	1.0	1.5	2.0	1	9	7	6.25
4	Gp5-12-17	1.5	1.5	1.5	1.5	1.0	1.0	1.5	0.5	0	9	8	7.25
TRAT C 200000 HuNoV GII.4/2003, PPD34													
Cerdo #	Cerdo ID	PID0	PID1	PID2	PID3	PID4	PID5	PID6	PID7	duración	Score acumulado	Inicio	AUCD
1	Gp6-9-17	1.5	0.0	0.0	2.0	1.5	1.0	1.5	3.0	2	9	3	5.25
2	Gp6-10-17	1.5	0.5	1.0	1.5	2.0	2.0	1.0	3.0	3	11	4	7.5
3	Gp6-11-17	0	1.0	3.0	3.0	1.5	2.5	2.0	3.0	5	16	2	12
4	Gp7-8-17	0	1.0	3.0	2.0	3.0	2.0	3.0	3.0	6	17	2	12.5
TRAT D 80000 HuNoV GII.4/2003, PPD34													
Cerdo #	Cerdo ID	PID0	PID1	PID2	PID3	PID4	PID5	PID6	PID7	duración	Score acumulado	Inicio	AUCD
1	Gp5-13-17	1	0.5	2.5	1.5	2.0	2.5	2.0	2.0	5	13	2	10
2	Gp5-14-17	0	1.0	2.0	1.5	0.5	2.0	1.0	1.5	2	10	2	7.5
3	Gp5-15-17	1	2.0	2.5	2.0	2.0	2.0	2.5	2.0	7	15	1	11.75
4	Gp5-16-17	1	0.5	2.0	2.0	2.0	2.0	2.0	2.0	6	13	2	9.5
5	Gp3-11-18	1	0.0	0.5	0.0	1.0	1.0	2.0	1.0	1	6	6	3.5
6	Gp3-12-18	1	0.0	0.0	1.0	0.0	0.5	2.0	2.0	2	6	6	2.5
TRAT E 20000 HuNoV GII.4/2003, PPD34													
Cerdo #	Cerdo ID	PID0	PID1	PID2	PID3	PID4	PID5	PID6	PID7	duración	Score acumulado	Inicio	AUCD
1	Gp5-4-18	1	0.5	2.5	1.0	3.0	2.0	2.0	0.5	4	12	2	11
2	Gp5-5-18	1	0.5	2.5	1.0	3.0	2.0	2.0	2.0	5	13	2	10
3	Gp5-8-18	0	0.0	0.5	3.0	0.5	0.0	1.0	2.0	1	7	3	6
TRAT F 3200 HuNoV GII.4/2003, PPD34													
Cerdo #	Cerdo ID	PID0	PID1	PID2	PID3	PID4	PID5	PID6	PID7	duración	Score acumulado	Inicio	AUCD
1	Gp7-5-17	1	1.5	2.5	0.5	0.0	1.0	1.0	0.0	1	7	2	3.5
2	Gp7-6-17	1	0.0	1.5	0.5	0.5	0.5	2.0	0.0	1	5	6	3
3	Gp7-7-17	1	2.0	1.0	0.0	0.0	0.0	0.0	0.5	1	4	1	3
TRAT G 800 HuNoV GII.4/2003, PPD35													
Cerdo #	Cerdo ID	PID0	PID1	PID2	PID3	PID4	PID5	PID6	PID7	duración	Score acumulado	Inicio	AUCD
	Gp1-5-18	1	0.0	1.0	1.0	0.5	0.0	1.5	0.0	0	4	8	3.25
1	Gp1-6-18	0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0	4	8	3.5
2	Gp1-10-18	0	0.0	1.0	0.0	1.0	0.0	1.5	1.5	0	5	8	2.75
3	Gp1-11-18	0	0.0	1.0	0.5	0.0	0.0	0.0	0.0	0	2	8	1.5

Tabla R 9. Estadística descriptiva. Severidad del cuadro clínico: AUCD.

TRAT	DOSIS	n	Media	D.E.	CV	Mín	Máy	Mediana
A	2E+06	4	9.06	0.96	9	8.25	10	9
B	400000	4	7.04	1.91	19	6.25	9.75	7.87
C	200000	4	9.31	2.8	38	5.75	12.50	9.75
D	80000	6	7.46	3.66	49	2.5	11.75	8.5
E	20000	3	1.50	3.06	0	1.5	1.5	1.5
F	3200	3	3.17	2.1	9	3	3.5	3
G	800	4	2.75	1.35	32	1.5	3.5	3

Referencias: n=número de cerdos en el grupo de dosis;

D.E. = desvío estándar de la muestra;

CV =coeficiente de variación

La severidad del cuadro clínico disminuye acorde a la reducción de la dosis de virus administrada. Los cerdos del grupo G no presentaron diarrea, llegando a una dosis 0% (Tabla R8, R9 y Figura 11 panel de la izquierda).

4.2. Comparación de medias para las variables asociadas a la infección y enfermedad causadas por el NoV humano en el ensayo con cerdos gnotobióticos

El primer análisis que se realizó fue la comparación de las medias entre los grupos experimentales para investigar si existieron diferencias estadísticamente significativas en la intensidad de la infección y la severidad del cuadro clínico en el rango de dosis elegidas para el ensayo de dosis-respuesta. En la **Tabla R10** se resumen los resultados de la comparación de medias entre grupos para las variables asociadas a la infección y diarrea en los grupos de cerdos experimentalmente infectados con dosis crecientes de NoV humano Cin-2. Medias en la misma columna con diferente letra difirieron significativamente.

Se observó que todos los cerdos pertenecientes a los grupos de dosis de A a D se infectaron y excretaron virus en sus heces, correspondiendo a la zona de límite o *plateau* superior de la curva dosis-respuesta. El 67% de los cerdos eliminaron el virus cuando se inocularon en dosis entre 2×10^4 y 3.2×10^3 copias de genoma viral, mientras que sólo el 25% de los cerdos que fueron inoculados con 8×10^2 copias excretaron virus (**Tabla R10**). Una limitación importante del ensayo fue que no se alcanzó una dosis lo suficientemente baja como para no causar infección en el total de cerdos de ese grupo. Otra limitación fue el salto no equidistante entre dosis y el número desbalanceado de unidades experimentales. Estos factores, en teoría, limitan los métodos a emplear para estimar las dosis de infección y enfermedad 50%.

El aumento en la dosis administrada se asoció positivamente con un período de incubación de la infección más corto, que coincidió con un aumento en la duración de la excreción viral, así como con un mayor título de virus excretado en heces (**Tabla R10, Figura 10**). El análisis de regresión ($R^2 = 0.46$, $p\text{-val} = 0.014$) indicó que el aumento de la duración de la excreción viral en función de la dosis fue estadísticamente significativo. Llamativamente, los cerdos del grupo C a los que se les administró una dosis de 2×10^5 copias de genoma de NoV excretaron una cantidad significativamente mayor de virus en sus heces ($p < 0,0001$) y durante un período significativamente más largo que los cerdos en los otros grupos ($p < 0,0001$), aún de aquellos que recibieron dosis más elevadas de virus (**Tabla R10, Figura 14**). Este resultado llama la atención y podría sugerir que a mayor dosis las partículas virales comienzan a agregarse y ese efecto reduciría su capacidad infecciosa. También es importante destacar que el proceso mecánico de realizar las diluciones posee un grado de error difícil de controlar en estos ensayos, por lo que es común observar estos efectos al titular un virus. Por último, es también frecuente al alcanzar dosis límite muy bajas de virus en las que se observan unidades experimentales infectadas y otras que no.

Con relación al cuadro clínico, el 100% de los cerdos de los grupos A, C y D desarrollaron diarrea, mientras que tres de cuatro lo hicieron en el grupo B. Por su parte, sólo un individuo manifestó el síntoma en el grupo F, mientras que los cerdos de los grupos E y G no desarrollaron la enfermedad. En este caso, el rango de dosis utilizadas sí cubrió el rango completo de respuesta del 100% al 0%, lo que indica que se necesitó una cantidad diez veces mayor de virus para producir enfermedad que para producir infección. Se observó una relación lineal estadísticamente significativa entre el período de incubación y la dosis, en escala logarítmica: el período fue más corto con el aumento de la dosis viral (pendiente = -1.61). Por su parte, el período de duración del cuadro clínico mostró un aumento de 1.4 días por cada aumento unitario de la dosis viral en escala logarítmica (**Figura 15**). Los cerdos del grupo C que fueron inoculados con 2×10^5 copias de genoma de NoV Cin-2 y que excretaron la mayor cantidad de virus en sus heces también sufrieron el cuadro clínico más severo con el desarrollo de diarrea

rápidamente, dentro de los 1.5 días posteriores a la inoculación y durante un período de tiempo más prolongado (entre 5 y 7 días) que los cerdos inoculados con las otras dosis, lo que indica la reproducción de un cuadro clínico severo que imita el de los humanos, donde la gastroenteritis asociada a los NoV suele durar de 3 a 5 días. Los cerdos en los grupos de dosis E y F que recibieron 2×10^4 y 3.2×10^3 copias de genoma viral tuvieron un período de incubación de la diarrea más prolongado, junto con scores de consistencia fecal más bajos que demuestran un cuadro clínico menos severo. Por su parte, los cerdos del grupo G que recibieron 800 copias de genoma viral, ya no desarrollaron diarrea (**Tabla R10, Figura 16**).

Tabla R 10. Comparación de medias entre grupos para las variables asociadas a la infección y diarrea causadas por la inoculación de NoV humano en los cerdos gnotobióticos.

Grupo	Dosis de NoV (copias de genoma)	Log10 (dosis)	n	Excreción viral				Diarrea			
				Número (%) de anim afectado	Inicio excreción viral (días)	Duración excreción viral (días)	log10 AUC excreción viral (log copias genoma)	Número (%) de anim afectado	Inicio diarrea (días)	Duración diarrea (días)	AUC diarrea score
A	2000000	6.301	4	4 (100%)	1.83 A	2.389 B	3.94 B	4 (100%)	1.5 B	5.00 A	9.06 A
B	400000	5.602	4	4 (100%)	3.83 AB	1.139 BC	3.64 B	3 (75%)	5.5 AB	1.50 ABC	7.04 A
C	200000	5.301	4	4 (100%)	1.08 A	6.139 A	5.05 A	4 (100%)	2.75 B	4.00 AB	9.31 A
D	80000	4.903	6	6 (100%)	1.50 A	2.833 B	4.22 B	6 (100%)	3.17 B	3.83 AB	7.46 A
E	20000	4.301	3	2 (67%)	3.18 AB	0.805 BC	1.30 C	0 (0%)	8.00 A	3.33 ABC	1.50 B
F	3200	3.505	3	2 (67%)	3.84 AB	1.139BC	2.84 C	1(33%)	5.00 AB	1.00 BC	3.17 AB
G	800	2.903	4	1 (25%)	6.58 B	0.389 C	1.02 C	0 (0%)	8.00 A	0.00 C	2.75 AB
Tipo de Análisis estadístico aplicado				LMM	LMM	LMM		AOV	AOV	LMM	
p-valor asociado al LMM o AOV realizado				0.002	<.0001	<.0001		0.0003	0.002	0.0003	
W* Shapiro Wilks sobre los residuos				0.95	0.95	0.90		0.95	0.97	0.94	
p-valor asociado a W*				0.157	0.2683	0.012		0.955	0.612	0.18	

Media en la misma columna con diferente letra difiere significativamente (Comparaciones múltiples de Tukey-Kramer HSD, p<0.05 con p ajustado para comparación de 7 grupos).

LMM: modelo lineal mixto anidado; AOV: anova de una vía

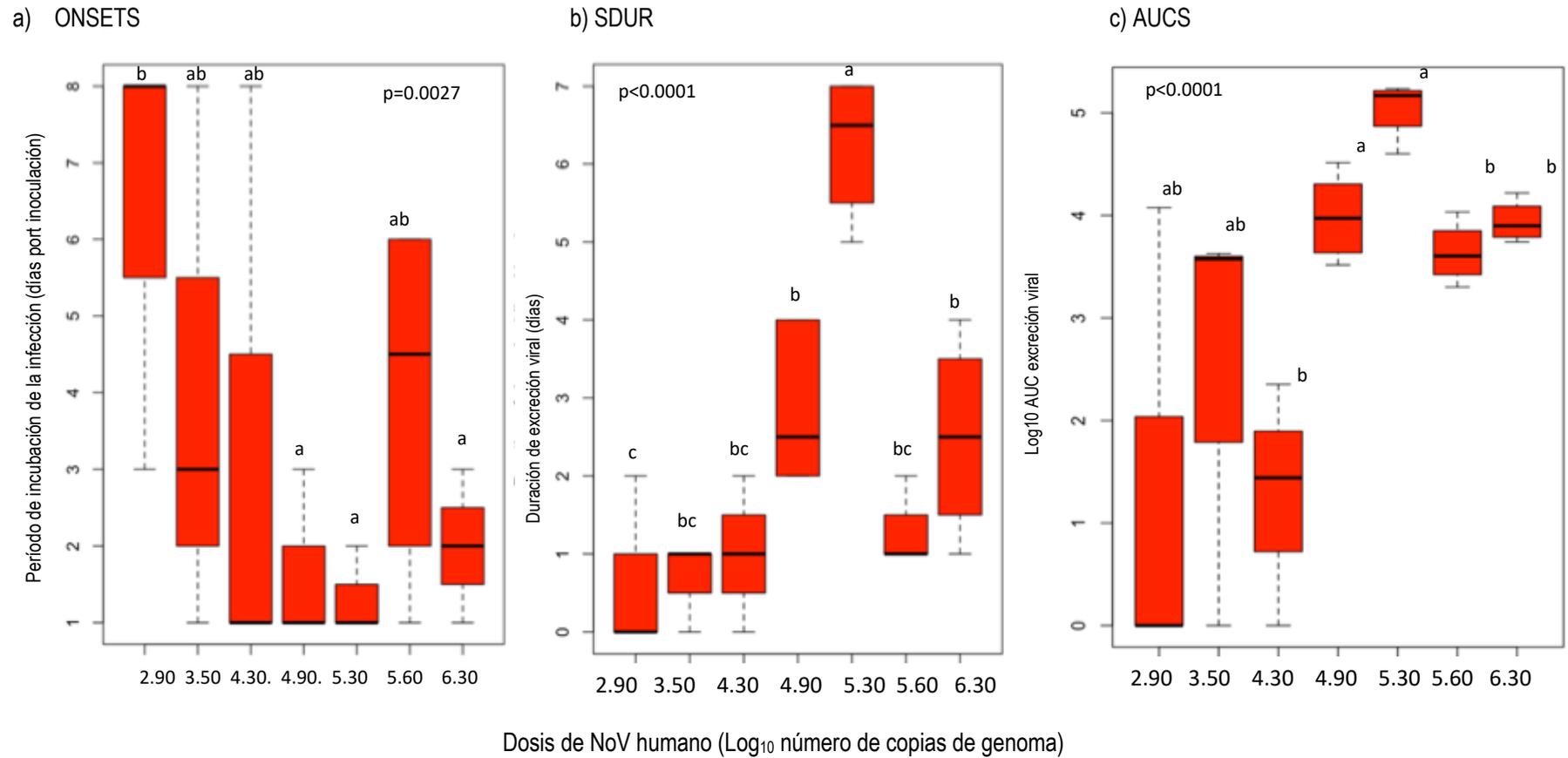
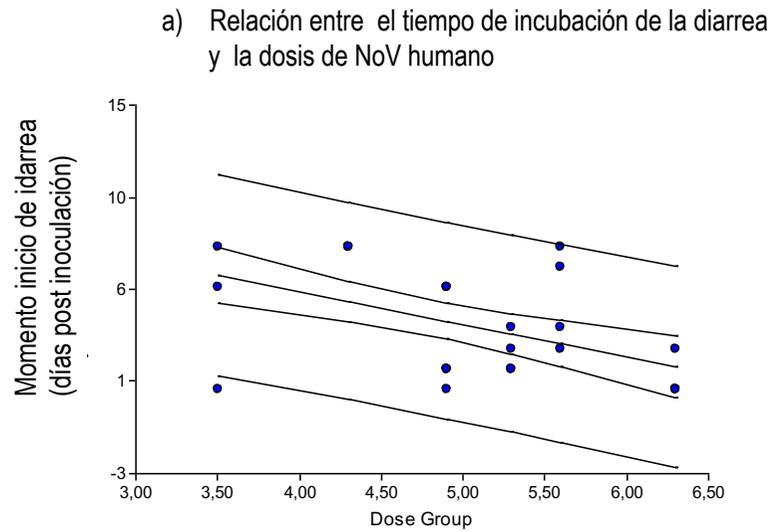
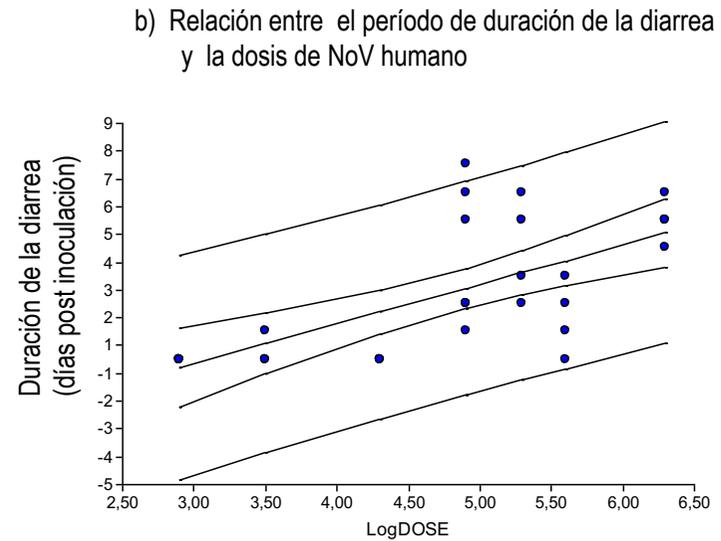


Figura 14. Box & Whisker ANOVA. A. Período de incubación de la infección (momento de inicio de la excreción viral en días post inoculación). B. Duración de la excreción viral. C. Magnitud de la excreción viral en términos de área bajo la curva de cantidad de virus excretado en los 7 días que duró el ensayo. Barras con diferente letra difieren significativamente (ver resumen del análisis estadístico en la Tabla R10. Los scripts y salidas de R e infostat están disponibles a solicitud)



Ecuación
 $ONSETD = 12.27 - 1.61 \log \text{ dosis}$
 $R^2=0.38$, $p\text{-value}=0.005$
 , $AIC= 130.08$ $BIC= 134.07$, $p=0.0005$
 Normalidad Residuos: Test de Shapiro Wilks
 $W^* = 0.98$, $p= 0.9829$ -> no se rechaza

Los resultados obtenidos indican una relación lineal estadísticamente significativa entre la dosis de virus inoculada y la reducción del período de incubación de la diarrea



Ecuación
 $DDUR = - 4.41 + 1.42 \log \text{ dosis}$
 $R^2=0.43$, $p=0.014$, $AIC=117$ $BIC= 121$ $p=0.0002$
 Normalidad Residuos: Test de Shapiro Wilks
 $W^* = 0.95$, $p= 0.428$ -> no se rechaza

Los resultados obtenidos indican una relación lineal estadísticamente significativa entre la dosis de virus inoculada y el número de días de diarrea

Figura 15. Análisis de regresión lineal del inicio y duración de la diarrea en función de la dosis de NoV inoculada en los cerdos gnotobióticos. (detalle de Scripts y salidas disponibles a solicitud).

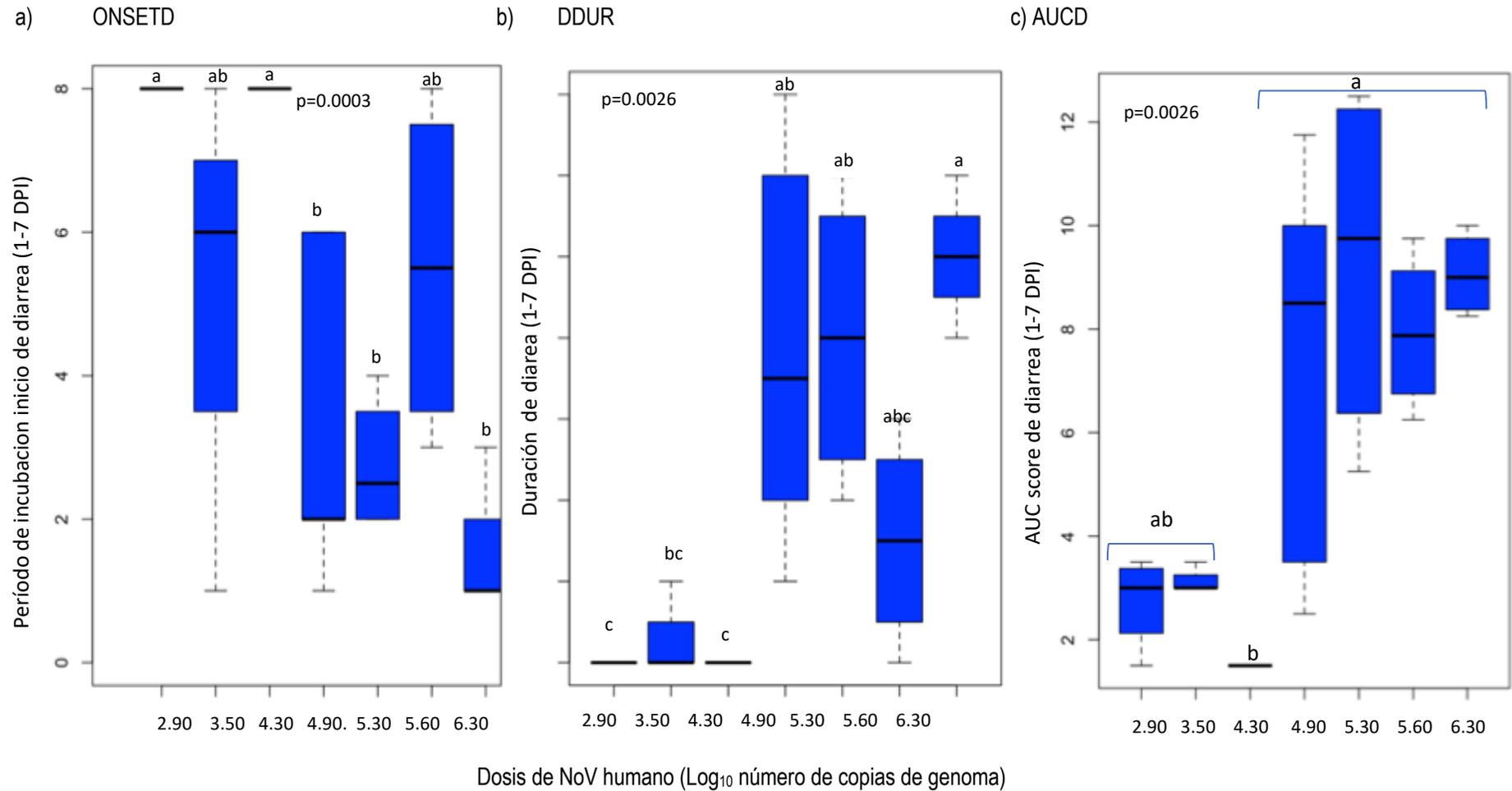


Figura 16. Box & Whisker. A. Período de incubación de la enfermedad (momento de inicio de la diarrea en días post inoculación). B. Duración de la diarrea en días. C. Severidad de la diarrea en términos de área bajo la curva de score fecal durante los 7 días que duró el ensayo. Barras con diferente letra difieren significativamente (ver resumen de análisis estadístico en Tabla R10, scripts y salidas de Infostat y R disponibles a solicitud).

4.3. Análisis del grado de agregación del inóculo

Al estudiar la infección y diarrea de los grupos experimentales incluidos en el ensayo pre-clínico de dosis respuesta, se observó que en los dos grupos que recibieron las dosis más bajas (grupos F y G, con 3200 y 800 copias de genoma, respectivamente) se detectaron animales que no se infectaron ni se enfermaron, mientras que también se detectaron animales que excretaron títulos muy elevados de virus (**Figura 17**; datos destacados en elipses). Por otro lado, también se observó que animales del grupo A que recibieron la dosis más elevada presentaron una menor excreción viral que los animales que pertenecían al grupo B y C. Este hallazgo hizo pensar que el inóculo podría haber sufrido cierto grado de agregación durante su conservación, efecto que ya había sido reportado en la literatura de NoV (Teunis et al. 2008).

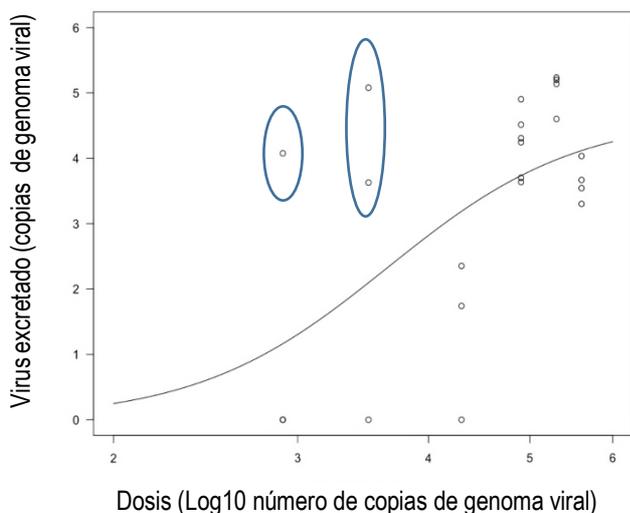


Figura 17. Se indican los tres cerdos, dos del grupo F (inóculo: 3200 copias de genoma viral) y uno del grupo G (inóculo: 800) que excretan títulos de virus muy elevados en sus heces.

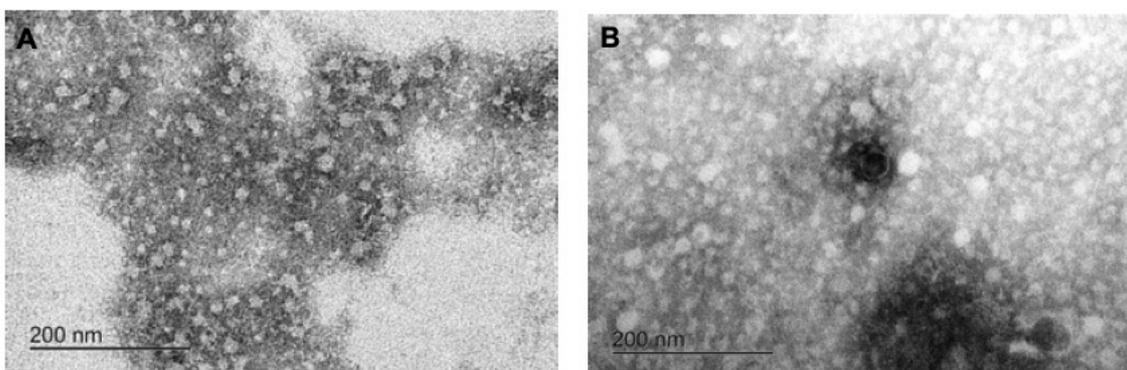


Figura 18. Fotografía de microscopía electrónica del inóculo de NoV Cin-2. a) Dosis: 8×10^4 , el recuento de partículas virales indicó la presencia de agregados con un tamaño de alrededor de 175 partículas; b) Dosis: 2×10^5 , agregados con alrededor de 390 partíc

Es importante destacar que los análisis estadísticos propuestos para estimar la $DI_{50\%}$ utilizando los modelos mecanicistas Exponencial y Beta-Poisson se basan en el supuesto de la teoría *single hit* para la cual se debe partir de una suspensión viral desagregada y distribuida homogéneamente. Por lo tanto, era importante investigar si el inóculo administrado a los cerdos cumplía dicho supuesto o presentaba algún efecto de agregación que interfiriera con la dilución. En dicho caso, sería necesario poder medir la magnitud del efecto. Es decir, si el inóculo sufría de un efecto muy fuerte de agregación, el ensayo no podría utilizarse para estimar la $DI_{50\%}$

aplicando estos métodos o debía considerarse en la ecuación para medir el fenómeno y estimarlo.

Para estudiar este efecto, se realizaron nuevas diluciones del stock viral de la cepa de NoV Cin-2 para preparar las mismas dosis que habían recibido los cerdos, se controlaron por qRT-PCR y paralelamente se prepararon grillas que fueron observadas al microscopio electrónico de transmisión para realizar el recuento de partículas virales (**Figura 18**). En la **Tabla R11** se detalla el tamaño de los acúmulos virales en 10 cuadrantes de cada grilla para cada dosis inóculo.

Tabla R 11. Resultados del tamaño de los acúmulos virales en cada dosis.

Dosis	Log dosis	Cuadrante Grilla	Tamaño del acúmulo (nro viriones)	Vol de stock en 5 ml de inóculo
400000	5.6	1	168	40
400000	5.6	2	192	40
400000	5.6	3	228	40
400000	5.6	4	171	40
400000	5.6	5	236	40
400000	5.6	6	199	40
400000	5.6	7	256	40
400000	5.6	8	228	40
400000	5.6	9	296	40
400000	5.6	10	275	40
200000	5.3	1	57	20
200000	5.3	2	197	20
200000	5.3	3	144	20
200000	5.3	4	301	20
200000	5.3	5	147	20
200000	5.3	6	40	20
200000	5.3	7	157	20
200000	5.3	8	192	20
200000	5.3	9	108	20
200000	5.3	10	92	20
80000	4.9	1	76	8
80000	4.9	2	10	8
80000	4.9	3	134	8
80000	4.9	4	17	8
80000	4.9	5	18	8
80000	4.9	6	41	8
80000	4.9	7	43	8
80000	4.9	8	34	8
80000	4.9	9	83	8
20000	4.3	1	63	2

20000	4.3	2	27	2
20000	4.3	3	12	2
20000	4.3	4	7	2
20000	4.3	5	32	2
20000	4.3	6	35	2
20000	4.3	7	29	2
20000	4.3	8	35	2
3200	3.51	1	15	0.32
3200	3.51	2	17	0.32
3200	3.51	3	34	0.32
3200	3.51	4	15	0.32
3200	3.51	5	4	0.32
3200	3.51	6	20	0.32
3200	3.51	7	17	0.32
3200	3.51	8	13	0.32
3200	3.51	9	16	0.32
800	2.9	1	3	0.08
800	2.9	2	3	0.08
800	2.9	3	18	0.08
800	2.9	4	14	0.08
800	2.9	5	3	0.08
800	2.9	6	1	0.08
800	2.9	7	1	0.08
800	2.9	8	1	0.08
800	2.9	9	5	0.08
800	2.9	10	3	0.08
800	2.9	11	3	0.08
800	2.9	12	2	0.08
800	2.9	13	2	0.08

Inicialmente se exploraron los datos. La distribución observada en el *scatterplot* sugirió que existía una relación entre el tamaño de los acúmulos virales, medido en número de viriones de NoV, y la concentración viral de los inóculos, medida en número de copias de genoma (**Figura 19**). Se observa que aumentan tanto la media como la varianza del tamaño de acúmulos virales en función de la dosis. Asimismo, y dado el carácter particulado de la muestra, se ajustó una regresión Poisson (Lopez-Gonzalez and Ruiz-Soler 2011).

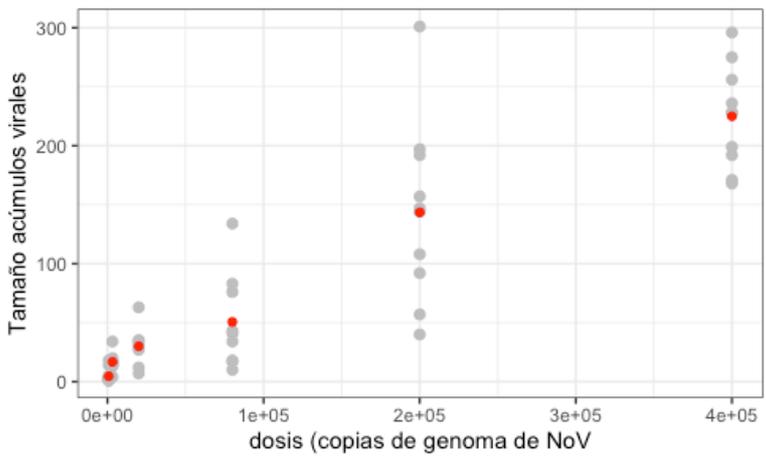


Figura 19. Scatter plot del tamaño de acúmulos virales en función del número de copias de genoma viral en cada dosis de inóculo. Los puntos grises indican cada dato y los puntos rojos la media en cada dosis.

El análisis de regresión Poisson arrojó una óptima bondad de ajuste con la inclusión de un término lineal y un término cuadrático para la dosis (scripts y salidas disponibles a solicitud). La curva de regresión se presenta en la **Figura 20**. Estos resultados, si bien no indican una ausencia total de un fenómeno de agregación, sí demuestran claramente que la cantidad de partículas virales disminuye conforme se diluye el inóculo y que existe una relación entre el número de partículas virales y las copias de genoma. Por lo tanto, continuamos con nuestro objetivo de estimar la dosis infecciosa 50% y la dosis diarrea 50% considerando que los inóculos administrados a los cerdos gnotobióticos representaban suspensiones virales homogéneamente dispersas, supuesto necesario para aplicar los modelos dosis-respuesta mecanicistas.

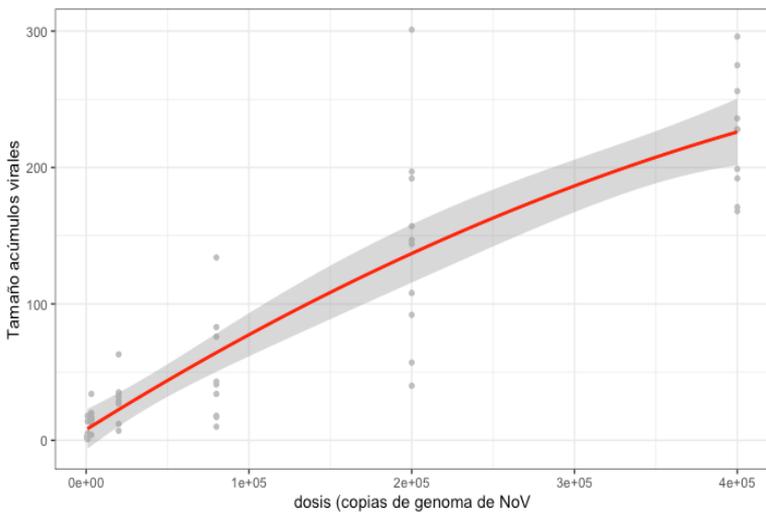


Figura 20. Análisis de Regresión de Poisson

Ecuación:

$$\ln(y) = \alpha + \beta_1 x + \beta_2 x^2$$

$\ln(\text{Tamaño de acúmulo viral}) = 2.616 + 1.685 \times 10^{-5} \text{ dosis} - 2.466 \times 10^{-11} \times \text{dosis}^2$

P-valor < 2e-16 *** para todos los coeficientes

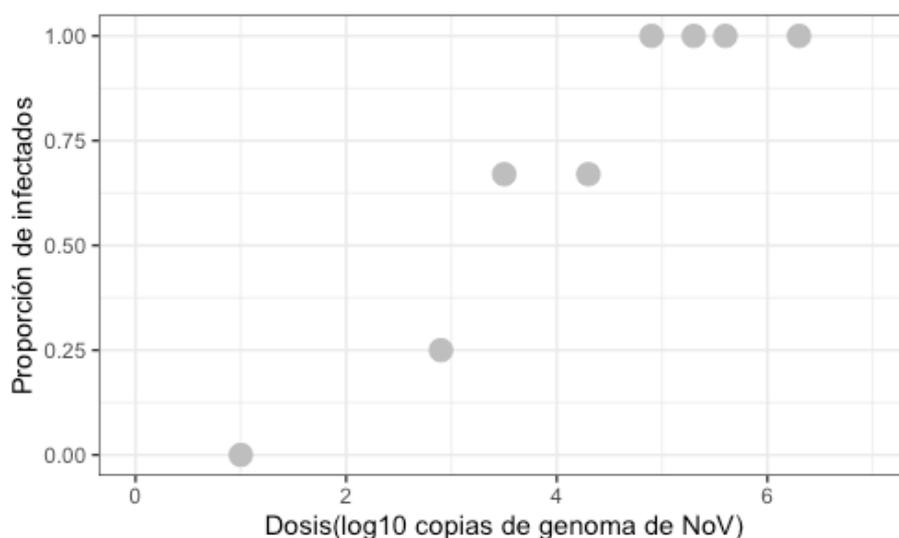
AIC 1313.6 Null deviance= 5719. Residual deviance=999.

(Scripts y salidas de R disponibles a solicitud)

4.4. Cálculo $DI_{50\%}$ y $DD_{50\%}$ por métodos clásicos

Inicialmente las $DI_{50\%}$ y $DD_{50\%}$ se estimaron utilizando los métodos clásicos de Reed & Muench, de Dragstedt-Behrens y el método no paramétrico de Spearman Karber. En la **Figura 21** se observan los gráficos de dispersión de la proporción de cerdos infectados y enfermos para cada nivel de dosis de inóculo y para el grupo control no desafiado que representa el valor de dosis 0 y respuesta 0.

a) Infección



b) Enfermedad

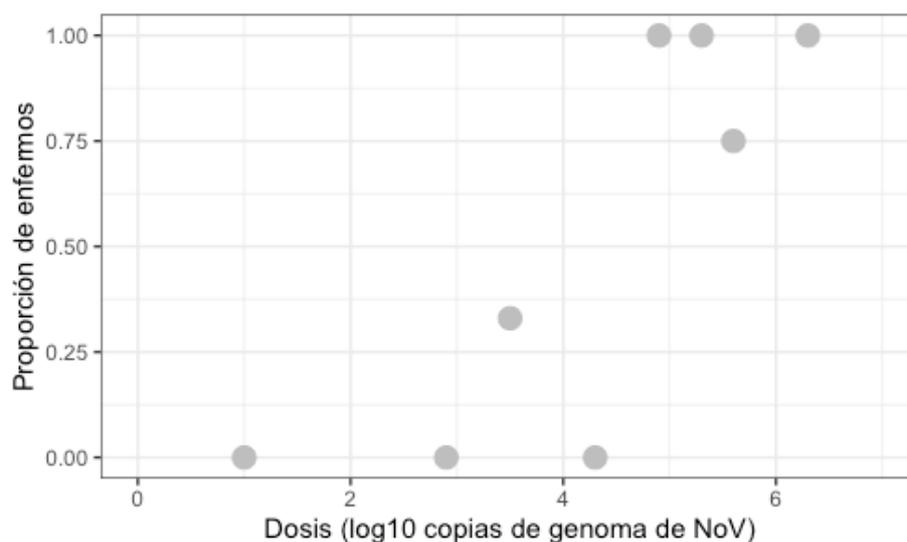


Figura 21. Distribución datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano y un octavo grupo control negativo que recibe solución salina.

En la **Tabla R12** se muestran los resultados de la estimación utilizando el **paquete skrmdb** publicado por David Siev en 2018. En todos los casos el cálculo se realizó incorporando el grupo control de cerdos no desafiados para robustecer el límite inferior de la curva llegando a 0%, especialmente para los datos de infección donde las dosis probadas no llegaron a dicho límite. En el uso del paquete skrmdb o la herramienta on line de NCBI, para evitar ceros en la

escala logarítmica el grupo control recibió un valor arbitrario de 10 copias de genoma para iniciar en $\log_{10}(10)=1$. A modo ejemplo, también se realizó el cálculo de R&M sin recurrir al paquete estadístico para estimar la $DI_{50\%}$ (ver ANEXO).

Tabla R 12. $DI_{50\%}$ y $DD_{50\%}$ por métodos clásicos.

Método	Herramienta	$\log_{10} DI_{50\%}$	SD	IC 95%	$\log_{10} DD_{50\%}$	D.E.	IC 95%
R&M	skrmdb	3.394	nd	nd	4.575	nd	nd
DB	skrmdb	3.389	nd	nd	4.575	nd	nd
SK	skrmdb 8G	3.354	0.3367	nd	4.492	0.046	nd
SK	On line Dil:10 8G	3.342	0.382	2.59- 4.09	2.205	0.219	1.776- 2.634
SK	On line Dil:10 7G On line	3.520	0.299	2.39- 4.1	4.759	0.219	4.330- 5.188

nd= dato no determinado por el método.

R&M: Reed y Muench, DB: Dragstedt-Behrens, SK: Spearman Karber. D.E.: desvío estándar.

Los resultados obtenidos por los métodos de interpolación de R&M y DB fueron idénticos si se redondea a dos decimales para ambos sets de datos: $\log_{10} DI_{50\%} = 3.39$ y $\log_{10} DD_{50\%} = 4.58$, respectivamente. Por su parte, el resultado estimado por el método de Spearman-Karber en el caso de la infección fue similar a los otros métodos: $DI_{50\%}=3.35$ y 3.34 , según se utilizara el paquete skrmdb de R o el software online de NCBI. Para los datos de enfermedad, sin embargo, el $\log_{10} DD_{50\%}$ estimado con la herramienta on line fue alrededor de la mitad ($\log_{10} DD_{50\%}= 2.2$) del obtenido con los métodos de interpolación e incluso con el mismo método, pero empleando el paquete skrmdb de R. En relación con este resultado no esperado, cuando se realizó el cálculo sin incluir el grupo control con el software online (ver última fila en la Tabla R12), éste arrojó un resultado en el mismo orden de valores.

4.5. Cálculo de las DI50% y DD50% utilizando modelos lineales generalizados

a. Regresión logística

En la **Tabla R13**, se detallan los parámetros estimados para infección y enfermedad luego de ajustar el modelo de regresión logística. En ambos casos el ajuste fue significativo (Scripts y salidas de R disponibles a solicitud).

Tabla R 13. Regresión logística.

Datos	Parámetros del modelo						Bondad de Ajuste		
	α	IC95%	p-valor	β	IC95%	p-valor	AIC	BIC	p-valor
Infección	-7.76	-16.1;2.8	0.018*	2.23	0.98;4.66	0.0105*	10.72	10.6	0.887
Enfermedad	-9.80	-19.7;-4.16	0.009*	2.26	1.04;4.37	0.005*	17.95	17.8	0.166

Ecuaciones

Infeción:

$$\text{Logit}(\pi) = \ln(\pi / 1-\pi) = -7.5756 + 2.23 \cdot \text{dosis}$$

Enfermedad:

$$\text{Logit}(\pi) = \ln(\pi / 1-\pi) = -9.8016 + 2.26 \cdot \text{dosis}$$

En ambas ecuaciones, el parámetro β representa la velocidad de crecimiento de la curva. Para interpretar mejor este parámetro es conveniente salir del entorno de la función de enlace. Aplicando la función inversa se obtiene el cociente de odds, es decir el cociente del odds de infectarse/enfermarse a dosis x y el odds de no infectarse/enfermarse a la misma dosis x . Cada uno de esos odds es el cociente de la probabilidad de infectarse/enfermarse al estar expuesto y la probabilidad de no infectarse/enfermarse al estar expuesto. En este caso, la chance de infección aumenta $e^{2.23} = 9,299$ veces con cada aumento de un orden de magnitud (\log_{10}) de la dosis, mientras que la chance de sufrir diarrea aumenta $e^{2.26} = 9,583$ veces.

El \log_{10} de las $DI_{50\%}$ y $DD_{50\%}$ se obtienen evaluando la función logit en $\pi = 0.5$ y despejando. En particular:

$$\text{Infeción: } \pi=0.5 \Rightarrow \log DI_{50\%} = -\alpha / \beta = 7.57/2.22 = \mathbf{3.40}$$

$$\text{Enfermedad: } \pi=0.5 \Rightarrow \log DD_{50\%} = -\alpha / \beta = 9.80/2.26 = \mathbf{4.33}$$

El valor de $DI_{50\%}$ es similar con los obtenidos por los métodos clásicos de la sección precedente. Por su parte, la estimación de la $DD_{50\%}$ se encuentra en el mismo orden (sería idéntico si se redondea a un decimal).

Paralelamente, con la herramienta *online* del NCBI, bajo la opción “*curve fit*” y con una función de enlace logit, también se pueden obtener las estimaciones de las dosis 50% y sus correspondientes intervalos de confianza 95% (Tabla R14).

Tabla R 14. Estimaciones por regresión logística: diferentes métodos.

Método	logDI50%	IC95%	logDD50%	IC95%
Estimado glm R	3.40	(2.78;4.03)	4.33	(3.79;4.88)
Estimado online <i>curve fit</i>	3.41	(2.78;4.03)	4.28	(3.80;4.75)

El agregado de un grupo de dosis 0 y respuesta 0 en la herramienta *online*, tal como se mencionó anteriormente, da un error, por lo que es necesario colocar un valor arbitrario de dosis=10 para el grupo control negativo. Los resultados fueron comparables a los obtenidos por R. En todos los casos la bondad de ajuste del modelo de regresión logística a los datos fue significativa ($p>0.05$). Las salidas de la implementación de la herramienta *online* se presentan en el ANEXO 5. Los *scripts* y salidas de R están disponibles a solicitud.

En la **Figura 22** se presenta las curvas de dosis-respuesta estimada para infección y enfermedad y sus correspondientes intervalos de confianza. Es importante destacar que los intervalos no se obtienen con la función *predict* directamente, que los calcula en la escala de la función de enlace, sino que hay que transformarlos a la escala de la variable respuesta (script disponible a solicitud).

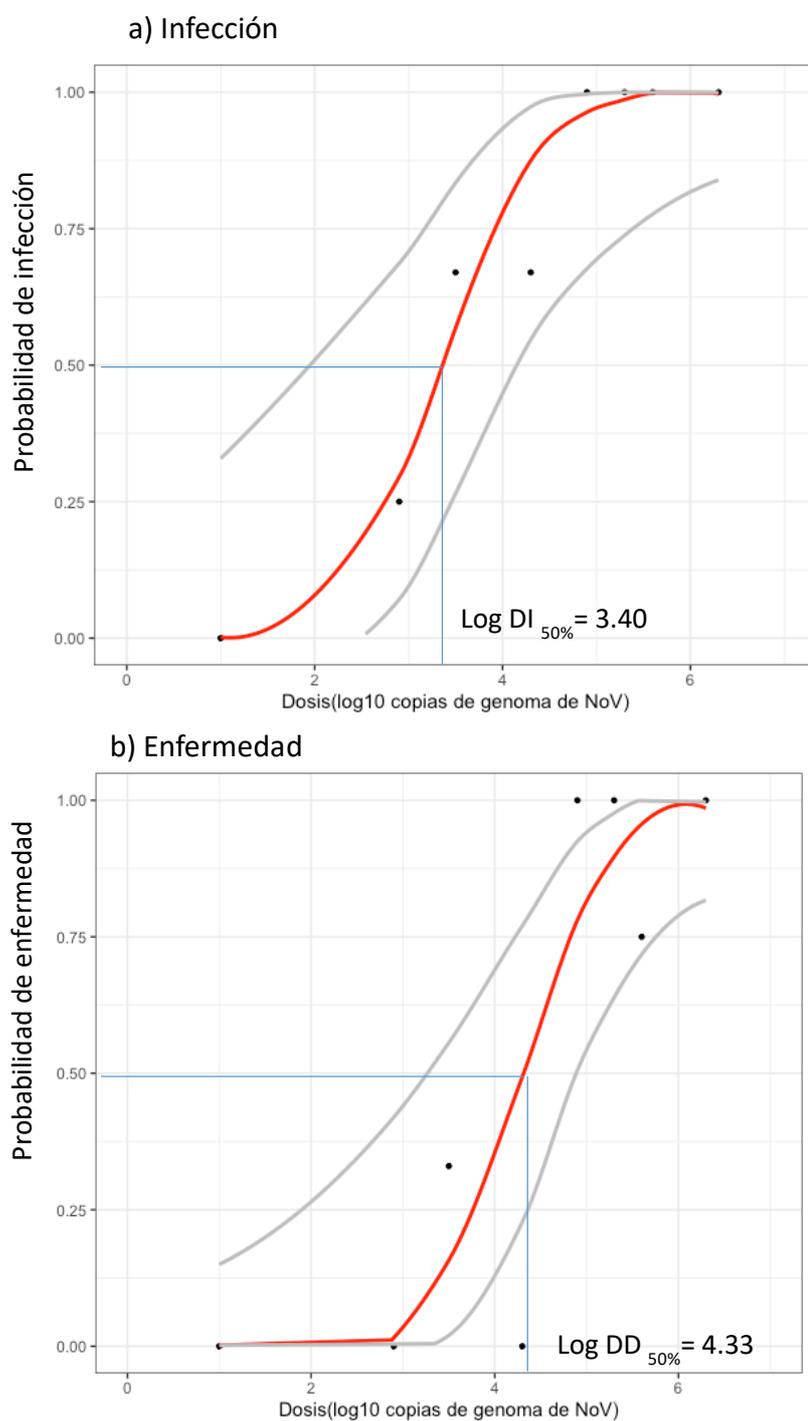


Figura 22. Ajuste de curva de modelo de regresión logística a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano. Línea roja: respuesta media; líneas grises: IC95%; puntos: datos observados proporción por grupo de dosis).

Finalmente, en la **Tabla R15** se detallan las dosis de NoV utilizadas y el resultado de probabilidad de infección y diarrea estimados por el modelo de regresión logística con sus correspondientes IC95% en correspondencia con los ilustrados en la **Figura 22**.

Tabla R 15. Probabilidad de infección y enfermedad estimado a partir del modelo de regresión logística.

NoV Log (dosis)	% infección observado	% infección predicho Regresión Logística (8 Grupos)	IC 95%		% diarrea observado	% diarrea predicho Regresión Logística	IC 95%	
			límite inf.	límite sup.			límite inf.	límite sup.
6.3	100	99.84	84	99.9	100	98.83	82	100
5.6	100	99.25	77	99.9	75	94.57	71	99
5.3	100	98.55	74	99	100	89.83	64	98
4.9	100	96.54	68	99	100	78.14	51	93
4.3	67	88.00	54	97	0	47.92	21	76
3.5	67	55.32	0.24	82	33	13.1	18	55
2.9	25	24.56	0	66	0	3.73	0	42
0	0	0	0	33	0	0	0	15

Los IC95% se calcularon manualmente aplicando la fórmula de IC95% para proporciones $IC\ 95\% = \text{predicho} \pm 1.96 \cdot SE$ utilizando los predichos y el S.E. solicitados con la inversa de la función de enlace, se corroboró que el resultado fuera el mismo al realizarlo dentro de la función de enlace y luego exponenciar. Los intervalos no son equidistantes.

b. Modelo de Log-log (herramienta *online* del NCBI)

Finalmente, como se describe en Materiales y Métodos, se utilizó también la herramienta *online* del NCB bajo la opción “*model fit*” para estimar la $DI_{50\%}$ y $DD_{50\%}$ ajustando un modelo Log-log. Para ambos sets de datos el programa rechazó el ajuste al incluir el grupo control. Sin embargo, al ejecutarlos sólo con los 7 grupos experimentales, el ajuste fue estadísticamente significativo para los datos de infección y arrojó un valor de $\log_{10} DI_{50\%}$ de 3.39 con un IC95 entre 2.76 y 4.04. Para el caso de los datos de enfermedad tanto con 7 u 8 grupos el modelo no fue adecuado para ajustar a los datos ($p < 0.05$), salidas disponibles a solicitud.

4.6. Cálculo de las $DI_{50\%}$ y $DD_{50\%}$ utilizando modelos de regresión no lineal

a. Regresión log-logística de dos parámetros 2PL.L.

En la **Figura 23** se ilustra el ajuste del modelo de regresión log-logística de dos parámetros (2PL.L.) con el paquete *drc* de R a cada set de datos. En la **Tabla R16**, por su parte, se presentan los parámetros estimados. La bondad de ajuste del modelo resultó adecuada para ambos conjuntos de datos, al igual que para el parámetro b , que presentó un p-valor significativo. Pero hay que destacar que el valor estimado del parámetro e , que precisamente representa a la dosis infecciosa y de enfermedad 50%, arrojó a un p-valor no significativo con intervalos de confianza muy amplios.

Aplicando logaritmo al parámetro e para expresarlo en la misma escala que con los métodos anteriormente descritos, el $\log_{10} DI_{50\%}$ y el $\log_{10} DD_{50\%}$ fue de 3.40 y 4.33, respectivamente. Estos valores son muy similares a los obtenidos con los métodos clásicos y con la regresión logística. En la **Tabla R17** también se presentan los valores predichos por el modelo para cada dosis y sus intervalos de confianza calculados con el método “delta”, que utiliza por defecto el programa.

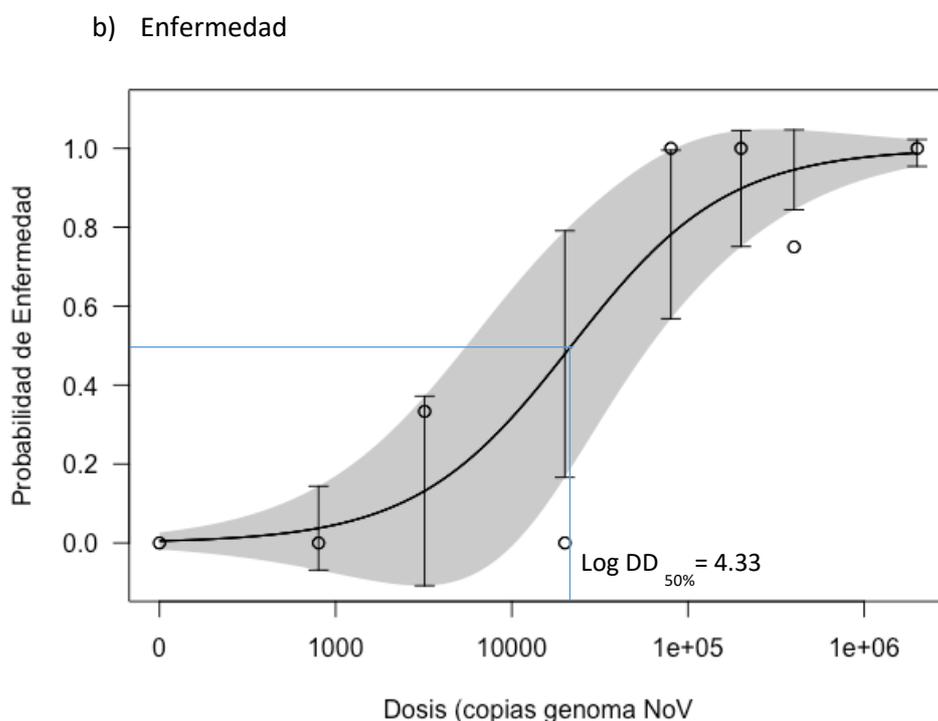
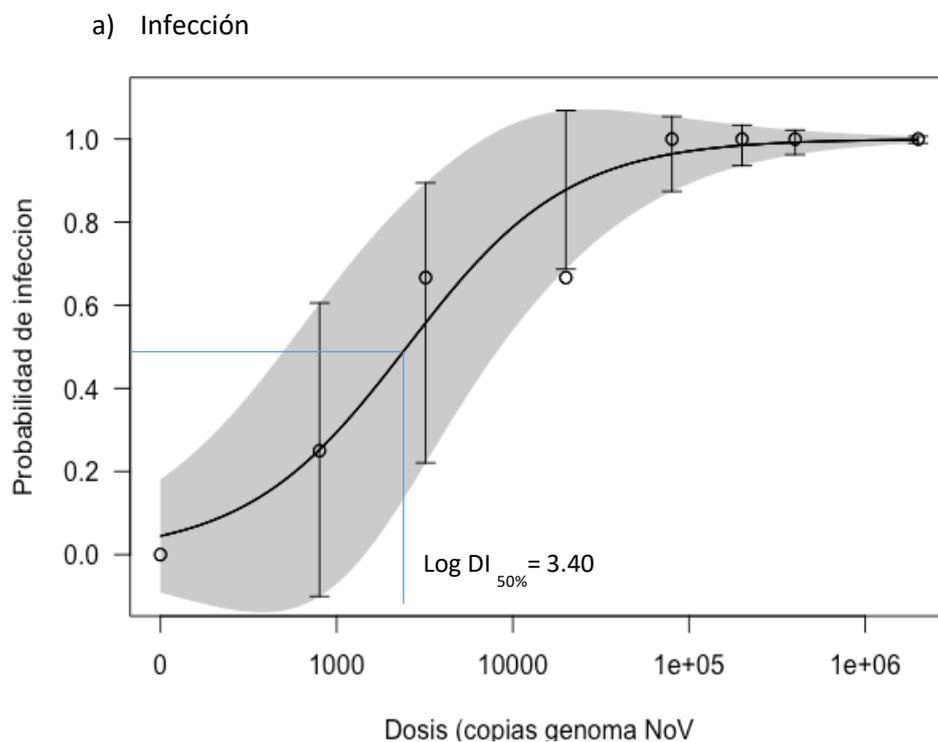


Figura 23. Curva de de dosis-respuesta ajustando el modelo de regresión 2PL.L. a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con 7 dosis crecientes de NoV humano y un octavo grupo control negativo que recibe solución salina en lugar de virus. Los círculos representan el valor predicho promedio. El sombreado y las barras los IC95%.

Tabla R 16. Resultado del ajuste de los datos de infección y diarrea mediante regresión log-logística 2P L.L.

Datos	Parámetros del modelo						Bondad de Ajuste				Normalidad Residuos deviance	
	b	IC95%	p-valor	e	IC95%	p-valor	AIC	BIC	χ^2 Crítico	p-valor	W*	p-valor
Infección	-0.95	(-1.71; -0.18)	0.015*	2508	(-1215; 6232)	0.187	10.67	10.86	1.72	0.887	0.79	0.03
Enfermedad	-0.98	(-1.68;-0.289)	0.005*	22615	(-5502; 49132)	0.118	17.95	18.09	9.133	0.166	0.90	0.295

Tabla R17. Probabilidad de infección y enfermedad estimada a partir del ajuste de modelo de regresión 2PL.L.

NoV Log (dosis)	% infección observado	% infección predicho 2P LL	IC 95%		% diarrea observado	%diarrea Predicho 2P L.L.	IC 95%	
			límite inferior	límite superior			límite inferior	límite superior
6.3	100	99.83	98.97	99.83	100	98.83	95.42	102
5.6	100	99.20	96.25	99.20	75	94.42	84.44	105
5.3	100	98.46	93.59	98.46	100	89.8	75.11	105
4.9	100	96.41	87.40	96.41	100	78.18	56.83	99.52
4.3	67	87.79	68.71	87.79	0	47.87	17	79.1
3.5	67	55.76	22.06	55.76	33	13.18	0	37.2
2.9	25	25.24	0	25.24	0	3.75	0	14.38
0	0	0	0	0	0	0	0	0

IC95% obtenidos con el comando confidence con el método delta que es el cálculo que realiza por defecto el paquete drc de R

b. Modelos mecanicistas: Exponencial y Beta Poisson

Como se explicó en las secciones previas, los modelos Exponencial y Beta-Poisson aproximado cumplen con supuestos mecanicistas. Ambos modelos de dosis-respuesta se utilizan de forma rutinaria para determinar la mediana de las dosis infecciosas y de enfermedad para diferentes patógenos en análisis de riesgo asociados al consumo de agua y alimentos contaminados, debido a su plausibilidad biológica. El análisis se realizó utilizando el script de R publicado por Weir et al. (2017) (script disponible a solicitud). Al igual que para los otros métodos, se ajustaron los modelos a los dos conjuntos de datos, de infección y enfermedad, incluyendo el grupo control. En este último caso, dado que el script de Weir, 2017 no admite una dosis cero, nuevamente se decidió utilizar el valor arbitrario de diez copias de genomas para la respuesta nula.

En este punto, vale la pena recordar que ambos métodos se denominan mecanicistas porque asumen que las partículas virales del inóculo siguen una distribución de Poisson alrededor de la media de la dosis. Mientras que el modelo Exponencial asume una probabilidad binomial constante de que una unidad infecciosa de virus sobreviva para generar infección y diarrea, el modelo Beta-Poisson asume que esta probabilidad es variable según una distribución beta.

En las **Tablas R18 y R19** se presentan los resultados del ajuste de los modelos Exponencial y Beta Poisson, respectivamente, y los valores estimados de $DI_{50\%}$ y $DD_{50\%}$. En el caso del modelo Beta-Poisson el valor de la dosis mediana coincide con el parámetro del $N50$. Como se puede observar, ambos modelos arrojaron una adecuada bondad de ajuste para los datos de infección. Sin embargo, para los datos de enfermedad, el modelo Beta-Poisson logra un ajuste significativo mientras que el modelo Exponencial no. Al observar los $\log_{10} DI_{50\%}$ y $\log_{10} DD_{50\%}$ obtenidos, ambos modelos mecanicistas arrojan valores en el mismo orden que los métodos clásicos y los modelos estadísticos previamente utilizados. En particular, el modelo Beta-Poisson es el que presenta los valores más cercanos a los otros modelos.

Tabla R 17. Parámetros y bondad de ajuste del modelo Exponencial para estimar la DI50% y la DD50% de NoV humano en cerdos gnotobióticos.

Datos	Modelo Exponencial				Estimación	
	Parámetro	Bondad de ajuste			Log ₁₀ DI/DD 50%	DI50% / DD50%(IC95%)
r=k (IC95%)	Y	p-valor	AIC			
Infeción	1.20x 10 ⁻⁴ (5E-05;7E-04)	3.75	0.808	5.75	3.76	5.78E+03 (1043;13707)
Enfermedad	1.20x 10 ⁻⁵ (7E-06;4E-05)	16.11	0.0241*	18.11	4.76	5.78E+04 (111802;122982)

* El modelo exponencial no ajusta a los datos de diarrea

H0: hipótesis nula. Si p>0.05 -> El modelo posee bondad de ajuste adecuada

Modelo exponencial p:r: un parámetro).

Y: deviance. El código de R estima por MLE la deviance (-2 ln Likelihood) del modelo

AIC: Criterio de información de Akaike: 2p -2 ln (Likelihood del modelo)

Tabla R 18. Parámetros y bondad de ajuste del modelo mecanicista Beta-Poisson para estimar DI50% y la DD50% de NoV humano en cerdos gnotobióticos.

Modelo Beta-Poisson aproximado						
Datos	Parámetros del modelo		Bondad de Ajuste			Estimación
	α (IC95%)	N50 (IC95%)	Y	p-valor	AIC	Log ₁₀ DI/DD50%
Infección	1.0074 (0.508;2355)	2603 (676;10533)	1.74	0.942	5.75	3.41
Enfermedad	0.929 (0.38;5936)	21360 (10605;39404)	10.58	0.102	14.58	4.33

H0: hipótesis nula. Si $p > 0.05$ -> El modelo posee óptima bondad de ajuste.

Beta Poisson, p: α y N50, dos parámetros.

Y: *deviance*. El código de R estima por MLE la deviance (-2 ln Likelihood) del modelo

AIC: Criterio de información de Akaike: $2p - 2 \ln(\text{Likelihood del modelo})$

Reemplazando los valores obtenidos para los parámetros de ambos modelos (**Tablas R18 y R19**) para los datos de infección y diarrea, se obtienen las siguientes ecuaciones de probabilidades de infección y enfermedad en función de la dosis del patógeno:

Modelo Exponencial:

Infección

$$p_{inf}(D|r) = 1 - e^{(-0.000120 * D)}$$

Enfermedad

$$p_{diarrea}(D|r) = 1 - e^{(-0.0000120 * D)}$$

Modelo beta Poisson aproximado

Infección

$$p_{inf}(D|\alpha, N_{50}) = 1 - \left[1 + \frac{D \times \left(2^{\frac{1}{0.998}} - 1 \right)}{2572.50} \right]^{-0.998}$$

Enfermedad

$$p_{diarrea}(D|\alpha, N_{50}) = 1 - \left[1 + \frac{D \times \left(2^{\frac{1}{0.929}} - 1 \right)}{21360} \right]^{-0.929}$$

La incertidumbre de los parámetros estimados fue evaluada por bootstrap. En la **Figura 24** se presenta el histograma de las estimaciones del parámetro r del modelo Exponencial, obtenido a partir de 10.000 pseudo-réplicas. Para los datos de infección, se observa centralidad y una dispersión reducida, en correspondencia con una adecuada bondad de ajuste. No sucede lo mismo con los datos de diarrea, cuyas estimaciones bootstrap del parámetro se dividen en dos poblaciones, en correspondencia a la falta de ajuste de este modelo a los datos.

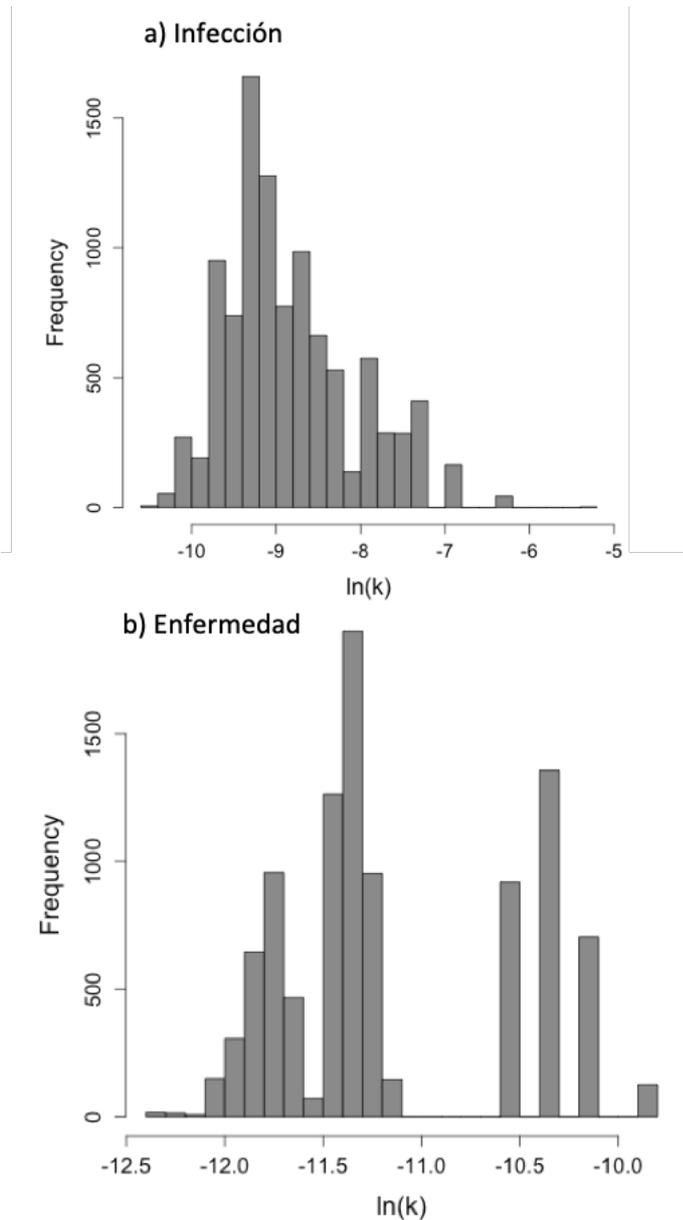


Figura 24. Histograma de las estimaciones del parámetro r del modelo Exponencial, con 10.000 pseudo-réplicas.

En forma análoga, en la **Figura 25** se presenta el scatter plot del Ln de los parámetros α y N50 del modelo Beta-Poisson. Para interpretar el gráfico se presenta una superficie de respuesta. El punto en gris con la x corresponde al par de valores estimados más probable de ambos parámetros N50 y α es escala logarítmica natural, para cada set de datos. Las elipses marcan los intervalos de confianza.

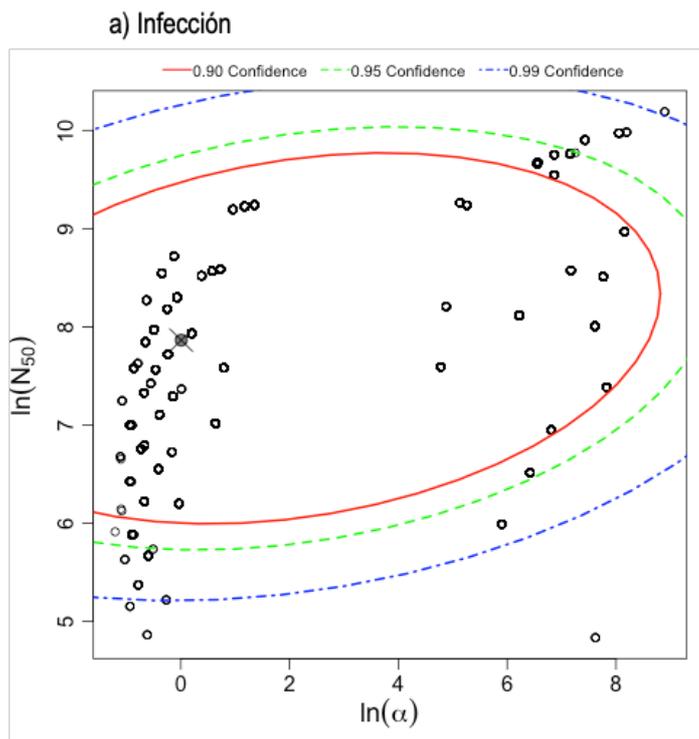
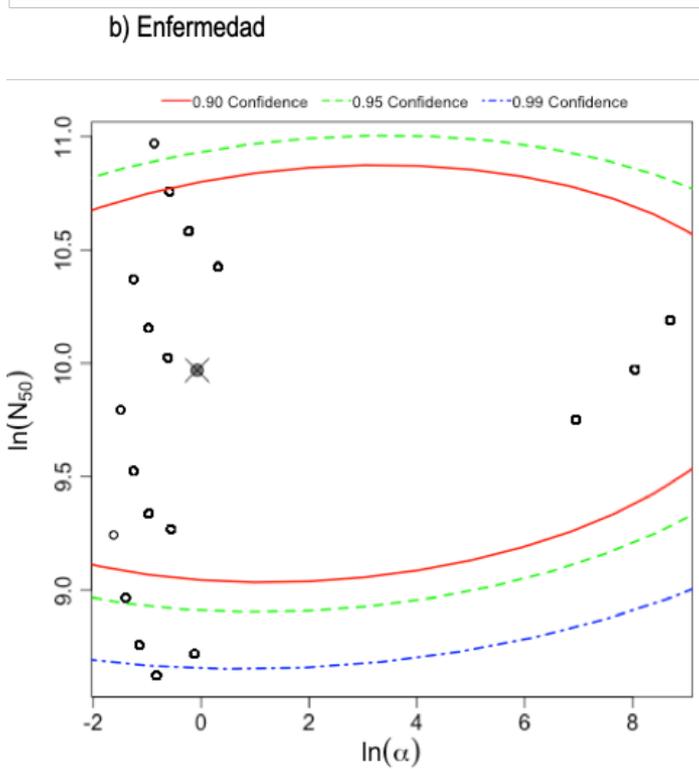


Figura 25. Estimación bootstrap de la incertidumbre de los parámetros α y N_{50} del modelo Beta-Poisson al ajustarse a los datos de (a) infección y (b) enfermedad. Elipses de confianza para los percentiles 90, 95 y 99 (10.000 pseudo réplicas), scripts de Weir, 2017

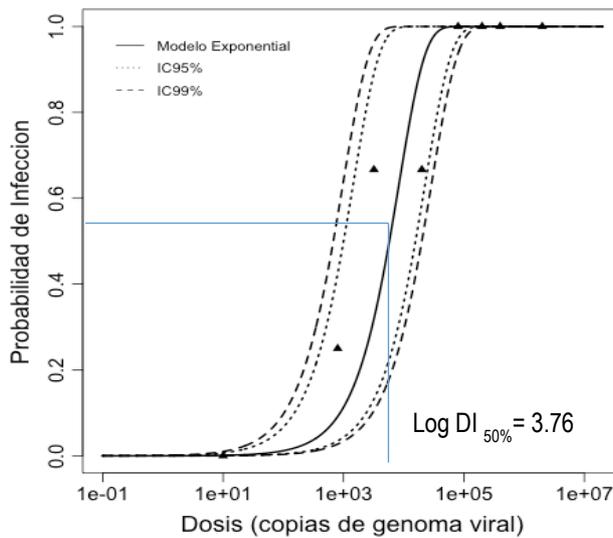
$\alpha = 1.007$
 $N_{50} = 2603$
 $\text{Log}_{10} \text{Log}_{10} (N_{50}) =$
 $\text{Log}_{10} \text{DI}50\% = 3.41$



$\alpha = 0.93$
 $N_{50} = 21360$
 $\text{Log}_{10} (N_{50}) = \text{Log}_{10} \text{DD}50\% = 4.33$

En la **Figura 26** se presentan las curvas las probabilidades de infección y enfermedad estimadas para el modelo Exponencial con sus respectivas bandas de de confianza del 95% y 99%, obtenidas por el método de bootstrap según el *script* de Weir, 2017.

a) Infección



b) Enfermedad

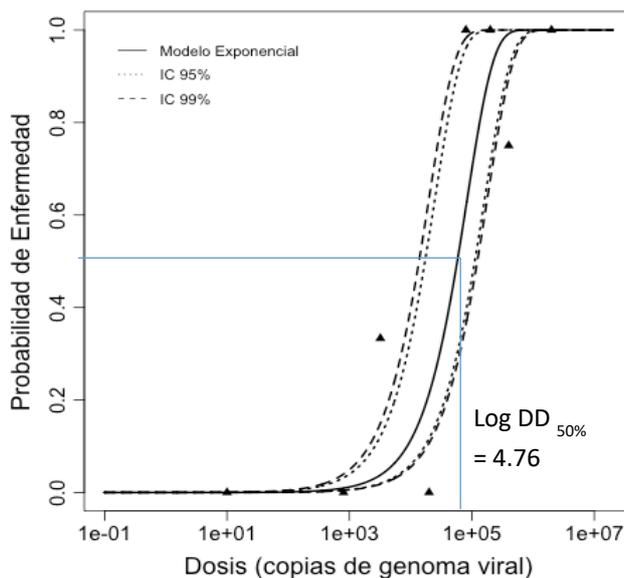


Figura 26. Curvas de probabilidad estimada e intervalos de confianza del modelo Exponencial ajustado a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano. Los intervalos de confianza fueron calculados por bootstrap implementado en el script de Weir et al. (2017).

El modelo exponencial ajusta muy bien a los datos de infección y a pesar de no ser significativo el p-valor para los datos de diarrea, se observan intervalos de confianza menores que los que se estiman para los datos de infección. Es importante resaltar, sin embargo, que arroja los valores de dosis infecciosa y diarrea 50% mas elevados en comparación con los otros modelos utilizados.

En forma análoga, en la **Figura 27** se presentan las curvas e IC para el modelo Beta-Poisson. Los valores de Log_{10} de $\text{DI}_{50\%}$ y $\text{DD}_{50\%}$ determinados por el modelo de Beta-Poisson aproximado fueron 3.41 y 4.33 copias de genoma viral, respectivamente. Estos valores son más

cercanos a los obtenidos por los métodos clásicos y las regresiones logística y 2PL.L. que los estimados por el modelo Exponencial (3.76 y 4.76, respectivamente).

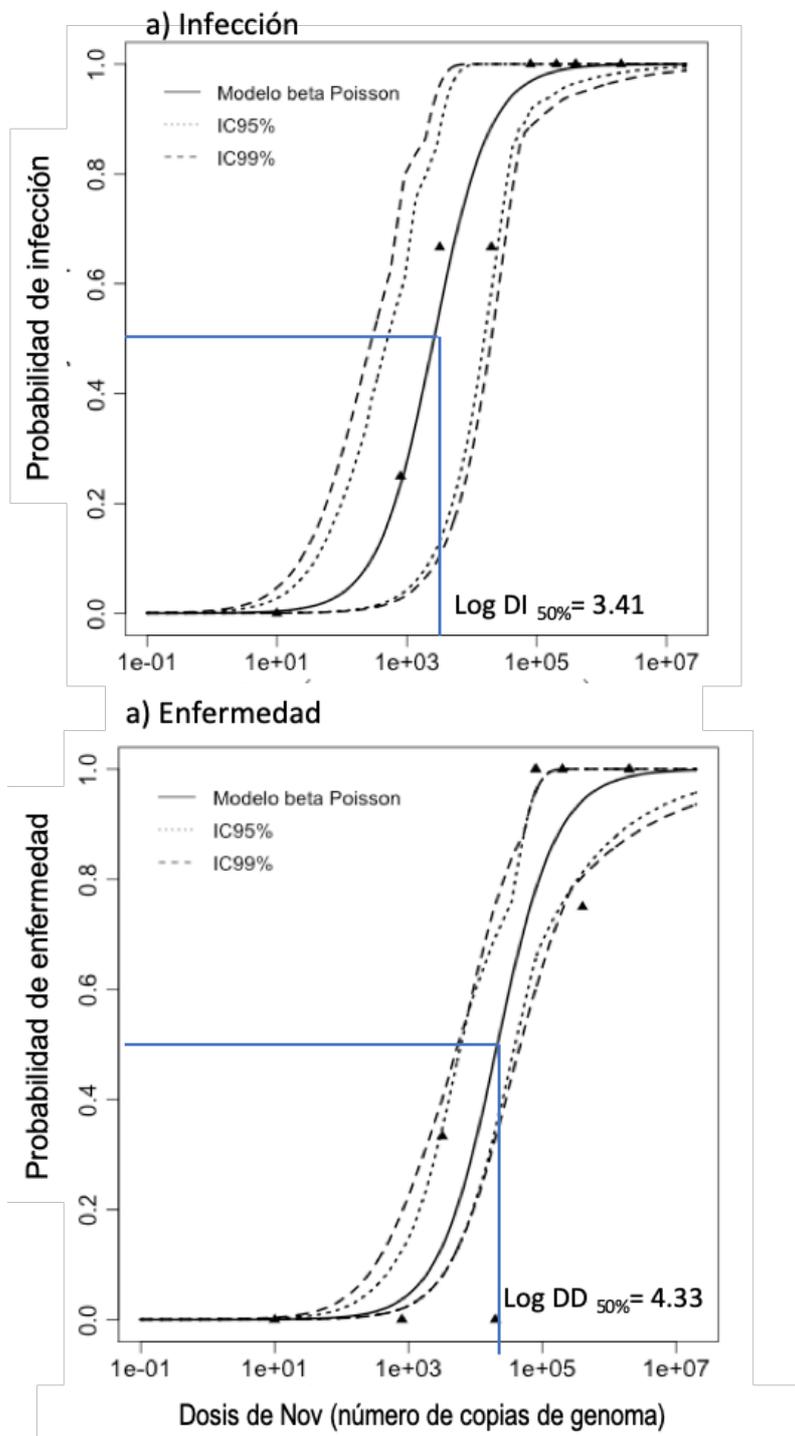


Figura 27. Curvas de probabilidad estimada e intervalos de confianza del modelo Beta-Poisson ajustado a los datos de cerdos gnotobióticos infectados (a) y enfermos (b) luego del desafío con dosis crecientes de NoV humano. Los intervalos de confianza fueron calculados por el método de bootstrap implementado en el script de Weir et al. (2017).

Como método alternativo al bootstrap, los intervalos de confianza también fueron calculados mediante un método de muestreo de cadenas de Markov Montecarlo basado en un enfoque Bayesiano (Schmidt, 2015) e implementado a través de un script en R del mismo autor. Los resultados fueron muy similares y, por ese motivo, no se incluyen en el cuerpo de la tesis, aunque pueden consultarse en el ANEXO.

En la **Tabla R19** se detallan los valores predichos con sus correspondientes intervalos de confianza del 95% para los modelos Exponencial y Beta-Poisson para cada dosis viral utilizada obtenidos a partir de la base de datos que da origen a las Figuras 26 y 27.

Tabla R 19. Probabilidad de infección a partir de los modelos Exponencial y Beta-Poisson aproximado.

Dosis NoV	% infección observado	% infección Predicho Exponencial	IC 95% bootstrap	% infección predicho Beta Poisson	IC 95% bootstrap
2000000	100	100	100	100	98;100
400000	100	100	100	99.18	96;100
200000	100	100	99.9;100	98.00	94;100
80000	100	99.99	97.1;100	95.94	91;100
20000	66.67	90.95	58.7; 99	61.74	58.6;99
3200	66.67	31.91	13;89	21.35	13;88
800	25.00	9.16	3;42	3.77	3;58
0	0	0	0	0	0

En forma análoga, en la **Tabla R20** se describen los valores predichos y sus correspondientes IC95% del modelo beta Poisson para los datos de enfermedad.

Tabla R 20. Probabilidad de diarrea a partir de los modelos Exponencial y Beta-Poisson aproximado.

Dosis NoV	% diarrea observado	% predicho Exponencial	IC95% Límite inferior	IC95% Límite superior	% Predicho Beta Poisson	IC95% Límite inferior	IC95% Límite superior
2000000	100	100	99	100	98.67	89	100
400000	100	100	99	100	94.28	81.32	100
200000	100	100	99	100	89.56	75.89	99.97
80000	100	99.99	96	100	78.21	65.72	96
20000	0	90.93	75	96	48.42	34.57	69.45
3200	33	31.89	22	50	13.32	7.89	38.90

800	0	0.091	0	34	3.71	2	17.53
0	0	0	0	0	0	0	0.003

En la **Tabla R21**, finalmente, se compara ambos métodos mecanicistas, el Exponencial y el Beta-Poisson en términos de su bondad de ajuste. Para los datos de infección, no hubo diferencia significativa entre los modelos. En cambio, al comparar los modelos para los datos de enfermedad, el modelo Beta-Poisson aproximado, se vio favorecido, como también lo confirma el criterio de información de Akaike (AIC).

Tabla R 21. Prueba de comparación entre modelos mecanicistas.

Datos	Modelo	Deviance Minimizada Y	diferencia $\Delta=Y1-Y2$	χ^2 valor crítico	p-valor	AIC	Conclusión
Infección	Exponential	3.7519				5.75	Ambos modelos ajustan adecuadamente
	Beta Poisson	1.7438	2.0081	3.8415	0.1565	5.75	
Enfermedad	Exponential	16.1142				18.11	Beta Poisson mejor ajuste
	Beta Poisson	10.5796	5.5346	3.8415	0.0186	14.58	

4.7. Cálculo $DI_{50\%}$ y $DD_{50\%}$ utilizando variables continuas

En esta sección se describe, por último, la estimación de la $DI_{50\%}$ y $DD_{50\%}$ a partir de los datos del \log_{10} del área bajo la curva (AUC) de excreción viral de cada cerdo, AUCS, y del score fecal, AUCD (ver Tabla R1). El seguimiento realizado a los cerdos gnotobióticos día por día permitió contar con el dato preciso de la cantidad de virus excretada y la severidad del cuadro clínico en el curso de la reproducción de la infección por NoV humano. En este caso, ambas variables son continuas (ver estadísticas descriptivas en Tablas R5 y R9) y se analizaron ajustando un modelo de regresión log-logístico de tres parámetros (3P L.L.). Al igual que para el ajuste de los modelos utilizados sobre la proporción de animales afectados, se comparó el ajuste y el valor de los parámetros estimados al incluir en el análisis 7 grupos experimentales y el grupo control negativo.

En la **Tabla R22** se presentan los resultados del análisis. Se observa que el modelo cumplió con los supuestos de normalidad y homocedasticidad de los residuos de *deviance* para ambos conjuntos de datos y que dos de los tres parámetros estimados fueron estadísticamente significativos. El parámetro e , que corresponde a la dosis efectiva 50% que buscamos estimar, no resultó significativo.

Las $DI_{50\%}$ y $DD_{50\%}$ estimadas arrojaron valores en el mismo rango que los obtenidos a partir de los datos de proporción de cerdos infectados y enfermos. Para infección el \log_{10} de la $DI_{50\%}$ fue de 3.85, mientras que para enfermedad el \log_{10} de $DD_{50\%}$ fue 4.47.

En la **Tabla R23** se detallan los valores correspondientes a cada dosis en estudio y en la **Figura 28** se grafican las curvas de dosis-respuesta en función de las copias de genoma viral (dosis) acompañadas de sus intervalos de confianza. A partir del análisis de las variables continuas de AUC de excreción viral y de severidad de diarrea se arribó a resultados de dosis medianas 50% muy similares a las obtenidas a partir de los datos de categóricos, representando otra forma de construcción de curvas dosis respuesta para la validación de modelos animales de valor preclínico.

Tabla R 22. Resultado del ajuste de la regresión log-logística de tres parámetros a los datos de excreción viral (Log AUCS) y de score fecal de diarrea (AUCD).

DATOS	Parámetros del modelo						Bondad de Ajuste				Normalidad Residuos de deviance				
	b	IC95%	p-valor	d	IC95%	p-valor	e	IC95%	p-valor	AIC	BIC	χ ² * crítico	p-valor	W*	p
Infección Log AUCS	-0.57	(-1;-0.11)	0.015*	4.52	(3.22;5.81)	0.000*	7203	(-6120; 20501)	0.277	108	113	31.7	0.000	0.97	0.482
Enfermedad AUCD	-0.48	(-9.57;-122.8)	0.044*	10.85	(5.7;16)	0.0002*	29464	(-42868; 1017957)	0.412	153	159	33.1	0.000	0.96	0.231

(Scripts y salidas de R disponibles a solicitud).

Reemplazando en la ecuación de la curva de la regresión log-logística los valores de los parámetros estimados en el análisis con los ocho grupos se obtienen las siguientes ecuaciones para la respuesta media en función de la dosis:

Infección:

$$AUCS\ media = \frac{4.52}{1 + e^{[-0.57(\log(dosis) - \log(7203))]}}$$

Parámetro e del modelo corresponde a la dosis efectiva 50%

$$DI50\% = 7203 = 7.2 \times 10^3\ copias\ de\ genoma$$

$$\log_{10}\ DI50\% : 3.80$$

Enfermedad:

$$AUCD\ media = \frac{10.85}{1 + e^{[-0.48(\log(dosis) - \log(29464))]}}$$

$$DD50\% = 29464 = 2.9 \times 10^4\ copias\ de\ genoma$$

$$\log_{10}\ DD50\% : 4.47$$

Tabla R 23. Predichos e intervalos de confianza y predicción del 95% para el AUCS de excreción viral (infección) y el AUCD de severidad de la diarrea (enfermedad).

NoV dosis	n	Media Log AUCS observada	Media log AUCS predicha 3P L.L.	Intervalo Confianza 95%	Intervalo Predicción 95%	Media AUCD observada	Media log AUCD predicha 3P L.L.	Intervalo Confianza 95%	Intervalo Predicción 95%
2000000	4	3.94	4.34	3.40-4.34	1.59-7.09	9.06	9.6	7.56;11.64	4.19;15
400000	4	3.64	4.1	3.44-4.76	1.43-6.77	9.31	8.45	7.13;9.78	3.27;14
200000	4	5.05	3.93	3.33-4.53	1.28-6.59	7.04	7.77	6.49;9.05	2.60;13
80000	4	4.22	3.62	2.98-4.26	0.96-6.59	7.46	6.71	5.42;8.00	1.54;12
20000	6	1.3	2.96	2.17-3.74	0.25-5.66	1.5	4.91	3.41;6.41	-0.32-10
3200	3	2.84	1.86	0.92-2.80	-0.89-4.61	3.17	2.75	0.63;4.88	-2.7;8.19
800	3	1.02	1.12	0.09-2.13	-1.66-3.89	2.75	1.6	-0.53;3.74	-3.83;7
1	4	0	0	0.052	-0.05-1.05	0	0.073	-0.25-0.40	-5;5

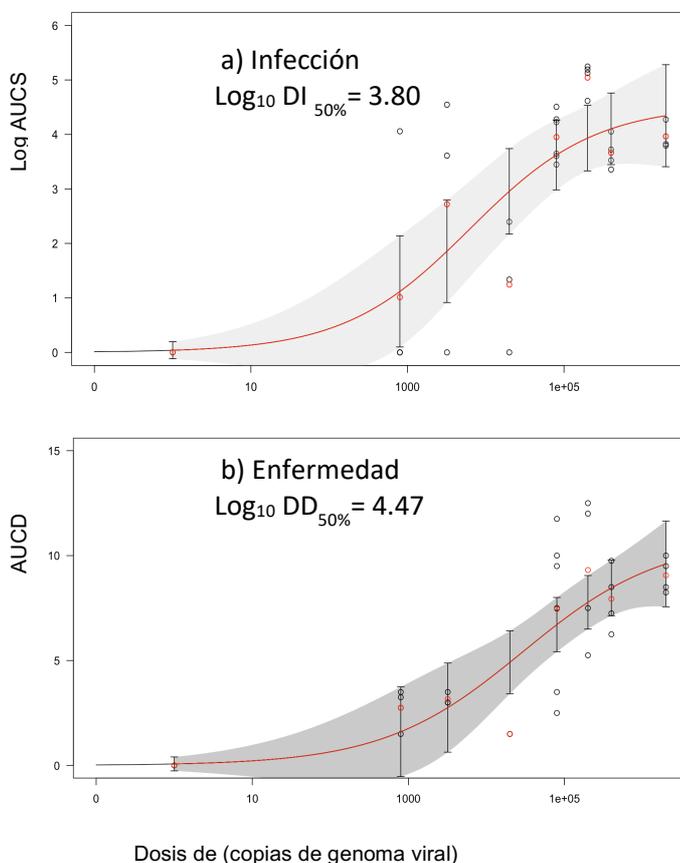


Figura 28. Ajuste del modelo de regresión log-logística de tres parámetros a los datos del área bajo la curva de excreción viral (a) y de los scores fecales AUCD. Los intervalos de confianza 95% fueron calculados por el método delta del paquete drc. Ritz, 2015.

4.7. Resumen de resultados $DI_{50\%}$ and $DD_{50\%}$

En la Tabla R24 se resumen todos los resultados obtenidos con respecto a la estimación de la $DI_{50\%}$ y $DD_{50\%}$ a partir de los diferentes métodos matemáticos y modelos estadísticos empleados en este trabajo.

Tabla R 24. Resumen de resultados $DI_{50\%}$ and $DD_{50\%}$

Método / modelo (herramienta)	Log ₁₀ $DI_{50\%}$	$DI_{50\%}$		Log ₁₀ $DD_{50\%}$	$DD_{50\%}$	
Métodos aplicados sobre los datos de proporciones de animales afectados sobre expuestos						
Métodos de interpolación matemática						
Reed-Muench						
R skrmdb	3.39	2461		4.58	37584	
Dragstedt-Behrens						
R skrmdb	3.39	2449		4.58	37584	
Método no paramétrico						
Spearman-Karber						
R skrmdb	3.35	2259		4.49	31017	
Online tool SK	3.34	2198		4.76 [#]	57544 [#]	
Modelos lineales generalizados						
	Log ₁₀ $DI_{50\%}$	$DI_{50\%}$	AIC	Log ₁₀ $DD_{50\%}$	$DD_{50\%}$	AIC
Regresión logística						
GLM. Binomial. Logit	3.40	2508		4.33	21434	
Online tool Binomial. Logit <i>Curve fit</i>	3.41	2559	10.67	4.28	18836	
R. L. Bootstrap paramétrico	3.40	2486		4.32	21047	17.95
Modelos de Regresión no lineal						
Regresión Log.-Logística 2P L.L.						
R drc - Binomial 2P.L.L.	3.40	2508	10.67	4.34	21821	17.95
Exponencial Mecanicista						
Script Weir 2017	3.76	5707	5.75	4.76	57544	18.11
Beta-Poisson aproximado Mecanicista						
Script Weir 2017	3.41	2603	5.75	4.33	21360	14.58
Online tool Log-log. Model	3.76 [#]	5700 [#]		rechazo		
Métodos aplicados sobre lossobre variable continuas AUC excreción viral y severidad de la diarrea						
Modelos de Regresión no lineal						
Log-logística 3P L.L.						
R drc variable logAUC S ó AUC D ~ DOSIS	3.80	7203	108	4.47	29464	153

[#] calculado con 7 grupos

En resumen, se destaca que todos los métodos utilizados arrojaron valores en el mismo orden. La dosis infecciosa 50% estimada a partir de los datos de proporciones de animales afectados sobre expuestos estuvo entre 2198 y 5707 copias de genoma, con una moda de 2500 copias. El valor obtenido a partir de la variable continua fue aún más extremo, con 7203 copias, pero en el mismo orden en la escala logarítmica, que es la adecuada para comparar dosis virales. Por su parte, la dosis diarrea 50% fue siempre un orden mayor que la de infección y estuvo alrededor de las 21000 copias de genoma de NoV, con valores en el mismo orden para la estimación basada en la variable continua y con solo un valor muy extremo para el modelo Exponencial, que tuvo una baja bondad de ajuste, y que coincide con el valor de SK obtenido con la herramienta *online* del NCBI. En función del valor de AIC, los modelos con mejor ajuste fueron los modelos no lineales mecanicistas, Exponencial y Beta-Poisson para infección, y Beta-Poisson para enfermedad.

Al estudiar comparativamente el ajuste de cada una de las curvas a los datos de proporciones de animales afectados sobre expuestos (**Figuras 29 y 30**) se observa que, para ambos conjuntos de datos, las curvas de dosis-respuesta obtenidas por la regresión logística y la regresión 2P L.L. se superponen completamente, mientras que las curvas del modelo mecanicista Beta-Poisson y más aún del modelo Exponencial se alejan hacia la derecha (Figura 29). En el caso de los datos de enfermedad todos los modelos se superponen, excepto el modelo exponencial (Figura 30).

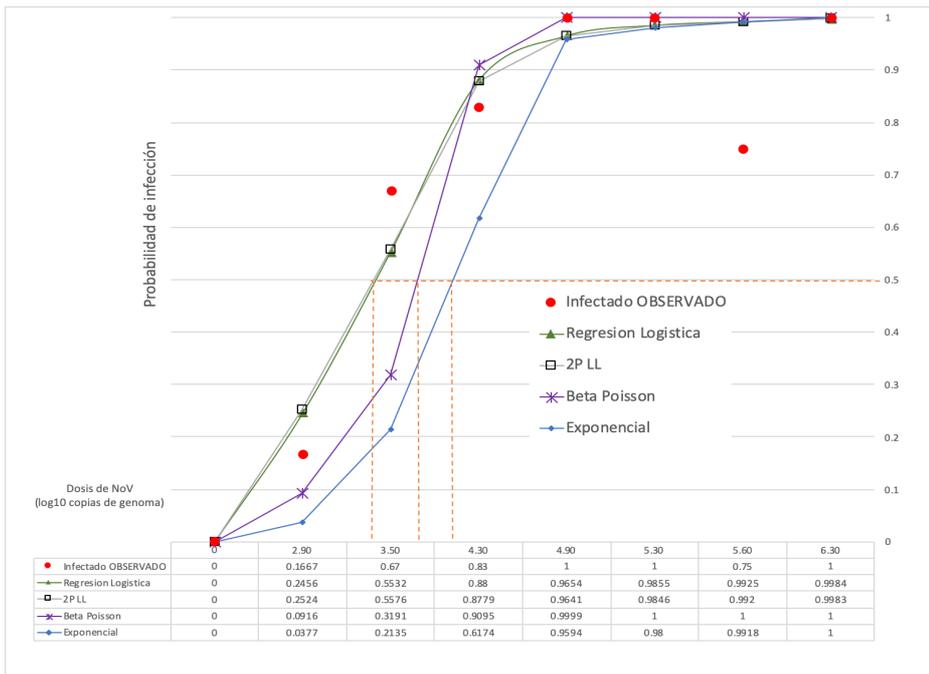


Figura 29. Comparación de los diferentes modelos para los datos de infección.

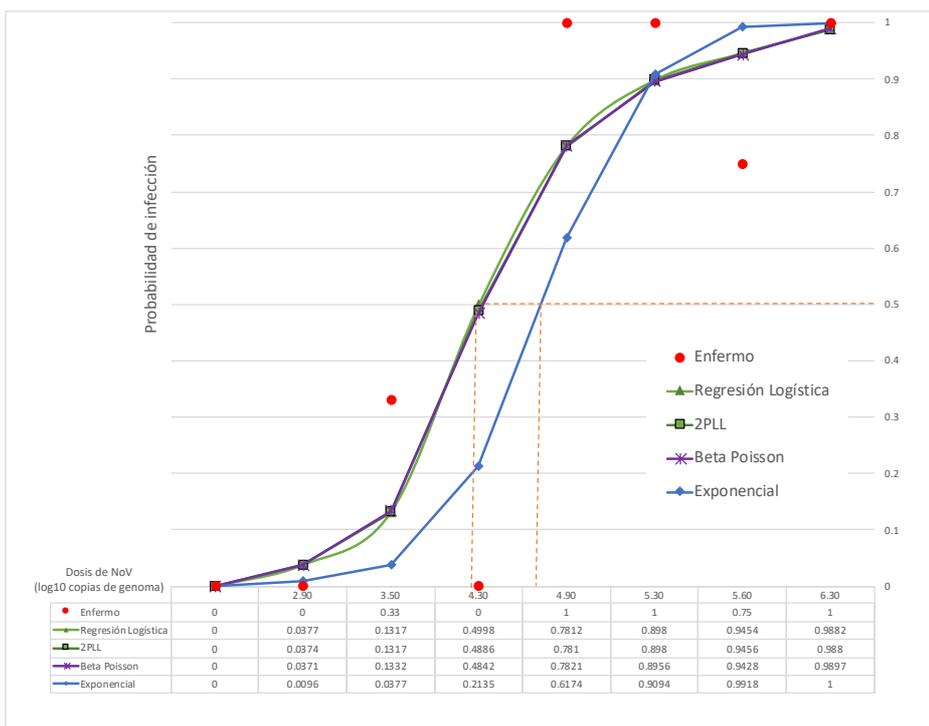


Figura 30. Comparación de los diferentes modelos para los datos de enfermedad.

4.8. Aplicación práctica: dosis óptima de desafío del modelo cerdo gnotobiótico para NoV humano

En esta sección se interpretan los resultados obtenidos luego del análisis de los datos del ensayo dosis-respuesta como punto de partida para estandarización de la dosis inóculo a utilizar en el modelo cerdo gnotobiótico para que el mismo posea validez preclínica. En este sentido, se debe comparar los resultados con los obtenidos en otros ensayos en voluntarios humanos y con el mismo modelo animal.

Los resultados obtenidos en el ensayo que se analizó en este trabajo de tesis muestran que los cerdos gnotobióticos infectados con una dosis de Cin-2 de $2,5 \times 10^3$ copias de genoma viral (grupo C) comienzan la excreción viral a los 1.3 días post inoculación. El total de los cerdos enrolados en el grupo se infectan y excretan grandes cantidades de virus en heces (media \log_{10} AUCS: 5.05) durante casi todo el período de estudio (en promedio, de 6.3 de 7 días). Además, los cerdos en este grupo también experimentaron la diarrea más severa comenzando a partir de los 2.8 días y por 5 a 7 días.

Si bien no se cuenta con los datos crudos de los ensayos realizados en voluntarios humanos (Frenck et al. 2012; Bernstein et al. 2015), cuando se compara el porcentaje de individuos afectados, la duración media y el día de excreción pico reportados en tres de ellos, desafiados con las mismas dosis que la utilizadas en el ensayo en cerdos gnotobióticos, se observan resultados semejantes. En particular, el grupo de cerdos infectados con 2×10^5 de la cepa Cin-2 se asemeja a los resultados obtenidos en humanos en cuanto a lograr una duración promedio similar (6.3 días vs 5.2 días), un día pico también similar (4 vs 3 días). Esta dosis afectó al 100% de los animales vs 70% de las personas. Para las dosis siguientes, 8×10^4 y 2×10^4 copias de genoma viral, si bien se mantiene el porcentaje de individuos afectados en valores similares a lo que ocurre con humanos, la duración promedio de la excreción de 3 y 1 días, respectivamente, podría hacer peligrar los ensayos, siendo un desafío leve para el caso de pretender evaluar una vacuna (**Tabla R25**).

Tabla R 25. Comparación de infección y enfermedad en por NoV en cerdos gnotobióticos versus voluntarios humanos.

Hospedador	Edad (años/días)	Dosis desafío	n	% Excreción viral	Duración promedio días (rango) ^c	Día pico de excreción (PID)
Humano ^a	19-48 años	5.00E+04	23	70	5.2 (2-30)	3
Humano ^b	18-50 años	4.40E+03	34	76.5	-	-
Humano ^b	18-50 años	4.40E+03	48	62.5	-	-
Cerdo Gn ^c	33-34 días	2.00E+04	3	67	1.0 (1-2)	2
Cerdo Gn ^c	33-34 días	8.00E+04	6	100	2.8 (2-4)	3
Cerdo Gn ^c	33-34 días	2.00E+05	4	100	6.3 (5-7)	4

- a. Datos reportados por Frenck et al. 2012 [32].
- b. Datos reportados Bernstein et al. 2015 [7].
- c. Excreción viral en muestras de los pacientes controlados por 30 días; y en cerdos Gn por 7 días post inoculación. En este trabajo.

En base a las proporciones de animales infectados y enfermos para cada dosis de inóculo evaluada en este ensayo dosis-respuesta se recomienda que para reproducir la infección y diarrea por NoV humanos en cerdos gnotobióticos con la cepa de NoV humano Cin-2, una variante de la cepa Farmington Hills de NoV, la dosis óptima de desafío a elegir sería de 2×10^5 copias de genoma viral, dado que esta dosis logra períodos de excreción y diarrea homologables a lo que ocurre en voluntarios humanos. En términos de la dosis infecciosa 50% que acabamos de calcular, se necesitan inocular a los cerdos con 78 veces la $DI_{50\%}$ (**Tabla R26**).

Tabla R 26. Comparación de la dosis desafío de cerdos gnotobióticos de dos cepas de NoV humano.

Cepa de NoV humano	Dosis óptima de desafío oral de cerdos gn (copias de genoma viral)	$DI_{50\%}$	Desafío en $DI_{50\%}$	Referencia
GII.4/2003 Cin-2	2.0×10^5	2.57×10^3 (Beta Poisson)	78 x $DI_{50\%}$	Este estudio
GII.4/2006b Cin-1	6.43×10^5	6.4×10^4 (Reed Muench)	10 x $DI_{50\%}$	Bui, 2013
Norwalk GI. 1 Humanos adultos (18-50 años de edad)	1.32×10^3	2.8×10^3	0.47 x $DI_{50\%}$	Atmar, 2014

Cin-2 ajustada en este estudio vs. la cepa previamente utilizada en la reproducción de la infección por NoV Cin-1, en el laboratorio de la Dra. Yuan. Politécnico de Virginia, USA.

La dosis 50% de la cepa Cin-2 fue menor a la calculada para la cepa utilizada previamente Cin-1, GII.4/2006b, otra variante del virus Farmington Hills, cuya mediana de la dosis infecciosa fue de 6.43×10^4 copias de genoma viral utilizando el método RM (**Tabla R26**) (Bui et al. 2013). La variante Cin-1 de HuNoV utilizada en el modelo cerdo gnotobiótico del estudio de Bui et al. (2013) causaba enfermedades entre la población pediátrica en guarderías y hospitales (Chen, 2015; Kumazaki, 2015), y fue responsable de alrededor del 64% de los brotes en cruceros y del 45% de los brotes en tierra, la mayoría de los cuales ocurrieron en centros de atención a largo plazo, escuelas y restaurantes en los EE. UU. en 2002 (Isakbaeva, 2005; Bull, 2006; Johnston, 2007).

La comparación de ambas cepas indica que se requiere un título de virus al menos 25 veces mayor para establecer la infección en el 50% de los cerdos inoculados con el virus Cin-1 en comparación con el Cin-2. De lo expuesto se deduce que Cin-2 es una cepa más infecciosa en

cerdos gnotobióticos según la información disponible sobre las dos variantes de NoV humano. Este es un dato muy importante a la hora de llevar a adelante los experimentos en cerdos gnotobióticos, ya que permite que el stock de materia fecal humana sea utilizado en un número mayor de animales permitiendo alcanzar un número de unidades experimentales estadísticamente significativos, así como la repetición de ensayos que garanticen la reproducibilidad de los estudios de eficacia que se realicen.

A su vez, la dosis infecciosa 50% de Cin-2 en cerdos resultó muy similar a la calculada en voluntarios humanos desafiados con la cepa de NoV Norwalk (**Tabla R26**). Este resultado es alentador respecto de la correcta estandarización del modelo cerdo como un óptimo modelo preclínico considerando que NoV humano es un virus muy difícil de replicar en dosis elevadas en modelos animales (Bui, 2013; Souza et al. 2008, 2007; Todd and Tripp 2019).

5. DISCUSIÓN

El presente trabajo de tesis abordó la estimación de la dosis infecciosa y dosis enfermedad 50% para reproducir la infección y cuadro clínico de diarrea causada por Norovirus humano en el modelo cerdo gnotobiótico, único modelo animal capaz de reproducir la infección y enfermedad por este agente viral y, por lo tanto, con un elevado valor preclínico para el estudio de Norovirus (Lei et al. 2019; Yang and Yuan 2014). A partir de estas estimaciones, se estableció la dosis óptima de desafío viral para reproducir la infección y el cuadro clínico de diarrea para emular los resultados que se obtienen en estudios con voluntarios humanos (Frenck et al. 2012; Atmar and Estes 2012; Bernstein et al. 2015).

Desde la primera visualización de NoV por microscopía electrónica (ME) realizadas por Albert Kapikian en la década del '70, se observó que en las muestras obtenidas de filtrados de heces sin ningún otro procesamiento, las partículas virales forman acúmulos (Teunis et al. 2008; McBride 2014). Un estudio reciente mostró que grandes fracciones de norovirus presentes en muestras fecales residían en el interior vesículas envueltas por una membrana que era difícil de disgregar aun por sonicación. Aparentemente, estas vesículas aumentan la estabilidad del virus dentro del tracto gastrointestinal y aumentan por lo tanto su infectividad (Teunis et al. 2008; Messner, Berger, and Nappier 2014). En este sentido, en nuestro estudio, el análisis de excreción viral de los cerdos desafiados, día por día, permitió detectar tres animales desafiados con las dosis virales más bajas que excretaron elevadas cantidades de virus. Asimismo, en las dosis más elevadas se registraron cargas virales menores que en los cerdos que recibieron la tercera dosis. La variabilidad observada en estos resultados podría ser tan sólo un efecto aleatorio o deberse al efecto de agregación que ha sido previamente mencionado en la bibliografía de NoV. Por lo tanto, antes de realizar los análisis propuestos en este trabajo de tesis era de suma importancia corroborar el grado de agregación del virus en nuestro inóculo, dado que todos los análisis se basaban en cumplir el supuesto de que el virus presente en el inóculo administrado a los cerdos se encuentra desagregado y sus partículas virales debían encontrarse homogéneamente distribuidas siguiendo una distribución de Poisson. Esta hipótesis fue corroborada al observar un ajuste significativo al realizar una regresión de Poisson entre el número de viriones contados por microscopía electrónica en relación con el título viral medido en copias de genoma en cada dosis. Dicho ensayo demostró que el tamaño de los acúmulos virales se redujo con el proceso de dilución. Por lo tanto, y si bien no podemos descartar completamente el fenómeno de agregación, que tal vez esté presente y será profundizado en estudios futuros en el modelo cerdo gnotobiótico, consideramos que en este ensayo dosis respuesta se cumplen los supuestos necesarios para la aplicación de los modelos mecanicistas. Con base en este método de cálculo, estimamos aquí que, una dosis de 2×10^5 copias de ARN contenía alrededor de 513 partículas de virus infecciosos en 100 μ l de inóculo de NoV Cin-2, lo que se traduce en $5,128 \times 10^3$ partículas infecciosas /ml (Messner, 2014; Schmidt, 2015).

La estimación de la dosis infecciosa y dosis enfermedad 50% partió entonces de datos de la proporción de animales infectados y enfermos para cada dosis y comparó diferentes métodos matemáticos y modelos estadísticos de estimación. Se inició con los métodos clásicos de interpolación matemática, pasando por métodos de regresión lineal generalizada, para luego ir aumentando el grado de complejidad del análisis hasta llegar a los modelos mecanicistas aplicados en los estudios de análisis de riesgo para Norovirus y otros enteropatógenos en humanos (Philip J. Schmidt and Emelko 2010; Haas, Rose, and Gerba 2014; Philip J. Schmidt 2015; Chaudhry et al. 2017). Para llevar a cabo los análisis se recurrió a un gran abanico de herramientas estadísticas: paquetes del lenguaje R, el software Infostat y herramientas *on line*. Como novedad, se presentó también la estimación de la dosis infecciosa y de diarrea 50% a partir del análisis estadístico de las variables continuas de área bajo la curva de excreción viral y score fecal, bajo un modelo de regresión log-logística de 3 parámetros, obteniendo resultados en el mismo orden de los obtenidos a partir de las variables binarias.

Antes de comenzar la discusión de cada método es importante destacar que al momento de realizar el análisis el experimento ya estaba concluido y no se podían incluir más animales. El número de cerdos gnotobióticos empleados en los ensayos de dosis-respuesta como el analizado en la presente tesis es ciertamente acotado hubiera sido deseable al menos balancear los grupos completando a 4 animales por dosis. En términos generales, cuando el número de unidades experimentales es acotado puede resentirse la precisión de la estimación de los parámetros del modelo. Por otro, la naturaleza de estos estudios impide en general aumentar el tamaño muestral dado que se solicita un equilibrio entre la significancia estadística del ensayo, la potencia y respetar la reducción al mínimo de animales experimentales para cumplir con las normas de bienestar animal requeridas por los comités de ética. Por otro lado, el costo de cada cerdo asciende a 12000 dolares cada cerda preñada y la mantención de los animales en las burbujas estériles se estima entre 1700 y 2000 dolares por cerdo, lo que hace que estos ensayos sean altamente costosos (Yuan, comunicación personal).

La repetición sistemática de estas experiencias es la que permitirá, a más largo plazo, confirmar, o no, la validez de las estimaciones.” Pero los múltiples ensayos consumen el inoculo que debe usarse para hacer los desafíos de pruebas preclínicas de vacunas, antivirales y anticuerpos.

Comenzando la discusión de los métodos clásicos, independientemente de su uso universal, los métodos RM y DB dependen en gran medida de una progresión secuencial de respuestas, ya que utilizan sólo información de dos puntos alrededor del título potencial pasando por alto por completo cualquier distribución aleatoria ó valor atípico dentro de los datos (Spearman 1908; Atkinson 1961; Kä r ber 1931; Ramakrishnan 2016; Siev 2018). Si bien, ha habido un debate continuo entre la comunidad de investigadores y estadísticos respecto de la confiabilidad del uso de estos métodos, está bien demostrado que, para rangos de dosis igualmente espaciados en escala logarítmica, arrojan resultados confiables. Esto los convierte en herramientas valiosas (particularmente debido a la facilidad de la metodología y la disponibilidad de tablas) que permiten realizar un cálculo rápido y / o aproximado de los títulos virales y las dosis infecciosas, letales o enfermedad 50% para profesionales no estadísticos que trabajan habitualmente en un laboratorio virológico (Siev 2018; Ramakrishnan 2016). Cuando se manejan niveles de dosis espaciados secuencialmente como los de este estudio, se ha descrito que RM y DB son altamente confiables y se sabe que brindan una alta tasa de cobertura (> IC del 95%), pero por otro lado cuando los datos están distribuidos más aleatoriamente, la realización del método RM y DB podría complicar la inferencia por inexactitudes, cuestionando la validez interna de los datos y resultados (Siev, 2018).

Por su parte, el método de Spearman-Kärber (SK) es un método no paramétrico que se puede calcular fácilmente a mano y produce estimaciones más precisas de la dosis infecciosa medias calculadas por los métodos RM y DB. Se basa en la naturaleza simétrica de los datos de dosis-respuesta para calcular la dosis infecciosa media (no la mediana), lo que significa que requiere una dosis con respuestas que abarque todo el rango de resultados del 0% al 100% y una curva cuya simetría permita que la mediana coincida con la media (Ramakrishnan 2016; Siev 2018; Atkinson 1961). Algunos estudios de dosis-respuesta no pueden generar datos que cumplan con ese criterio. En el presente estudio, esto se puede corroborar claramente con los datos de infección donde todos los grupos excretaron virus. Al utilizar sólo los 7 grupos desafiados, el valor estimado del $\log_{10} DI_{50\%}$ fue 3.520 y al agregar el grupo control negativo este valor se redujo a 3.342, que fue el tomado por correcto e informado en esta tesis. Por su parte, el método SK sí fue apropiado para calcular la $DD_{50\%}$, ya que los grupos de dosis contenían el rango completo de respuesta (100% a 0%) de los cerdos enfermos y el agregado del grupo control no modificó el resultado ($\log_{10} DD_{50\%}$: 4.492).

En décadas recientes, la disponibilidad de mayor capacidad de cómputo ha permitido la aplicación de modelos estadísticos para la estimación de las curvas de dosis-respuesta. Estos modelos, de base probabilística, son más precisos y robustos que los métodos no paramétricos (Bolker et al. 2009; Lopez-Gonzalez and Ruiz-Soler 2011). Nuestro estudio resalta el valor de las diferentes regresiones con forma sigmoideas como la regresión logística y log-logística de dos parámetros, que logran una adecuada bondad de ajuste a los datos de proporciones de animales afectados sobre expuestos, resultando en dosis estimadas del mismo orden que los modelos clásicos y los mecanicistas (Beckon et al. 2008; Ritz et al. 2015).

Los modelos basados en regresión logística han sido muy utilizados para los cálculos de dosis-respuesta por su precisión y robustez en comparación con los métodos no paramétricos (Han et al. 2018). En relación a Norovirus específicamente, este modelo fue utilizado para el cálculo de dosis infecciosa en voluntarios expuestos a la cepa Norwalk original (cepa GI.1, 8fIIa), donde se obtuvo una $DI_{50\%}$ de 2.8×10^3 copias de genoma viral (Philip J. Schmidt 2015). De manera similar, en el estudio actual determinamos que la $DI_{50\%}$ de la cepa Cin-2 es de alrededor de 2.5×10^3 copias de genoma. Por su parte, a partir de una dosis con 8×10^4 copias de genoma ya se manifestó un cuadro significativo de severidad y duración de la diarrea y la dosis con 2×10^5 copias de genoma de Cin-2 fue la dosis que generó mayor respuesta en términos de intensidad y duración de la infección y sería, por lo tanto, la dosis de elección para la evaluación de eficacia de vacunas y otros tratamientos. Estos datos muestran un potencial de infección similar entre la cepa prototipo original de NoV GI.1 Norwalk y la cepa Cin-2, variante Farmington Hills del genogrupo II.4, aquí estudiada.

En otro estudio clínico en humanos, la administración de una dosis de 5×10^4 copias de genoma viral de la cepa de Nov de Cin-1 causó infección en el 70% de los individuos secretores positivos (Frenck et al. 2012). Un estudio piloto en voluntarios humanos llevado a cabo por Bernstein y colegas utilizando una dosis de desafío de 4.4×10^3 copias de genoma mostró tasas de infección (26/34 [76,5%]) y enfermedad (19 / 34 [55,9%]) similares a las obtenidas por (Frenck et al. 2012) al inocular 5×10^4 copias de genoma. Sin embargo, en un estudio posterior Fase I de prueba de eficacia de vacunas esta dosis (4.4×10^3 copias de genoma) fue demasiado baja y causó infección solo en el 62.5% y enfermedad (vómitos o diarrea de gravedad leve o mayor) en el 37.5% de los voluntarios del grupo de control con placebo (Bernstein et al. 2015). Estos datos fueron comparables con la incidencia ligeramente reducida de infección (54%) y enfermedad (20%) entre los receptores de la vacuna, lo que condujo a un efecto protector no concluyente de la vacuna debido a la falta de significación estadística. De manera similar, en el presente estudio, en el grupo de dosis E (2×10^4 copias de genoma), solo el 67% de los cerdos excretaron el virus y el 33% tuvieron diarrea, mientras que el 100% de los cerdos en los grupos de dosis entre 2×10^6 y 8×10^4 copias de genoma excretaron virus y entre el 75% y el 100% sufrió diarrea. Estos resultados en cerdos gnotobióticos coinciden con los resultados de Frenck y Bernstein en humanos. En conjunto, todos los datos expuestos destacan la importancia de identificar la dosis infecciosa 50% de cada inóculo, así como la dosis óptima de desafío viral para la evaluación de la eficacia protectora de las vacunas en modelos animales y en humanos.

Nuestro trabajo también deja de manifiesto que actualmente se cuenta con una gran cantidad de recursos biológicos, matemáticos y computacionales; y que en estos ensayos de dosis-respuesta se necesita un equipo multidisciplinario de virólogos, biólogos moleculares, veterinarios, médicos y profesionales expertos en biometría para una correcta supervisión del diseño experimental antes de realizar un ensayo complejo y costoso que involucre modelos animales y, más aún, voluntarios humanos (Philip J. Schmidt 2015; Philip J. Schmidt, Emelko, and Thompson 2020). A futuro, se resalta la necesidad de contar con estudios más profundos del grado de agregación de los inóculos y ensayos dosis-respuesta paralelos en cerdos gnotobióticos y en humanos, lo que permitirá robustecer la estandarización del modelo preclínico y llegar a

ensayos en humanos con cepas virales mejor caracterizadas, de manera de reducir al mínimo el riesgo de fracaso de las fases clínicas por falta de significancia estadística entre grupos tratados y controles.

Volviendo a la discusión del uso de diferentes modelos estadísticos, una gran ventaja de los modelos de dosis-respuesta de regresión logística, log-logística, Exponencial y Beta-Poisson respecto a los métodos clásicos utilizados en el cálculo de dosis efectivas, es que estos modelos estadísticos, más que generar una estimación puntual de las $DI_{50\%}$ y $DD_{50\%}$, pueden describir el rango completo de probabilidad de respuesta ($DI_{90\%}$ o $DI_{10\%}$ por ejemplo), también son precisos en dosis bajas (Thakur and Fezio 1981), poseen una mayor flexibilidad y una mayor comprensión de la probabilidad estimada de infección permitiendo analizar datos no equidistantes y desbalanceados. Tanto el modelo Exponencial como el Beta-Poisson se utilizan habitualmente para la determinación cuantitativa de las relaciones dosis-respuesta (Lei et al. 2019; Weir et al. 2017; Philip J. Schmidt 2015; Philip J. Schmidt et al. 2013). La ecuación exponencial tiene en cuenta la naturaleza simétrica de la curva alrededor del punto de inflexión al calcular la dosis mediana, al tiempo que reduce los efectos de variabilidad, debido a las grandes diferencias entre las dosis utilizadas. El estudio aquí realizado confirma el alto grado de ajuste de los modelos mecanicistas Exponencial y Beta-Poisson, previamente utilizado para el cálculo de la $DI_{50\%}$ de diferentes patógenos entéricos sobre nuestros datos de infección (Weir et al. 2017) (<http://qmrawiki.org>).

El criterio de información de Akaike, AIC, que puede ser calculado para todos los modelos estadísticos empleados, es una estimación de la distancia relativa esperada entre el modelo ajustado y el mecanismo verdadero desconocido que generó realmente los datos observados (Akaike 1974) y permite comparar los diferentes modelos utilizados para un mismo conjunto de datos. Es el presente estudio nos permitió demostrar la bondad de ajuste de los dos modelos mecanicistas como los más adecuados para estimar la $DI_{50\%}$, mientras que el modelo Beta-Poisson se impuso frente al resto de las estimaciones resultando el modelo de elección para la estimación de $DD_{50\%}$.

Asimismo, el uso de bootstrap paramétrico se destaca como un método adecuado para la estimación de la incertidumbre de las estimaciones de los parámetros de los modelos, así como de los valores predichos para toda la curva dosis-respuesta (Weir, 2017). En nuestro estudio, arrojó intervalos de similar amplitud que los generados por un método de muestreo basado en cadenas de Markov Montecarlo.

Por último, el ajuste de un modelo log-logístico de 3 parámetros a los datos de AUC bajo la excreción viral y el *score fecal* de diarrea, permitió estimar la $DI_{50\%}$ y $DD_{50\%}$ y arrojó resultados en el mismo orden que los modelos aplicados a las variables discretas. La cercanía de los valores obtenidos sugiere la fiabilidad de esta alternativa como un enfoque novedoso para abordar el cálculo de dosis infecciosas y enfermedad 50%. Los datos aquí analizados poseen la limitante de que el parámetro a estimar correspondiente al punto de inflexión de la curva, y que corresponde biológicamente a la dosis 50%, no resultó estadísticamente significativo. Sin embargo, los valores obtenidos fueron comparables a los obtenidos a partir de las variables discretas y se corroboró el cumplimiento de los supuestos del modelo. Este es un enfoque innovador que abre un camino de análisis interesante para continuar explorando con nuevas bases de datos de ensayos dosis-respuesta. Se espera que en próximos experimentos con NoV y otros agentes entéricos se confirme el valor de su utilización (Ritz, comunicación personal).

Es importante destacar que, si bien los enfoques de análisis de datos presentan diferentes supuestos, limitaciones y grados de complejidad estadística, la obtención de resultados dentro de un rango relativamente acotado de valores para ambos parámetros ($\log_{10} DI_{50\%}$: 3.39 a 3.35; $\log_{10} DD_{50\%}$: 4.32 a 4.76) indicaría que todos los métodos utilizados funcionan

adecuadamente para este conjunto de datos en particular, a pesar de la falta de equidistancia entre dosis y grupos con número de unidades experimentales desbalanceadas. Este resultado, si bien debe ser corroborado con nuevos conjuntos de datos y otros agentes virales, para explorar escenarios en los que los enfoques clásicos pueden volverse poco confiables y compararlos con enfoques contemporáneos estadísticamente más rigurosos. Sin embargo, le otorga cierto grado de validez a todos los enfoques, en especial a los métodos clásicos que aún se siguen utilizando de rutina para titular suspensiones virales en los laboratorios de virología, sin necesitar cálculos complejos ni recurrir a un análisis estadístico. Asimismo, en este trabajo de tesis se destaca el adecuado desempeño de la herramienta *online* desarrollada por el NCBI, que es muy versátil y arrojó muy buenos resultados a pesar que haber trabajado con grupos desbalanceados y no equidistantes. Es importante destacar al utilizar este tipo de herramienta seguir las recomendaciones del diseñador en cuanto a la carga de datos y el cumplimiento de supuesto (por ejemplo, el ingreso de dosis crudas, contar con datos completo de 0 a 100% en caso de utilizar la opción SK, introducir los datos de menor a mayor, no introducir ceros, y seleccionar de las opciones correctas de análisis). El programa posee una función de prueba escalonada de modelos e indica cuando los datos cumplen los supuestos o cuando hay inconsistencia rechazando su ajuste con p-valores menores a 0.05.

En general, ambos enfoques a partir de variables discretas y continuas aquí aplicados demuestran su valor para los estudios de dosis-respuesta y pueden utilizarse por separado para simplemente determinar la mediana de las dosis infecciosas, o conjuntamente, para lograr una comprensión más amplia de la curva completa de dosis-respuesta y la severidad de la infección y enfermedad causada por cada dosis bajo estudio. Desde este punto de vista del estado del arte del tema abordado, consideramos que los resultados obtenidos en esta tesis representan un aporte valioso a los estudios dosis-respuesta de Norovirus en particular y de virus en general aplicados a la validación de modelos animales de importancia preclínica.

6. CONCLUSIONES

En este trabajo de tesis se ha realizado un análisis estadístico exhaustivo para estimar la dosis infecciosa 50% así como la dosis diarrea 50% a partir de los animales afectados de un ensayo dosis-respuesta. Proporcionamos aquí un ejemplo de cómo y dónde se pueden aplicar mejor los métodos clásicos de interpolación matemática y cómo emplear y comparar entre diferentes modelos estadísticos a partir de un criterio de bondad de ajuste para interpretar los fenómenos de dosis-respuesta y determinar con precisión la $DI_{50\%}$ y $DD_{50\%}$ de NoV humano.

Es de destacar como una contribución novedosa el análisis de las variables continuas de área bajo la curva de excreción y área bajo la curva de severidad de diarrea, que arrojaron valores similares a los obtenidos a partir de las variables binarias.

Este análisis en su conjunto permite estimar en forma más rigurosa la dosis a utilizar para la reproducción correcta de la infección y la enfermedad por NoV en el modelo cerdo gnotobiótico. El cerdo es una especie que por sus similitudes fisiológicas con los humanos son susceptibles a la infección por NoV humano. La infección desarrollada es de curso completo y enfermedad es muy similar a la que se presenta en los humanos. Nuestros estudios han demostrado que los cerdos gnotobióticos son una herramienta preclínica confiable para estudiar la infectividad de las cepas GII.4 de NoV humano, realizar ensayos de evaluación de la eficacia vacunas, estrategias de inmunidad pasiva y antivirales como candidatos profilácticos y terapéuticos antes de pasar a los ensayos clínicos en humano.

7. ANEXO

1. Resultados 4.4. Cálculo $DI_{50\%}$ y $DD_{50\%}$ métodos clásicos. Ejemplo de desarrollo del cálculo por R&M

En las Tablas A1 y A2 se detallan los cálculos previos necesarios para aplicar el método de R&M a los datos de cerdos infectados y con diarrea, donde previamente se realizan las sumas acumuladas de los animales afectados y no afectados y luego se aplican las formulas (1) y (2). El cálculo a mano se realizó a modo de ejemplo solo para este método.

Tabla A1. Infección con NoV: Cálculos previos para aplicar el método de Reed & Muench.

Grupo	Log (dosis)	N	INFECT.	NO INFECT.	INFECT. ACUM.	NO INFECT. ACUM.	PROP. ACUM.	% ACUM.
A	6.30	4	4	0	23	0	23/23	100.00
B	5.60	4	4	0	19	0	19/19	100.00
C	5.30	4	4	0	15	0	15/15	100.00
D	4.90	6	6	0	11	0	11/11	100.00
E	4.30	3	2	1	5	1	5/6	83.33
F	3.50	3	2	1	3	2	3/5	60.00
G	2.90	4	1	3	1	5	1/6	16.67

Referencias: **Acum.:** número acumulado de cerdos que excretaron virus (infectados);

Prop acum. proporción acumulada con excreción viral;

%ACUM.: porcentaje acumulado de cerdos con excreción viral.

$$DP = \frac{60 - 50}{60 - 16.67} = 0.230$$

$$\text{Log(punto final}_{50\%}) = 3.50 - 0.230 \times 0.303 = 3.43$$

Tabla A2. Diarrea por NoV: Cálculos previos para aplicar el método de Reed & Muench.

Grupo	NoV dosis	LOG dosis	Cerdos enfermos	Cerdos sanos	Enfermos acumulados	Sanos Acumulados	Proporción Enfermos acumulado	% enfermos ACUM.
A	2000000	6.301	4	0	18	0	18/18	100.00
B	400000	5.6021	3	1	14	1	14/15	93.33
C	200000	5.301	4	0	11	1	11/12	91.67
D	80000	4.9031	6	0	7	1	7/8	87.50
E	20000	4.301	0	3	1	4	1/5	20.00
F	3200	3.5051	1	2	1	6	1/7	14.29
G	800	2.9031	0	4	0	10	0/10	0.00

Referencias: **Acum.:** número acumulado de cerdos que presentan síntomas de diarrea (enfermos);

Prop acum. proporción acumulada con diarrea;

%ACUM.: porcentaje acumulado de cerdos con diarrea.

$$DP = \frac{87.50 - 50}{87.50 - 14.29} = 0.512$$

$$\text{Log}(\text{punto final}_{50\%}) = 4.9031 - 0.512 \times 0.303 = 4.747$$

2. Resultado 4.5. Intervalos de confianza para las curvas de dosis-respuesta de los modelos mecanicistas por el método de Cadenas de Markov Montecarlo, MCM

En este ANEXO se presentan los intervalos de confianza para las curvas de dosis-respuestas de los modelos mecanicistas Exponencial y Beta-Poisson obtenidas por muestreo de Cadenas de Markov Montecarlo (Dr Schmidt, comunación personal).

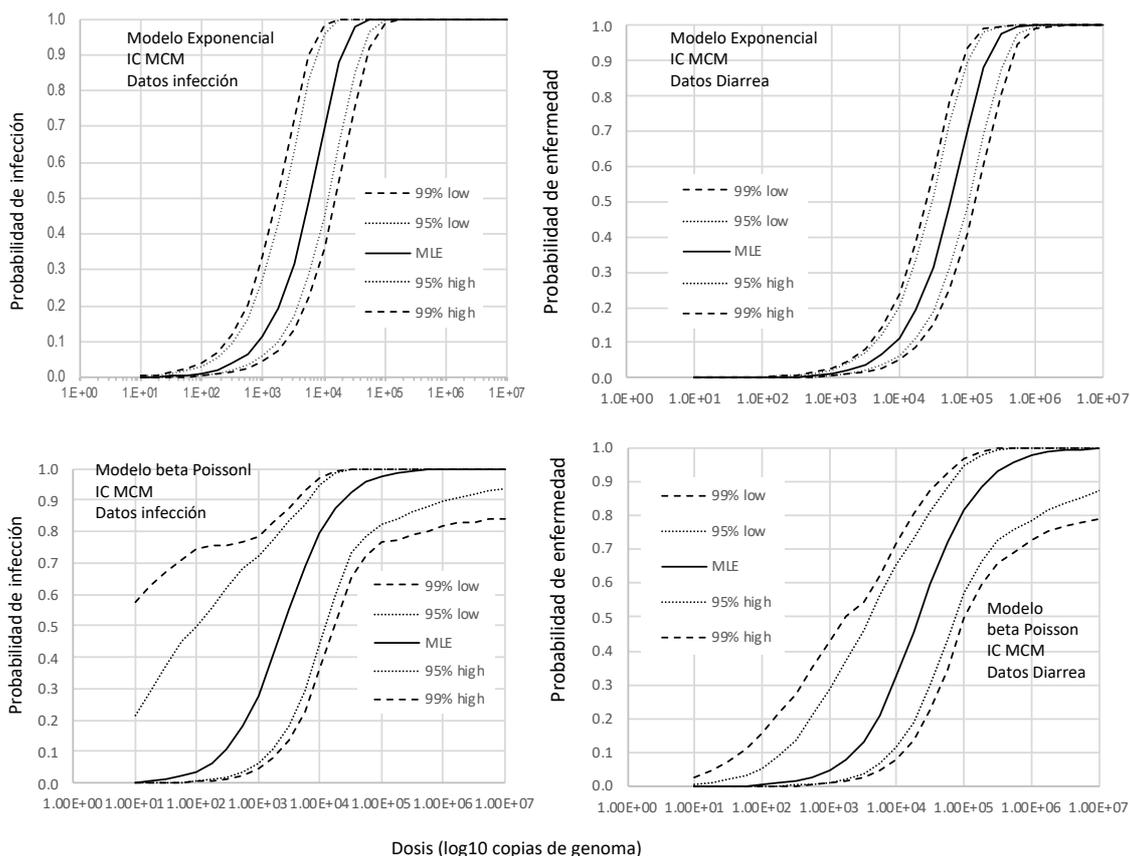


Figura A1. Curvas de dosis-respuesta ajustadas por los modelos Exponencial (a y b) y Beta-Poisson aproximado (c y d) a los datos de cerdos gnotobióticos infectados (a y c) y enfermos (b y d) luego del desafío con dosis crecientes de NoV humano. Los intervalos de confianza fueron calculados por el método de MCM (Dr Schmidt, comunación personal).

Al comparar con los intervalos obtenidos por la técnica de Bootstrap y los presentados en este ANEXO, se observa aquellos obtenidos por MCM son levemente más estrechos para los datos de infección pero más amplios en el caso de los datos de enfermedad. Ambos métodos resultan igualmente apropiados y, por lo tanto, en el cuerpo de la tesis se decidió presentar los intervalos de Bootstrap.

8. Bibliografía

- Akaike, Hirotugu. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control*. <https://doi.org/10.1109/TAC.1974.1100705>.
- . 1981. "Likelihood of a Model and Information Criteria." *Journal of Econometrics*. [https://doi.org/10.1016/0304-4076\(81\)90071-3](https://doi.org/10.1016/0304-4076(81)90071-3).
- Atkinson, G F. 1961. "The Spearman-Kärber Method of Estimating 50% Endpoints." *Biometrics Unit Technical Reports*.
- Atmar, Robert L., and Mary K. Estes. 2012. "Norovirus Vaccine Development: Next Steps." *Expert Review of Vaccines*. <https://doi.org/10.1586/erv.12.78>.
- Atmar, Robert L., Sasirekha Ramani, and Mary K. Estes. 2018. "Human Noroviruses: Recent Advances in a 50-Year History." *Current Opinion in Infectious Diseases*. <https://doi.org/10.1097/QCO.0000000000000476>.
- Bartsch, Sarah M., Benjamin A. Lopman, Sachiko Ozawa, Aron J. Hall, and Bruce Y. Lee. 2016. "Global Economic Burden of Norovirus Gastroenteritis." *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0151219>.
- Bates, Douglas, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v067.i01>.
- Beckon, William N., Cary Parkins, Alexey Maximovich, and Angela V. Beckon. 2008. "A General Approach to Modeling Biphasic Relationships." *Environmental Science and Technology*. <https://doi.org/10.1021/es071148m>.
- Bernstein, David I., Robert L. Atmar, G. Marshall Lyon, John J. Treanor, Wilbur H. Chen, Xi Jiang, Jan Vinjé, et al. 2015. "Norovirus Vaccine against Experimental Human GII.4 Virus Illness: A Challenge Study in Healthy Adults." *Journal of Infectious Diseases*. <https://doi.org/10.1093/infdis/jiu497>.
- Bok, Karin, Gabriel I Parra, Tanaji Mitra, Eugenio Abente, Charlene K Shaver, Denali Boon, Ronald Engle, et al. 2011. "Chimpanzees as an Animal Model for Human Norovirus Infection and Vaccine Development." *Proceedings of the National Academy of Sciences of the United States of America* 108 (1): 325–30. <https://doi.org/10.1073/pnas.1014577107>.
- Bolker, Benjamin M., Mollie E. Brooks, Connie J. Clark, Shane W. Geange, John R. Poulsen, M. Henry H. Stevens, and Jada-Simone S. White. 2009. "Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution." *Trends in Ecology & Evolution*. <https://doi.org/10.1016/j.tree.2008.10.008>.
- Bui, Tammy, Jacob Kocher, Yanru Li, Ke Wen, Guohua Li, Fangning Liu, Xingdong Yang, et al. 2013. "Median Infectious Dose of Human Norovirus GII.4 in Gnotobiotic Pigs Is Decreased by Simvastatin Treatment and Increased by Age." *Journal of General Virology*. <https://doi.org/10.1099/vir.0.054080-0>.
- Chambers, J., T. Hastie, and D. Pregibon. 1990. "Statistical Models in S." In *Compstat*. https://doi.org/10.1007/978-3-642-50096-1_48.
- Chaudhry, Rabia M., Kerry A. Hamilton, Charles N. Haas, and Kara L. Nelson. 2017. "Drivers of

- Microbial Risk for Direct Potable Reuse and de Facto Reuse Treatment Schemes: The Impacts of Source Water Quality and Blending." *International Journal of Environmental Research and Public Health*. <https://doi.org/10.3390/ijerph14060635>.
- Cheetham, S., M. Souza, R. McGregor, T. Meulia, Q. Wang, and L. J. Saif. 2007. "Binding Patterns of Human Norovirus-Like Particles to Buccal and Intestinal Tissues of Gnotobiotic Pigs in Relation to A/H Histo-Blood Group Antigen Expression." *Journal of Virology*. <https://doi.org/10.1128/jvi.01306-06>.
- Dragastedt, C.A., and V.F. Lang. 1928. "No Title." *Journal of Pharmacology and Experimental Therapeutics* 32 (3): 215–22.
- Frenck, Robert, David I. Bernstein, Ming Xia, Pengwei Huang, Weiming Zhong, Susan Parker, Michelle Dickey, Monica McNeal, and Xi Jiang. 2012. "Predicting Susceptibility to Norovirus GII.4 by Use of a Challenge Model Involving Humans." *Journal of Infectious Diseases*. <https://doi.org/10.1093/infdis/jis514>.
- Grytdal, Scott P., Emilio DeBess, Lore E. Lee, David Blythe, Patricia Ryan, Christianne Biggs, Miriam Cameron, Mark Schmidt, Umesh D. Parashar, and Aron J. Hall. 2016. "Incidence of Norovirus and Other Viral Pathogens That Cause Acute Gastroenteritis (AGE) among Kaiser Permanente Member Populations in the United States, 2012-2013." *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0148395>.
- Haas, Charles N., Joan B. Rose, and Charles P. Gerba. 2014. *Quantitative Microbial Risk Assessment: Second Edition*. *Quantitative Microbial Risk Assessment: Second Edition*. <https://doi.org/10.1002/9781118910030>.
- Hammersley, J.M., and D.C. Handscomb. 1975. *Monte Carlo Methods*. Edited by (John Wiley. New York.
- Han, Ningning, Zishu Chen, Hu Wan, Guohua Huang, Jianhong Li, and Byung Rae Jin. 2018. "A Rapid Method for Titration of Ascovirus Infectivity." *Journal of Virological Methods*. <https://doi.org/10.1016/j.jviromet.2018.02.011>.
- Iturriza-Gómara, Miren, and Benjamin Lopman. 2014. "Norovirus in Healthcare Settings." *Current Opinion in Infectious Diseases*. <https://doi.org/10.1097/QCO.0000000000000094>.
- Kä r ber, G. 1931. "Beitrag Zur Kollektiven Behandlung Pharmakologis- Cher Reihenversuche [A Contribution to the Collective Treatment of A." *Pharmacological Experimental Series*. *Archiv Für Experimentelle Pathologie Und Pharmakologie* 162: 480–83.
- Lee, Y., and J. A. Nelder. 1996. "Hierarchical Generalized Linear Models." *Journal of the Royal Statistical Society: Series B (Methodological)*. <https://doi.org/10.1111/j.2517-6161.1996.tb02105.x>.
- Lei, Shaohua, Junghyun Ryu, Ke Wen, Erica Twitchell, Tammy Bui, Ashwin Ramesh, Mariah Weiss, et al. 2016a. "Increased and Prolonged Human Norovirus Infection in RAG2/IL2RG Deficient Gnotobiotic Pigs with Severe Combined Immunodeficiency." *Scientific Reports*. <https://doi.org/10.1038/srep25222>.
- . 2016b. "Increased and Prolonged Human Norovirus Infection in RAG2/IL2RG Deficient Gnotobiotic Pigs with Severe Combined Immunodeficiency." *Scientific Reports* 6 (April): 25222.
- Lei, Shaohua, Erica L. Twitchell, Ashwin K. Ramesh, Tammy Bui, Elizabeth Majette, Christine M. Tin, Roger Avery, et al. 2019. "Enhanced GII.4 Human Norovirus Infection in Gnotobiotic Pigs Transplanted with a Human Gut Microbiota." *The Journal of General Virology*.

- <https://doi.org/10.1099/jgv.0.001336>.
- Literatur, :- Thema, B Efron, and R J Tibshirani. 2003. "Einführung in Bootstrap." *Statistical Science*.
- Lopez-Gonzalez, Emelina;, and Marcos Ruiz-Soler. 2011. "Análisis de Datos Con El Modelo Lineal Generalizado. Una Aplicación Con R." *Revista Española de Pedagogía* 248: 59–80.
- McBride, Graham B. 2014. "Norovirus Dose-Response in Sewage-Related QMRA: The Importance of Virus Aggregation." In *Proceedings - 7th International Congress on Environmental Modelling and Software: Bold Visions for Environmental Modeling, IEMSS 2014*.
- Messner, Michael J., Philip Berger, and Sharon P. Nappier. 2014. "Fractional Poisson - A Simple Dose-Response Model for Human Norovirus." *Risk Analysis*. <https://doi.org/10.1111/risa.12207>.
- Nelder, John;, and Robert Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society. Series A (General)*, no. 3: 370–384.
- Payne, Daniel C., Jan Vinjé, Peter G. Szilagyi, Kathryn M. Edwards, Mary Allen Staat, Geoffrey A. Weinberg, Caroline B. Hall, et al. 2013. "Norovirus and Medically Attended Gastroenteritis in U.S. Children." *New England Journal of Medicine*. <https://doi.org/10.1056/NEJMs1206589>.
- Ramakrishnan, Muthannan Andavar. 2016. "Determination of 50% Endpoint Titer Using a Simple Formula." *World Journal of Virology*. <https://doi.org/10.5501/wjv.v5.i2.85>.
- Reed, L. J., and H. Muench. 1938. "A Simple Method of Estimating Fifty per Cent Endpoints." *American Journal of Epidemiology*. <https://doi.org/10.1093/oxfordjournals.aje.a118408>.
- Ritz, Christian, Florent Baty, Jens C. Streibig, and Daniel Gerhard. 2015. "Dose-Response Analysis Using R." *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0146021>.
- Ritz, Christian, and Jens C Streibig. 2016. "Package 'Drc': Analysis of Dose-Response Curves." *R Project*.
- Ryan, Michael O., Charles N. Haas, Patrick L. Gurian, Charles P. Gerba, Brian M. Panzl, and Joan B. Rose. 2014. "Application of Quantitative Microbial Risk Assessment for Selection of Microbial Reduction Targets for Hard Surface Disinfectants." *American Journal of Infection Control*. <https://doi.org/10.1016/j.ajic.2014.07.024>.
- Schmidt, P. J., and M. B. Emelko. 2011. "QMRA and Decision-Making: Are We Handling Measurement Errors Associated with Pathogen Concentration Data Correctly?" *Water Research*. <https://doi.org/10.1016/j.watres.2010.08.042>.
- Schmidt, P. J., M. B. Emelko, and M. E. Thompson. 2013. "Analytical Recovery of Protozoan Enumeration Methods: Have Drinking Water QMRA Models Corrected or Created Bias?" *Water Research*. <https://doi.org/10.1016/j.watres.2013.02.001>.
- Schmidt, Philip J. 2015. "Norovirus Dose-Response: Are Currently Available Data Informative Enough to Determine How Susceptible Humans Are to Infection from a Single Virus?" *Risk Analysis*. <https://doi.org/10.1111/risa.12323>.
- Schmidt, Philip J., and Monica B. Emelko. 2010. "Using Pathogen Occurrence Data for Risk Analysis and Treatment Decisions: Preventing Bias by Addressing Measurement Error." In *Water Quality Technology Conference and Exposition 2010*.
- Schmidt, Philip J., Monica B. Emelko, and Park M. Reilly. 2010. "Quantification of Analytical

- Recovery in Particle and Microorganism Enumeration Methods." *Environmental Science and Technology*. <https://doi.org/10.1021/es902237f>.
- Schmidt, Philip J., Monica B. Emelko, and Mary E. Thompson. 2020. "Recognizing Structural Nonidentifiability: When Experiments Do Not Provide Information About Important Parameters and Misleading Models Can Still Have Great Fit." *Risk Analysis*. <https://doi.org/10.1111/risa.13386>.
- Schmidt, Philip J., Katarina D.M. Pintar, Aamir M. Fazil, and Edward Topp. 2013. "Harnessing the Theoretical Foundations of the Exponential and Beta-Poisson Dose-Response Models to Quantify Parameter Uncertainty Using Markov Chain Monte Carlo." *Risk Analysis*. <https://doi.org/10.1111/risa.12006>.
- Siev, D. 2018. "Package to Estimate ED50 by the Methods of Spearman-Kärber, Reed-Muench, Dragstedt-Behrens."
- Souza, M., S. M. Cheetham, M. S. P. Azevedo, V. Costantini, and L. J. Saif. 2007. "Cytokine and Antibody Responses in Gnotobiotic Pigs after Infection with Human Norovirus Genogroup II.4 (HS66 Strain)." *Journal of Virology*. <https://doi.org/10.1128/jvi.00558-07>.
- Souza, M, M S P Azevedo, K Jung, S Cheetham, and L J Saif. 2008. "Pathogenesis and Immune Responses in Gnotobiotic Calves after Infection with the Genogroup II.4-HS66 Strain of Human Norovirus ." *Journal of Virology* 82 (4): 1777–86. <https://doi.org/10.1128/JVI.01347-07>.
- Spearman, C. 1908. "The method of 'Right and wrong cases' ('Constant Stimuli') without Gauss's Formulae." *British Journal of Psychology* 2: 227–42.
- Spiegel, M.R. 1975. *Schaum's Outline Series: Probability and Statistics*. Edited by McGraw-Hill. New York.
- Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. <https://doi.org/10.1111/1467-9868.00353>.
- Spouge, J. L. 1992. "Statistical Analysis of Sparse Infection Data and Its Implications for Retroviral Treatment Trials in Primates." *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.89.16.7581>.
- Taube, Stefan, Abimbola O. Kolawole, Marina Höhne, John E. Wilkinson, Scott A. Handley, Jeffrey W. Perry, Larissa B. Thackray, Ramesh Akkina, and Christiane E. Wobus. 2013. "A Mouse Model for Human Norovirus." *MBio*. <https://doi.org/10.1128/mBio.00450-13>.
- Teunis, Peter F.M., Christine L. Moe, Pengbo Liu, Sara E. Miller, Lisa Lindesmith, Ralph S. Baric, Jacques Le Pendu, and Rebecca L. Calderon. 2008. "Norwalk Virus: How Infectious Is It?" *Journal of Medical Virology*. <https://doi.org/10.1002/jmv.21237>.
- Thakur, Ajit K., and W. Lane Fezio. 1981. "A Computer Program for Estimating Ld50 and Its Confidence Limits Using Modified Behrens-Reed-Muench Cumulant Method." *Drug and Chemical Toxicology*. <https://doi.org/10.3109/01480548109018136>.
- Todd, Kyle V., and Ralph A. Tripp. 2019. "Human Norovirus: Experimental Models of Infection." *Viruses*. <https://doi.org/10.3390/v11020151>.
- Vaart, A. W. van der. 2013. "Delta Method." In *Asymptotic Statistics*. <https://doi.org/10.1017/cbo9780511802256.004>.
- Weir, Mark H., Jade Mitchell, William Flynn, and Joanna M. Pope. 2017. "Development of a

Microbial Dose Response Visualization and Modelling Application for QMRA Modelers and Educators.” *Environmental Modelling and Software*.
<https://doi.org/10.1016/j.envsoft.2016.11.011>.

World Health Organization. 2016. *Quantitative Microbial Risk Assessment: Application for Water Safety Management*. WHO Press. <https://apps.who.int/iris/handle/10665/246195>.

Yang, Xingdong, and Lijuan Yuan. 2014. “Neonatal Gnotobiotic Pig Models for Studying Viral Pathogenesis, Immune Responses, and for Vaccine Evaluation.” *British Journal of Virology*.