



# Dose-Response Mixed Models for Repeated Measures – a New Method for Assessment of Dose-Response

Gustaf J. Wellhagen<sup>1,2</sup> • Bengt Hamrén<sup>1</sup> • Maria C. Kjellsson<sup>2</sup> • Magnus Åstrand<sup>1</sup>

Received: 20 March 2020 / Accepted: 14 July 2020 / Published online: 31 July 2020  
© The Author(s) 2020

## ABSTRACT

**Purpose** In this paper we investigated a new method for dose-response analysis of longitudinal data in terms of precision and accuracy using simulations.

**Methods** The new method, called Dose-Response Mixed Models for Repeated Measures (DR-MMRM), combines conventional Mixed Models for Repeated Measures (MMRM) and dose-response modeling. Conventional MMRM can be applied for highly variable repeated measure data and is a way to estimate the drug effect at each visit and dose, however without any assumptions regarding the dose-response shape. Dose-response modeling, on the other hand, utilizes information across dose arms and describes the drug effect as a function of dose. Drug development in chronic kidney disease (CKD) is complicated by many factors, primarily by the slow progression of the disease and lack of predictive biomarkers. Recently, new approaches and biomarkers are being explored to improve efficiency in CKD drug development. Proteinuria, i.e. urinary albumin-to-creatinine ratio (UACR) is increasingly used in dose finding trials in patients with CKD. We use proteinuria to illustrate the benefits of DR-MMRM.

**Results** The DR-MMRM had higher precision than conventional MMRM and less bias than a dose-response model on UACR change from baseline to end-of-study (DR-EOS).

**Conclusions** DR-MMRM is a promising method for dose-response analysis.

**KEY WORDS** Mixed models for repeated measures (MMRM) · dose-response analysis · dose-response mixed models for repeated measures (DR-MMRM) · chronic kidney disease (CKD) · urinary albumin-to-creatinine ratio (UACR)

## BACKGROUND

Traditional dose-response analyses are strongly dependent on the choice of model when the response is highly variable due to unexplained variability. Model-based analyses give higher statistical power than group-wise comparisons [1]. However, many alternative models might provide similar predictions. To select the most appropriate model can be challenging when the signal-to-noise ratio (SNR) is low, since the knowledge that can be gained is proportional to the SNR [2, 3]. The model uncertainty results in large sample size and/or uncertainty in the dose selection for the following study.

Mixed Models for Repeated Measures (MMRM) is an approach to model data with high unexplained variability, making few/no assumptions regarding the response. Instead, at each visit a placebo and treatment response is estimated independent of other visits. This approach has proved superior in terms of precision and accuracy to analysis of (co)variance (ANOVA/ANCOVA) with the end-of-study data in cases with dropout, where the traditional alternative is to use last observation carried forward (LOCF) [4–11].

As an ANCOVA is only based on the information from the last visit, it does not make use of the totality of data. The conventional MMRM technique also ignores some of the information in the data, since each visit and dose level is modeled independently as a factor. By assuming a dose-response relationship where the dose is handled as a continuous variable, information can be shared between dose arms to improve prediction

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11095-020-02882-0>) contains supplementary material, which is available to authorized users.

✉ Gustaf J. Wellhagen  
gustaf.wellhagen@farmbio.uu.se

<sup>1</sup> Clinical and Quantitative Pharmacology, Clinical Pharmacology and Safety Sciences, R&D, AstraZeneca, Gothenburg, Sweden

<sup>2</sup> Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

accuracy. Dose-response analysis is most commonly applied to end-of-study data, thus ignoring shared information across visits. In this manuscript we present a new method for dose-response analysis of longitudinal data. The new method, Dose-Response Mixed Models for Repeated Measures (DR-MMRM), combines conventional MMRM and dose-response modeling, with few assumptions regarding the response but sharing information across doses and visits. In DR-MMRM each visit has a separate placebo and  $E_{\max}$  estimate, while  $ED_{50}$  is a global parameter. Similar methods have been described previously, but then focused on exposure-response analysis of QT interval prolongation [12, 13]. These studies also handled  $E_{\max}$  as a global parameter, while DR-MMRM can account for different time-courses of the drug effect.

Chronic kidney disease (CKD) is a global burden, which was estimated to cause 1.2 million deaths in 2015, an increase of over 30% compared to 2005 [14]. The disease progression of CKD is often slow with few initial symptoms. Traditionally, drug development in CKD has focused on estimated glomerular filtration rate (eGFR). It is recognized that clinical trials within the CKD field face many challenges, related to the slow progression of the disease but also the selection of sensitive clinical endpoint, patient recruitment and influence of comorbidities [15]. Recently, the National Kidney Foundation (NKF), held a workshop together with the FDA and EMA where new approaches to improve efficiency in CKD drug development were discussed [16]. New opportunities were highlighted including the validity of urinary albumin-to-creatinine ratio (UACR) as an important clinical endpoint [16]. UACR is commonly used as a dose-finding clinical endpoint in CKD, instead of eGFR, because it typically has a faster response that allows for shorter studies [16, 17]. However, UACR is highly variable both between and within subjects. The new method, DR-MMRM, was applied to simulated UACR as a clinical endpoint to demonstrate its usefulness.

## OBJECTIVES

To investigate the precision and accuracy of placebo-adjusted change from baseline of simulated UACR ( $\Delta\Delta$ UACR) with different methods: dose-response on end-of-study data (DR-EOS), conventional MMRM and dose-response MMRM (DR-MMRM).

## METHODS

### Previous Studies

Inspiration for the simulation study setup and the estimates of variability and correlation between residuals were taken from a previous study of dapagliflozin where UACR was measured over time (data on file,  $n = 251$ ). This was a three-arm, double-

blinded, placebo-controlled, parallel group, randomized clinical trial in a population with type 2 diabetes and mild renal impairment. Scheduled visits occurred at baseline and weeks 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24. Several models for the correlation of residuals (autoregressive of different order, independence, etc.) in the placebo arm of this study ( $n = 84$ ) were evaluated and compared based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as well as by inspecting the model population predictions. As all models provided similar fit (not shown), a 1st order autoregressive model was chosen for simulations.

### Simulations

The model for the previous data was used as a basis for the simulations. A true dose-response model following an  $E_{\max}$  shape with a range of  $ED_{50}$  (2–128 mg) was assumed, see Fig. 1. The  $E_{\max}$  parameter was set so that the highest dose achieved a reduction of 40% in UACR at the last visit. The baseline UACR was set to a mean of 5.63 log(mg/g) and the baseline variability ( $\omega$ ) was set to 0.3716 log(mg/g), as found in the previous study. As the simulated values were change from baseline, the actual value of the baseline had no importance for the results. A 1st order autoregressive model (AR [1]) with  $\rho = 0.226/\text{visit}$  and a residual error magnitude ( $\sigma$ ) of 0.50 log(mg/g) was assumed for the correlation of residuals, again based on previous findings.

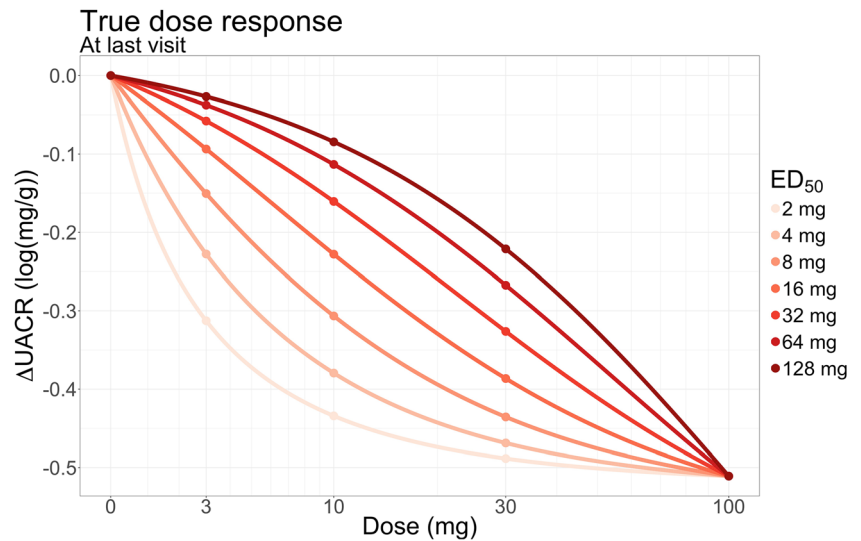
The simulated values of UACR were log-transformed changes from baseline ( $\Delta$ UACR). Three different time-courses of the drug effects were investigated: direct, exponential or linear. For each combination of  $ED_{50}$  and drug effect time-course, a number ( $n = 1000$ ) of 16-week studies were simulated, totaling 21,000 studies. Samples were taken at week -2, -1, 0, 2, 4, 6, 8, 10, 12, 14, 15 and 16. There was 1 placebo arm and 3 or 4 dose arms (10, 30 and 100 mg or 3, 10, 30 and 100 mg). The baseline was averaged across the visits at week -2, -1 and 0.

The sample size was set to provide 95% statistical power for detecting a 40% reduction in  $\Delta$ UACR between the highest dose arm and placebo at end-of-study using a  $t$ -test, given the variability that was previously assumed. This led to the same sample size ( $n = 39$  per arm) in all trials, while varying the  $E_{\max}$  parameter at the end-of-study visit.

The  $E_{\max}$  parameter also changed with respect to different time-courses of the drug effect. For the exponential and linear time-courses the 40% reduction was reached at the last visit with the highest dose. The direct effect-profile reached the 40% reduction already at the first visit post baseline with the highest dose. The following equations (Eq. 1–2) were used to calculate the  $E_{\max}$  at each visit for each  $ED_{50}$  and time-course:

$$E_{\max,k} = \log(0.6) \times \frac{100 + ED_{50,k}}{100} \quad (1)$$

**Fig. 1** True dose-response relationship used in the simulations. The effect is shown at the last visit



$$E_{max,k,t} = \begin{cases} E_{max,k} \times 1, & \text{Direct effect} \\ E_{max,k} \times \left(1 - e^{-\frac{\log(2)}{1.75} \times t}\right), & \text{Exponential effect} \\ E_{max,k} \times \frac{t}{16}, & \text{Linear effect} \end{cases} \quad (2)$$

where  $E_{max,k,t}$  is the change in UACR with  $ED_{50,k}$  (2, 4, 8, 16, 32, 64 or 128 mg) at  $t$  weeks after start of treatment.

**Models**

Three methods were investigated. First, the dose-response relationship on change from baseline at end-of-study data was fitted only on change from baseline data at the last visit (week 16). Then, the placebo response,  $ED_{50}$  and  $E_{max}$  were estimated. Secondly, in the conventional MMRM analysis, each visit and dose arm had a separate estimate of both the placebo response and the response in each of the different dose arms. No dose-response through  $ED_{50}$  or  $E_{max}$  was estimated then. Lastly, in the dose-response MMRM analysis, each visit had a separate estimate of the placebo response and  $E_{max}$  but all visits shared the  $ED_{50}$  parameter. This saved 2 or 3 parameters per visit (depending on number of dose arms) but added 1 global. The MMRM methods were estimated with a 1st order autoregressive covariance matrix for the correlation of residuals.

A schematic of the simulation and estimation setup is shown in Appendix Fig. 1.

**Evaluation Metrics**

The precision was assessed through the magnitude of the uncertainty of the estimated  $\Delta\Delta UACR$ . Precision was also evaluated in the same way for the  $E_{max}$  and  $ED_{50}$  estimates in the dose-response informed methods that contained these parameters (DR-EOS and DR-MMRM).

The accuracy of  $\Delta\Delta UACR$  was assessed through relative bias, computed as the absolute bias at the last visit compared to the true values, relative to the maximal effect (40% reduction of UACR). The bias for  $ED_{50}$  and  $E_{max}$  was assessed through absolute bias for the dose-response informed methods.

For  $\Delta\Delta UACR$ , the root mean squared error (RMSE), which is a weighted measure of precision and accuracy, was also calculated at the last visit through Eq. 3:

$$RMSE = \sqrt{sd^2 + bias^2} \quad (3)$$

**Software and Implementation**

Both simulations and estimations were performed in R version 3.2.4 [18]. The autocorrelation models of previous studies were fit using the nlme() package with the lme() function [19]. The nlme() package was also used to fit the dose-response MMRM relationship and the MCPMod() package was used to fit the dose-response relationship on end-of-study data [20].

In order to directly get the standard deviation and partly to stabilize the model, the dose-response MMRM was parameterized so that the estimates from each dose arm were obtained separately and the standard  $E_{max}$  equation was changed so the effect was given at dose = 1 – here called  $E_{dose}$  – instead of  $E_{max}$ , reformulating the model as in Eq. 4:

$$y = Plc + E_{dose} \times \frac{dose}{ED_{50} + dose} \times \frac{ED_{50} + dose_m}{dose_m} \quad (4)$$

where  $y$  are the responses,  $Plc$  the placebo responses and  $dose_m$  is the dose in question (3, 10, 30 or 100 mg), collapsing to Eq. 5 when dose =  $dose_m$ :

$$y = Plc + E_{dose} \quad (5)$$

The estimates were only saved when dose = dose<sub>m</sub> was fulfilled. This meant fitting the same data 3 or 4 times to get estimates for each dose arm, while obtaining the same estimate for ED<sub>50</sub> – and also E<sub>max</sub>, which could be back-calculated. The number of times the models could not converge was recorded. Since the dose-response MMRM was performed with several attempts, a complete convergence failure was defined as when all parameterizations for one study failed simultaneously.

For full details on the implementation, R code with a working example is provided in Appendix File 1.

## RESULTS

The true and estimated  $\Delta\Delta$ UACR with 95% confidence intervals for 3 example studies with linear drug effect time-course and ED<sub>50</sub> = 32 mg are shown in Fig. 2. Due to the large number of simulated scenarios, the  $\Delta\Delta$ UACR for more cases are not shown. The illustrational example in Fig. 2 shows how all 3 methods are comparable in precision for the highest dose at the end-of-study visit. The dose-response informed methods have shorter confidence intervals for the lower doses, and the DR-MMRM has the highest precision (lowest standard errors) in all cases. The same figure, but for the simulation scenario where only 3 doses and placebo were simulated is shown in Appendix Fig. 2.

In Fig. 3, the bias at the last visit is shown in relation to the maximal effect (a reduction of 40% in UACR), stratified by time-course of the effect and the dose levels. The gray area indicates the expected variability ( $\pm 2$  SD) of estimates known to be unbiased given the simulation setup. The conventional MMRM is unbiased (if any dropout is at random) [4, 8] and generally showed estimated bias within the gray area which was the expected outcome: only 3 of 84 estimates (4%) fell outside this range. The expected outcome was 5%. All methods had low bias for the highest dose. For the dose-response informed methods, the bias increased with increasing ED<sub>50</sub> which may be due to that the studied dose range (3–100 mg) covered a smaller part of the true dose-response relationship. The DR-MMRM had 18 of 84 estimates (21%) outside the expected theoretical range. The DR-EOS always had the strongest bias, which at most was around 6% of the maximal UACR effect, and 53 of 84 estimates (63%) had larger bias than the expected theoretical range. For DR-MMRM the highest bias was around 3% of the maximal UACR effect.

In Fig. 4, the RMSE for the last visit is shown, stratified by time-course of the drug effect and the dose levels. As stated in Eq. 3, this is a composite measure of variance and bias, and all methods had constant RMSE across different values of ED<sub>50</sub>

for the highest dose, where conventional MMRM was slightly higher than the other methods. The conventional MMRM always had a relatively constant RMSE for different time-courses of the drug effect as well as dose levels, which was higher than the dose-response informed methods and on par with the theoretical expected RMSE,  $\frac{\sqrt{\omega^2 + \sigma^2}}{\sqrt{n}} \times \sqrt{2} \approx 0.14 \log\left(\frac{mg}{g}\right)$ . The dose-response informed methods had decreasing RMSE with increasing ED<sub>50</sub> for the lower doses, dose-response MMRM RMSE was lowest at a dose of 3 mg and ED<sub>50</sub> of 128 mg – the lowest dose and highest ED<sub>50</sub>. The DR-MMRM had the lowest RMSE in all cases.

In Fig. 5, the median estimates of ED<sub>50</sub> from DR-EOS and DR-MMRM are shown, exemplified with direct time-course of the drug effect. The median estimates were generally well aligned with the true value that was used for simulation, except for the two highest ED<sub>50</sub> (64 and 128 mg), where a slight bias could be seen and the uncertainty of the estimates were higher. The median was used since the mean was heavily influenced by estimates that were on the upper boundary. The estimates for all time-courses of the drug effect are shown in Appendix Fig. 3.

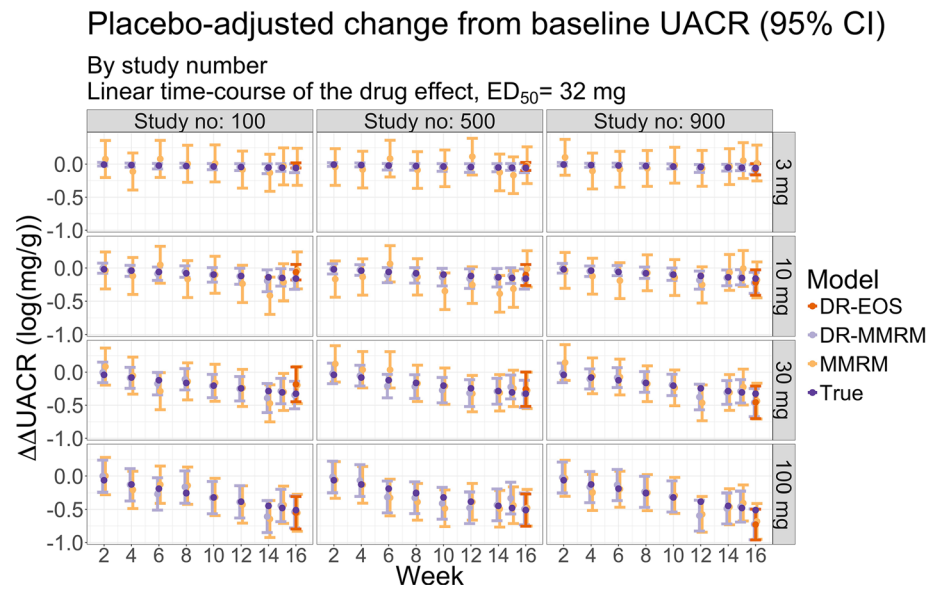
In Fig. 6, the median estimates of the E<sub>max</sub> parameter for the last visit from DR-EOS and DR-MMRM are shown, exemplified with direct time-course of the drug effect. The 2.5th and 97.5th percentiles of the estimates are shown with error bars, which indicate that the uncertainty of the estimates was increasing with increasing ED<sub>50</sub>. The median was used for plotting since some extreme values influenced the mean to overestimate the E<sub>max</sub> parameter for both methods. Instead, the median E<sub>max</sub> was overestimated by DR-EOS but appeared unbiased for DR-MMRM. The percentiles were always wider for DR-EOS than DR-MMRM, however that may be a feature of the constraints in the settings as many values were at the upper boundary. The estimates for all time-courses of the drug effect are shown in Appendix Fig. 4.

In the simulation scenario with 3 doses, at least one of the 3 parameterizations converged successfully for all 21,000 simulated studies and provided an estimate of ED<sub>50</sub> and E<sub>max</sub>. When 4 dose arms were used, it was possible to get an estimate of ED<sub>50</sub> and E<sub>max</sub> for all cases, except 2 of the 21,000 simulated studies where none of the 4 parameterizations converged. The failed attempts both occurred for an ED<sub>50</sub> of 128 mg, with either exponential or direct time-course of the drug effect. For DR-EOS and conventional MMRM, estimation was always successful. A summary of the pros and cons of the investigated methods is shown in Table 1.

## DISCUSSION

Conventional MMRM analyses are inherently unbiased if dropout is at random, which is not the case for the dose-

**Fig. 2** Placebo-adjusted  $\Delta$ UACR and 95% CI for 3 studies with linear time-course of the drug effect where  $ED_{50} = 32$  mg, stratified by dose. The true  $\Delta\Delta$ UACR is also shown

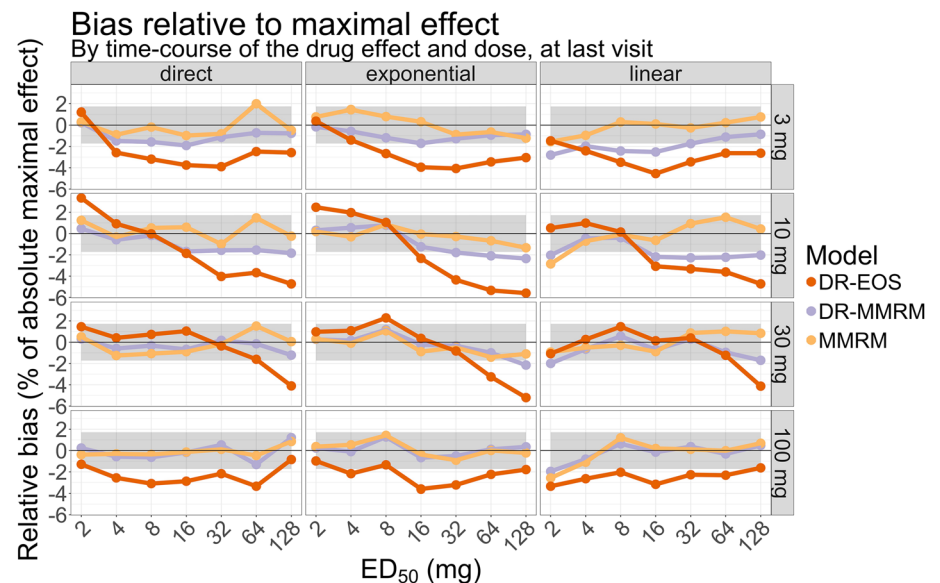


response informed methods. None of the investigated methods had a large bias at any point – it was never more than 6% of the maximal drug effect on UACR. The DR-MMRM had lower bias than the DR-EOS, which can be explained by the fact that they use different amounts of data. The MMRM methods were also mostly within the expected range of bias following the study design and simulation setup, while the DR-EOS had a clear trend of increasing bias when  $ED_{50}$  increased.

Dose-response at end-of-study or dose-response MMRM offers an improvement in precision over conventional MMRM analyses. The improvement is mostly seen for lower doses. The precision for the highest dose is comparable for all

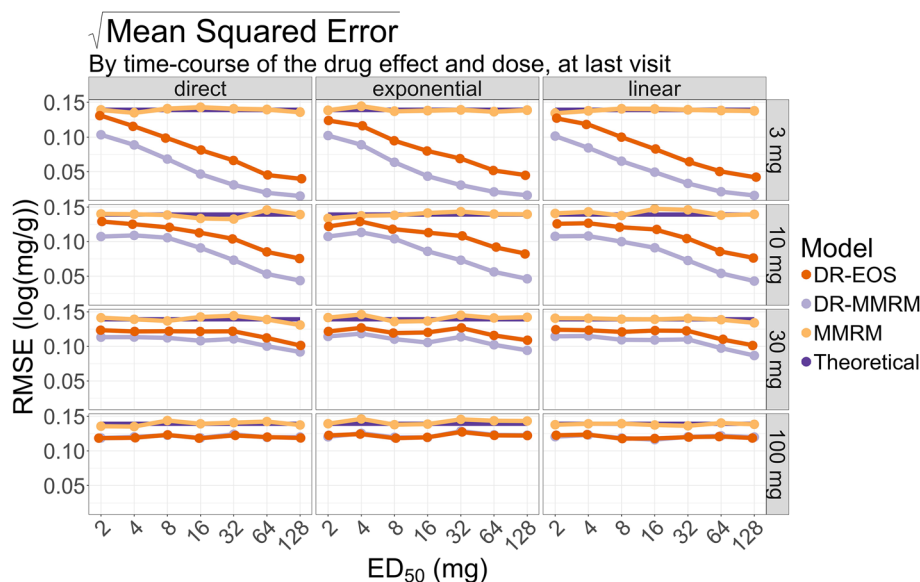
methods, but slightly higher for MMRM. Since the precision for the highest dose level was not improved to any large degree by adding the 4th dose (3 mg), it seems that the highest dose contains most information regarding the dose-response, and certainly so with increasing  $ED_{50}$ , as lower doses are not informing as much about the  $E_{max}$  and  $ED_{50}$  parameters. As expected, DR-EOS performs well (on par with DR-MMRM in terms of precision for the highest dose) since there was no dropout in the simulated data. Should we have included dropout, we know an ANCOVA with LOCF, which resembles the DR-EOS, will be biased [4, 5]. The relatively frequent observations in this study favors the DR-MMRM method as more information regarding  $ED_{50}$  is utilized.

**Fig. 3** Bias relative to the absolute maximal effect at last visit, stratified by time-course of the drug effect and dose level. The gray area ( $\pm 1.7\%$ ) indicates the expected variability ( $\pm 2$  SD) of estimates known to be unbiased given the simulation setup of 1000 simulations for each scenario





**Fig. 4** The RMSE for the investigated models with varying ED<sub>50</sub>, stratified by time-course of the drug effect and the dose levels. The theoretical RMSE following the study design is also shown



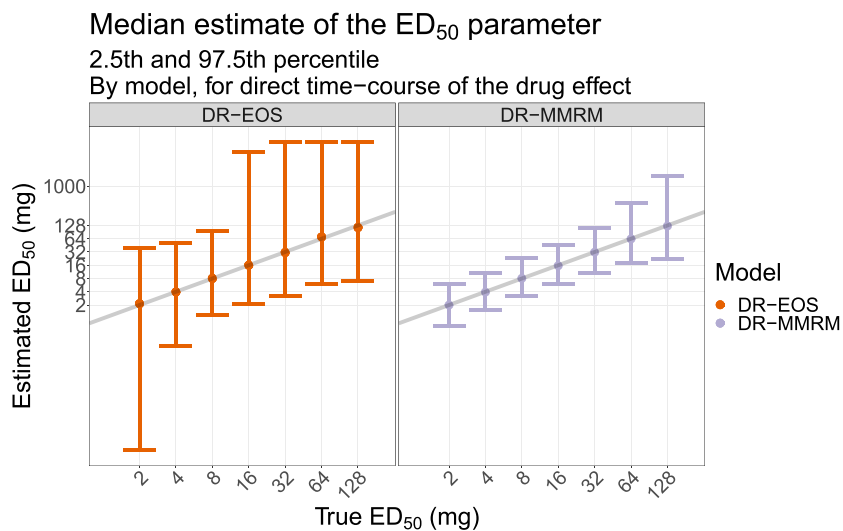
For the correlation of residuals most conventional MMRM analyses estimate an unconstrained covariance matrix whereas we used a 1st order autoregressive covariance matrix in this work. However, again since we simulate without drop-out, this does not alter the comparisons presented here unless the variances would have varied substantially across visits.

DR-MMRM always had lower RMSE than MMRM. RMSE is the weighted measure of both precision and bias but in this example the bias was very low for all methods investigated, so RMSE was effectively a measure of the precision. Higher precision, especially in the lower dose range, allows for better predictions and interpolations to doses that have not been studied. Phase 3 clinical trials failures are predominantly due to lack of efficacy or safety concerns [21]. Choosing the correct dose could increase the success rates, which is why proper characterization of dose-response relationships are of high importance. Phase 2 trials typically aim to

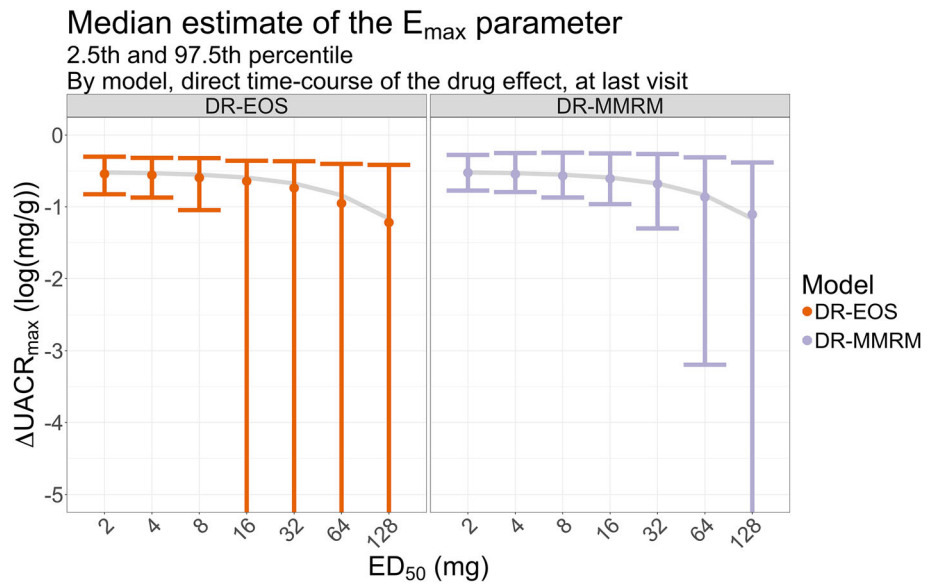
both establish proof-of-concept (PoC) as well as to characterize the dose-response relationship, but the number of dose arms is often limited so that the latter part is difficult. Uncertainty in the dose-response means that assumptions regarding the response in phase 3 need to be made, which is a risk that is preferably minimized. The positive trade-off between precision and bias (lowered RMSE) is encouraging and shows that the DR-MMRM methodology should be considered when evaluating dose-response in dose-finding trials when the endpoint is a repeated measure. Even when the simulated ED<sub>50</sub> was low there was a marked reduction in RMSE in favor of DR-MMRM, which only grew with increasing ED<sub>50</sub>.

In this work the highest simulated ED<sub>50</sub> was 128 mg, while the highest dose was 100 mg, i.e. well below the ED<sub>50</sub>. For this scenario, the dose range was not sufficient to characterize an E<sub>max</sub> relationship in dose-response. This is evident from Figs. 5

**Fig. 5** The median estimated ED<sub>50</sub> with 2.5th and 97.5th percentiles vs. true ED<sub>50</sub> for dose-response on end-of-study data and dose-response MMRM, exemplified by a direct time-course of the drug effect



**Fig. 6** The median estimate of the  $E_{\max}$  parameter with 2.5th and 97.5th percentiles for dose-response on end-of-study data and dose-response MMRM, stratified by time-course of the drug effect and  $ED_{50}$ . The true  $E_{\max}$  is also shown in the gray line. The y axis was cut at  $-5$  for visibility, error bars for higher true  $ED_{50}$  extend well below the range of the graph



and 6, where the uncertainty of the estimated  $ED_{50}$  and  $E_{\max}$  parameters increases as the simulated  $ED_{50}$  increases. The limited dose range in combination with the selected model ( $E_{\max}$ ) is the root cause for this bias when the simulated  $ED_{50}$  is high. A linear or perhaps log-linear model would likely have performed equally well in these cases. Model averaging could have been applied to avoid selection bias, which uses a weighted average of several proposed structural models [22]. That could be a future extension to this work, together with further investigations of the impact of model misspecification.

As previously discussed, it is important to choose doses wisely in the design of phase 2 studies, so that the dose-response relationship can be properly characterized, i.e. studying a sufficient number and wide enough range of doses to avoid bias. In this example there was roughly a 3-fold difference between 3 or 4 consecutive doses (4 or 5 including placebo). As clinical trials are usually optimized for sample size, this fold difference could be increased to explore a larger range with fewer doses, depending on earlier knowledge of a

compound's effectiveness. It is worth to mention that the ratio between  $ED_{90}$  and  $ED_{10}$  is 81 under an  $E_{\max}$  function without sigmoidicity, a range that is seldom covered by phase 2 trials. If the confidence in a proposed  $ED_{50}$  is high, the fold difference may not need to be as wide. In this exercise we simulated without sigmoidicity in the  $E_{\max}$  equation so that practically any dose larger than 0 would illicit a detectable effect, meaning that all doses carried some information on the dose-response relationship. When the true relationship is sigmoidal, further doses might be warranted to estimate  $ED_{50}$  at any meaningful certainty if there is no/little previous knowledge to learn from. Also,  $E_{\max}$  and  $ED_{50}$  are correlated parameters, and it can be seen that they are biased in the same direction. Again, the setup without sigmoid  $E_{\max}$  relationship makes this finding expected.

The time-course of the effect is not affecting the conclusions, because the same trends are visible in all the scenarios. The simulations were performed so that the full effect was reached at least by the end of the study (at visit 12).

**Table 1** Summary of pros and cons for the investigated methods

	Pros	Cons
DR-EOS	Simple, can be used for extrapolation or interpolation	Disregards information from all other visits
MMRM	Always unbiased, uses information from all visits	Parameter heavy, cannot extrapolate or interpolate
DR-MMRM	Makes use of shared information from all visits and dose arms, higher $ED_{50}$ precision than DR-EOS, can be used for extrapolation or interpolation, overall best precision of $\Delta\Delta UACR$	Assumes the same $ED_{50}$ for all visits

DR-EOS, Dose-Response at End-of-Study; DR-MMRM, Dose-Response Mixed Models for Repeated Measures;  $ED_{50}$ , effective dose giving 50% of maximal effect; MMRM, Mixed Models for Repeated Measures;  $\Delta\Delta UACR$ , Placebo-adjusted change from baseline of Urinary Albumin-to-Creatinine Ratio

One potential caveat of the MMRM methods is that they handle time as discrete values/fixed effects – each visit is independent of all other visits. If the study protocol is not followed, e.g. if a sample is taken on the wrong date, this can lead to a bias. However, in phase 2 trials such as those we have investigated in this work, subjects are typically followed closely and the study protocol adhered to. But when planned and actual visits deviate from one another there will be benefits of modeling time continuously. For the conventional MMRM method, the information about a dose level in relation to other doses is also not considered, which DR-MMRM amends.

As the same model ( $E_{\max}$  without sigmoidicity) was used for both simulation and estimation, the results are possibly different in terms of power, precision and accuracy than they would be if another model had been used for simulation (e.g. simulating with sigmoidicity but estimating without, or an even more misspecified model). It should be noted that if a sigmoid model had been used for both simulation and estimation, the comparison with conventional MMRM probably would be less favorable. However, this does not affect the conclusions regarding DR-MMRM versus DR-EOS. Also, both the DR-EOS and DR-MMRM were implemented with the same dose-response model to make a fair comparison. The same argument as before can be reiterated regarding the number of doses – more doses could have been studied but would likely not alter this comparison. The dose-response was equally well characterized already with three doses and, besides, it is uncommon to have a large number of doses in phase 2 trials.

The drug effect was simulated to give a rather large (40%) reduction in UACR, which is more than the 30% reduction recently suggested as a meaningful clinical improvement [16, 23]. This affects the sample size in this setup so that studies are smaller than they would be in reality, but again, would not alter the comparison of the investigated methods.

## CONCLUSIONS

All three methods (dose-response on end-of-study data, conventional MMRM and dose-response MMRM) predicted the outcome at the end of the study with good accuracy. When MMRM methods are warranted, for instance when there is a highly variable placebo response, the precision of MMRM models could be improved by adding a dose-response relationship. The precision was mostly improved for the lower dose arms. This method can aid in the general understanding of the dose-response characteristics of a compound and increase confidence in the dose(s) to bring forward to the next stage in drug development.

## ACKNOWLEDGEMENTS

The work was funded by AstraZeneca.

## FUNDING INFORMATION

Open access funding provided by Uppsala University.

## COMPLIANCE WITH ETHICAL STANDARDS

**Conflict of Interest** Gustaf Wellhagen (formerly), Magnus Åstrand and Bengt Hamrén are employees of AstraZeneca and also hold stocks/shares in AstraZeneca. Maria Kjellsson declares no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

1. Karlsson KE, Grahnén A, Karlsson MO, Jonsson EN. Randomized exposure-controlled trials; impact of randomization and analysis strategies. *Br J Clin Pharmacol*. 2007 Sep;64(3):266–77.
2. Williams PJ, Ette EI. Pharmacometrics: Impacting Drug Development and Pharmacotherapy. In: *Pharmacometrics [Internet]*. John Wiley & Sons, Ltd; 2006 [cited 2020 May 28]. p. 1–21. Available from: <http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470087978.ch1>
3. Mould DR. Developing Models of Disease Progression. In: *Pharmacometrics [Internet]*. John Wiley & Sons, Ltd; 2006 [cited 2020 May 28]. p. 547–81. Available from: <http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470087978.ch21>
4. Mallinckrodt CH, Clark WS, David SR. Accounting for drop out Bias using mixed-effects models. *J Biopharm Stat*. 2001 Jan 1;11(1–2):9–21.
5. Mallinckrodt CH, Sanger TM, Dubé S, DeBrota DJ, Molenberghs G, Carroll RJ, et al. Assessing and interpreting treatment effects in



- longitudinal clinical trials with missing data. *Biol Psychiatry*. 2003;53(8):754–60.
6. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Inf J*. 2008 Jul 1;42(4):303–19.
  7. Prakash A, Risser RC, Mallinckrodt CH. The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. *Int J Clin Pract*. 2008 Aug;62(8):1147–58.
  8. Siddiqui O, Hung HMJ, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *J Biopharm Stat*. 2009;19(2):227–46.
  9. Andersen SW, Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013 Feb;12(1):7–16.
  10. Gomeni R, Goyal N, Bressolle F, Fava M. A novel methodology to estimate the treatment effect in presence of highly variable placebo response. *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol*. 2015 Oct;40(11):2588–95.
  11. Gosho M, Hirakawa A, Noma H, Maruo K, Sato Y. Comparison of bias-corrected covariance estimators for MMRM analysis in longitudinal data with dropouts. *Stat Methods Med Res*. 2017 Oct;26(5):2389–406.
  12. Conrado DJ, Chen D, Denney WS. Cardiovascular safety assessment in early-phase clinical studies: a meta-analytical comparison of exposure-response models. *CPT Pharmacomet Syst Pharmacol*. 2016;5(6):324–35.
  13. Garnett C, Bonate PL, Dang Q, Ferber G, Huang D, Liu J, et al. Scientific white paper on concentration-QTc modeling. *J Pharmacokinet Pharmacodyn*. 2018 Jun 1;45(3):383–97.
  14. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015. *Lancet*. 2016 Oct 8;388(10053):1459–544.
  15. Baigent C, Herrington WG, Coresh J, Landray MJ, Levin A, Perkovic V, et al. Challenges in conducting clinical trials in nephrology: conclusions from a kidney disease—improving global outcomes (KDIGO) controversies conference. *Kidney Int*. 2017 Aug 1;92(2):297–305.
  16. Levey AS, Gansevoort RT, Coresh J, Inker LA, Heerspink HL, Grams ME, et al. Change in albuminuria and GFR as end points for clinical trials in early stages of CKD: a scientific workshop sponsored by the National Kidney Foundation in collaboration with the US Food and Drug Administration and European medicines agency. *Am J Kidney Dis Off J Natl Kidney Found*. 2019 Aug;23.
  17. Wajjer SW, Gansevoort RT, Heerspink HJL. *Curr Opin Nephrol Hypertens: Change in albuminuria as a surrogate endpoint*; 2019 Aug 22.
  18. Core Team R. R: a language and environment for statistical computing. In: R Foundation for statistical computing, Vienna: Austria. URL; 2018. <https://www.R-project.org/>.
  19. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models [Internet]*. 2019. Available from: <https://CRAN.R-project.org/package=nlme>
  20. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*. 2005 Sep;61(3):738–48.
  21. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern Med*. 2016 Dec 1;176(12):1826–33.
  22. Aoki Y, Röshammar D, Hamrén B, Hooker AC. Model selection and averaging of nonlinear mixed-effect models for robust phase III dose selection. *J Pharmacokinet Pharmacodyn*. 2017 Dec 1;44(6): 581–97.
  23. Heerspink HJL, Greene T, Tighiouart H, Gansevoort RT, Coresh J, Simon AL, et al. Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomised clinical trials. *Lancet Diabetes Endocrinol*. 2019 Feb;7(2):128–39.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.