Trends in Parasitology

Models for measuring anthelmintic drug efficacy for parasitologists

--Manuscript Draft--

Manuscript Number:	TREPAR-D-14-00129R1
Article Type:	Review
Corresponding Author:	Martin Walker, Ph.D. Imperial College London London, UNITED KINGDOM
First Author:	Martin Walker, Ph.D.
Order of Authors:	Martin Walker, Ph.D.
	Thomas S Churcher, Ph.D.
	María-Gloria Basáñez, Ph.D.
Abstract:	Anthelmintic drug efficacy (ADE) is generally estimated as a population average effect, despite drug responses varying among individuals according to a variety of measurable and non-measurable factors. Model-based and/or individual-level analyses are scarce and often methodologically frail. We propose that wider application of marginal and mixed models would offer benefits to the evaluation of ADE. We demonstrate, with a worked example, how model-based analyses: (a) capture the effects of correlation among hierarchically structured longitudinal data on estimates of ADE; (b) permit robust inference on the association of measurable factors with ADE, and (c) enable estimation of variation among individual-level estimates of ADE. The application of modelling approaches is discussed in the context of mass drug administration-based control of human helminthiases.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Imperial College London

Department of Infectious Disease Epidemiology
Imperial College London
School of Public Health
Room G27, Medical School Building
St Mary's campus
Norfolk Place
Paddington
London, UK
W2 1PG
Tel: +44 (0)20 7594 3229 Fax: +44 (0) 20 7402 3927

E-mail: m.walker06@imperial.ac.uk
www.imperial.ac.uk/medicine/people/m.walker06

Martin Walker PhD, MSc

3rd July, 2014

The Editor
Trends in Parasitology

Dear Dr Sherrer,

Please find enclosed an invited manuscript entitled "Models for measuring anthelmintic drug efficacy for parasitologists", which my co-authors and I would like you to consider as a review article for publication in *Trends in Parasitology*. Please also accept our apologies for the delay in this submission.

In this manuscript we review the current statistical approaches used to estimate anthelmintic drug efficacy in the contexts of treating and controlling helminth infections of livestock and of monitoring and evaluating large-scale control programs of human helminthiases. We describe how these methods are: (a) ill-suited for identifying individual and population level explanatory variables; (b) poorly adapted to capture the effects of correlated repeated measures (before and after treatment on the same individual) data on parasite transmission stages (eggs/larvae), and (c) incompatible with individual-level analyses. We illustrate—with an analysis of individual host data on hookworm infection before and one week after treatment with albendazole—how contemporary statistical modelling techniques can overcome these shortcomings, offering superior inference over the current predominating approaches that were developed in the early 1990s. We discuss their application for identifying subtle changes in the distribution of drug responses among individuals indicative of changing drug efficacy, possibly caused by emerging drug resistance, during mass drug administration control of human helminthiases.

We concentrate on reviewing, applying and extending statistical techniques in a didactic manner, inspired by the long-running series of methodological articles published in *Trends in Parasitology* that include: Wilson and Grenfell (1997) *Generalized linear modelling for parasitologists*; Paterson and Lello (2003) *Mixed models: getting the best use of parasitological data*, and Basáñez *et al.* (2004) *Bayesian statistics for parasitologists*. Hence, we believe that our manuscript represents a natural continuation of this series, applying many of the previously advocated methods to address an issue that has been identified as a research priority by the academic community, and by public health organisations and initiatives, notably the Mectizan Donation Program and the World Health Organization's Special Programme for Research and Training in Tropical Diseases.

We look forward to hearing from you,

Martin Walker (first and corresponding author on behalf of all co-authors)

Models for measuring anthelmintic drug efficacy for parasitologists

Martin Walker, Thomas S. Churcher and María-Gloria Basáñez

Department of Infectious Disease Epidemiology, London Centre for Neglected Tropical Disease Research, School of Public Health, Faculty of Medicine (St Mary's Campus), Imperial College London, Norfolk Place, London, W2 1PG, UK

Corresponding author: Walker, M (m.walker06@imperial.ac.uk)

Response to Editor

Editor's comment 0.1: Highlights: Reviews are limited to 4 highlights of 85 characters in length including spaces. Currently you have 5 highlights, and one of these (#4) is too long. Please revise the highlights within these guidelines.

Authors' response 0.1: We have amended this mistake by deleting the first highlight and shortening the fourth to the character limit.

Editor's comment 0.2: Boxes are generally limited to 400 words; while we can be somewhat flexible on this, please try to revise Box 1 in order to reduce the word count (do not exceed 500 words).

Authors' response 0.2: We have shortened the text in Box 1 from 574 words to 499 words.

Editor's comment 0.3: A Glossary would be a very useful addition to the paper. The Glossary allows you to present ideas more succinctly within the main text while adding information to aid non-specialists. For this review, it would be beneficial to define both statistical terminology and concepts, as well as the specifics of some of the equations.

Authors' response 0.3: We appreciate the Editor's helpful suggestion and we have included a glossary in the revised manuscript.

Editor's comment 0.4: Both reviewers have requested the annotated code used in the different methods as supplemental information. This is not required during revision, but if the authors wish to include the code, please do so according to the instructions for authors for supplementary information, which can be found at: http://www.cell.com/trends/parasitology/authors , under the Reviews & Opinions tab.

Authors' response 0.4: We have created a supplementary material document giving the R code, complete with annotation and a didactic description of the critical steps in implementing the models described in the main text.

Editor's comment 0.5: The editor has made only minor changes to text, which can be found in the marked-up copy of the text. Most of these changes are formatting, or for sentences that were too long and lost clarity. Some additional minor changes are requested in the text.

Authors' response 0.5: We thank the Editor for making these changes which improve the flow of the manuscript.

Response to Reviewers

Authors' remarks

We thank the Reviewers for their complimentary comments and thoughtful suggestions which we have endeavoured to incorporate to improve the manuscript. A detail point-by point response follows.

Reviewer 1

Reviewer's comment 1.1: An excellent review that will no doubt prove very useful as a reference to statisticians or statistically minded biologists.

Authors' response 1.1: We thank the referee for these encouraging words.

Reviewer's comment 1.2: More emphasis could be made of the orders of magnitude of difference between different proposed methods and current paradigms in their estimations of intervention efficacy.

Authors' response 1.2: The purpose of taking a modelling approach is primarily to permit greater levels of statistical insight into factors associated with anthelmintic drug efficacy (i.e. using marginal models) or variation among individual responses to anthelmintic drugs (i.e. using mixed models). For simple analyses, such as a population estimate of drug efficacy in a relatively homogenous population, one would expect the magnitude of drug efficacy estimated by a model or simple sample based approach to be very similar, probably with modest differences in estimated magnitudes of uncertainty. Thus, the merit in using a model-based analysis really depends on the question at hand and the degree of desired

statistical insight. We have endeavoured to clarify this in the revised manuscript. For example:

> "Anthelmintic drug efficacy can be estimated by two contrasting approaches. Sample estimates (statistics), or sample efficacies, are calculated directly from data using simple arithmetic operations, without invocation of distributional assumptions. Sample efficacies are easy to calculate (without statistical software) and straightforward to interpret as population average effects. By contrast, model estimated efficacies are derived from fitted statistical models which necessarily apply distributional assumptions to the data and require somewhat more expertise to implement and interpret."
>
> "The most appropriate modelling approach to estimate ADE will depend on the question under consideration and the R&D or M&E context. If interest lies only in the average ADE among population subgroups, within a population as a whole or among different populations, then a marginal model will suffice in giving robust point estimates of (average) ADE and accompanying uncertainties within a well-defined framework that facilitates high quality statistical inference. By contrast, mixed models are suitable for inference at the level of the individual, including the degree of variation among similar individuals (i.e. individuals sharing measured covariates). Yet, rather than representing a dichotomous choice, mixed models complement marginal model analyses, adding a depth of insight from the population to the individual level."

Reviewer's comment 1.3: The authors could release annotated code (ideally script for free software e.g. R) for their different proposed methods along with this article as supporting material. Without this, I fear that these authors will miss out on much of their targeted audience.

Authors' response 1.3: We include the R code in a new supplementary material document (see **Authors' response 0.4).**

Reviewer 2

Reviewer's comment 2.1: This paper provides a helpful overview of the different methods that can be used to estimate anthelmintic drug efficacy (ADE). The argument sometimes feels a bit one-sided in favour of multi-level modelling. Although there are advantages to

this approach, it makes assumptions about the distribution of the data and the random effects that could make it less reliable than simple statistics, such as the risk ratio or ratio geometric means, when these assumptions are not met.

Authors' response 2.1: In the revised manuscript we have balanced the argument by placing more emphasis on the drawbacks of modelling approaches in terms of the need to make more assumptions than their sample based counterparts. For example, our conclusion now starts with:

> "Model-based approaches to estimating ADE hold advantages over sample approaches in terms of power, versatility and strength of statistical inference, albeit requiring more assumptions and a somewhat more complicated formulation."

Reviewer's comment 2.2: It was nice to have a worked example in the paper. I think it would be useful to have the code as an appendix for researchers who have not used these methods before.

Authors' response 2.2: We include the R code in a new supplementary material document (see **Authors' response 0.4).**

Reviewer's comment 2.3: p.2 l.48, Definition of cure rate. Presumably only people with parasites are treated, but this should be made clear in the definition e.g. "the proportion of those positive for parasites that become negative after treatment"?

Authors' response 2.3: We have modified the definition of cure rate in accordance with the Reviewer's suggestion. The definition is also included in the glossary of the revised manuscript (see **Authors' response 0.3**).

Reviewer's comment 2.4: p.3 l.53. The risk ratio and relative risk imply the measure is binary, but for intensity isn't the measure the ratio of means or geometric mean?

Authors' response 2.4: We agree that the statement to which the Reviewer refers is somewhat confusing and so we have removed it from the revised manuscript.

Reviewer comment 2.5: p.4 Eqn 1. It is a bit odd to write that the first formula equals the second since risk of infection is meaningless for the intensity measure.

Authors' response 2.5: We have amended this in the revised manuscript, splitting the original formula into two new formulae, one for the sample cure rate and one for the sample intensity reduction rate. The relevant section now reads:

> "In the context of longitudinal cohort data, frequently used sample estimates of ADE are given by the following generic formulae,

Sample CR = 1 – (prevalence of infection after treatment) /
(prevalence of infection before treatment), Eqn. I

Sample IRR = 1 – (mean intensity of infection after treatment) /
(mean intensity of infection before treatment), Eqn. II

although there are variations of Eqn. II for calculating IRRs [33, 34]."

We have also revised the proceeding section, distinguishing between the terms risk ratio and rate ratio:

"The quotient in Eqn. I represents a risk ratio (RR) of infection after treatment compared to that before treatment. The quotient in Eqn. II is a ratio of means, a rate ratio (assuming equal follow-up times among individuals), which is also abbreviated to RR."

Reviewer comment 2.6: p.5 l.122. The arguments made in this paragraph are not very clear and appear to be inaccurate in places. The difficulties in calculating confidence intervals for the effect measures (risk ratio or ratio of geometric means) are related to the sample size and validity of the distributional assumptions rather than the standard error per se. For example, if on the log scale the data follow a normal distribution with equal variance before and after treatment then a confidence interval based on the t-distribution will be valid irrespective of the magnitude of the variance, sample size or magnitude of SE. For large sample sizes the distribution is less important because of the central limit theorem. I think it would help if you distinguish between risk ratio and ratio of means (or geometric means).

Authors' response 2.6: We have endeavoured to clarify our arguments in this paragraph which we has been facilitated by the Reviewer's suggestion of distinguishing between the risk ratio and the ratio of means (see **Reviewer's comment 2.5** and **Authors' response 2.5**). The revised paragraph now reads:

"The primary limitation of sample estimates is their incompatibility with measured explanatory variables (also referred to as covariates) that may systematically affect estimates of ADE. Sample estimates can be calculated in different population strata, but this is at the expense of sample size. Sample estimates also typically lack accompanying mathematical expressions for calculating uncertainty metrics such as standard errors (SEs) and confidence intervals (CIs) (although see [25] for an exception). There are at least two reasons for this. First, the exact sampling distribution of the ratio of two estimated quantities (the

prevalence of infection, Eqn. I, or the mean intensity of infection, Eqn. II) has a number of awkward statistical properties [35, 36], including undefined statistical moments. Second, the validity of approximations to the sampling distribution that are based on the logarithm of the quotients in Eqn. I and Eqn II [37] are questionable in conjunction with hyper-variable repeated measures parasitological data. Consequently, SEs and CIs are often calculated numerically (e.g. [15, 38-40]) without invoking distributional assumptions, typically using bootstrap methods [41] based on case resampling from the empirically observed distribution of the data. The CIs of the sample IRRs and CRs calculated from the KSD (Table 2) are estimated using one such bootstrap resampling method."

Reviewer's comment 2.7: p.6 l.143. You could mention the difficulties of log transforming the data when there are zeros (they have to be replaced with an arbitrary positive number)

Authors' response 2.7: We thank the Reviewer for this suggestion which we have incorporated into the amended manuscript. The revised section now reads:

"While transformations can be useful, they are seldom undertaken using the most appropriate arithmetic operations [47, 48], which can lead to biased results [49]. More specifically, logarithmic transformations require zero counts to be replaced by an arbitrary positive number, which is often performed in a rather subjective and equivocal manner."

Reviewer's comment 2.8: p.6 l.159. It is confusing to describe the data as both cross-sectional and longitudinal. I would replace "cross-sectionally and longitudinally repeated measures" with "repeated measures".

Authors' response 2.8: We have revised the text in accordance with the Reviewer's suggestion.

Reviewer's comment 2.9: p.7 l.191 I couldn't find any discussion of the sandwich estimator. This is usually used when fitting generalised estimating equations because it ensures that the confidence interval for a parameter estimate is valid even if the correlation structure is incorrectly specified.

Authors' response 2.9: We thank the reviewer for noting this omission. We have included a glossary in the revised manuscript (see **Authors' response 0.3**) that includes a description of sandwich estimators and their importance in conjunction with parameters estimated using the marginal model framework. We also include a footnote in Table 1 and Table 2 stating

that sandwich estimators were used to calculate the confidence intervals of relevant parameter estimates.

Reviewer's comment 2.10: p.8 l.225. You might want to mention in this section that you can be used mixed models to quantify the variability between individuals in the treatment effect, and also to make predictions for the size of this effect in particular individuals.

Authors' response 2.10: We have incorporated this welcome suggestion into the revised manuscript. The relevant section now reads:

> "Mixed models (Box 3) yield individual estimates of treatment responses (regression lines, Figure 1a, 1b) and corresponding individual estimates of ADE (Figure 1c) by considering the means of repeated measures conditional on individual-specific random effects terms. They also permit quantification of the degree of variation in treatment responses among individuals."

Reviewer's comment 2.11: p.8 l.229. This sentence is unclear. You should clarify that it is the parameters of the model rather than covariates that are fixed or random across individuals.

Authors' response 2.11: We thank the Reviewer for highlighting this ambiguity which we have clarified in the revised text:

> "The mixed model formulation divides covariate coefficients specified as exerting random effects into two components"

Reviewer's comment 2.12: p.8 l.234. For a log-link the population average effect is the same as the parameter in the random effects model (see p.137 "Analysis of longitudinal data" by Diggle, Zeger and Liang). The logit link is more problematic.

Authors' response 2.12: We agree with the Reviewer's assertion which applies to the special case of the log-linear random intercepts model. The models we describe include an additional random slope parameter which crucially permits drug efficacy to vary among individuals and thus, in this more general case, the population average effect is not the same the fixed component of a random effects parameter. We have amended the relevant sentence to reflect these subtleties:

> "In particular, and unlike marginal GLMs, fixed effects generally only equate to marginal (population) mean effects on the scale of the linear predictor (Box 1, albeit with the exception of some special cases [51])"

Reviewer's comment 2.13: p.9 l.262. You might consider dropping the technical aspects of this section such as the discussion of the iteratively re-weighted least squares algorithm.

Authors' response 2.13: We thank the Reviewer for this suggestion. We have removed specific reference to the iteratively re-weighted least squares algorithm in our overview of the generalized estimating equation approach. We have also removed similarly specific references to the Laplace and Gauss-Hermite algorithms in the discussion of fitting mixed models.

Reviewer's comment 2.14: p.12 l.350. You could mention the issue of regression to the mean when discussing the association between pre-treatment infection and ADE. I think regression to the mean implies a positive association.

Authors' response 2.14: We have considered this issue in terms of models that include pre-treatment data on infection levels as a covariate of the post-treatment responses. In such a formulation, the estimated drug efficacy would be related to the gradient of the regression line with respect to the pre-treatment data. This is in contrast to our longitudinal formulation where efficacy is a function of the 'gradient' with respect to time (before or after treatment). In the former, (considerable) measurement error in the pre-treatment data would bias the estimated gradient and consequently also the estimated efficacy by the regression to the mean phenomenon. Pre-treatment density-dependent effects on drug efficacy would presumably manifest in a non-linear relationship between the post-treatment response and the pre-treatment covariate. Difficulties in detecting such density dependencies would be confounded by regression towards the mean effects. While these issues are interesting, we have not discussed them in the manuscript because we feel they would detract from the clarity of the arguments already presented by introducing a model with a fundamentally different regression structure.

Reviewer's comment 2.15: p.14 I think your conclusion that model-based approached are superior to sample statistics such as the risk ratio or ratio of geometric means is too strong.

Authors' response 2.15: We have tempered our conclusion in the revised manuscript (see **Authors' response 2.1**).

Reviewer's comment 2.16: Table 1 You could describe the data in a bit more detail. Children came from how many schools? Were there equal numbers of girls and boys? What was the average egg count before treatment and after treatment? How variable were egg counts before and after treatment?

Authors' response 2.16: In the revised manuscript we have included the additional information requested by the Reviewer.

Reviewer's comment 2.17: Table 2 I find it surprising that the model-based confidence intervals are so narrow compared to the "sample" estimates. It would be good if this could be checked somehow e.g. maybe the model could be fitted using a different statistical package?

Authors' response 2.17: We thank the Reviewer for highlighting this surprising result. There was a mistake in calculating the confidence intervals in the original manuscript which we have now corrected. Regarding comparing results from other statistical packages, the models were fitted using the geepack package (Halekoh *et al.* 2006 *J. Stat. Softw.* 15) for R, primarily because this package permits userdefined correlation structures. This allowed different correlations among observations made on the same individual at the same time point (before *or* after treatment) and observations made on the same individual at *different* time points (before *and* after treatment). We are not aware of other packages that offer such functionality. However, we did note that the estimated—and now corrected—uncertainties were very similar between models fitted using the customized correlation structure and those fitted using a simpler exchangeable correlation assumption. This is unsrurprising given the use of sandwich estimators which are robust to misspeciicatiob of the correlation structure (see **Reviewer's comment 2.9** and **Authors' response 2.9**). Further, we were able to confirm that an alternative package, gee (Carey 2002 gee: Generalized Estimation Equation Solver), which has been ported to R from S-PLUS, yielded identical results to geepack under the exchangeable correlation assumption.

Reviewer's comment 2.18: Box 2. If Y_ij is a binary variable (i.e., Y_ij is a Binomail with n=1) then the over-dispersion parameter cannot be estimated. To estimate the parameter Y_ij must be a binomial with n>1.

Authors' response 2.18: We thank the Reviewer for noticing this oversight. We have removed the relevant sentence from the revised manuscript.

Reviewer's comment 2.19: Box 3 I find it a bit misleading to say that e_ij permits extra-Poisson variation when Y_ij follows a Poisson distribution in the model that includes e_ij

Authors' response 2.19: We have revised the sentence to read:

"The term, e_{ij}, is an observation-specific, normally distributed, random effects error term, permitting extra-Poisson variation (overdispersion) among Y_{ij} sharing a common set of covariates [70, 85, 86]."

Reviewer's comment 2.20: Figure 1 I think you need to give more detail about how you fitted the model. What statistical package was used? What Bayesian MCMC techniques were used? What priors were used? How did you check for convergence?

Authors' response 2.20: We agree with the Reviewer's suggestion and have included the additional information in the legend to Figure 1. We have also included our Markov chains as supplementary figures in the supplementary material document (see **Authors' response 0.4**). The relevant section of the legend to Figure 1 now reads:

"The model was fitted using Bayesian Markov chain Monte Carlo (MCMC) techniques implemented with the MCMCglmm package [85] for R [6]. Fixed effects were assigned uninformative normal prior distributions (priors), covariance terms of random effects were assigned uninformative inverse-gamma priors. Three starting values for the MCMC algorithm were assigned in order to assess convergence on the parameter posterior distributions and to check that our conclusions were not sensitive to the choice of starting values."

Models for measuring anthelmintic drug efficacy for parasitologists

Martin Walker, Thomas S. Churcher and María-Gloria Basáñez

Highlights

- Modern statistical methods offer robust ways to estimate anthelmintic drug efficacy
- Marginal models permit robust inference on population covariates of drug efficacy
- Mixed models allow inference on individuals and quantify inter-individual variation
- Models may be used to detect changing drug efficacy during mass drug administration

Manuscript
Click here to download Manuscript: Walker_et_al_Manusc_editor_MW_clean.docx

Models for measuring anthelmintic drug efficacy for parasitologists

Martin Walker, Thomas S. Churcher and María-Gloria Basáñez

Department of Infectious Disease Epidemiology, London Centre for Neglected Tropical Disease Research, School of Public Health, Faculty of Medicine (St Mary's Campus), Imperial College London, Norfolk Place, London, W2 1PG, UK

Corresponding author: Walker, M (m.walker06@imperial.ac.uk)

Anthelmintic drug efficacy (ADE) is generally estimated as a population average effect, despite drug responses varying among individuals according to a variety of measurable and non-measurable factors. Model-based and/or individual-level analyses are scarce and often methodologically frail. We propose that wider application of marginal and mixed models would offer benefits to the evaluation of ADE. We demonstrate, with a worked example, how model-based analyses: (a) capture the effects of correlation among hierarchically structured longitudinal data on estimates of ADE; (b) permit robust inference on the association of measurable factors with ADE, and (c) enable estimation of variation among individual-level estimates of ADE. The application of modelling approaches is discussed in the context of mass drug administration-based control of human helminthiases.

Keywords*:* marginal models; mixed models; mass drug administration; longitudinal data; hookworm; albendazole

The imperative to measure anthelmintic drug efficacy

The effectiveness of treating and controlling human and livestock helminthiases critically depends on the efficacy of anthelmintic drugs. In livestock, the utility of anthelmintics has been severely diminished by the rapid evolution and spread of anthelmintic resistance [1-3]. Despite the lessons learnt from 50 years of somewhat indiscriminate livestock treatment strategies [4], there is due concern that resistance, or at least sub-optimal drug efficacy, could derail the burgeoning global onslaught against helminth infections of humans. The current strategic intervention is based principally on anthelmintic mass drug administration (MDA) [5-8], and is endorsed by the World Health Organization (WHO) in their roadmap on the control and elimination of neglected tropical diseases (NTDs) by 2015 and 2020 [9]. Application of appropriate and powerful statistical methods that enable accurate estimation of anthelmintic drug efficacy (ADE) is a high priority, both for monitoring and evaluation (M&E) of control programmes [10] and for analysing outcomes from clinical trials of the next generation of new [11-13] or repurposed anthelmintics [14-16].

In this article, we show that established extensions of generalized linear models (GLMs) [17] offer a versatile and practical way of estimating ADE both at the population level, in terms of an average effect, but also at the level of the individual host. We review the predominant currently used methods for estimating ADE, many of which were developed in a veterinary context and are based on sample statistics. These are contrasted with modelling approaches using previously published data on hookworm egg counts collected from Kenyan schoolchildren before and one week after treatment with albendazole [18]. These data are summarized in Table 1 and are referred to as the Kenyan schoolchildren dataset, abbreviated to KSD. We discuss the application of modelling approaches, particularly in the context of the M&E of ADE during MDA-based control of human helminthiases.

Cure rates and intensity reduction rates

Anthelmintic drug efficacy is typically, but not exclusively, expressed as either a cure rate (CR) or an intensity reduction rate (IRR). Cure rates (the proportion of those positive for parasites that become parasitologically negative after treatment) are calculated from binary data on the presence or absence of infection; IRRs (the proportional reduction of infection

load effected by the treatment) are calculated from (typically count) data on the intensity of infection. Both quantify reduction in infection levels after treatment (the drug response) as a percentage of infection levels before treatment using longitudinal data from cohort studies.

Cure rates and IRRs can be calculated, in principle, using parasitological, molecular or any other type of data that measure, respectively, infection status or infection intensity. Currently there are few quantitative molecular methods which yield estimates of infection intensity [19], albeit with some notable exceptions. One exception is measurement of circulating filarial antigen (CFA) for *Wuchereria bancrofti* (causing Bancroftian lymphatic filariasis) infection, although because of difficulties in counting adult worms, quantities of CFA have not been calibrated to worm burden, although this has been achieved in animal models [20]. Other examples include measurement of circulating anodic antigen (CAA), which have been correlated with *Schistosoma mansoni* egg output [21], and quantitative polymerase chain reaction (qPCR) for *Ascaris lumbricoides* [22, 23] and hookworm [24] infections. Consequently, molecular diagnostics are mostly used for measuring infection status. Indeed, even for infections where molecular diagnostics do exist, or are undergoing field testing, ADE remains overwhelmingly assessed using data on parasite transmission stages (eggs or larvae). Therefore, we focus on modelling approaches for such parasitological data, although in principle the methods are readily adaptable to other types of data.

Intensity reduction rates are more informative and generally more desirable than CRs, and have been used extensively for assessing anthelmintic efficacy in livestock. Perhaps the most well-known IRR is the faecal egg count reduction (FECR) [25], which is calculated from data on egg counts in faeces. More recently, the WHO has endorsed IRRs for the M&E of human schistosomiasis and soil-transmitted helminthiasis (STH) MDA-based control programmes [26]. Cure rates are often criticised because some anthelmintics are never truly curative (e.g. ivermectin only affects the microfilarial progeny of adult female *Onchocerca volvulus* and exerts only temporary deleterious effects on worm fertility [27]). Also, CRs are less relevant to morbidity reduction since morbidity is, by and large, associated with infection intensity [28], and do not adequately reflect the impact of repeated rounds of treatment in the context of MDA interventions in human populations [29]. In a research and development context (R&D; e.g. clinical trials and epidemiological studies), intensity should

always be measured, permitting calculation of IRRs. Yet for M&E, logistical complexities and the availability of field-ready quantitative diagnostic tools [19] means that data on the presence of absence of infection is common, guaranteeing the continued usefulness of CRs, or other metrics based on binary data.

Contrasting methods of estimation

Anthelmintic drug efficacy can be estimated by two contrasting approaches. Sample estimates (statistics), or sample efficacies, are calculated directly from data using simple arithmetic operations, without invocation of distributional assumptions. Sample efficacies are easy to calculate (without statistical software) and straightforward to interpret as population average effects. By contrast, model estimated efficacies are derived from fitted statistical models which necessarily apply distributional assumptions to the data and require somewhat more expertise to implement and interpret. Yet, the invocation of valid distributional assumptions makes modelling approaches substantially more powerful than the sample approach. Despite this, modelling approaches are seldom used for evaluating ADE, with some exceptions [30-32].

Sample efficacies

In the context of longitudinal cohort data, frequently used sample estimates of ADE are given by the following generic formulae,

Sample CR = 1 – (prevalence of infection after treatment) /
(prevalence of infection before treatment), Eqn. I

Sample IRR = 1 – (mean intensity of infection after treatment) /
(mean intensity of infection before treatment), Eqn. II

although there are variations of Eqn. II for calculating IRRs [33, 34]. The quotient in Eqn. I represents a risk ratio (RR) of infection after treatment compared to that before treatment. The quotient in Eqn. II is a ratio of means, a rate ratio (assuming equal follow-up times among individuals), which is also abbreviated to RR. Both equations can be adapted to incorporate data from untreated controls in case-control study designs [33]. However, because of the ethical considerations of withholding effective treatment from infected humans, and an unwillingness of farmers to leave livestock untreated, control groups are rarely used outside of R&D, particularly trials for novel anthelmintics. Thus, the majority of

data used in estimating ADE for M&E do not comprise treatment and control groups. For example, in the KSD, two simultaneously obtained samples by the Kato-Katz thick smear method were used to obtain two hookworm egg counts per individual before and one week after treatment (Table 1). Thus, the measurements are repeated (repeated measures) both cross-sectionally (two measures per time point) and longitudinally (two time points) per individual. The sample estimates of ADE (sample CRs and sample IRRs) from these data are given in Table 2 and are calculated by applying Eqn. I and Eqn II respectively.

The primary limitation of sample estimates is their incompatibility with measured explanatory variables (also referred to as covariates) that may systematically affect estimates of ADE. Sample estimates can be calculated in different population strata, but this is at the expense of sample size. Sample estimates also typically lack accompanying mathematical expressions for calculating uncertainty metrics such as standard errors (SEs) and confidence intervals (CIs) (although see [25] for an exception). There are at least two reasons for this. First, the exact sampling distribution of the ratio of two estimated quantities (the prevalence of infection, Eqn. I, or the mean intensity of infection, Eqn. II) has a number of awkward statistical properties [35, 36], including undefined statistical moments. Second, the validity of approximations to the sampling distribution that are based on the logarithm of the quotients in Eqn. I and Eqn. II [37] are questionable in conjunction with hyper-variable repeated measures parasitological data. Consequently, SEs and CIs are often calculated numerically (e.g. [15, 38-40]) without invoking distributional assumptions, typically using bootstrap methods [41] based on case resampling from the empirically observed distribution of the data. The CIs of the sample IRRs and CRs calculated from the KSD (Table 2) are estimated using one such bootstrap resampling method.

The lack of straightforward procedures for quantifying uncertainties associated with sample efficacies, combined with the ubiquity of extra-Poisson variation (overdispersion) in parasitological data on infection intensity, has probably contributed to a prevailing focus on robust and invariant estimates of central tendency, i.e. *mean intensity of infection* in Eqn I. This is particularly the case for parasite transmission stages which is compounded (at least) by the overdispersion of adult parasites among hosts [42, 43] and the high degree of sampling variation resulting from the most widely used quantification methods [44-46]. A popular approach is to transform individual intensity data prior to calculation of an IRR. For example, helminth egg count data are frequently log transformed to calculate geometric

means, which are then used to estimate sample IRRs (i.e. replace *mean* by *geometric mean* in Eqn. II) [33, 34]. While transformations can be useful, they are seldom undertaken using the most appropriate arithmetic operations [47, 48], which can lead to biased results [49]. More specifically, logarithmic transformations require zero counts to be replaced by an arbitrary positive number, which is often performed in a rather subjective and equivocal manner. Moreover, inconsistencies in measures of 'average' (arithmetic mean, geometric mean) make it difficult to compare efficacies among populations [50], through time, or against reference values of expected efficacies [26].

Modelling approaches

Modelling approaches to estimating ADE are more powerful than methods based on sample statistics. This means that, subject to a well-specified and judiciously formulated model, model-derived efficacies offer substantially more insight and robust inference than their sample-based counterparts. Modelling approaches naturally accommodate measured covariates, permitting suitable adjustments for the effects of confounders, but also enabling comparison of efficacies among individuals or population sub-groups. Moreover, when formulated using an underlying GLM, but with adaptations to account for the repeated measures [51], models offer a natural way to estimate uncertainties, such as SEs and CIs, and related indicators of statistical significance (*p*-values) as part of computationally efficient and accessible [52, 53] fitting procedures. We first review some of the important properties of GLMs, before describing two specific extensions suitable for capturing the effects of correlations among repeated measures data: marginal models, for population-level inference; and mixed models for inference at the level of the individual host.

Generalized linear models relate, via the so-called link function or link, the expected value (mean) of each observation of the response or dependent variable (either a binary measure of presence or absence of infection or a count measure of infection intensity) to a linear combination of covariates and accompanying regression coefficients. Covariates must include, at least, an indicator of whether an observation is made before or after treatment (hereafter referred to as the observation time) because the coefficient of this variable captures the effect of treatment on infection. In the example analysis of the KSD, we include as explanatory variables, in addition to observation time, the village of observation

(Chiramani or Kidimu), and the interaction between village and the observation time (Box 1).

The link function determines how the coefficient of observation time, or the sum of coefficients of covariates involving observation time if interaction terms are included (Box 1), is interpreted as a measure of ADE. A logarithmic link, which yields a log-linear regression structure, is ideal because the exponent of the relevant coefficient(s) translates directly into an IRR (for count data) or a CR (for binary data). The log-linear structure is a natural choice for count data. This is not so for binary data, although it can be applied [54], albeit with potential problems of model convergence because the modelled expected value of the response variable (the prevalence) is not constrained between 0 and 1. The logit link, which yields a logistic regression structure, is more practicable for binary data. Although at extremely low infection probabilities associated with high CRs, the logit link is non-linear on the scale of the linear predictor [17], which may potentially give misleading results. However, in this formulation, the natural metric of ADE is the cure odds (CO); not the CR (Box 1). Odds ratios can be converted to risk ratios [55and hence COs can be converted to CRs, although uncertainties resulting from this non-linear conversion cannot be adequately approximated by closed formulae [and should be evaluated numerically.

Marginal models for measuring population-level drug efficacy

Marginal models yield population estimates of ADE, analogous to sample efficacies, but estimated within a modelling framework that can efficiently incorporate the effects of covariates (Table 2). An attractive property of marginal models is that their regression coefficients have exactly the same interpretation as the corresponding GLM, albeit with suitably adjusted (inflated) SEs to reflect correlation among repeated measures. The term *marginal* refers to individuals (or more generally, units of observation) sharing a common set of covariates, and reflects that these models consider the marginal means of repeated observations (measures) per individual. For example, in the KSD, the four marginal means are the mean egg counts in children from Chiramani or Kidimu, before or after treatment.

Unlike GLMs for independent (uncorrelated) data, marginal models require explicit specification of the correlation structure among repeated measures. For example, a popular correlation structure for normally distributed time series data is to model the correlation as an exponentially declining function of their degree of temporal separation [56]. Non-

normally distributed data are more restricted in the range of permissible correlation structures [51]. The so-called exchangeable structure is simplest, assuming that the correlation among repeated measures is constant among individuals. However, in the KSD, and more generally in data collected for the purposes of estimating ADE, a more plausible correlation structure may comprise two exchangeable coefficients; one describing the cross-sectional correlation among repeated measures at the same time point (before or after treatment), the other describing the longitudinal correlation among repeated measures made at different time points, but with constant (among individuals) temporal separation (Box 2).

In addition to the correlation structure, the two other components of a marginal model are the same as a standard GLM, namely, a regression structure relating the mean response to a linear combination of covariates and coefficients (the linear predictor) via a link function, and a model for the marginal variance among observations from individuals sharing covariates. Typically, the marginal variance is derived from the variance of a suitable distribution; for count data, the default choice is the Poisson distribution. However, in the KSD, and typically for parasitological intensity data, especially on egg or larval transmission stages, the variance among observations (sharing common covariates) is much greater than the Poisson variance (variance-to-mean ratio >> 1, not least because of the hierarchical structuring of the data). A typical solution is to include a scale parameter which inflates the variance relative to the marginal mean (Box 2).

Mixed models for estimating individual-level drug efficacy

Mixed models (Box 3) yield individual estimates of treatment responses (regression lines, Figure 1a, 1b) and corresponding individual estimates of ADE (Figure 1c) by considering the means of repeated measures conditional on individual-specific random effects terms. They also permit quantification of the degree of variation in treatment responses among individuals. The mixed model formulation divides covariate coefficients specified as exerting random effects into two components; a fixed component which is constant among individuals, and a random component which varies among individuals. The random component reflects the stochastic effect of unmeasured or unmeasurable differences among individuals. Subtleties in the interpretation of these fixed and random components can prove troublesome for mixed models defined within the GLM framework, so-called

generalized linear mixed models (GLMMs). In particular, and unlike marginal GLMs, fixed effects generally only equate to marginal (population) mean effects on the scale of the linear predictor (Box 1, albeit with the exception of some special cases [51]); so, for log-linear mixed models, on the logarithmic scale. This interpretation is lost on the scale of the response because of the non-linearity induced by the link function (Box 1).

In a mixed (or random effects) model, the correlation among repeated measures made on the same individual is implicitly captured by a random effects intercept term which permits individual-specific random deviations from the fixed component of the intercept. For example, in the KSD, one would expect *a priori* that repeated egg counts from the same individual would be correlated and consistently higher or lower than the population average egg count before or after treatment; hence the inclusion of a random intercept term (Box 3). Because assessment of ADE is made using data collected soon after treatment (one week in the KSD) prior to substantive re-infection, this correlation is presumably predominantly driven by individuals' unknown underlying adult worm burden combined with an ADE less than 100%, rather than by predisposition to (re-)infection [57, 58].

A mixed model must treat (at least some) coefficients of covariates that include observation time (before or after treatment) as exerting random effects to enable estimation of individual ADEs; a mixed model that includes only a random intercept term would not possess this quality. In the mixed model analysis of the KSD, the coefficient of observation time is treated as a random effect, while the interaction between observation time and village (Chiramani or Kidimu) is considered as a fixed effect. This means that ADE can vary among individuals within a village, but that any systematic difference in ADE between villages exerts a constant (fixed effect) among individuals (Box 3). Despite remaining limitations in our understanding of anthelmintic pharmacology [59], plausible explanations of why ADE may vary among individuals are drawn readily from the manifestations of physiological and pharmacokinetic-pharmacodynamic (PK-PD) variation on individual responses to antimicrobial agents [60]. Indeed, random effects are ideally suited to capture such differences which are difficult to measure directly.

Parameter estimation

Marginal models

Generalized estimating equations (GEEs) [61, 62] are used to estimate the parameters of marginal models. The principles of GEE estimation are closely tied to methods used for estimating parameters in GLMs [17]. In both cases, estimation involves minimising the weighted sum of squared residuals; the weights being functions of the modelled variance which depends on the nominal choice of parametric distribution (Box 2). In GEE approaches, the weights assigned to individual observations not only depend on the nominal variance, but also on the correlation among observations within a cluster, or in the KSD, within an individual. For perfectly correlated data, the weight per observation is the reciprocal of the number of observations. This gives an effective sample size that equals the number of individuals, not the number of observations. By contrast, for uncorrelated data, each observation within an individual is treated as independent, assigned a relative weight of one, and the effective sample size is equal to the total number of observations. Hence, the value of the GEE approach lies in estimating unbiased (suitably weighted) coefficients and accompanying SEs that adequately reflect the effective sample size of the data [53, 63]. Routines for implementing GEEs are widely available in statistical software packages [64], including, proc genmod in SAS [65]; the gee procedure in SPSS [66]; the xtgee command in Stata [67], and the geepack package [68] in R [69].

Mixed models

There are numerous approaches to mixed model parameter estimation [52]. Unlike GEEs for parameter estimation in marginal models, the (log-) likelihood of the data is integral to all mixed model estimation approaches. Maximum likelihood (ML) estimation involves evaluating and maximising the likelihood directly, although this is slow and computationally unfeasible for more than two or three random effects [70]. A more versatile approach is to maximise an approximation of the likelihood, although it noteworthy that one of the most popular algorithms for this purpose, the penalized quasi-likelihood method, is unsuitable for ADE estimation because it works poorly for binary data or when the mean count is less than about 5 [52]. Generalized linear mixed models are readily implemented in statistical

software packages [52], including proc glimmix in SAS; the meglm command in Stata and the lme4 package in R.

Maximum likelihood optimization methods yield biased estimates of the random effects parameters. This occurs because estimation is conditional on the fixed effects, which are treated as precisely known. Restricted maximum likelihood (REML) estimation is a solution which works by averaging over some of the uncertainty in the fixed effects. However, by either ML or REML optimization, there is no consensus on estimating uncertainty for the variance components, nor the random effects parameters [71]. In the context of estimating individual ADEs, this represents a significant drawback because uncertainty in the variance components is of direct interest for quantifying the distribution of ADE among individuals, and the precision of individual estimates of ADE.

Bayesian Markov chain Monte Carlo (MCMC) estimation approaches offer a powerful and flexible alternative to ML or REML optimization, albeit with the requirement to assign prior distributions to unknown parameters. Markov chain Monte Carlo algorithms generate random samples from the posterior distribution of parameter values for fixed and random effects enabling construction of uncertainty intervals (often so-called Bayesian credible intervals, BCIs) on all parameters and any other derived quantities. Moreover, the approach offers flexibility to estimate parameters of bespoke and highly complex models suitable for analysing highly (hierarchically) structured datasets [52, 72], and a means to incorporate external information on diagnostic performance [73] as informative priors [74]. A variety of software packages are available for implementing MCMC methods within a Bayesian framework including WinBUGS [75], OpenBUGS [76], JAGS [77], and the MCMCglmm package [78] for R.

Which model to use?

The most appropriate modelling approach to estimate ADE will depend on the question under consideration and the R&D or M&E context. If interest lies only in the average ADE among population subgroups, within a population as a whole or among different populations, then a marginal model will suffice in giving robust point estimates of (average) ADE and accompanying uncertainties within a well-defined framework that facilitates high quality statistical inference. By contrast, mixed models are suitable for inference at the level of the individual, including the degree of variation among similar individuals (i.e. individuals

sharing measured covariates). Yet, rather than representing a dichotomous choice, mixed models complement marginal model analyses, adding a depth of insight from the population to the individual level.

The example analysis of the KSD offers a pertinent illustration of the complementarity of the two approaches. The marginal model analysis indicates that albendazole is a highly efficacious treatment of hookworm within the village populations of Chiramani and Kidimu; the IRR in both villages is between 97% and 99% (Table 2). Yet this overlooks the substantial variation among individuals (Figure 1c). More generally, whether such variation constitutes normal variability among individual drug responses or something more insidious, such as reduced parasite susceptibility in certain individuals, is of particular concern to the M&E of chemotherapeutic control, particularly MDA control of human helminthiases. This cannot be determined purely from isolated statistical analyses, but one can envisage an early warning analytical framework based on mixed models whereby: (a) the distribution of estimates among individual responses are tracked through repeated rounds of MDA, looking for shifts that may be indicative of changing ADE; and (b) sub-optimally responding individuals are followed-up and assessed further to identify the cause of their poor response. Such a framework would benefit from characterising baseline distributions of normal responses to anthelmintics using contemporary or historical data from communities predominately naïve to MDA, and thus presumably harbouring maximally susceptible parasites, before any possible emergence of drug resistance. These analytical advances would enhance current WHO recommendations for the M&E of ADE during MDA control of human STHs and schistosomiasis [26]. These recommendations are based on sample estimates of average ADE and therefore are ill-suited for identifying subtle changes in the distribution of drug responses among individuals [79], which may mark the early stages of decreased ADE. In particular, by the time changes in average ADE are detectable, parasite genotypes with relatively low drug susceptibility will likely be at relatively high frequencies [80].

In addition to M&E applications, mixed models are also suited to testing hypotheses in R&D contexts, over and above what is achievable with marginal models alone. For example, ADE is (negatively) associated with pre-treatment levels of infection, both at an individual [81] and population level (i.e. the level of endemicity) [34], and generally increasing with *decreasing* level of infection. This is postulated to occur either via the (non-

linear) relaxation of density-dependent constraints on female worm fecundity or other parasite population parameters following anthelmintic treatment [82], or on interference with the bioavailability of anthelmintics in high intensity infections [7]. Yet the opposite association of increasing efficacy with *increasing* infection level has also been reported [83]. These conflicting conclusions perhaps reflect the difficulty in robustly testing the hypothesis of density-dependent ADE using either sample-based or marginal model-based approaches. In the mixed model approach, however, inference is achieved via the off-diagonal elements of the variance-covariance matrix, which describes the correlation between an individuals' infection level before treatment and their response to treatment (Box 3). In the analysis of the KSD, the mean of this parameter's posterior distribution was -2.45, but with uncertainty intervals (Bayesian credible intervals) which cross 0 (-6.90 to 2.02), indicating no statistically significant association in either direction. This is also reflected visually in in Figure 1b where the intercept and the gradient of the individual fitted regression lines (on the logarithmic scale) show not obvious relationship.

Diagnostic sensitivity

An important enhancement to the modelling approaches described here will be to capture the effects of diagnostic performance on estimates of ADE, particularly on CRs which are believed to be particularly sensitive to imperfect diagnostic sensitivity [84]. A recent meta-analysis [85] concluded that the use of single or (cross-sectional) duplicate Kato-Katz egg counts makes little difference to sample estimates of IRRs, but improves the *accuracy* of prevalence estimates and resulting (sample) CRs. This suggests that diagnostic performance may only influence CRs, but this conclusion is confounded by the fundamentally different ways in which prevalence and intensity are calculated. The customary procedure to estimate helminth prevalence from repeated cross-sectional diagnostic tests is to treat individuals contributing at least on positive test result as infected, discarding other negative results. Therefore, an individual is deemed uninfected only if the diagnostic produces a string of negative results; while an individual is deemed infected if a single positive result is returned. By contrast, intensity is typically calculated either by pooling all measured counts or by taking an average of the mean of repeated counts measured per individual. Therefore, it is unsurprising that the number of identified infected individuals increases with increasing diagnostic repeats (repeated measures), producing seemingly more *accurate* estimates of

prevalence and CRs (assuming 100% diagnostic specificity). However, this fails to capture uncertainty arising from imperfect sensitivity (and possible specificity) on each individual measurement or observation, whether a count or a binary indicator of infection status, which rather requires adaptation of the modelling approaches outlined here to incorporate prior information on diagnostic performance [73, 74].

Concluding remarks

Model-based approaches to estimating ADE hold advantages over sample approaches in terms of power, versatility and strength of statistical inference, albeit requiring more assumptions and a somewhat more complicated formulation. Marginal models are suitable for inference at a population level, suitably inflating SEs of coefficients estimated from longitudinal and cross-sectional repeated measures while simultaneously allowing ADE to vary with individual- or community-level covariates. Mixed models permit greater depth of insight at the individual level, particularly on variation among individual drug responses and estimated ADE. This has an important potential M&E application for detecting shifts in the distribution of drug responses among individuals that may be indicative of decreasing ADE, possibly caused by emerging drug resistance, during chemotherapeutic control of human and livestock helminthiases.

References

1 Wolstenholme, A.J.*, et al.* (2004) Drug resistance in veterinary helminths. *Trends Parasitol.* 20, 469-476

2 Kaplan, R.M. (2004) Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol.* 20, 477-481

3 Hoste, H. and Torres-Acosta, J.F. (2011) Non chemical control of helminths in ruminants: adapting solutions for changing worms in a changing world. *Vet. Parasitol.* 180, 144-154

4 Geerts, S. and Gryseels, B. (2001) Anthelmintic resistance in human helminths: a review. *Trop. Med. Int. Health* 6, 915-921

5 Fenwick, A. and Webster, J.P. (2006) Schistosomiasis: challenges for control, treatment and drug resistance. *Curr. Opin. Infect. Dis.* 19, 577-582

6 McCarthy, J. (2005) Is anthelmintic resistance a threat to the program to eliminate lymphatic filariasis? *Am. J. Trop. Med. Hyg.* 73, 232-233

7 Vercruysse, J.*, et al.* (2011) Is anthelmintic resistance a concern for the control of human soil-transmitted helminths? *Int. J. Parasitol. Drugs Drug Resist.* 1, 14-27

8 Osei-Atweneboana, M.Y.*, et al.* (2011) Phenotypic evidence of emerging ivermectin resistance in *Onchocerca volvulus*. *PLoS Negl. Trop. Dis.* 5, e998

9 World Health Organization (2012) Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation. World Health Organization

10 Albonico, M.*, et al.* (2004) Monitoring drug efficacy and early detection of drug resistance in human soil-transmitted nematodes: a pressing public health agenda for helminth control. *Int. J. Parasitol.* 34, 1205-1210

11 Keiser, J.*, et al.* (2007) Evaluation of the in vivo activity of tribendimidine against Schistosoma mansoni, *Fasciola hepatica*, *Clonorchis sinensis*, and *Opisthorchis viverrini*. *Antmcirob. Agents Ch.* 51, 1096-1098

12 Kaminsky, R.*, et al.* (2008) A new class of anthelmintics effective against drug-resistant nematodes. *Nature* 452, 176-180

13 Steinmann, P.*, et al.* (2008) Tribendimidine and albendazole for treating soil-transmitted helminths, *Strongyloides stercoralis* and *Taenia* spp.: open-label randomized trial. *PLoS Negl. Trop. Dis.* 2, e322

14 Basra, A.*, et al.* (2013) Efficacy of mefloquine intermittent preventive treatment in pregnancy against *Schistosoma haematobium* infection in Gabon: a nested randomized controlled assessor-blinded clinical trial. *Clin. Infect. Dis.* 56, e68-75

15 Speich, B.*, et al.* (2012) Efficacy and safety of nitazoxanide, albendazole, and nitazoxanide-albendazole against *Trichuris trichiura* infection: a randomized controlled trial. *PLoS Negl. Trop. Dis.* 6, e1685

16 Hoerauf, A. (2008) Filariasis: new drugs and new opportunities for lymphatic filariasis and onchocerciasis. *Curr. Opin. Infect. Dis.* 21, 673-681

17 McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models* Chapman & Hall

18 Diawara, A.*, et al.* (2013) Association between response to albendazole treatment and beta-tubulin genotype frequencies in soil-transmitted helminths. *PLoS Negl. Trop. Dis.* 7, e2247

19 McCarthy, J.S.*, et al.* (2012) A research agenda for helminth diseases of humans: diagnostics for control and elimination programmes. *PLoS Negl. Trop. Dis.* 6, e1601

20 Weil, G.J.*, et al.* (1986) Use of parasite antigen-detection to monitor the success of drug-therapy in *Dirofilaria immitis*-infected dogs. *J. Parasitol.* 72, 737-740

21 Van Lieshout, L.*, et al.* (1995) Analysis of worm burden variation in human *Schistosoma mansoni* infections by determination of serum levels of circulating anodic antigen and circulating cathodic antigen. *J. Infect. Dis.* 172, 1336-1342

22 Pecson, B.M.*, et al.* (2006) A real-time PCR method for quantifying viable *Ascaris* eggs using the first internally transcribed spacer region of ribosomal DNA. *Appl. Environ. Microb.* 72, 7864-7872

23 Leles, D.*, et al.* (2009) Molecular diagnosis of ascariasis from human feces and description of a new *Ascaris* sp. genotype in Brazil. *Vet. Parasitol.* 163, 167-170

24 Verweij, J.J.*, et al.* (2007) Simultaneous detection and quantification of *Ancylostoma duodenale*, *Necator americanus*, and *Oesophagostomum bifurcum* in fecal samples using multiplex real-time PCR. *Am. J. Trop. Med. Hyg.* 77, 685-690

25 Coles, G.C.*, et al.* (1992) World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. *Vet. Parasitol.* 44, 35-44

26 World Health Organization (2013) Assessing the efficacy of anthelmintic drugs against schistosomiasis and soil-tranmsitted helminths. World Health Organization

27 Basáñez, M.-G.*, et al.* (2008) Effect of single-dose ivermectin on *Onchocerca volvulus*: a systematic review and meta-analysis. *Lancet Infect. Dis.* 8, 310-322

28 Walker, M.*, et al.* (2012) Density-dependent mortality of the human host in onchocerciasis: relationships between microfilarial load and excess mortality. *PLoS Negl. Trop. Dis.* 6, e1578

29 Montresor, A. (2011) Cure rate is not a valid indicator for assessing drug efficacy and impact of preventive chemotherapy interventions against schistosomiasis and soil-transmitted helminthiasis. *Trans. R. Soc. Trop. Med. Hyg.* 105, 361-363

30 Vidyashankar, A.N.*, et al.* (2007) Statistical approach to measure the efficacy of anthelmintic treatment on horse farms. *Parasitology* 134, 2027-2039

31 Nielsen, M.K.*, et al.* (2013) Hierarchical model for evaluating pyrantel efficacy against strongyle parasites in horses. *Vet. Parasitol.* 197, 614-622

32 Mejia, M.E.*, et al.* (2003) Multispecies and multiple anthelmintic resistance on cattle nematodes in a farm in Argentina: the beginning of high resistance? *Vet. Res.* 34, 461-467

33 Cabaret, J. and Berrag, B. (2004) Faecal egg count reduction test for assessing anthelmintic efficacy: average versus individually based estimations. *Vet Parasitol* 121, 105-113

34 Vercruysse, J.*, et al.* (2011) Assessment of the anthelmintic efficacy of albendazole in school children in seven countries where soil-transmitted helminths are endemic. *PLoS Negl. Trop. Dis.* 5, e948

35 Hinkley, D.V. (1969) On the ratio of two correlated normal random variables. *Biometrika* 56, 635-639

36 Marsaglia, G. (2006) Ratios of Normal Variables. *J Stat Softw* 16

37 Kirkwood, B.R. and Sterne, J.A.C. (2003) *Essential Medical Stastistics*. Blackwell Science

38 Steinmann, P.*, et al.* (2011) Efficacy of single-dose and triple-dose albendazole and mebendazole against soil-transmitted helminths and *Taenia* spp.: a randomized controlled trial. *PloS ONE* 6, e25003

39 Knopp, S.*, et al.* (2010) Albendazole and mebendazole administered alone or in combination with ivermectin against *Trichuris trichiura*: a randomized controlled trial. *Clin. Infect. Dis.* 51, 1420-1428

40 Vidyashankar, A.N.*, et al.* (2012) Statistical and biological considerations in evaluating drug efficacy in equine strongyle parasites using fecal egg count data. *Vet. Parasitol.* 185, 45-56

41 Davison, A.C. and Hinkley, D.V. (1997) *Bootstrap methods and their application*. Cambridge University Press

42 Shaw, D.J. and Dobson, A.P. (1995) Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* 111, S111-S133

43 Anderson, R.M. and May, R.M. (1992) *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press

44 Anderson, R.M. and Schad, G.A. (1985) Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. *Trans. R. Soc. Trop. Med. Hyg.* 79, 812-825

45 Hall, A. (1981) Quantitative variability of nematode egg counts in feces - a study among rural Kenyans. *Trans. R. Soc. Trop. Med. Hyg.* 75, 682-687

46 Utzinger, J.*, et al.* (2001) Relative contribution of day-to-day and intra-specimen variation in faecal egg counts of *Schistosoma mansoni* before and after treatment with praziquantel. *Parasitology* 122, 537-544

513 47 O'Hara, R.B. and Kotze, D.J. (2010) Do not log-transform count data. *Methods Ecol. Evol.*
514 1, 118-122
515 48 Anscombe, F.J. (1948) The transformation of the Poisson, binomial and negative binomial
516 data. *Biometrika* 35, 246-254
517 49 Dobson, R.J*., et al.* (2009) Geometric means provide a biased efficacy result when
518 conducting a gaecal egg count reduction test (FECRT). *Vet. Parasitol.* 161, 162-167
519 50 Keiser, J. and Utzinger, J. (2008) Efficacy of current drugs against soil-transmitted
520 helminth infections: systematic review and meta-analysis. *JAMA* 299, 1937-1948
521 51 Diggle, P.J*., et al.* (2002) *Analysis of Longitudinal Data*. Oxford University Press
522 52 Bolker, B.M*., et al.* (2009) Generalized linear mixed models: a practical guide for ecology
523 and evolution. *Trends Ecol. Evol.* 24, 127-135
524 53 Hanley, J.A*., et al.* (2003) Statistical analysis of correlated data using generalized
525 estimating equations: an orientation. *Am. J. Epidemiol.* 157, 364-375
526 54 Greenland, S. (2004) Model-based estimation of relative risks and other epidemiologic
527 measures in studies of common outcomes and in case-control studies. *Am. J. Epidemiol.*
528 160, 301-305
529 55 Zhang, J. and Yu, K.F. (1998) What's the relative risk? A method of correcting the odds
530 ratios in cohort studies of common outcomes. *J Am. Med. Assoc.* 280, 1690-1691
531 56 Diggle, P.J. (1990) *Times Series: A Biostatistical Introduction*. Oxford University Press
532 57 Quinnell, R.J*., et al.* (2010) Genetic and household determinants of predisposition to
533 human hookworm infection in a Brazilian community. *J. Infect. Dis.* 202, 954-96158 Schad,
534 G.A. and Anderson, R.M. (1985) Predisposition to hookworm infection in humans. *Science*
535 228, 1537-1540
536 59 Geary, T.G*., et al.* (1999) Frontiers in anthelmintic pharmacology. *Vet. Parasitol.* 84, 275-
537 295
538 60 Drusano, G.L. (2004) Antimicrobial pharmacodynamics: critical interactions of 'bug and
539 drug'. *Nat Rev. Microbiol.* 2, 289-300
540 61 Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear
541 models. *Biometrika* 73, 13-22
542 62 Zeger, S.L. and Liang, K.-Y. (1986) Longitudinal data analysis for discrete and continuous
543 outcomes. *Biometrics* 42, 121-130

544 63 Rosner, B. and Milton, R.C. (1988) Significance testing for correlated binary outcome
545 data. *Biometrics* 44, 505-512
546 64 Horton, N.J. and Lipsitz, S.R. (1999) Review of software to fit generalized estimating
547 equation regression models. *Am. Stat.* 53, 160-169
548 65 SAS Institute Inc. (2008) *SAS/STAT 9.2 User's Guide*. SAS Institute Inc.
549 66 SPSS Inc. SPSS Advanced Statistics 17.0. SPSS Inc.
550 67 StataCorp LP (2013) *Stata 13 Longitudinal/Panel Data Reference Manual*. StatCorp LP
551 68 Halekoh, U.*, et al.* (2006) The R package geepack for generalized estimating equations. *J.*
552 *Stat. Softw.* 15
553 69 R Core Team (2013) R: A Language and Environment for Statistical Computing,. R
554 Foundation for Statistical Computing
555 70 Bolker, B.M. (2008) *Ecological Models and Data in R*. Princeton University Press
556 71 Jiang, J. (2007) *Linear and Generalized Linear Mixed Models and Their Applications*.
557 Springer
558 72 Gelman, A.*, et al.* (2004) *Bayesian Data Analysis* Chapman & Hall
559 73 Greenland, S. (1996) Basic methods for sensitivity analysis of biases. *Int. J. Epidemiol.* 25,
560 1107-1116
561 74 Diggle, P.J. (2011) Estimating prevalence using an imperfect test. *Epidemiol. Res. Int.*
562 608719
563 75 Lunn, D.J.*, et al.* (2000) WinBUGS - a Bayesian modelling framework: concepts, structure,
564 and extensibility. *Stat. Comput.* 10, 325-337
565 76 Lunn, D.*, et al.* (2009) The BUGS project: Evolution, critique and future directions. *Stat.*
566 *Med.* 28, 3049-3067
567 77 Plummer, M. (2003) JAGS: A program for analysis of Bayesian graphical models using
568 Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical*
569 *Computing*. Vienna, Austria
570 78 Hadfield, J.D. (2010) MCMC methods for multi-response generalized linear mixed
571 models: the MCMCglmm R package. *J Stat. Softw.* 33, 1-22
572 79 Olliaro, P.*, et al.* (in press) Comparing methods to assess *schistosoma* response to
573 praziquantel treatment: a pooled patient data analysis. *PLoS Negl. Trop. Dis.*
574 80 Churcher, T.S. and Basáñez, M.-G. (2008) Density dependence and the spread of
575 anthelmintic resistance. *Evolution* 62, 528-537

81 Utzinger, J.*, et al.* (2000) Efficacy of praziquantel agaianst *Schistosoma mansoni* with particular consideration for intensity of infection. *Trop. Med. Int. Health*, 771-778

82 Kotze, A.C. and Kopp, S.R. (2008) The potential impact of density dependent fecundity on the use of the faecal egg count reduction test for detecting drug resistance in human hookworms. *PLoS Negl. Trop. Dis.* 2, e297

83 Albonico, M.*, et al.* (2003) Efficacy of mebendazole and levamisole alone or in combination against intestinal nematode infections after repeated targeted mebendazole treatment in Zanzibar. *B. World Health Organ.* 81, 343-352

84 Levecke, B.*, et al.* (2011) A comparison of the sensitivity and fecal egg counts of the McMaster egg counting and Kato-Katz thick smear methods for soil-transmitted helminths. *PLoS Negl. Trop. Dis.* 5, e1201

85 Levecke, B.*, et al.* (2014) Effect of sampling and diagnostic effort on the assessment of schistosomiasis and soil-transmitted helminthiasis and drug efficacy: a meta-analysis of six drug efficacy trials and one epidemiological survey. *Parasitology*, 1-15

86 Davies, H.T.O. (1998) When can odds ratios mislead. *BMJ* 316, 989

87 Wedderburn, R.W.M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61, 439-447

88 Elston, D.A.*, et al.* (2001) Analysis of aggregation, a worked example: numbers of ticks on red grouse chicks. *Parasitology* 122, 563-569

89. Gelman A. (2005) Analysis of variance: why it is more important than ever. *Ann. Stat.* 33 1-53

Boxes

Box 1 Regression model notation, structure and interpretation

Response variable

The response variable is denoted Y_{ij}, where $j = 1, 2, \ldots, m_i$ indicates an observation on individual $i = 1, 2, \ldots, n$. Hence a dataset comprises $\Sigma_i m_i$ realisations of Y_{ij}, denoted y_{ij}. The Kenyan schoolchildren dataset (KSD) comprises $m_i = 4$ Kato-Katz hookworm egg count observations of Y_{ij}—two before treatment, y_{i1} and y_{i2}, and two after treatment, y_{i3} and y_{i4}—from each of $n = 78$ individuals, giving $78 \times (2 + 2) = 312$ observations. The expected value and variance of Y_{ij} are denoted μ_{ij} and v_{ij} respectively. In the analysis of the KSD, μ_{ij} naturally represents a mean egg count. However, for illustration, if Y_{ij} denotes a binary variable indicating the presence or absence of hookworm eggs, then μ_{ij} represents prevalence.

Covariates

Covariates are collected in a $p + 1$ vector of $\mathbf{x}_{ij} = (1, x_{ij1}, x_{ij2}, \ldots, x_{ijp})$ with accompanying coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$, the intercept denoted β_0. This notation is adapted to focus on the time of observation covariate which permits estimation of anthelmintic drug efficacy (ADE). Hence, in the KSD analysis, we indicate observations made before and after treatment by $t_{ij} = 0$ and $t_{ij} = 1$ respectively. The level of infection before treatment is allowed to vary among villages by setting $x_{ij} = 0$ and $x_{ij} = 1$ for Chiramani and Kidimu respectively. We include the interaction $x_{ij} \times t_{ij}$, such that observations made after treatment in Chiramani and Kidimu are indicated by $x_{ij} \times t_{ij} = 0$ and $x_{ij} \times t_{ij} = 1$ respectively. This allows ADE to vary among villages. The complete covariate vector is $\mathbf{x}_{ij} = (1, t_{ij}, x_{ij}, x_{ij} \times t_{ij})$.

Regression structure

The linear combination (linear predictor), $\eta_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}$, relates to μ_{ij} by a link function, $g(\mu_{ij})$. Binary and count data are typically modelled using links that convert μ_{ij} to the logarithmic scale; often $\eta_{ij} = g(\mu_{ij}) = \ln(\mu_{ij})$ and $\eta_{ij} = g(\mu_{ij}) = \text{logit}\,(\mu_{ij}) = \ln[\mu_{ij}/(1 - \mu_{ij})]$, yielding log-linear and logistic regression structures respectively. The logarithmic scale permits relative rather than absolute comparison of covariate effects (i.e. the relative change in infection levels

before and after treatment). Risk or rate ratios (RRs) and odds ratios (ORs) are natural outputs from log-linear and logistic models respectively.

Coefficient interpretation

Anthelmintic drug efficacy is expressed as the proportional *reduction* in infection levels after treatment compared to before treatment; one minus the relative *change*. For log-linear structures, ADE = 1 – RR, representing either an intensity reduction rate (applied to count data) or a cure rate (CR, applied to binary data). For logistic structures applied to binary data, ADE = 1 – OR, but here ADE represents *cure odds* (CO) not CR. Since the OR is less intuitive than the RR (prone to misinterpretation [86]) one can either convert CO to CR [55], or preferably, estimate CR directly using a log-linear regression [54]. The latter approach was used to estimate the model-derived CRs (Table 2) from the KSD.

Box 2 Marginal models for measuring population anthelmintic drug efficacy

A marginal model appropriate for estimating anthelmintic drug efficacy (ADE) is defined by

$g(\mu_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta}$,

$v_{ij} = f(\mu_{ij})$,

$\text{Corr}(Y_{ij}, Y_{ik}) = R$, Eqn. I

where R is a correlation matrix comprising coefficients r_1 and r_2 which indicate, respectively, cross-sectional and longitudinal correlations among Y_{ij} and other notation is as defined in Box 1. Note that the simplest exchangeable correlation structure would comprise only a single correlation coefficient and could be written succinctly as $\text{Corr}(Y_{ij}, Y_{ik}) = r$ for $j \neq k$. Marginal models are fitted to data using generalized estimating equations (GEEs) [61, 62], a quasi-likelihood approach [87] that does not require assumptions on the distribution of Y_{ij} beyond its first two moments [51, 53]; here defined by μ_{ij}, v_{ij} and R.

In the example analysis of the Kenyan schoolchildren dataset (KSD) (Table 2) extra-Poisson variation among marginal observations (observations sharing a set of covariates) was modelled by incorporating a scale parameter φ, such that $f(\mu_{ij}) = \varphi\mu_{ij}$. To facilitate interpretation of the estimated coefficients as a measure of ADE (Box 1), a log-link function, $g(\mu_{ij})$, was used both for when Y_{ij} was considered as a count variable (to estimate intensity reduction rate, IRR) and when it was considered as binary variable (to estimate cure rate, CR). Specifically,

$\text{ADE}_i = 1 - \exp[\beta_1 + \beta_3(x_{ij} \times t_{ij})]$, Eqn. II

noting that the interaction between village and observation time ($x_{ij} \times t_{ij}$, Box 1) allows ADE_i to be different among individuals from different villages; but within a village ADE_i is constant.

Box 3 Mixed models for measuring individual anthelmintic drug efficacy

A mixed model appropriate for estimating individual anthelmintic drug efficacy (ADE) is defined by

$Y_{ij} \sim D(\mu_{ij})$,

$g(\mu_{ij}) = \boldsymbol{\beta}\mathbf{x}_{ij} + \boldsymbol{b}_i\mathbf{z}_{ij} + e_{ij}$,

$\mathbf{b}_i \sim MVN(0, \boldsymbol{\Sigma})$. Eqn I

Here D denotes a parametric distribution for Y_{ij}; typically either a Poisson distribution if Y_{ij} is a count, or a Bernoulli distribution if Y_{ij} is binary. Vector $\mathbf{z}_{ij}$ comprises covariates which are treated as having coefficients that exert random effects; vector $\mathbf{b}_i$ comprises accompanying coefficients. An ubiquitous assumption is that $\mathbf{b}_i$ follows a multivariate normal (MVN) distribution with means 0 and variance-covariance matrix (VCM) $\boldsymbol{\Sigma}$. The term, e_{ij}, is an observation-specific, normally distributed, random effects error term, permitting extra-Poisson variation (overdispersion) among Y_{ij} sharing a common set of covariates [70, 78, 88]. Other notation is defined in Box 1.

In the example mixed model analysis of the Kenyan schoolchildren dataset (KSD) (Figure 1), the intercept, β_0, and coefficient of observation time t_{ij}, β_1, are treated as exerting random effects. Hence, $\mathbf{z}_{ij} = (1, t_{ij})$, $\mathbf{b}_i = (b_{i0}, b_{i1})$ and $\boldsymbol{\Sigma}$ is a 2-by-2 VCM. The coefficient b_{i0} captures correlation among repeated observations per individual; the coefficient b_{i1} permits ADE to vary among individuals. The variance components in the main diagonal of $\boldsymbol{\Sigma}$ (top left to bottom right) quantify, respectively, variation in β_0 and β_1 among individuals; the anti-diagonal elements (top right to bottom left) quantify the covariance. Defining a log-link function, $g(\mu_{ij})$, ensures that the measure of ADE is an intensity reduction rate (IRR) or a cure rate (CR) when fitting to count or binary data respectively (Box 1). Specifically,

$ADE_i = 1 - \exp[\beta_1 + \beta_3(x_{ij} \times t_{ij}) + b_{i1}]$, Eqn II

noting that b_{i1} allows ADE_i to vary among individuals, even among those sharing the same x_{ij}, i.e. within the same village.

Glossary

Arithmetic mean: the sum of a collection of numbers divided by the number in the collection, often simply called the mean or average.

Cure rate (CR): the proportion of individual hosts positive for parasites who become parasitologically negative after treatment.

Exchangeable correlation: correlation among observations measured from a single unit (e.g. multiple parasite counts measured from a single host) that is assumed constant among units (i.e. among hosts).

Fixed and random effects: definitions of fixed and random effects vary with the specific context [89]. Here, a covariate coefficients (parameter) specified as fixed exert a constant effect among individuals while coefficients specified as random effects exert a variable effect among individuals. Parameters exerting random effects include a fixed component which represents the hypothetical effect on the 'average individual' but not necessarily the average effect among individuals.

Generalized estimating equation (GEE): a technique for estimating the parameters of a marginal model fitted to correlated repeated measures (observations). The GEE approach is semi-parametric because it relies on the first two moment of the observed data, but not on the full likelihood.

Generalized linear model (GLM): an extension of the simple linear regression model that is compatible with error distributions from any of the exponential family of probability distributions, including the normal, Poisson, binomial and gamma distributions. The simple linear regression model is a GLM with normally distributed errors.

Generalized linear mixed model (GLMM): an extended GLM that includes a linear predictor comprised of covariate and accompanying coefficients that exert both fixed and random effects.

Geometric mean: a type of mean or average which quantifies the central tendency of a set of numerical observations using the product, rather than the sum, of their values. Typically, geometric means are calculated by first taking the arithmetic mean of the log-transformed values before taking the exponent of the result to transform back onto the original scale.

Hierarchical structure: observations that are nested within units to define a natural hierarchy. Examples are multiple parasite counts measured within a host; multiple hosts

living within a single household; multiple households within a single community. Such structure typically produces correlations among repeated measures (observations) made on the same unit. Thus repeated measures cannot be assumed statistically independent.

Intensity reduction rate (IRR): the infection load after treatment expressed as a proportion of the infection load before treatment.

Linear predictor: the linear combination of covariates and coefficients within a statistical model.

Link function: a function that relates the expected value of a probability distribution to the linear predictor within a statistical model. The natural logarithmic link function is typically used within statistical models for count data. For binomial models, where p is the probability of 'success', the logit link is often used, $\text{logit}(p) = \ln[p / (1 - p)]$.

Longitudinal data: measurements or observations made repeatedly on the same unit (repeated measures) through time, e.g. multiple hookworm egg counts made from the same individual host at different times.

Marginal model: an adaptation of a GLM for use with correlated repeated measures (observations). Marginal refers to the marginal mean of observations from individuals (units) sharing a set of covariates. A marginal model comprises three model components; a marginal mean which depends on covariates; a marginal variance which is typically a function of the marginal mean, and a correlation structure for the repeated measures.

Markov chain Monte Carlo (MCMC): a stochastic algorithm central to Bayesian statistical inference which samples parameter values from the posterior probability distribution by combining information from the likelihood of the observed data and the prior probability distribution of the parameters.

Maximum likelihood (ML) estimation: a framework for estimating parameters of a statistical model by conditioning the probability of the observed data (the likelihood) on unknown parameter values using a probability distribution.

Odds ratio (OR): the ratio of the odds that an outcome occurs given a set of covariates compared to the odds that the outcome occurs in their absence. E.g., the odds of observing (by Kato-Katz) hookworm eggs after one week treatment with albendazole divided by the odds of observing hookworm eggs before treatment.

Overdispersion: the occurrence of variance that is greater than expected based on a simple probability distribution. Extra-Poisson variation is an example of overdispersed count data;

where the variance is greater than expected if the data were Poisson distributed (i.e. variance greater than the mean, $v > \mu$).

Posterior probability distribution (posterior): the probability distribution of a random variable conditional on relevant observed data and prior information. The posterior probability is proportional to the likelihood of the data conditional on a set of parameter values multiplied by prior probability of the parameters. That is, posterior probability $\propto$ likelihood $\times$ prior probability.

Prior probability distribution (prior): the probability distribution of a random variable that captures one's uncertainty before (prior to) observing relevant data. An uninformative or vague prior expresses a high degree of prior uncertainty. This results in a posterior distribution which is dominated by the likelihood of observed data. Conversely, an informative prior will dominate the posterior if the data holds little information on the variable of interest.

Rate ratio (RR): the ratio of the rate of occurrence of an event given a set of covariates compared to the rate of occurrence in their absence. E.g., the average number of hookworm eggs counted (by Kato-Katz) one week after treatment with albendazole divided by the average number of hookworm eggs counted before treatment.

Risk ratio (RR): the ratio of the probability of an event occurring given a set of covariates compared to the probability of the event occurring in their absence. E.g., the probability of observing hookworm eggs (by Kato-Katz) one week after treatment with albendazole divided by the probability of observing hookworm eggs before treatment.

Repeated measures: measurements or observations made repeatedly on the same unit, e.g. multiple hookworm counts measured from the same individual host.

Restricted maximum likelihood (REML) estimation: an alternative to ML estimation for models that include random effects. In REML estimation, the dispersion of the random effects is estimated having averaged over some of the uncertainty in the fixed effects. By contrast, in ML estimation, the fixed effects estimates are treated as precisely correct.

Sample statistic: a quantity calculated from a sample of data using simple mathematical functions which are independent of the sample's distribution.

Sampling distribution: the hypothetical expected distribution of a quantity estimated from a random sample of observations.

Sandwich estimator: a standard error (SE) of an estimated quantity that is robust to misspecifications in the variance-covariance of the error distribution in a statistical model. Sandwich estimators are typically used with marginal models so that SEs (and confidence intervals) are invariant to inaccuracies in the specification of the repeated measures correlation structure. In this context, sandwich estimators are based on the empirically observed variation among unit-level statistics rather than on the model-derived variance-covariance matrix which depends on the assumed correlation structure [53].

Standard error (SE): the standard deviation of the sampling distribution of an estimated statistic (e.g. an arithmetic or geometric mean).

Figure legend

Figure 1 Mixed model estimation of the efficacy of albendazole against hookworm in individual Kenyan schoolchildren

The outputs depicted in panels (a)-(c) are derived from the log-linear mixed model described in Box 3, fitted to the longitudinal hookworm egg count data from the Kenyan schoolchildren dataset (KSD) summarized in Table 1. The model was fitted using Bayesian Markov chain Monte Carlo (MCMC) techniques implemented with the MCMCglmm package [85] for R [6]. Fixed effects were assigned uninformative normal prior distributions (priors), covariance terms of random effects were assigned uninformative inverse-gamma priors. Three starting values for the MCMC algorithm were assigned in order to assess convergence on the parameter posterior distributions and to check that our conclusions were not sensitive to the choice of starting values (see Supplementary material). In panel (a) black circles denote means of the two Kato-Katz egg counts observed per individual, before (BT) and seven days after (AT) treatment with albendazole in the villages of Chiramani and Kidimu. The grey lines join the estimated posterior means of the two Kato-Katz egg counts per individual. The thick black lines join the (population) posterior means marginalised over all individuals. These marginal means are indistinguishable from the sample mean egg counts BT and AT given in Table 1. In panel (b) thin grey lines and thick black lines join, respectively, the estimated individual and (population) marginal posterior means on the natural logarithmic scale of the linear predictor. The error bars are 95% Bayesian credible intervals (BCIs) (analogous to classical confidence intervals, CIs) about the posterior means. Note that after transformation onto the count scale depicted in panel (a), the BCIs become narrower than the thickness of the plotted black lines. In panel (c) the small black circles represent the estimated posterior medians of the individual intensity reduction rates (IRRs; also egg reduction rates, ERRs) and thin horizontal lines are accompanying 95% BCIs. Note the scale of the *x*-axis which indicates that the vast majority of individual IRRs are above 90%, with a minority of substantive 'outliers'.

Tables

Table 1 The Kenyan schoolchildren dataset

Village	Mean (range) age in years	Males / females	*n*[a]	Time[b]	Data[c]	Mean (±95% CI[d])	Variance
Chiramani	10 (7, 13)	15 / 13	56	BT	Binary	0.75 (0.62, 0.85)	0.19
					Count	12.75 (6.50, 25.02)	869
				AT	Binary	0.16 (0.07, 0.32)	0.14
					Count	0.20 (0.09, 0.42)	0.27
Kidimu	13 (9, 18)	20 / 30	200	BT	Binary	0.46 (0.34, 0.58)	0.25
					Count	9.81 (5.05, 19.05)	696
				AT	Binary	0.05 (0.02, 0.13)	0.05
					Count	0.28 (0.06, 1.30)	4.5

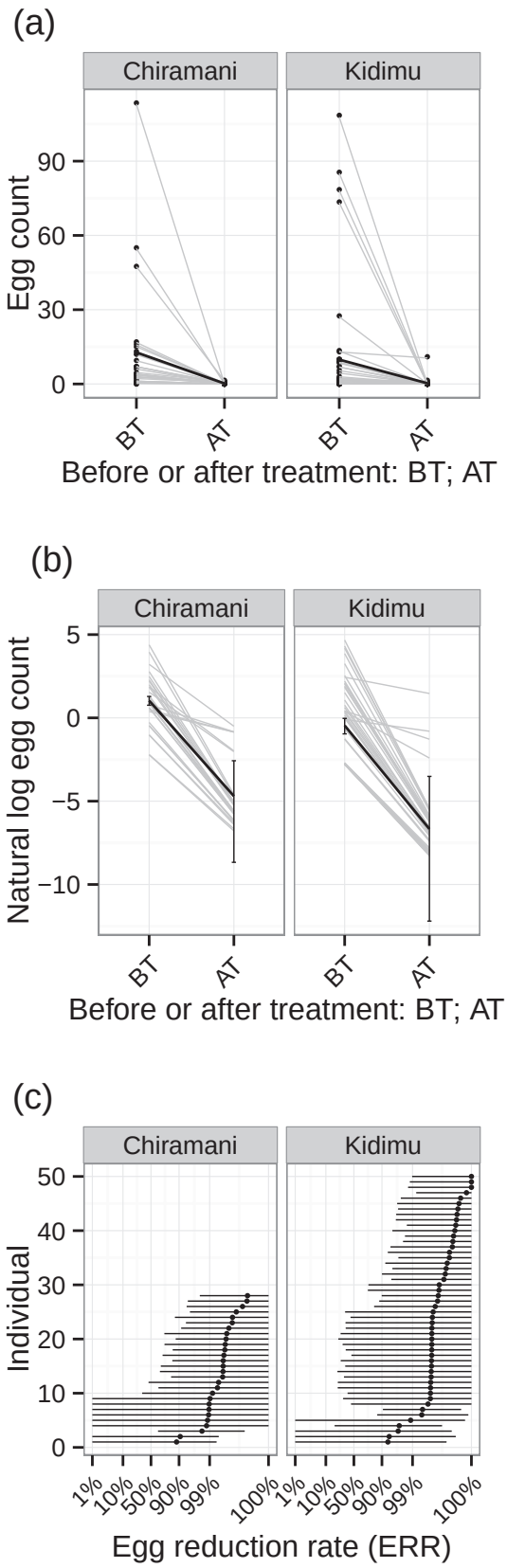
[a] Sample size, two observations per individual at each time point. The effective sample size is smaller because of the correlation among repeated measures.
[b] Before treatment, BT, after treatment, AT.
[c] Data on hookworm eggs in faeces measured by Kato-Katz, either recorded as a count or a binary (presence or absence) variable.
[d] Confidence intervals calculated using a robust 'sandwich' estimator [53].

Table 2 Efficacy of albendazole against hookworm in the Kenya schoolchildren dataset: sample and marginal model efficacies

Village	Method	Efficacy (±95% CI[a])	
		Intensity reduction rate[b]	Cure rate
Chiramani	Sample	98% (88%, 100%)	79% (66%, 97%)
	Model	99% (96%, 99%)	79% (53%, 90%)
Kidimu	Sample	97% (92%, 100%)	89% (70%, 95%)
	Model	97% (85%, 100%)	89% (72%, 96%)

[a] Confidence interval. For sample estimate, calculated by Monte Carlo block resampling percentile bootstrap [41]; for model estimate, calculated using a robust sandwich estimator [53].
[b] Also an egg reduction rate.

Figure


(a)
Chiramani
Kidimu
Egg count
90
60
30
0
BT
AT
BT
AT
Before or after treatment: BT; AT
(b)
Chiramani
Kidimu
Natural log egg count
5
0
−5
−10
BT
AT
BT
AT
Before or after treatment: BT; AT
(c)
Chiramani
Kidimu
Individual
50
40
30
20
10
0
1%
10%
50%
90%
99%
100%
1%
10%
50%
90%
99%
100%
Egg reduction rate (ERR)

Author Supplementary Material

Click here to download Author Supplementary Material: Walker et al Supplementary Material.docx