Veterinary Parasitology 272 (2019) 83–94

Contents lists available at ScienceDirect

Veterinary Parasitology

journal homepage: www.elsevier.com/locate/vetpar

Research paper

Statistical analyses of chicken intestinal lesion scores in battery cage studies of anti-coccidial drugs

Qing Kang[a], Christopher I. Vahl[b,*], Huihao Fan[c], Thomas Geurden[d], Keith A. Ameiss[c], Lucas P. Taylor[c]

[a] The Statistical Intelligence Group LLC, Manhattan, KS, 66503, United States
[b] Department of Statistics, Kansas State University, Manhattan, KS, 66506, United States
[c] Zoetis, Veterinary Medicine Research and Development, Kalamazoo, MI, 49007, United States
[d] Zoetis, Veterinary Medicine Research and Development, Mercuriusstraat 20, 1930 Zaventem, Belgium

ARTICLE INFO

Keywords:
Battery cage study
Coccidiosis
Generalized linear mixed model
Lesion score

ABSTRACT

Establishing the efficacy of an anti-coccidial drug in poultry begins with conducting multiple battery cage studies, where the target animals are challenged with single and mixed *Eimeria* species inoculum under controlled laboratory conditions. One of the primary outcomes in a battery cage study is the intestinal lesion score defined on a discrete ordinal scale of 0 to 4. So far, the statistical analysis of lesion scores has routinely employed the linear mixed model (**LMM**). This present work proposes to apply the generalized linear mixed model (**GLMM**) with the cumulative logit link to statistically analyze coccidial lesion scores collected from battery cage studies. Upon applying this new approach on 9 datasets generated by challenging battery-cage-housed broilers with various mixtures of *Eimeria* species, it is observed that the GLMM fitted adequately to the data, produced variance component estimates that agreed with the experimental setup, and, at the 0.05 significance level, generated statistical results in complete concordance with the LMM approach. Advantages of the proposed GLMM over the LMM are discussed from several standpoints. Parallel to the regulatory requirement of a $\geq$1-unit reduction in the mean lesion score for clinical relevant efficacy under the LMM, the clinical relevancy criterion under the GLMM could be set as a $\geq$10-fold increase in the odds of having low lesion scores. That is, the effect of an anti-coccidial drug product would be deemed clinically relevant in battery-cage studies when the odds of having low lesion scores with the medication is 10 times or more than the odds without the medication.

1. Introduction

Coccidiosis is a major disease in poultry caused by the protozoan parasite *Eimeria.* Infection with *Eimeria* damages tissues and causes lesions in the intestines, leading to a reduction in body weight gain and feed efficiency. The annual loss to the poultry industry as a result of coccidiosis was estimated at over US$ 3 billion worldwide (Williams, 1999; Dalloul and Lillehoj, 2006). About 80% of the cost is due to impaired growth performance and about 20% is due to expenses associated with prophylaxis and treatment. Nine *Eimeria* species are known to infect chickens, but most of the economic impact is due to *E. acervulina*, *E. maxima*, and *E. tenella* (McDougald et al., 2008). These three species differ in infection sites and gross lesions: *E. acervulina* is found mainly in the upper small intestine and is characterized by white plaques on the mucosal surface; *E. maxima* is found primarily in the middle small intestine and results in red petechiae on the serosal surface, orange mucus and blood in the lumen; *E. tenella* infects the ceca and causes hemorrhagic lesions.

Prevention and treatment of poultry coccidiosis largely depends on the in-feed administration of anti-coccidial drugs and, to a lesser extent, vaccination (Dalloul and Lillehoj, 2006). Establishing the efficacy of an anti-coccidial drug is a multi-stage process that calls for battery cage studies and floor pen studies (Holdsworth et al., 2004). Battery cage studies are conducted to determine the effective dose range of the investigational new drug and/or to compare against existing drug(s) under controlled laboratory conditions. Floor pen studies, conducted at

Abbreviations: DOF, degrees of freedom; FDA-CVM, Food and Drug Administration Center of Veterinary Medicine; GLMM, generalized linear mixed model; INC, inoculated, non-medicated control; KR, Kenward-Roger; LMM, linear mixed model; NNC, non-inoculated, non-medicated control; OR, odds ratio

* Corresponding author at: Department of Statistics, Kansas State University, 101 Dickens Hall, Manhattan, KS, 66506-0802, United States.
E-mail addresses: kangqing@hotmail.com (Q. Kang), vahl@ksu.edu (C.I. Vahl), huihao.fan@zoetis.com (H. Fan), thomas.geurden@zoetis.com (T. Geurden), keith.ameiss@Zoetis.com (K.A. Ameiss), lucas.p.taylor@zoetis.com (L.P. Taylor).

https://doi.org/10.1016/j.vetpar.2018.12.002
Received 18 June 2018; Received in revised form 1 December 2018; Accepted 5 December 2018

single sites, serve to confirm the efficacy of the selected dose(s) under conditions more analogous to the commercial production facility. One of the primary response variables in these 2 types of studies is the intestinal lesion score described by Johnson and Reid (1970). This scoring system grades macroscopic lesions observed at necropsy on a scale of 0–4. Score 0 denotes no gross lesions. Scores 1–3 correspond to increasing severity of lesions. Score 4 represents the most severe lesions, including coccidiosis-related death. Separate scores are determined for upper, middle, lower areas of the small intestine and the ceca.

The Johnson and Reid lesion scores have routinely been statistically analyzed using the linear mixed model (LMM) that assumes continuous, bell-shaped (i.e. normal) distributions with homogeneous variances (Bafundo et al., 2008; Wang et al., 2008; Cox et al., 2010; Jordan et al., 2011; Lee et al., 2011; Lensing et al., 2012; Abdelrahman et al., 2014; Ritzi et al., 2014). Regulatory evaluation of anti-coccidial drugs has also relied on comparing mean lesion scores under the LMM framework. For examples, see the Freedom of Information Summary for New Animal Drug Application 138–952, 140–940, 140–951, and 141–281 (FDA-CVM, 1989, 1994, 1999, 2007). The LMM approach benefits from its easy execution and convenient summary of analysis results. However, lesion scores are measured on an ordinal scale with discrete values over a limited range and rarely have a symmetric distribution. Summarizing treatment efficacy by the mean score is less than ideal since the mean could result from infinitely many layouts of lesion severity (e.g. the average of 0 and 4 is the same as the average of 1 and 3). Practitioners have attempted to address these issues by checking bar charts of lesion scores with no further statistical analyses (Casterlow et al., 2011) or employing conventional statistical methods, such as the non-parametric Mann-Whitney/Kruskal-Wallis tests based on ranks (Elmusharaf et al., 2007; Parker et al., 2007; Lan et al., 2016), the contingency-table-based categorical data analysis (Conway et al., 1986, 1993; Amerah and Ravindran, 2015), and the logistic regression for cumulative probabilities of an ordinal response (da Costa et al., 2017). A basic assumption shared by these approaches is that the experimental units for treatments also serve as the observation units. For battery cage and floor pen studies, medicated treatments are applied to cages or pens; lesions are scored on individual birds; even when birds are individually challenged, the parasite's fecal-to-oral route of transmission causes persistent infection at the cage or pen level, making it inappropriate to ignore the effect of cage or pen. Suitability of these statistical approaches is therefore questionable.

Variance heterogeneity in the LMM for lesion scores is an issue that is often overlooked. Summary statistics published in research journals show that the non-inoculated, non-medicated control (NNC) typically had all 0 scores with no variability; depending on the infection level and efficacy of the medication, scores of the inoculated, non-medicated control (INC) and the inoculated medicated treatment groups scattered from 0 to 4 with mostly non-zero variances (Bafundo et al., 2008; Wang et al., 2008; Lee et al., 2011; Jordan et al., 2011; Lensing et al., 2012; Ritzi et al., 2014). There, statistical analysis was performed via the GLM procedure in SAS (SAS Institute Inc., Cary, NC), the ANOVA and GLM procedures in SPSS (SPSS Inc., Chicago, IL), the ANOVA procedure in Genstat (VSN International, Hemel Hempstead, UK), and the Fit Model platform in JMP Pro (SAS Institute Inc., Cary, NC). Unfortunately, these common statistical analysis tools could only fit statistical models with homogeneous variance.

The guidance of the United States Food and Drug Administration Center of Veterinary Medicine (FDA-CVM) (2012) states that "response variables that are at the interval/ratio level of measurement can be described using a linear model; for those that are nominal/ordinal, a generalized linear model is appropriate". The generalized linear model automatically accounts for the relationship between the mean and the variance. Hence variance heterogeneity is of less concern as compared with the LMM analysis. To account for the study design, the statistical model should contain not only fixed effects, i.e. treatment, but also random effects, including replicate (cage or pen) and other relevant design variables for broad inference (e.g. block). Such a statistical model is referred to as the generalized linear mixed model (GLMM). The logit function for cumulative probabilities is the most common link function that associates linear fixed and/or random effects to the probability distribution of an ordinal response (Section 7.2 of Agresti, 2002). Recently da Costa et al. (2017) applied a general linear model with cumulative logit link in analyzing coccidial lesion scores of a floor pen study. This present work applies the GLMM with the cumulative logit link, alongside the LMM, on 9 datasets of chicken coccidial lesion scores collected from 4 battery cage studies. The purpose of this work is to examine the fitness of LMM and GLMM, to review estimates of variance components in the 2 statistical models, and to compare the 2 approaches with respect to their utility in regulatory decision making. For the sake of fairness, the LMM with heterogeneous inter-bird variances was employed when there was statistical evidence against variance homogeneity.

2. Material and methods

2.1. Study procedure

Four battery cage studies were conducted to compare efficacy of 4 existing anti-coccidial drug products and several investigational new drug products for use in broiler chicken feed. Information regarding drug names and the exact dosages does not pertain to the purpose of this investigation and is concealed for proprietary reasons. In each study, day-old male broilers were obtained from a local commercial hatchery and placed in heated coccidia-free battery brooders. (Heritage birds were used in studies Ia and Ib; Cobb 500 birds were used in studies IIa and IIb.) At day 12, birds were weighed, transferred into battery grower cages and began receiving in-feed medication according to the randomized arrangement of the experimental design. At day 14, birds were challenged via oral gavage with mixed *Eimeria* species (Table 1). According to the timing of peak lesion development for combination of

Table 1
Eimeria species in the mixed inoculum and the virulence level.

Study	Species in the mixed inoculum	Intestinal tracts scored	Level of virulence	
			INC birds with lesion score ≤2	NNC birds with lesion score 0
Ia	*E. maxima*	Middle small intestine	56%	100%
	E. tenella	Ceca	19%	100%
Ib	*E. acervulina*	Upper small intestine	9%	100%
	E. tenella	Ceca	22%	100%
IIa	*E. maxima*	Middle small intestine	0%	100%
	E. tenella	Ceca	34%	100%
IIb	*E. acervulina*	Upper small intestine	0%	100%
	E. maxima	Middle small intestine	94%	100%
	E. tenella	Ceca	34%	100%

INC = inoculated, non-medicated control; NNC = non-inoculated, non-medicated control.

the *Eimeria* species being evaluated, birds were euthanized by cervical dislocation for lesion scoring 6 days post challenge. The battery cages were housed in an environment with controlled temperature (27 °C) and continuous illumination. Birds were fed ad libitum with commercial broiler starter ration as mash for non-medicated treatments. All studies were conducted in the United States in compliance with the guidelines of the Federation of Animal Science Societies (1999) as well as local, state, and federal regulations.

2.2. Coccidial lesion score measurement

Mortality was checked daily. Coccidiosis-related lesions were assessed by the Johnson and Reid scoring system for any bird that died post challenge as well as all birds euthanized on day 20. The person who scored lesions was masked to the treatment allocation and relied only on the challenge species to facilitate proper grading (at relevant sections of the intestinal track). There were a total of 9 lesion score datasets for statistical analyses: studies Ia, Ib and IIa each have 2 datasets; study IIb generated 3 datasets. Details on the inoculation species and the actual virulence level are given in Table 1.

2.3. Experimental design

Each study included 2 controls: the NNC and the INC. There were either 8 or 13 inoculated, medicated treatments corresponding to existing and investigational drug products at various dose levels. Virulence of the inoculation was evaluated by comparing the INC to the NNC. Efficacy of medicated treatments was evaluated against the INC. For the purpose of efficacy assessment, the experimental designs were considered to have a one-way treatment structure (Section 4.2.2 of Milliken and Johnson, 1992). Per FDA-CVM (2012) guidance, all studies utilized cages of 8 birds as the experiment unit. Studies Ia and Ib implemented the randomized incomplete block design, where a battery column of 5 cages formed the block for assigning treatments. Studies IIa and IIb employed the randomized complete block design, where a battery unit of 15 cages served as the block. There were 4 replicates per treatment in every study. The experimental designs are summarized in Table 2.

2.4. Statistical models

2.4.1. Liner mixed model

In battery cage studies, the observational unit is the bird. Let Y_{ijk} be the lesion score of the k^{th} bird in the cage receiving treatment i in block j. The following LMM reflects the study design and allows cage to be the experimental unit for treatment comparisons.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

μ: overall mean;

α_i: fixed effect of treatment$i$;

β_j: random effect of block$j$;

γ_{ij}: *random*effect of the cage receiving treatment$i$in block$j$;

ε_{ijk}: random effect of bird$k$under treatment$i$in block$j$.

Random effects due to block, cage and bird were assumed to be normally distributed with 0 means and certain variances. The LMM with homogenous variances assumed these variances to be the same across treatments and, consequently, the statistical model fitting yields 3 estimated variance components. The LMM with variance heterogeneity to treatments does not impose the constant variance restriction and demands more variance components to be estimated. A χ^2 likelihood ratio statistic was used to test variance heterogeneity. The inoculated, medicated treatment i was compared with the INC via testing whether their mean difference is 0, i.e.

$$H_0: \alpha_i - \alpha_{INC} = 0 \quad vs. \quad H_1: \alpha_i - \alpha_{INC} \neq 0$$

The t statistic was used to test the above hypotheses and its degrees of freedom (DOF) were approximated by the Kenward-Roger (KR) method (Kenward and Roger, 1997, 2009). Statistical analysis was conducted via SAS v9.3 (SAS Institute Inc., 2011) throughout. Appendix A. contains the sample code of SAS PROC GLIMMIX for the LMM with homogeneous and heterogeneous inter-bird variances with respect to treatments. On account of the limited number of cages per treatment, the inter-cage variance was assumed homogeneous. Fixed effects were estimated from the statistical model with homogenous variances unless the test of inter-bird variance heterogeneity was significant at the 0.05 level.

2.4.2. Generalized linear mixed model

The response variable Y_{ijk} takes 5 possible values and follows a multinomial distribution of size 1 with cumulative probability $\Pr(Y_{ijk} \leq l)$, $l = 0, \ldots, 3$. Parameter $\Pr(Y_{ijk} \leq l)$ represents the success rate that the lesion score is no greater than l. (Note: $\Pr(Y_{ijk} \leq 4) = 1$.) Similar to da Costa et al. (2017), this present work uses the logit link function to model $\Pr(Y_{ijk} \leq l)$. The GLMM is given as

$$\ln\left(\frac{\Pr(Y_{ijk} \leq l)}{1 - \Pr(Y_{ijk} \leq l)}\right) = \dot{\mu}l + \dot{\alpha}_i + \dot{\beta}_j + \dot{\gamma}_{ij}$$

$\dot{\mu}l$: overall log odds that the lesion score is no greater than$l$;

$\dot{\alpha}^i$: fixed effect of treatment$i$;

$\dot{\beta}^j$: random effect of block$j$;

$\dot{\gamma}^{ij}$: *random*effect of the cage receiving treatment$i$in block$j$.

Random effects due to block and cage were assumed to be normally distributed with mean 0 and certain variances. The probability

Table 2
Study designs.

Study	Number of medicated treatments[a]	Number of non-medicated controls[b]	Number of blocks	Number of treatments per block	Number of cages per treatment	Number of birds per cage
Ia	8	2	8	5	4	8
Ib	8	2	8	5	4	8
IIa	13	2	4	15	4	8[c]
IIb	13[d]	2	4	15	4	8[e]

[a] Birds in medicated treatments were inoculated with mixed *Eimeria* species.
[b] Two types of controls were included. One received inoculation and the other didn't.
[c] One cage in the analysis dataset only had 6 birds and 4 cages only had 7 birds.
[d] Two medicated treatments were excluded from the statistical analyses of lesion scores at the ceca due to the absence of variability (birds in these treatments all scored 0).
[e] Four cages in the analysis dataset only had 7 birds.

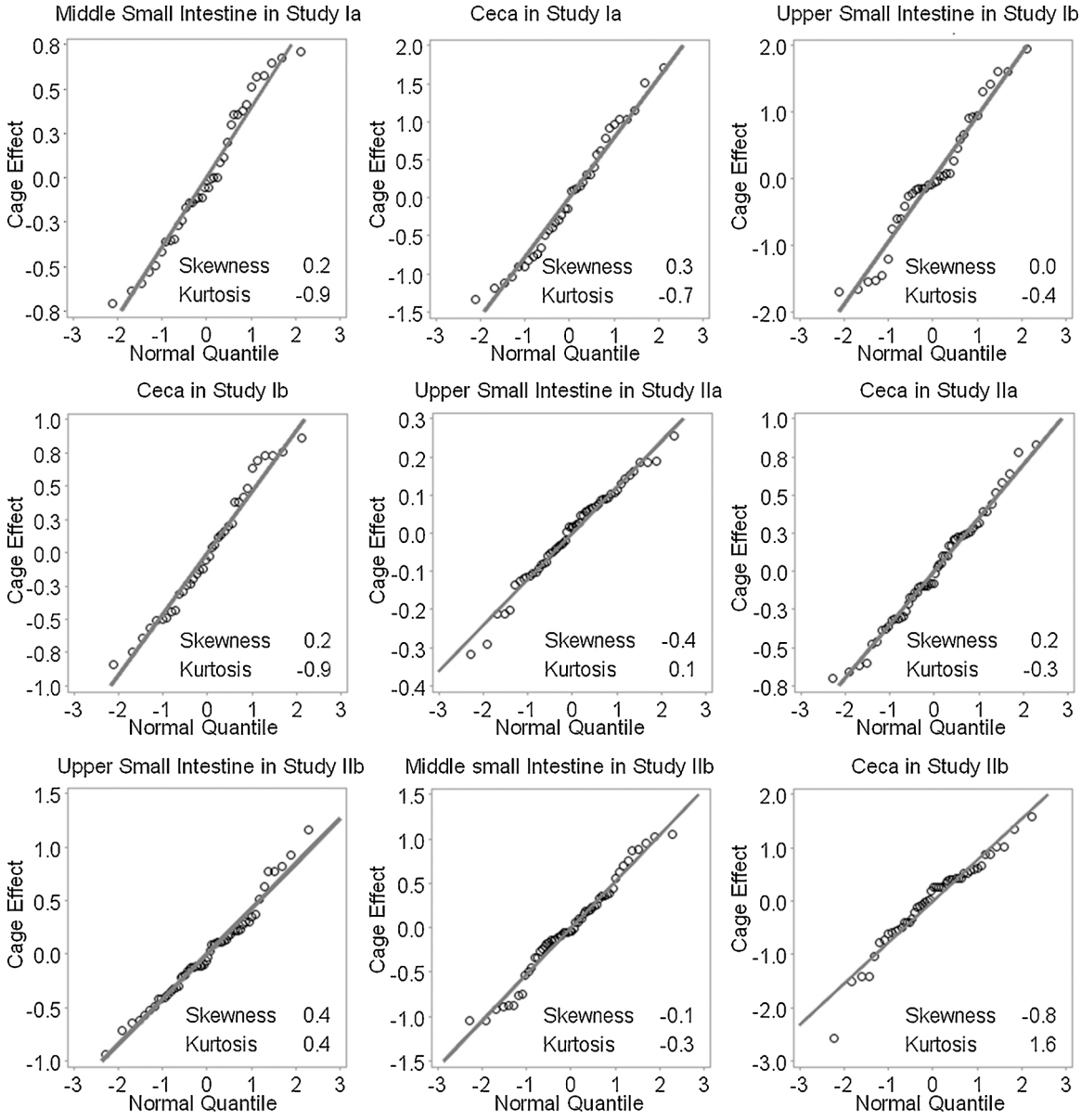

Fig. 1. Distributions of the generalized linear mixed model cage effect estimates with respect to normal distributions with the same means and variances.

parameters of a multinomial distribution determine both its mean and variance. As long as the overdispersion was not substantial, it was unnecessary for the GLMM to include the random effect of bird. Variance heterogeneity was of less concern in the GLMM and random effects here were assumed to be homogenous among treatments.

A key feature of the above statistical model is that its intercept term, $\dot{\mu}l$, increases as the lesion level increases from 0 to 3, i.e. $\dot{\mu}0 \le \dot{\mu}1 \le \dot{\mu}2 \le \dot{\mu}3$, whereas the other parameters in the statistical model are not indexed by the lesion level. Parameter $\dot{\mu}l + \dot{\alpha}_i$ represents the marginal mean of treatment i on the cumulative logit scale. Back-transforming it to the probability scale generates $1 - 1/[1 + \exp(\dot{\mu}l + \dot{\alpha}_i)]$, which is approximately the marginal cumulative probability for treatment i. In theory, the marginal distribution of lesion scores under treatment i is obtained by integrating $1 - 1/[1 + \exp(\dot{\mu}l + \dot{\alpha}_i + \dot{\beta}_j + \dot{\gamma}_{ij})]$ over the probability measures of $\dot{\beta}_j$ and $\dot{\gamma}_{ij}$. Such an intricate calculation is currently not supplied in SAS PROC GLIMMIX.

Efficacy of the inoculated, medicated treatment i relative to the INC is evaluated via testing

$$H_0: \dot{\alpha}_i - \dot{\alpha}_{INC} = 0 \quad vs. \quad H_1: \dot{\alpha}_i - \dot{\alpha}_{INC} \neq 0$$

On the probability scale, the above test evaluates whether the odds of success for treatment i equals that for the INC. (Note: the odds ratio (OR) under the cumulative logit model is constant with respect to the lesion level that defines success; this is oftentimes referred to as the proportional odds model). The t statistic was used to test the above hypotheses. Faes et al. (2009) caution against using the KR method in the GLMM. Simulation studies shows that the KR method had a problem in maintaining the desired type I error in the GLMM analysis of binary outcomes (Table 1 of Stroup, 2013; Li and Redden, 2015). The present work tentatively applied the containment method, which is the default option of SAS PROC GLIMMIX, to approximate the DOF for the t statistic. This method sets the DOF as the number of cages minus the number of treatments minus the number of blocks plus 1. A SAS code example for the proposed GLMM is given in Appendix A.

Although the range of the lesion score is 0∼4, the highest score observed in 4 out of 9 datasets was 3. The GLMM for these 4 datasets then assumed $\Pr(Y_{ijk} \le 3) = 1$ and applies exclusively to $\Pr(Y_{ijk} \le l)$,

Table 3
Counts of lesions scores for each treatment at the middle small intestine in study Ia.

	Lesion Score				
Treatment	0	1	2	3	Total
INC	0	0	18	14	32
T1	2	23	7	0	32
T2	1	13	18	0	32
T3	1	24	7	0	32
T4	0	13	18	1	32
T5	0	10	20	2	32
T6	1	19	12	0	32
T7	7	20	5	0	32
T8	5	23	4	0	32

$l = 0, \ldots, 2$. There was 1 statistical analysis dataset with the lowest lesion score being 1. In this case, the GLMM assumes $\Pr(Y_{ijk} \leq 0) = 0$ and pertains solely to $\Pr(Y_{ijk} \leq l)$, $l = 1, \ldots, 3$.

2.4.3. Criteria for excluding data from statistical modeling

Birds in the NNC all had a lesion score of 0. Following the FDA-CVM (2012) guidance, they were excluded from the LMM and GLMM for evaluating treatment efficacy. In study IIb, birds in 2 inoculated, medicated treatments responded unanimously with a lesion score of 0 at the ceca. Such treatments were excluded from the analyses of mixed models because (1) the absence of variability in them will bias the variance component estimates in the LMM; (2) their fixed effects in the GLMM are theoretically non-estimable (Section 9.8 of Agresti, 2002), causing non-convergence of the entire likelihood maximization process. These 2 treatments were separately evaluated against the INC via the Fisher's exact test, where the conclusion of efficacy was based on a significant increase in the proportion of birds scored 0.

3. Results

3.1. Goodness of statistical model fitting

Inspection of residuals in the LMM indicates that the normality assumption was achieved (results not shown). The assumption of proportional odds could easily be tested when the model has no random effect. Pan and Lin (2005) stated that "The developments of the model-checking procedures for GLMMs are challenging because the existence of random effects not only complicates the theoretical derivations, such as the proofs for the asymptotic distributions of the cumulative-sum processes and for the consistency of the tests, but also imposes computational challenges." Several goodness-of-fit tests have been developed for GLMM (Pan and Lin, 2005; Abad et al., 2010). To our knowledge, these tests have not been implemented by any commercial software. Regarding the analysis of multinomial responses, SAS PROC GLIMMIX does not compute residuals due to the lack of a clear

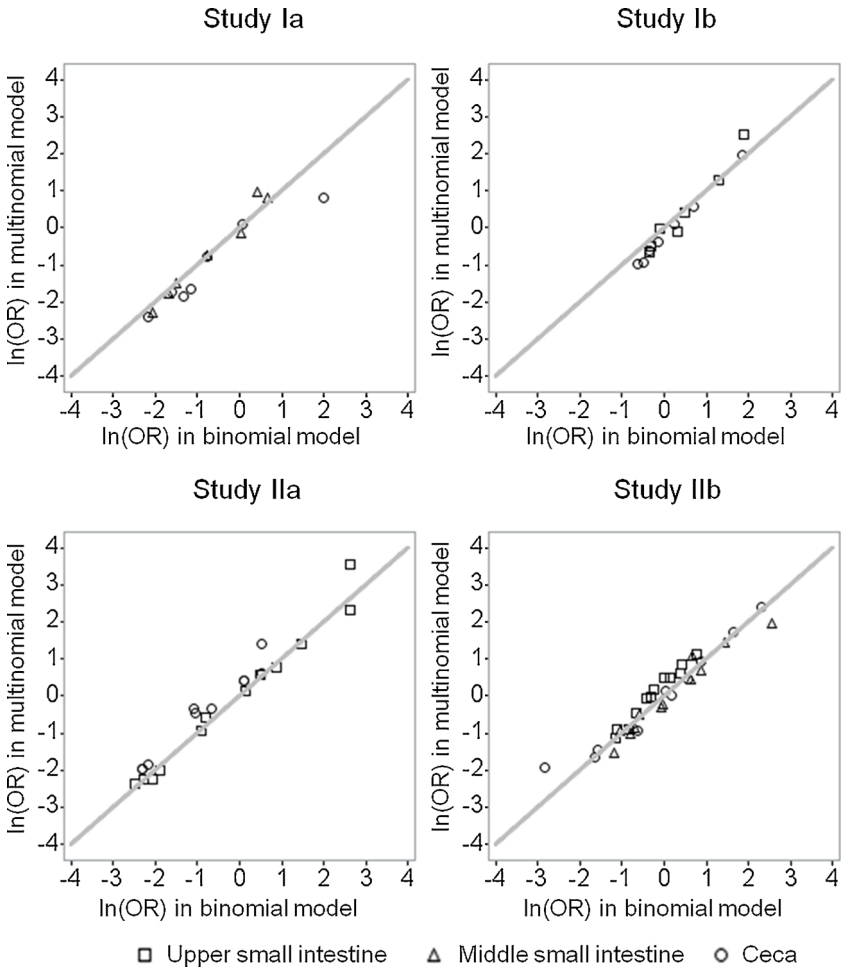

Fig. 2. The natural logarithm of odds ratio between 2 inoculated, medicated treatments. OR = odds ratio.

Table 4
Variance component estimates in LMM and GLMM.

		LMM				GLMM	
Dataset	Bird variance	Block variance	Cage variance	Bird variance	P-value	Block variance	Cage variance
Middle small intestine	Homogenous	0 (–)	0.03 (0.02)	0.28 (0.02)	0.922	0 (–)	0.46 (0.29)
in study Ia	Heterogeneous	0 (–)	0.03 (0.02)	0.20 (0.05) ~ 0.35 (0.09)			
Ceca	Homogenous	0 (–)	0.19 (0.07)	0.66 (0.06)	< 0.001	0 (–)	1.18 (0.47)
in study Ia	Heterogeneous	0 (–)	0.18 (0.07)	0.13 (0.04) ~ 1.00 (0.26)			
Upper small intestine	Homogenous	0 (–)	0.13 (0.05)	0.34 (0.03)	0.261	0 (–)	1.60 (0.62)
in study Ib	Heterogeneous	0 (–)	0.12 (0.05)	0.17 (0.05) ~ 0.50 (0.14)			
Ceca	Homogenous	0.03 (0.04)	0.09 (0.05)	0.57 (0.05)	< 0.001	0.19 (0.25)	0.58 (0.33)
in study Ib	Heterogeneous	0.03 (0.04)	0.09 (0.05)	0.19 (0.05) ~ 0.92 (0.24)			
Upper small intestine	Homogenous	< 0.01 (0.01)	0.01 (0.01)	0.40 (0.03)	0.272	0.01 (0.05)	0.11 (0.15)
in study IIa	Heterogeneous	< 0.01 (0.01)	0.01 (0.01)	0.25 (0.07) ~ 0.58 (0.15)			
Ceca	Homogenous	0 (–)	0.05 (0.02)	0.41 (0.03)	0.157	0 (–)	0.38 (0.20)
in study IIa	Heterogeneous	0 (–)	0.04 (0.02)	0.25 (0.07) ~ 0.70 (0.19)			
Upper small intestine	Homogenous	0 (–)	0.05 (0.02)	0.41 (0.03)	0.813	0 (–)	0.48 (0.22)
in study IIb	Heterogeneous	0 (–)	0.06 (0.02)	0.27 (0.07) ~ 0.59 (0.15)			
Middle small intestine	Homogenous	< 0.01 (0.01)	0.08 (0.03)	0.45 (0.03)	0.001	0.03 (0.10)	0.65 (0.26)
in study IIb	Heterogeneous	< 0.01 (0.01)	0.08 (0.03)	0.09 (0.02) ~ 0.78 (0.20)			
Ceca	Homogenous	0 (–)	0.21 (0.07)	0.58 (0.05)	< 0.001	0 (–)	1.33 (0.51)
in study IIb	Heterogeneous	0 (–)	0.16 (0.07)	0.03 (0.01) ~ 1.30 (0.35)			

GLMM = generalized linear mixed model; LMM = linear mixed model.
Notation: Values inside the parentheses are the standard error for the estimates; the standard error is not available for a variance estimate at its boundary value 0; the p-value pertains to testing homogeneity in the LMM inter-bird variances among treatments.

Table 5
Fixed effects estimated by the GLMM analysis of lesion scores at the middle small intestine in study Ia.

"Solution for Fixed Effects" outputted by SAS				
Effect	Lesion score	Treatment	Estimate	Parameter
Intercept	0		-8.23	$\dot{\mu}0$
Intercept	1		-4.31	$\dot{\mu}1$
Intercept	2		0.22	$\dot{\mu}2$
Trt		T1	5.60	$\dot{\alpha}_1$
Trt		T2	4.13	$\dot{\alpha}_2$
Trt		T3	5.48	$\dot{\alpha}_3$
Trt		T4	3.85	$\dot{\alpha}_4$
Trt		T5	3.32	$\dot{\alpha}_5$
Trt		T6	4.85	$\dot{\alpha}_6$
Trt		T7	6.57	$\dot{\alpha}_7$
Trt		T8	6.42	$\dot{\alpha}_8$
Trt		INC	0	$\dot{\alpha}_{INC}$

GLMM = Generalized liner mixed model; INC = inoculated, non-medicated control.
Notation: T1-T8 correspond to inoculated medicated treatments; "Trt" is the name of the variable in the dataset that records the treatment.

Table 6
Estimates of treatment effects at the middle small intestine in study Ia.

	LMM		GLMM			
Treatment	Mean score	Mean score reduction	Pr(score = 0)	Pr(score ≤ 1)	Pr(score ≤ 2)	OR of success
INC	2.44	–	0.0%	1.3%	55.5%	–
T1	1.16	1.28	6.7%	78.4%	99.7%	271
T2	1.53	0.91	1.6%	45.3%	98.7%	62
T3	1.19	1.25	6.0%	76.2%	99.7%	239
T4	1.62	0.81	1.2%	38.6%	98.3%	47
T5	1.75	0.69	0.7%	27.0%	97.2%	28
T6	1.34	1.09	3.3%	63.0%	99.4%	127
T7	0.94	1.50	16.0%	90.5%	99.9%	714
T8	0.97	1.47	14.0%	89.1%	99.9%	612

GLMM = generalized linear mixed model; INC = inoculated, non-medicated control; LMM = linear mixed model; OR = odds ratio.
Notation: T1-T8 represent inoculated, medicated treatments.

definition: the lesion score of an animal is observed at a single level; the model prediction places probability masses at ≥3 levels. A statistic for assessing overdispersion, comparable with deviance in generalized linear models, is not available in SAS PROC GLIMMIX, either. Although some statistical software does provide residuals and/or deviance under GLMM, the statistical properties of these outputs may not be comparable to those under the generalized linear model.

Estimates of the random effects in the GLMM are analogous to the best linear unbiased predictors in the LMM. These estimates were used, as a last resort, to check the GLMM distributional assumption. Fig. 1 presents quantile-quantile plots of the cage effects, estimated from each of the 9 dataset, with respect to the normal distributions with the same means and variances. The linear patterns seen in Fig. 1 support the normality assumption in the GLMM. Skewness and kurtosis are respective measures of symmetry and tail heaviness of a distribution. Their values for the estimated distributions of the GLMM cage effect are presented in the quantile-quantile plots, as well. Note that a perfect normal distribution has 0 skewness and 0 kurtosis. Eight lesion score datasets have -0.4 ~ 0.4 skewness and -0.9 ~ 0.4 kurtosis. One cage effect for the cecal lesion scores in study IIb was a noticeable outlier. Removing this cage from the analysis changed skewness from -0.8 to -0.4 and kurtosis from -1.6 to -0.1. A closer look at the original dataset reveals that the outlier corresponds to an inoculated, medicated treatment: this outlier cage had 6 out of 8 birds scored 3 whereas the other 3 cages under the same treatment had 5, 6 and 8 birds scored 0. These evidences suggest that the proposed GLMM fit adequately to the coccidial lesion scores collected from battery cage studies and is capable of identifying deviant observations. Given the limited number of cages per treatment, the outlier cage was not excluded in subsequent analyses.

For the analysis of ordinal responses in survey/observational studies, the assumption of proportional odds are routinely assessed by comparing ORs estimated separately from data dichotomized under various cutoffs (Bell and Dexter, 2000; O'Connell et al., 2008). The

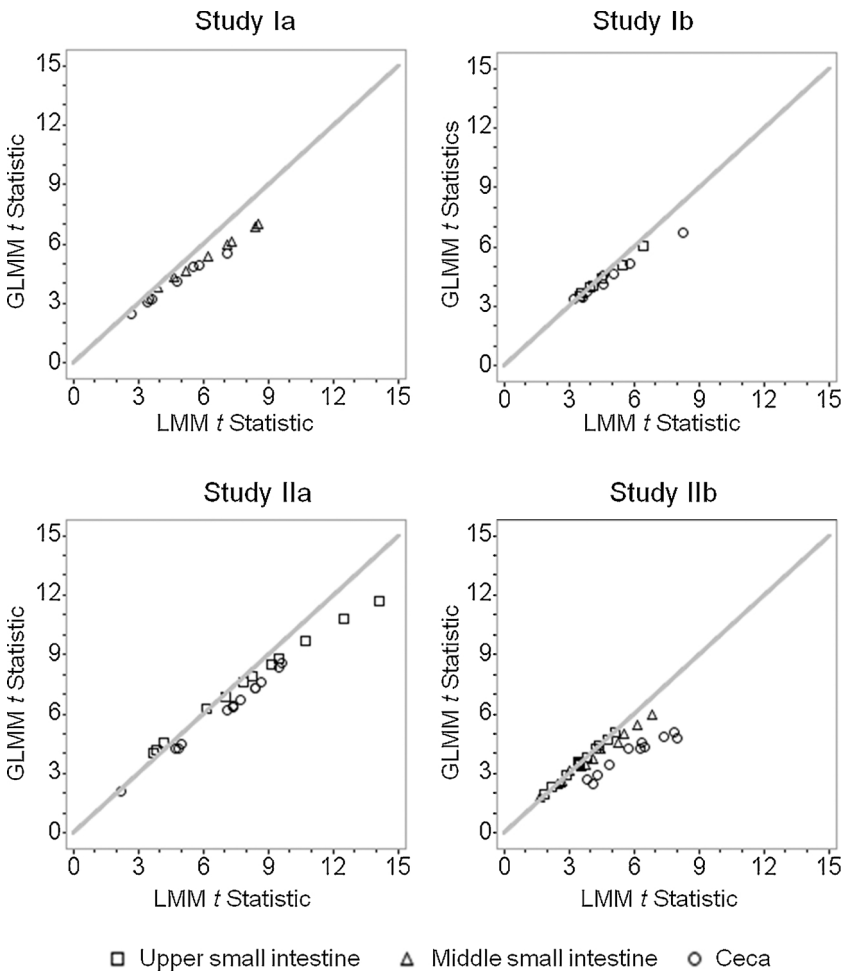

Fig. 3. The *t* statistic values used to compare the inoculated, medicated treatments with the inoculated, non-medicated control. GLMM = generalized linear mixed model; LMM = linear mixed model.

sample sizes there were in thousands or tens of thousands. In contrast, the sample sizes for battery cage studies were in hundreds (288 for Studies IA and IB, 442 for Study IIA, 444 for Study IIB, specifically). Table 3 presents the counts of lesion scores for inoculated treatment groups at the middle small intestine in Study IA. Cutting the data into 0 vs. 1 ~ 3 leads to a treatment × binary response contingency table with ~ 50% cell counts ≤ 5 where ORs could not be well estimated; cutting the data into 0–2 vs. 3 also leads to a contingency table with ~ 50% cell counts ≤ 5; when cutting the data into 0–1 vs. 2–3, ORs involving INC could not be well estimated; there were no 2 cutoff points allowing reliable comparison of ORs among all treatment groups. Similar data pattern were observed in the other 8 example datasets (results not shown). Modeling dichotomized data is therefore feasible only at a single cutoff point and upon the exclusion of INC. Because the 4 studies shared a common inoculated, medicated treatment, this present work uses the ORs involving this treatment to evaluate the assumption of proportional odds. Fig. 2 plots the natural logarithm of ORs estimated under the multinomial clogit GLMM against those estimated under the binomial logit GLMM (the cutoff of 0–1 vs. 2–4 was applied to each location in studies Ia and Ib, ceca in study IIa, and middle small intestine in study IIb; the cutoff of 0–2 vs. 3–4 was applied to the upper small intestine in studies IIa and IIb; the cutoff of 0 vs. 1–4 was used for ceca in study IIb). Estimates under the 2 GLMMs appear to be close in most cases, except for a ln(OR) for cecal lesion scores in study Ia (0.8 vs. 2.0), a ln(OR) for upper small intestinal lesion scores in Study IIa (3.5 vs. 2.6), a ln(OR) for cecal lesion scores in Study IIa (1.4 vs. 0.5), and a ln(OR) for cecal lesion scores in Study IIb (− 1.9 vs. − 2.8). Considering that the standard error of ln(OR) under the binomial logit GLMM was respective 1.3, 1.1, 1.0 and 1.1, there were no strong evidence against the assumption of proportional odds.

3.2. Analysis of variance components

Table 4 presents variance component estimates in the LMMs with homogenous and heterogeneous inter-bird variances. Heterogeneity in variance was significant at the 0.05 level in 4 out of 9 datasets. However, the setting for inter-bird variance of the statistical model has little effect on the estimation of inter-block and inter-cage variances. The inter-block variance in all LMM analyses was either 0 or negligible with respect to its estimation error. The inter-cage variance is non-trivial but it is oftentimes smaller than the inter-bird variance. Table 4 also lists the estimates of the GLMM variance components. Similar to the LMMs, the GLMM has negligible inter-block variance whereas the inter-cage variance is considerable. The trivial effect of blocking by cage location is consistent with the fact that these battery cage studies were conducted in a controlled laboratory environment. The sizable in-cage variances in LMM and GLMM also demonstrate the importance of including the cage effect in statistical modeling.

Table 7
DOF for the t distribution used to compare the inoculated, medicated treatments to the inoculated, non-medicated control.

Dataset	Number of comparisons	DOF of the t distribution		
		LMM KR method	GLMM Containment method	GLMM KR method
Middle small intestine in study Ia	8	27	20	55.8 ~ 73.0
Ceca in study Ia	8	16.1 ~ 22.2	20	31.7 ~ 36.5
Upper small intestine in study Ib	8	27	20	36.0 ~ 40.4
Ceca in study Ib	8	13.5 ~ 23.4	20	25.9 ~ 35.6
Upper small intestine in study IIa	13	41.2, 44.4[a]	39	51.1 ~ 94.5
Ceca in study IIa	13	43.2, 45.5[a]	39	43.8 ~ 62.3
Upper small intestine in study IIb	13	41.8, 42.4[a]	39	41.7 ~ 44.9
Middle small intestine in study IIb	13	22.1 ~ 31.8	39	45.3 ~ 52.7
Ceca in study IIb	11	18.1 ~ 33.5	33	26.1 ~ 106.8

DOF = degrees of freedom; GLMM = generalized linear mixed model; KR = Kenward-Roger; LMM = linear mixed model.
[a] Because cages in some treatments contain less than 8 birds, the DOF in the LMM with homogenous variances took 2 values.

3.3. Evaluation of the fixed effect

Interpretation of fixed effects in the GLMM is not as straightforward as in the LMM. To help readers become accustomed to the GLMM, Table 5 pairs parameters in the GLMM with the SAS "Solution for Fixed Effects" output for the GLMM analysis of lesion scores at the middle small intestine in study Ia. (Parameter $\dot{\mu}3$ was not estimable due to the absence of score 4 in this dataset.) The lesion score distribution for the INC is calculated below.

$$\begin{aligned}\Pr(score_{INC}=0) &= 1-1/[1+\exp(\dot{\mu}0+\dot{\alpha}_{INC})]\\ &= 1-1/[1+\exp(-8.23+0)]\approx 0.0\%\\ \Pr(score_{INC}\le 1) &= 1-1/[1+\exp(\dot{\mu}1+\dot{\alpha}_{INC})]\\ &= 1-1/[1+\exp(-4.31+0)]\approx 1.3\%\\ \Pr(score_{INC}\le 2) &= 1-1/[1+\exp(\dot{\mu}2+\dot{\alpha}_{INC})]\\ &= 1-1/[1+\exp(0.22+0)]\approx 55.5\%\end{aligned}$$

The lesion score distribution for an inoculated, medicated treatment, say T1, is given by

$$\begin{aligned}\Pr(score_{T1}=0) &= 1-1/[1+\exp(\dot{\mu}0+\dot{\alpha}_1)]\\ &= 1-1/[1+\exp(-8.23+5.60)]\approx 6.7\%\\ \Pr(score_{T1}\le 1) &= 1-1/[1+\exp(\dot{\mu}1+\dot{\alpha}_1)]\\ &= 1-1/[1+\exp(-4.31+5.60)]\approx 78.4\%\\ \Pr(score_{T1}\le 2) &= 1-1/[1+\exp(\dot{\mu}2+\dot{\alpha}_1)]\\ &= 1-1/[1+\exp(0.22+5.60)]\approx 99.7\%\end{aligned}$$

The OR of success between treatment T1 and the INC is

$$\begin{aligned}\frac{\Pr(score_{T1}=0)}{\Pr(score_{T1}>0)}\cdot\frac{\Pr(score_{INC}>0)}{\Pr(score_{INC}=0)} &= \frac{\Pr(score_{T1}\le 1)}{\Pr(score_{T1}>1)}\cdot\frac{\Pr(score_{INC}>1)}{\Pr(score_{INC}\le 1)}\\ =\frac{\Pr(score_{T1}\le 2)}{\Pr(score_{T1}>2)}\cdot\frac{\Pr(score_{INC}>2)}{\Pr(score_{INC}\le 2)} &= \exp(\dot{\alpha}_1-\dot{\alpha}_{INC})=\exp(5.6-0)\\ \approx 271\end{aligned}$$

The OR and the relative risk are two separate concepts in categorical data analysis: the OR is a ratio of odds; the relative risk is a ratio of probabilities. Between T1 and INC, the relative risks (with "risk" being the probability of having low lesion scores) are

$$\begin{aligned}&\Pr(score_{T1}=0)/\Pr(score_{INC}=0)\approx 252.3\\ &\Pr(score_{T1}\le 1)/\Pr(score_{INC}\le 1)\approx 59.2\\ &\Pr(score_{T1}\le 2)/\Pr(score_{INC}\le 2)\approx 1.8\end{aligned}$$

Table 6 reports all LMM and GLMM fixed effect estimates for this dataset.

Analyses of the 9 datasets produced a total of 95 statistical comparisons of inoculated, medicated treatments with the INC. Both LMM and GLMM approaches identified 93 comparisons significant at the 0.05 level. It is interesting to see from Fig. 3 that for each comparison, the t statistic in the GLMM was either close to or moderately smaller than the one in the LMM. The DOF given in Table 7 were all greater than 13. Calculating p-values for a t distribution is not sensitive to DOF over 10. These results explain concordance of the two approaches in statistical decision making. Table 7 also shows that having heterogeneous inter-bird variances in the LMM reduced the DOF from 27 to 13.5 ~ 23.4 in the design for studies Ia and Ib, and from 41.2 ~ 45.5 to 18.1 ~ 33.5 in the design for studies IIa and IIb. For demonstration purposes only, Table 7 gives the KR DOF calculated by SAS for the GLMM. The DOF were observed to vary widely for several datasets. The most extreme case occurred in the analysis of cecal lesion scores in study IIb, where the DOF for 11 comparisons ran from 26.1 to 106.8. This irregularity in DOF, to a certain extent, suggests that the KR method is unsuitable to the GLMM.

Efficacy of an anti-coccidial drug relative to the INC is quantified by the mean score reduction under the LMM and the OR of success under the GLMM. Fig. 4 illustrates the relationship between the two relative efficacy parameters estimated for each comparison. Note that the same mean score reduction could correspond to different ORs of success, and vice versa; but within a given dataset, the OR of success is almost monotonically related to the mean score reduction.

3.4. Criterion of clinical relevancy

In addition to statistical significance, the guidance of FDA-CVM (2012) requests that the investigational new drug lowers lesions by at least 1 score unit for clinical relevancy. This requirement might be too restrictive, as anti-coccidial drugs approved before this guidance do not always satisfy this criterion. For examples, see Table 11 of FDA-CVM (1989), Table 4.5 of FDA-CVM (1999), and Table 1 of FDA-CVM (2007). As to the 93 significant comparisons in the present work, 67 of them passed the clinical relevancy criterion for mean score reduction and their ORs of success were greater than 10; seven comparisons failed the regulatory requirement and their corresponding ORs were less than 10. There were 19 comparisons where the reduction of mean score was less than one but the OR was greater than 10. Seven out of these 19 comparisons correspond to existing anti-coccidial drug products at their recommended dosages; three comparisons correspond to investigational drug products containing at least one active pharmaceutical ingredient at the recommended dosage. Table 8 presents details of these 19 comparisons. Comparisons involving treatment T5 in study Ia provide some valuable insight on clinical relevance. This treatment corresponds to an existing drug product administered at its recommended dose level. At the middle small intestine, the percentage of birds scored 2 or less is 97.2% for T5 and 55.5% for the INC; the prevailing lesion scores in the INC were 2 (55.5%-1.3% = 54.2%) and 3 (100%-55.5% = 44.5%) whereas birds treated with T5 were mostly scored 1

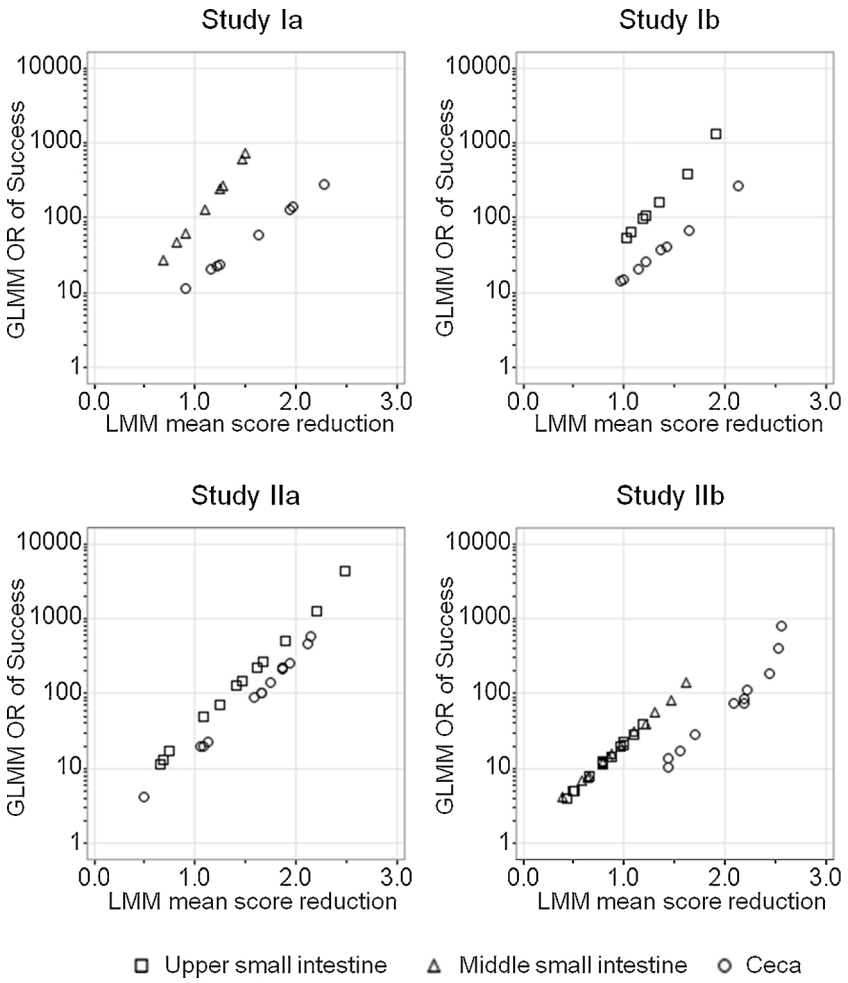

Fig. 4. Relative efficacy of the inoculated, medicated treatments against the inoculated, non-medicated control. GLMM = generalized linear mixed model; LMM = linear mixed model.
OR = odds ratio.

(27.0%-0.7% = 26.3%) and 2 (97.2%-27.0% = 70.2%). Here treatment T5 led to a mean score reduction of 0.69 and an increase in the odds of success of 28 fold. At the ceca, the percentage of birds scored 2 or less is 67.1% for T5 and 15.2% for the INC; birds in the INC were mostly scored 3 (100%–15.2% = 84.8%); treatment T5 shifted the distribution of lesion scores to 30.2% (35.3%–5.1%) at level 1, 31.8% (67.1%–35.3) at level 2 and 32.9% (100%–67.1%) at level 3. Here treatment T5 led to a 0.91-unit reduction of mean lesion score and an 11-fold increase in the odds of success. Given the apparent disparity of lesion score distributions between treatment T5 and the INC, a sensible criterion for clinically relevant efficacy under the GLMM could then be set as at least a 10-fold increase in the odds of success. That is, the effect of an anticoccidial drug product would be deemed clinically relevant when the odds of having low lesion scores in the medicated group is 10 times or more than the odds in the un-medicated group.

4. Discussion

This present work proposes to apply the GLMM with the cumulative logit link to analyze chicken coccidial lesion scores collected in battery cage studies. Fitness of the GLMM was supported by distributions of the (random) cage effect estimated from 9 datasets generated by challenging battery-cage-housed broilers with various mixtures of *Eimeria* species. Analyses of GLMM variance components reveal that blocking cage by its housing location has a trivial effect on the study outcome, which agrees with the study setup. Results of efficacy testing under the GLMM are in complete concordance with those under the LMM. Yet, the proposed GLMM approach has several advantages over the LMM approach. First, there is little concern on the variance heterogeneity of GLMM random effects. Second, the GLMM describes the entire distribution of lesion scores and offers flexibility in assessing relative efficacy based on user's preference, such as mean score reduction, change in success or failure rate at certain lesion level, prevented and mitigated fractions (Sievi, 2005). In contrast to the current regulatory requirement of a 1-unit reduction in the LMM mean score for battery cage studies, the clinical relevance criterion accompanying the GLMM (i.e. a 10-fold increase in the odds of success) is more capable of capturing the disparity in lesion score distributions and generates an assessment more consistent with historical determinations of efficacy. Nonetheless, progress made here in analyzing lesion scores from battery cage studies may not fully extend to floor pen studies given the disparity in their study design. In general, a battery cage study randomizes every treatment to 3 ~ 5 battery cages of same sex, each of which houses 4 ~ 10 chicken under controlled laboratory conditions. The experiment usually contains multiple inoculated medicated treatments. A single-site floor pen study is conducted under conditions more analogous to the commercial production facility and typically assigns every treatment to 4 ~ 12 pens, each of which holds 20 ~ 200 chickens. Besides, only a small portion of birds in a floor pen study are sampled for intestinal lesions whereas lesion scores are recorded for all birds in a battery cage

Table 8
Efficacy evaluation of inoculated, medicated treatments against the INC where comparison was significant, the OR of success was greater than 10, but the mean lesion score reduction was less than 1.

		LMM			GLMM					
Dataset	Treatment	Mean score	Mean score reduction	P-value	Pr(score = 0)	Pr(score ≤ 1)	Pr(score ≤ 2)	Pr(score ≤ 3)	OR of success	P-value
Middle small intestine	INC	2.44	–	–	0.0%	1.3%	55.5%	100%	–	–
in study Ia	T2 A	1.53	0.91	< 0.001	1.6%	45.3%	98.7%	100%	62	< 0.001
	T4	1.62	0.81	< 0.001	1.2%	38.6%	98.3%	100%	47	< 0.001
	T5 A	1.75	0.69	< 0.001	0.7%	27.0%	97.2%	100%	28	0.001
Ceca in study Ia	INC	2.81	–	–	0.5%	4.6%	15.2%	100%	–	–
	T5 A	1.91	0.91	0.020	5.1%	35.3%	67.1%	100%	11	0.022
Ceca in study Ib	INC	2.79	–	–	0.3%	3.6%	20.4%	100%	–	–
	T4	1.82	0.97	0.002	4.6%	34.8%	78.6%	100%	14	0.003
	T6 A/-	1.80	0.99	0.004	4.8%	35.8%	79.3%	100%	15	0.003
Upper small intestine	INC	3.59	–	–	0.0%	0.1%	1.7%	40.6%	–	–
in study IIa	T2	2.91	0.68	< 0.001	0.0%	1.7%	18.8%	90.0%	13	< 0.001
	T3	2.94	0.65	< 0.001	0.0%	1.5%	17.0%	88.8%	12	< 0.001
	T4	2.84	0.74	< 0.001	0.0%	2.2%	23.1%	92.1%	17	< 0.001
	T6 A	2.91	0.68	< 0.001	0.0%	1.7%	18.8%	90.0%	13	< 0.001
Upper small intestine	INC	3.44	–	–	–	0.3%	5.0%	53.4%	–	–
in study IIb	T1 A	2.66	0.78	0.001	–	3.3%	39.5%	93.4%	12	0.001
	T3	2.56	0.87	< 0.001	–	3.9%	43.6%	94.3%	15	< 0.001
	T7 A	2.64	0.80	0.001	–	3.3%	39.1%	93.3%	12	0.001
	T10 A/-	2.66	0.78	0.001	–	3.1%	38.0%	93.0%	12	0.001
	T13 A/-	2.47	0.97	< 0.001	–	5.3%	51.4%	95.8%	20	< 0.001
Middle small intestine	INC	2.03	–	–	1.4%	12.0%	85.1%	100%	–	–
in study IIb	T1 A	1.06	0.97	< 0.001	21.0%	72.6%	99.1%	100%	19	< 0.001
	T3	1.25	0.78	0.005	13.9%	61.6%	98.5%	100%	12	0.004
	T4	1.19	0.84	0.003	16.3%	66.0%	98.8%	100%	14	0.002
	T5	1.16	0.87	0.002	17.7%	68.1%	98.9%	100%	16	0.001

GLMM = generalized linear mixed model; INC = inoculated, non-medicated control; LMM = linear mixed model; OR = odds ratio.
Notation: T1-T13 correspond to inoculated medicated treatments; label "A" denotes existing drug products at their recommended dosages; label "A/-" denotes investigational drug products containing at least 1 active pharmaceutical ingredient at the recommended dosage; the p-value under the LMM pertains to testing whether the mean score difference is 0; the p-value under the GLMM pertains to testing whether the OR is 1.

study. Without fitting the GLMM to a few datasets generated from floor pen studies, it is difficult to foresee the convergence of the likelihood maximization process and stability of parameter estimation.

Alternatively, the ordinal lesion score could be converted into a binary response of "success" and "failure", which is then analyzed by the GLMM with logit link or by the LMM on an arc-sine-square-root-transformed success rate observed from each cage. The caveat of this approach is that data dichotomization must be determined *a priori* in the planning stage of the study. When birds are challenged with mixed *Eimeria* species, it is difficult to manage virulence of the inoculum so that the INC experiences a similar level of damage across every relevant section of the intestines (see Table 1 for examples). Consequently, the study may not be powered to detect the same strength of efficacy at every intestinal section examined. Suppose that the protocol of study IIb specifies "success" as a score of 2 or less. The success rate of the INC turned out to be $0/32 \approx 0\%$ at the upper small intestine, $30/32 \approx 94\%$ at the middle small intestine, and $11/32 \approx 34\%$ at the ceca (Table 1). Even if a medicated treatment had a 32/32 success rate, the statistical comparison to the INC would not be significant at the middle small intestine. Re-defining "success" as a score of 0 or 1, after the fact, lowers the INC success rate to $1/32 \approx 3\%$, making it possible to prove efficacy. In contrast, the proposed GLMM approach simultaneously evaluates treatment efficacy over every possible definition of success. It puts less stress on the experimenter in terms of gauging the virulence of the inoculum and avoids any controversial *post hoc* analysis/discussion.

The analyses of large-scale survey/observational data indicates that the odds related to ordinal responses were not always proportional to attributes in the generalized linear model (Bell and Dexter, 2000; O'Connell et al., 2008). The absence of well-established statistical methodology and the limitation in the sample size, unfortunately, prevent a more rigorous evaluation of the model assumption made here. It is not clear how robust the proposed method is in terms of type I error and power for anti-coccidial efficacy evaluation when the proportional odds assumption is violated. A thorough assessment of this issue will require an extensive simulation study which is beyond the scope of this work.

In summary, the proposed GLMM is anticipated to serve as a valuable tool for evaluating the efficacy of investigational new anti-coccidial drugs. With a statistical modeling framework that truly reflects the ordinal nature of lesion scores, it is now possible to examine the experimental design in conjunction with the proper DOF approximation method via simulation studies.

Funding

This research was supported by Zoetis Inc.

Declarations of interest

None

Acknowledgements

The authors thank Hima Bindu Vanimisetti of Zoetis for her valuable suggestions.

Appendix A. Sample SAS code for the mixed model analyses

```
data ceca_lesion;
input blk $ trt $ cage $ chick $ lesion;
cards;
1 T1 30 124 2
2 T2 17 408 3
...
;
run;

title 'LMM with homogeneous variances';
proc glimmix; where trt^='NNC';
class blk trt;
model lesion=trt/ddfm=kr ;
random blk blk*trt;/*blk*trt corresponds to cage effect*/
run;

title 'LMM with heterogeneous residual variances';
proc glimmix; where trt^='NNC';
class blk trt;
model lesion=trt/ddfm=kr;
random blk blk*trt;
random _residual_/grp=trt;
covtest 'Test of residual variance homogeneity' homogeneity;
run;

title 'GLMM with cumulative logit link';
proc glimmix; where trt^='NNC';
class blk trt;
model lesion=trt/dist=multinomial link=clogit s;
random blk blk*trt;
run;
```

References

Abad, A.A., Litière, S., Molenberghs, G., 2010. Testing for misspecification in generalized linear mixed models. Biostatistics 11, 771–786.

Abdelrahman, W., Mohnl, M., Teichmann, K., Doupovec, B., Schatzmayr, G., Lumpkins, B., Mathis, G., 2014. Comparative evaluation of probiotic and salinomycin effects on performance and coccidiosis control in broiler chickens. Poult. Sci. 93, 3002–3008.

Agresti, A., 2002. Categorical Data Analysis, 2nd ed. Wiley Interscience, Hoboken, NJ.

Amerah, A.M., Ravindran, V., 2015. Effect of coccidia challenge and natural betaine supplementation on performance, nutrient utilization, and intestinal lesion scores of broiler chickens fed suboptimal level of dietary methionine. Poult. Sci. 94, 673–680.

Bafundo, K.W., Cervantes, H.M., Mathis, G.F., 2008. Sensitivity of *Eimeria* field isolates in the United States: responses of nicarbazin-containing anticoccidials. Poul. Sci. 87, 1760–1767.

Bell, J.F., Dexter, T., 2000. Using ordinal multilevel models to assess the comparability of examinations. Multilevel Modell. Newslett. 12, 4–9.

Casterlow, S., Li, H., Gilbert, E.R., Dalloul, R.A., McElroy, A.P., Emmerson, D.A., Wong, E.A., 2011. An antimicrobial peptide is downregulated in the small intestine of *Eimeria maxima*-infected chickens. Poult. Sci. 90, 1212–1219.

Conway, D.P., Dayton, A.D., Hargis, J.W., 1986. Statistical analysis of coccidial lesion scores. Proc. Georgia Coccidiosis Conf. (Nov. 18-20, 1985) 239–245.

Conway, D.P., Sasai, K., Gaafar, S.M., Smothers, C.D., 1993. Effects of different levels of oocyst inocula of Eimeria acervulina, E. tenella, and E. maxima on plasma constituents, packed cell volume, lesion scores, and performance in chickens. Avian Dis. 37, 118–123.

Cox, C.M., Sumners, L.H., Kim, S., McElroy, A.P., Bedford, M.R., Dalloul, R.A., 2010. Immune responses to dietary β-glucan in broiler chicks during an *Eimeria* challenge. Poult. Sci. 89, 2597–2607.

Da Costa, M.J., Bafundo, K.W., Pesti, G.M., Kimminau, E.A., Cervantes, H.M., 2017. Performance and anticoccidial effects of nicarbazin-fed broilers reared at standard or reduced environmental temperatures. Poult. Sci. 96, 1615–1622.

Dalloul, R.A., Lillehoj, H.S., 2006. Poultry coccidiosis: recent advancements in control measures and vaccine development. Expert Rev. Vaccines 5, 143–163.

Elmusharaf, M.A., Cuppen, J.J., Grooten, H.N.A., Beynen, A.C., 2007. Antagonistic effect of electromagnetic field exposure on coccidiosis infection in broiler chickens. Poult. Sci. 86, 2139–2143.

Faes, C., Molenberghs, G., Aerts, M., Verbeke, G., Kenward, M.G., 2009. The effective sample size and an alternative small-sample degrees-of-freedom method. Am. Stat. 63, 389–399.

FDA-CVM, 1989. NADA 138-952 Original Approval for Narasin and Nicarbazin Type A Article. Available from. https://animaldrugsatfda.fda.gov/adafda/app/search/public/document/downloadFoi/439.

FDA-CVM, 1994. NADA 140-940 Original Approval for Semduramicin Sodium. Available from. https://animaldrugsatfda.fda.gov/adafda/app/search/public/document/downloadFoi/2061.

FDA-CVM, 1999. NADA 140-951 Original Approval for Diclazuril. Available from. https://animaldrugsatfda.fda.gov/adafda/app/search/public/document/downloadFoi/498).

FDA-CVM, 2007. NADA 141-281 Original Approval for Semduramicin. Available from. https://animaldrugsatfda.fda.gov/adafda/app/search/public/document/downloadFoi/844.

FDA-CVM, 2012. Guidance for Industry: Evaluating the Effectiveness of Anticoccidial Drugs in Food-Producing Animals. Available from. https://www.fda.gov/downloads/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/UCM280767.pdf.

Federation of Animal Science Societies, 1999. Guide for the Care and Use of Agricultural Animals in Research and Teaching, 2nd ed. Fed. Anim. Sci. Soc., Champaign, IL.

Holdsworth, P.A., Conway, D.P., McKenzie, M.E., Dayton, A.D., Chapmand, H.D., Mathis, G.F., Skinner, J.T., Mundt, H.C., Williams, R.B., 2004. WAAVP guidelines for evaluating the efficacy of anticoccidial drugs in chickens and turkeys. Vet. Parasitol. 121, 189–212.

Johnson, J., Reid, W.M., 1970. Anticoccidial drugs: lesion scoring techniques in battery and floor-pen experiments with chickens. Exp. Para. 28, 30–36.

Jordan, A., Caldwell, D.J., Klein, J., Coppedge, J., Pohl, S., Fitz-Coy, S., Lee, J.T., 2011. *Eimeria tenella* oocyst shedding and output in cecal or fecal contents following experimental challenge in broilers. Poult. Sci. 90, 990–995.

Kenward, M.G., Roger, J.H., 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53, 983–997.

Kenward, M.G., Roger, J.H., 2009. An improved approximation to the precision of fixed effects from restricted maximum likelihood. Comput. Stat. Data Anal. 53, 2583–2595.

Lan, L., Zuo, B., Ding, H., Huang, Y., Chen, X., Du, A., 2016. Anticoccidial evaluation of a traditional Chinese medicine—*Brucea javanica*—in broilers. Poult. Sci. 95, 811–818.

Lee, J.T., Eckert, N.H., Ameiss, K.A., Stevens, S.M., Anderson, P.N., Anderson, S.M., Barri, A., McElroy, A.P., Danforth, H.D., Caldwell, D.J., 2011. The effect of dietary protein level on performance characteristics of coccidiosis vaccinated and nonvaccinated broilers following mixed-species *Eimeria* challenge. Poult. Sci. 90, 1916–1925.

Lensing, M., van der Klis, J.D., Yoon, I., Moore, D.T., 2012. Efficacy of *Saccharomyces cerevisiae* fermentation product on intestinal health and productivity of coccidian-challenged laying hens. Poult. Sci. 91, 1590–1597.

Li, P., Redden, D.T., 2015. Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC Med. Res. Methodol. 15, 38.

McDougald, L.R., Fitz-Coy, S., 2008. Coccidiosis. In: Saif, Y.M., Fadly, A.M., Glisson, J.R., McDougald, L.R., Nolan, L.K., Swayne, D.E. (Eds.), Diseases of Poultry., 12th ed. Blackwell Publishing, Ames, IA, pp. 1068–1085.

Milliken, G.A., Johnson, D.E., 1992. Analysis of Messy Data. Volume I: Designed Experiments. Chapman & Hall, London.

O'Connell, A.A., Goldstein, J., Rogers, H.J., Peng, C.Y.J., 2008. Multilevel logistic models for dichotomous and ordinal data. In: O'Connell, A.A., McCoach, D.B. (Eds.),

Multilevel Modeling of Educational Data. IAP, Charlotte NC, pp. 199–242.
Pan, Z., Lin, D.Y., 2005. Goodness-of-fit methods for generalized linear mixed models. Biometrics 61, 1000–1009.
Parker, J., Oviedo-Rondón, E.O., Clack, B.A., Clemente-Hernández, S., Osborne, J., Remus, J.C., Kettunen, H., Mäkivuokko, H., Pierson, E.M., 2007. Enzymes as feed additive to aid in responses against *Eimeria* species in coccidia-vaccinated broilers fed corn-soybean meal diets with different protein levels. Poult. Sci. 86, 643–653.
Ritzi, M.M., Abdelrahman, W., Mohnl, M., Dalloul, R.A., 2014. Effects of probiotics and application methods on performance and response of broiler chickens to an *Eimeria* challenge. Poult. Sci. 93, 2772–2778.
SAS Institute Inc, 2011. SAS/STAT® 9.3 User's Guide. SAS Institute Inc., Cary, NC.
Siev, D., 2005. An estimator of intervention effect on disease severity. J. Mod. App. Stat. 4, 500–508.
Stroup, W.W., 2013. Non-normal data in agricultural experiments. Proc. 25th Ann. Conf. App. Stat. Agric., (Apr. 28-30, 2013) 103–135.
Wang, M.L., Suo, X., Gu, J.H., Zhang, W.W., Fang, Q., Wang, X., 2008. Influence of grape seed proanthocyanidin extract in broiler chickens: effect on chicken coccidiosis and antioxidant status. Poult. Sci. 87, 2273–2280.
Williams, R.B., 1999. A compartmentalised model for the estimation of the cost of coccidiosis to the world's chicken production industry. Int. J. Parasitol. 29, 1209–1229.