Describing the curves: uncertainty propagation and bias correction for predictor effects in simple and generalized linear (mixed) models

Steve Cygu[1], Benjamin M. Bolker[1,2], Jonathan Dushoff[1,2]

1 School of Computational Science and Engineering, McMaster University, Hamilton, Ontario, Canada
2 Department of Biology, McMaster University, Hamilton, Ontario, Canada

* cygu@aims.ac.za

Abstract

In many applications which use generalized linear (mixed) models, outcome predictions or predicted probabilities are often of interest. For models that involve complex multiplicative interactions, additional non-focal predictors or nonlinear link functions, the estimated coefficient are not readily interpretable. A general way to summarize these kind of models is through predictor effects, which are not only dependent on choice of representative values of focal predictor but also sensitive to which values of non-focal predictors are chosen. The most common approach is generating predictor effects at a “reference point”, usually the mean, of non-focal predictor, i.e., mean-based which could be considered as the effect of an “average case” in the population. In the presence of sources of bias such as additional non-focal predictors, nonlinear link functions, random effect terms, etc., mean-based approach generate predictor effects that are biased and may not be consistent with the observed quantities. An alternative is the whole-sample-based approach which estimates the average effect in the population. Moreover, isolated confidence intervals provide an alternative and a more clear way to describe uncertainty associated with the focal predictor of interest. In addition to theoretical and methodical comparison, using simulation, we illustrate the two approaches and show that they can produce substantially different results and that whole-sample-based approach can not only produce estimates consistent with the observed values, but also appropriate for bias correction. We also present an alternative way, isolated confidence intervals, to describe uncertainties associated with these predictions.

Definitions

- **Input variables**: Scientific variables underlying an inference or exploration. The focal predictor we use for effects or predictions is an input variable.
- **Model variables**: Variables that represent columns in the model matrix. Each input variable will correspond to one or more model variables. In particular, variables with more than two categories, or variables modeled with a spline or polynomial response, will correspond to more than one input variable.
- **Center:** A point corresponding to a column-wise mean of the model matrix (the mean of one or more model variables). The center point for a set of model

variables corresponding to an input variable may not represent a possible value of the input variable.

- **Reference point**: Value or values chosen for *non-focal predictors*, when estimating effects. Typically the center point, but can instead be a population of quantiles or observations.
- **Anchor**: The value chosen for the *focal predictor* when estimating effect confidence intervals (anchor choice does not affect the estimate). Typically chosen as the center point of the model variables corresponding to the focal input variable.

Author summary

FIXME

Introduction

In many applications which use simple or generalized linear (mixed) models, outcome predictions or probabilities, are often of interest. Predictions provide a great way to summarize what the regression model is telling us and very useful for interpreting and visualizing model estimates. For example, in logistic regression models, the coefficient estimates are usually not easy to interpret and less informative. Since logistic models are nonlinear (due to the nonlinear link function), in multivariate models with interactions, the magnitude of the effect of change in the outcome depends on the values of the predictor of interest and other predictors. Hence, the conclusions one can make about the estimated effects greatly depends on how well we choose the values of the other predictors, i.e., *non-focal* predictors.

Both simple and generalized linear (mixed) models (GL(M)Ms) can examine very complex relationships, including nonlinear relationships between response and predictors, interactions between predictors (and via splines for example), and nonlinear transformations via link functions due to their flexibility. This flexibility comes at a cost, for example, complex multivariate models may risk misinterpretation, and miscalculation of quantities of interest. Also, coefficient estimates of models involving nonlinear link functions or interactions lose their direct interpretation [1], meaning that interpretation of derived quantities from these estimates requires some understanding of the specified model. An alternative is to explore the *effects* of predictors on the predictions or probabilities of the outcome at various level of the predictor of interest – *predictor effects* [2–4]. For example, as a way to plan, public health officials may want to know the effects of household income, wealth index, etc, on the predicted probability of having improved water services among slum dwellers.

When visually presented, predictor effects provide a unified and intuitive way of describing relationships from a fitted model, especially complex models involving interaction terms or some kind of transformations on the predictors whose estimates are usually, but not always, a subject to less clarity of interpretation. Further, generating predictor effects together with the associated confidence intervals for regression models has a number of challenges. In particular:

1. choice of representative values of *focal* predictor and the *reference point* for the non-focal predictors especially in multivariate models
2. propagation of uncertainty – appropriate choice of *anchor* for computing confidence intervals; and how to incorporate the uncertainty due to non-focal predictors

3. bias in the expected mean prediction induced by the nonlinear transformation of the response variable (especially in GL(M)Ms)

The most common way of dealing with the first challenge is taking unique levels of the focal (predictor of interest) predictor if discrete or taking appropriately sized quantiles (or bins) if continuous, and then calculating the predictions while holding non-focal (other predictors other than the focal predictor) predictors at their reference point (e.g., means) [5]. This generates – *predictor effects* [2], *marginal predictions* [3] or *estimated marginal means* [4]. In this article, we refer to this quantity as predictor effects since it should, for example, tell us what we would expect the response to be at a particular value or level of the predictor, for an "average case". More specifically, predictor effects computes the expected outcome by meaningfully holding the non-focal predictors constant (or averaged in some meaningful way) while varying the focal predictor, with the goal that the outcome expected prediction represents how the model responds to the changes in the focal predictor.

The commonly used R software packages (**emmeans** and **effects**) for generating predictor effects, by default, use the average of the non-focal predictors as the reference point. However, there are a number of choices one can make when considering this approach – for example, in the presence of interaction, averaging the interactions (averaging product of interacting model variables) versus product of the averages of the interacting model variables (default for **emmeans** and **effects**). We claim that neither of these two approaches is the most appropriate but we show that averaging the interaction closely matches the observed values. To illustrate this, consider models 1 and 2 below, with x_1 as the focal predictor:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \tag{1}$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{23} x_2 x_3 + \epsilon \tag{2}$$

We first simulate data with the two models, such that $x_{1,2,3} \sim \text{Normal}(0, 1)$, $\beta_0 = 5$, $\beta_1 = -3$, $\beta_2 = 1$, $\beta_3 = 2$ and $\beta_{23} = 5$, and then compare the predictions from the **emmeans**, **effects** and our proposed alternative (**varpred**) to the "true" predictions and observed average i.e., $\bar{y}$, as shown in Fig 1.

In the absence of interaction (model 1), the three approaches produce similar estimates, which match the simulated values, Fig 1A. However, in the presence of interaction, even as simple as the one in model 2, the estimates starts to differ. In particular, **emmeans** and **effects** give similar estimates ($\bar{\hat{y}}$) but different from the **varpred**'s which, however, is very close to the simulated average ($\bar{y}$), Fig 1B. To generate Fig 1A, all the three packages averages the non-focal predictors x_2 and x_3. On other hand, to generate Fig 1B, the difference in the estimates lies on how each of the packages average the interaction term ($x_2 x_3$). In particular **emmeans** and **effects** computes $\bar{x_2}\bar{x_3}$ while **varpred** computes $\overline{x_2 x_3}$.

Sometimes, one may be interested in the uncertainties associated with the focal predictor of interest only, excluding other uncertainties due to other non-focal predictors – *isolated* confidence intervals. However, currently, the two packages do not provide straightforward way do achieve this. To illustrate this, we generate predictor effect of x_2 from model 1 together with the associated 95% confidence bands, as shown in Fig 2.

For **emmeans** and **effects**, the confidence bands are much wider because they include uncertainties associated with the non-focal predictors, but narrower and crosses at the mean of the focal predictor, i.e., model center, in **varpred**. In other words, with **varpred**, we are able to generate isolated confidence bands indicating zero uncertainty at the value of the focal predictor we are more certain about, i.e., mean of the model variable corresponding to the focal predictor. For simple models, the point where the

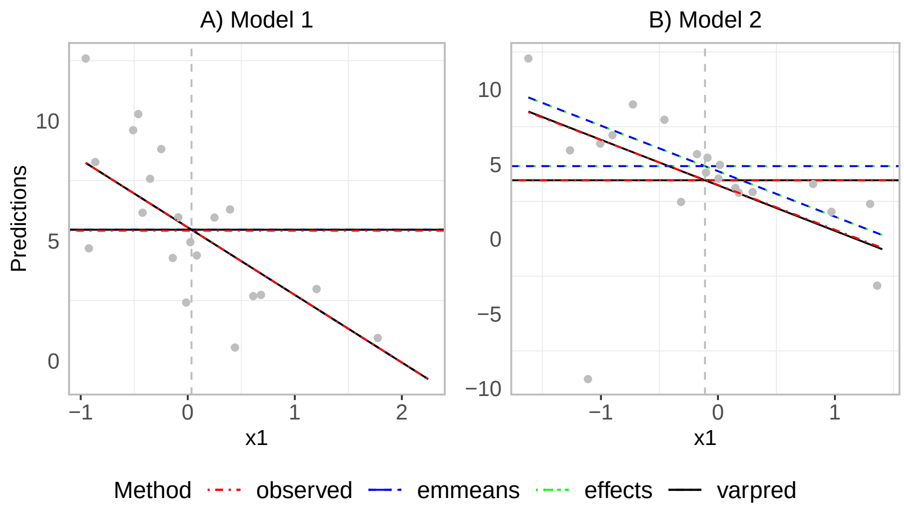

Fig 1. A comparison of emmeans, effects and varpred predictor effects for x_1 on y for models with and without interactions. The horizontal blue, green and black lines are the mean predictions, i.e., $\bar{\hat{y}}$, the red ones are the means of the simulated y, i.e., $\bar{y}$, while the vertical dotted grey lines are the means of the corresponding focal predictors. The grey points are the binned simulated y. The trend lines represents the corresponding $\hat{y}$ at various levels of x_1, while holding the other predictors at their average, based on observed values (red) and the three methods. A: In the absence of interaction, the predicted mean, $\bar{\hat{y}}$, closely matches the simulated in all the three approaches, i.e., $\bar{y} \approx \bar{\hat{y}}$. Similarly, the predictions based on the three approaches closely matches the "true" predictions. B: Even with the simple interaction between the non-focal predictors, we start seeing deviation in both predicted mean, $\bar{\hat{y}}$, and overall predictions from the simulated, $\bar{y}$, and "true" predictions, respectively, in two commonly used packages (**emmeans** and **effects**), but not, the proposed **varpred**.

confidence bands crosses is the model center and it corresponds to anchor we choose, in this case, the mean of the focal predictor.

When dealing with nonlinear link functions or models with multiple non-focal predictors, the correct predictions, for example, are even much harder to estimate. One approach is to make predictions on the transformed scale (linear predictor scale), and then back-transform to the original scale. However, the back-transformation may either result in biased predictions or requires some approximation. In particular, bias in expected mean prediction induced by nonlinear transformation of the response variable can lead inaccurate predictions. An alternative to mean-based reference point, i.e., averaging the non-focal predictors, is the *whole-sample-based* approach, discussed in detail later, which involves computing the prediction over the population of the non-focal predictors and then averaging across the values of the focal predictor [5].

The main purpose of this article is to discuss and implement various approaches for computing predictor effects and provide an alternative method for computing the associated confidence intervals. We further explore and demonstrate, using simulated data, approaches for correcting bias in predictions for GL(M)Ms involving nonlinear link functions.

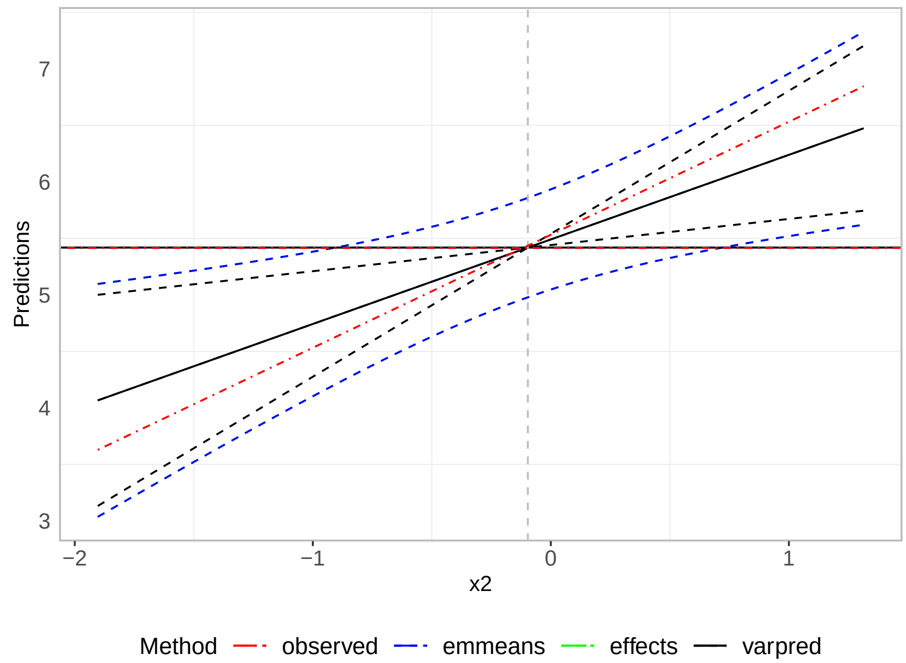

Fig 2. A comparison of traditional and isolated 95% **confidence bands.** The description of horizontal, vertical and trend lines remain the same as above. The wider dotted blue curves overlaying the green curves are the traditional confidence bands from **emmeans** and **effects**, while the narrower black curves crossing at the mean of the focal predictor, i.e., the model center, are the isolated confidence bands from **varpred**. For simple models, the isolated confidence bands crosses at the model center.

Quantities of interest

Several quantities of interest may be derived from regression models. The first one is the coefficient estimates. Others are *input and model variables*, *model matrix*, *model center*, *reference point*, *anchor*, *predictors values* and *marginal effects*. In particular: input variables are scientific variables underlying an inference or exploration. The focal predictor we use for effects or predictions is an input variable [6]; model matrix refers to the design matrix whose rows include all combination of variables appearing in the interaction terms, along with the typical values of the focal and non-focal predictors; model variables are variables that represent columns in the model matrix. Each input variable will correspond to one or more model variables. In particular, variables with more than two categories, or variables modeled with a spline or polynomial response, will correspond to more than one input variable; model center is a point corresponding to a column-wise mean of the model matrix (the mean of one or more model variables). The center point for a set of model variables corresponding to an input variable may not represent a possible value of the input variable; reference point is a (or are) value (or values) chosen for non-focal predictors, when estimating effects. Typically the center point, but can instead be a population of quantiles or observations; and lastly, an anchor is the value chosen for the focal predictor when estimating effect confidence intervals (anchor choice does not affect the estimate). Typically chosen as the center point of the model variables corresponding to the focal input variable.

In simple linear models with no interaction terms, the default output for the coefficient estimates are simple and directly interpretable as the expected change in outcome for a unit change in focal predictor. This is the *unconditional marginal effect* [1] and it is constant across all the observations and levels of all other predictors. Consider models 1 and 2. The marginal effect of x_2 in model 1 is $\frac{\partial y}{\partial x_2} = \beta_2$. On the other hand, the marginal effect of x_2 in model 2 is given by $\frac{\partial y}{\partial x_2} = \beta_2 + \beta_{23}x_3$. In other words, if there are no interactions, the marginal effect of x_2 on y is constant, while, if there are interactions in the model, the marginal effect of a change in x_2 on y depends on the value of the other *conditioning* predictor, x_3.

To distinguish between predictor effects and marginal effects, consider result from a hypothetical simulated example – regression of household size as function of household wealth index and age of household head, as shown in Fig 3. Since the model has no interactions, the relationship between the predicted household size and age is linear, hence the marginal effect of age is the slope ($\frac{\Delta \text{hhsize}}{\Delta \text{age}}$) of the predictor effect line and can be calculated irrespective of the values of wealth index.

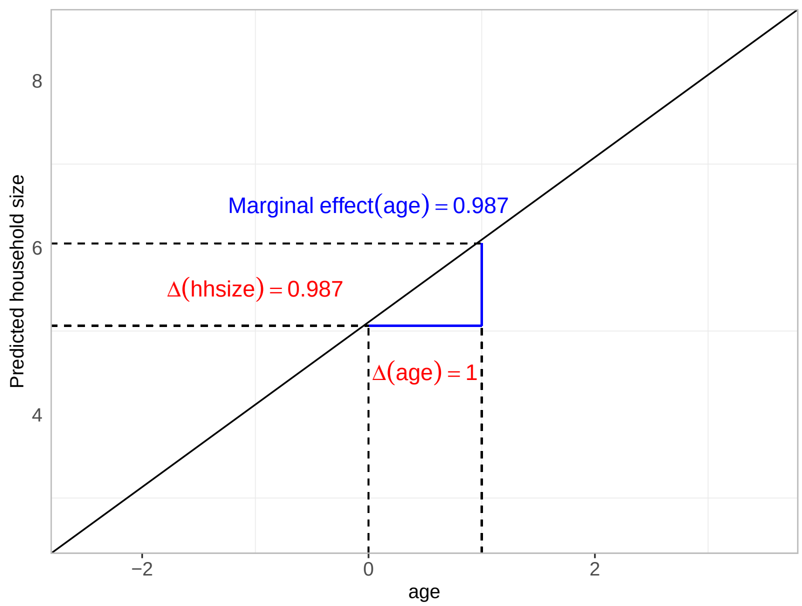

Fig 3. A comparison of predictor effect and unconditional marginal effect of age on household size. The black line is the predictor effect trend line, while the dotted lines indicate the changes along the `age` and `household size` axis. For a linear model with no interaction, the marginal effect is the slope of the predictor effect line.

Predictor effect, on the other hand, is the expected household size for a particular age, holding wealth index at its mean. In particular, the purpose and goal of a predictor effect seems fairly straightforward; for specified values of a focal predictor, we want to give a point estimate and confidence intervals for the prediction of the model for a "reference point" individual with those values of the predictors.

Statistical background

To get an intuition of how conditioning on the mean values of the non-focal predictors work, suppose we are interested in predictor effects of a particular predictor, i.e., focal, x_f, from the set of predictors. To keep it simple, assume that the model has no interaction terms. The idea is to choose a reference point for the values of non-focal predictors. For example, fixing the values of non-focal predictor(s) at some typical values – typically determined by averaging (for now) in some meaningful way, for example, arithmetic mean for continuous and average over the levels of the factors for categorical non-focal predictors. We refer to this as *mean-based* approach. The most convenient way to achieve this is by constructing $\mathbf{X}^\star$ by averaging the columns of non-focal predictors in model matrix $\mathbf{X}$, and together with appropriately chosen values of focal predictor.

Consider a simple linear model with linear predictor $\eta = \mathbf{X}\boldsymbol{\beta}$ and let $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$ be an identity link function (in the case of simple linear model), where $\boldsymbol{\mu}$ is the expected value of response variable y. Let $\hat{\boldsymbol{\beta}}$ be the estimate of $\boldsymbol{\beta}$, together with the estimated covariance matrix $\Sigma = V(\hat{\boldsymbol{\beta}})$ of $\hat{\boldsymbol{\beta}}$. Let $\mathbf{X}^*$ be the model matrix, inheriting most of its key properties, for example transformations on predictors and interactions from the model matrix, $\mathbf{X}$. Then the prediction $\hat{\boldsymbol{\eta}}^\star = \mathbf{X}^\star\hat{\boldsymbol{\beta}}$ is the predictor effect for the focal predictor in question [2].

An alternative formulation of predictor effect involves, expressing the linear predictor as the sum of the focal and non-focal predictors' linear predictor. In particular,

$$\eta^\star(x_f, \bar{x}^\star_{\{n\}}) = \beta_f x_f + \sum \beta_{\{n\}}\bar{x}^\star_{\{n\}} \tag{3}$$

$$\hat{y}_f = g^{-1}\left(\eta^\star(x_f, \bar{x}^\star_{\{n\}})\right) \tag{4}$$

where $\bar{x}^\star_{\{n\}}$ are the appropriately averaged entries of non-focal predictors and x_f is a vector of values of the focal predictors for a particular observation.

Dealing with higher order interactions

Higher order terms such as interactions, splines, polynomials, etc., can be within the focal predictor, within (between) non-focal predictor(s), between the focal and non-focal predictor(s). To distinguish the three, suppose the model which describes the hypothetical household size simulation described above is

$$\text{hh size}_i = \beta_0 + \beta_{\text{A}_1}\text{Age}_i + \beta_{\text{A}_2}\text{Age}_i^2 + \beta_{\text{A}_3}\text{Age}_i^3 + \beta_{\text{W}}\text{Wealthindex}_i + \epsilon_i \tag{5}$$

In the first case, with *Age* as the focal predictor, the interaction is within its polynomial terms. In this case, each of the polynomial terms of the focal predictor are evaluated independently across the chosen levels of the focal predictor, Age_i. Specifically, the higher order terms are treated as additional columns of the model matrix evaluated with the same values of the focal predictor, while the non-focal predictors are fixed at their reference point as discussed in the previous section. In this case, the predictor effect of *Age* on the linear scale becomes

$$\begin{aligned}\eta^\star(\text{Age}_i, \overline{\text{Wealthindex}}_{\{n\}}) = {} & \beta_0 + \beta_{\text{A}_1}\text{Age}_i + \beta_{\text{A}_2}\text{Age}_i^2 + \beta_{\text{A}_3}\text{Age}_i^3 \\ & + \beta_{\text{W}}\overline{\text{Wealthindex}}\end{aligned}$$

In the second case, with $Wealthindex$ as the focal predictor, the higher order interactions in the non-focal predictor, *Age*, are treated simply as additional columns in the variable space (of model matrix) and appropriate choice of reference point applies just like in the models without interactions. For instance, in mean-based approach, we simply average all the non-focal interaction terms. In this setup, the predictor effect of the previous model, on link scale, becomes

$$\eta^\star(\text{Wealthindex}_i, \{\overline{\text{Age}}, \overline{\text{Age}^2}, \overline{\text{Age}^2}\}_{\{n\}}) = \beta_0 + \beta_{\text{A}_1}\overline{\text{Age}} + \beta_{\text{A}_2}\overline{\text{Age}^2} + \beta_{\text{A}_3}\overline{\text{Age}^3} + \beta_{\text{W}}\text{Wealthindex}_i$$

Lastly, consider model 2, with the interactions between the focal predictor, x_2, and the non-focal predictor, x_3. In this case, the non-interacting predictors are treated as non-focal predictors, with appropriate reference points. However, for the interacting predictors, we first choose representative values for the focal predictor and some particular values of the interacting predictor similar to the way we choose values for the focal predictor or predetermined. In our example, suppose we pick i and j unique values of the focal predictor, x_2 and interacting predictor x_3, respectively, then the prediction on linear scale is given by

$$\eta^\star(x_{2i}, x_{3j}, \bar{x}^\star_{\{n\}}) = \beta_0 + \beta_1\bar{x}_1 + \beta_2 x_{2i} + \beta_3 x_{3j} + \beta_{23} x_{2i} x_{3j}.$$

In general, our formulation, even for more complicated interactions, follow these three basic principles – interaction within focal predictor, interaction between non-focal predictors and interaction between focal and non-focal predictors.

Uncertainty propagation

What about the confidence intervals (CI)? The limits of the confidence intervals are points, not mean values. In principle, every value of focal predictor has a different CI. The traditional way to compute variances for predictions is $\sigma^2 = \text{Diag}(\mathbf{X}^\star \Sigma \mathbf{X}^{\star\top})$ [2, 4], so that the confidence intervals are $\eta \pm q\sigma$, where q is an appropriate quantile of Normal or t distribution. This approach incorporates all the uncertainties – including the uncertainty due to non-focal predictors. But what if we are only interested in the uncertainty as a result of the focal predictor, so that the confidence intervals are $\eta \pm q\sigma_f$? We call this isolated confidence intervals.

Currently, commonly used R packages for constructing predictions, by default, do not exclude the uncertainties resulting from the non-focal predictors when computing the CIs. A non-trivial way to exclude uncertainties associated with non-focal predictors in some of these packages is to provide a user defined variance-covariance matrix with the covariances of non-focal terms set to 0 – *zeroing-out* variance-covariance matrix. This only works when the input predictors are *centered* prior to model fitting, in case of numerical predictors, and even much complicated when the predictors are categorical. We first describe the variance-covariance based approach and then discuss our proposed method which is based on *centering model matrix* and does not require input predictors to be scaled prior to model fitting.

Variance-covariance

The computation of $\hat{\eta}^\star$ remains the same as described above. However, to compute σ, Σ is modified by *zeroing-out* (the variance-covariance of all non-focal predictors are set to zero) variances of non-focal terms. Although this is the simplest approach, it requires centering of continuous, i.e., $x_c = x - \bar{x}$, predictors prior to model fitting and proper way to average categorical predictors.

Centered model matrix

Consider centered model matrix $\mathbf{X}^\star_c = \{\mathbf{X}^\star_f, \mathbf{X}^\star_{\{n\}} - \bar{\mathbf{X}}^\star_{\{n\}}\}$. It follows that the non-focal terms in $\mathbf{X}^\star_c$ are all zero in simple models without interactions but are isolated (narrower) around the model center in models involving complex interactions. Consequently the uncertainty due to non-focal predictors are isolated in the computation of $\sigma^2 = \mathrm{Diag}(\mathbf{X}_c^\star \Sigma \mathbf{X}_c^{\star\top})$. In addition, the computation of $\mathbf{X}_c^\star$ impacts only on the intercepts and the non-focal terms, i.e., the slopes and variance of the focal predictors are not affected. This means that we can still generate isolated CIs without necessarily centering the predictors prior to model fitting.

Bias correction

In many applications, it usually important to report the estimates that reflect the expected values of the untransformed response. However, when dealing with nonlinear link functions, it is even harder to generate correct predictions that reflect the untransformed response due to the bias in the expected mean induced by the nonlinear transformation of the response variable. In such cases, bias correction is needed when back-transforming the predictions to the original scales. Most common approach for bias-adjustment is second-order Taylor approximation [4,7]. Another potential source of bias comes from additional non-focal predictors, especially in models with nonlinear link functions. Here, we describe and implement a different approach, whole-sample-based approach for bias correction.

Whole-sample-based approach for bias correction

The most precise (although not necessarily accurate!) way to predict is to condition on values of the focal predictor and make predictions for all observations (members of the population) [5]. A key point is that the nonlinear transformation involved in these computations is always *one-dimensional*; all of the multivariate computations required are at the stage of collapsing the multidimensional set of predictors for some subset of the population to a one-dimensional distribution of $\eta^\star(x_f, x_{\{n\}})$, which is a function of focal predictor and the observed values of the non-focal predictors, as opposed to the definition in Equation 3. More specifically:

- compute linear predictor associated with the non-focal predictors, $\eta_{\{n\}} = \sum \beta_{\{n\}} x_{\{n\}}$
- compute linear predictor associated with the focal predictors, $\eta_{jf} = \sum \beta_f x_{jf}$
- for every value of the focal predictor, η_{jf}:

$$
\begin{aligned}
\eta_j^\star(\eta_{jf}, \eta_{\{n\}}) &= \eta_{jf} + \eta_{\{n\}} \\
&= \eta_j^\star(x_f, x_{\{n\}})
\end{aligned} \tag{6}
$$

Once Equation 6 is computed, one can back-transform the estimates to the original scale and the average over the levels of the focal predictors, j:

$$
\hat{y}_f = \text{mean } g^{-1}\left(\eta_j^\star(x_f, x_{\{n\}})\right) \tag{7}
$$

We make similar adjustments to compute the variances of the predictions at every level of the focal predictor:

$$
\sigma_{jf}^2 = \mathrm{Diag}(\mathbf{X}_{jc}^\star \Sigma \mathbf{X}_{jc}^{\star\top}) \tag{8}
$$

where $\mathbf{X}^\star_{jc} = \{\mathbf{X}^\star_{jf}, \mathbf{X}^\star_{\{n\}} - \bar{\mathbf{X}}^\star_{\{n\}}\}$ and

$$\mathrm{CI}_f = \text{mean } g^{-1}\left(\eta^\star_j(x_f, x_{\{n\}}) \pm q\sigma_{jf}\right) \tag{9}$$

For models with random effects components, we make further adjustment to correct for bias induced by the random effects terms. In the population approach, we treat the random effects terms as additional non-focal predictors and simply make adjustment to Equation 6. In particular

$$\begin{aligned} \tau &= \mathbf{Z}b \\ \eta^\star_j(x_f, x_{\{n\}}, \tau) &= \eta^\star_j(x_f, x_{\{n\}}) + \tau \end{aligned} \tag{10}$$

where $\mathbf{Z}$ and b are the design matrix and a vector of random effects, respectively.

Mean-based vs whole-sample-based

As mentioned above, the two common choices to generate predictor effects are: 1) setting the non-focal predictors to their mean – mean-based; and 2) using the entire population of the non-focal predictors – whole-sample-based. If the link function is nonlinear and/or the model has complex higher order interactions, the average of the predictions evaluated at the mean of the non-focal predictors and the average of predictions evaluated at the population level (and then averaged) of the non-focal predictor are not equivalent, i.e.,

$$g^{-1}\left(\eta^\star_j(\bar{x}_f, \bar{x}_{\{n\}})\right) \neq \frac{1}{n}\sum_{i=1}^{n} g^{-1}\left(\eta^\star_j(\bar{x}_f, x_{\{n\}})\right). \tag{11}$$

If $g()$ is strictly concave or strictly convex, one can use Jensen's inequality to determine which side of Equation 11 is smaller or larger. However, for non-fully convex or concave link function, one can use Taylor series expansion to approximate the range over which the two are smaller or smaller than the other [5].

As stated above, in simple linear models without interactions, the effect is constant, so both approaches yield similar results. However, picking a single value, e.g., mean of the predictor, on which to draw conclusions about the effect can be problematic, unrealistic or not contained in or representative of the population. In addition, the mean-based approach fails to use every values of the non-focal predictors hence not utilizing the full potential of the information contained in the data. This may limit the inferences we can make about the entire population. In general, the mean-based approach provides the predictor effect of an average case, whereas, the whole-sample-based approach, summarizes the predictor effect over the entire population – and in some applications, the effect of an average case might not be generalizable to the entire population, especially, if the average does not represent the population. The whole-sample-based focuses on specific observations since the prediction is first obtained for each observation and then averaged across the levels of the focal predictor.

Another potential concern with the mean-based approach arises in situations where direct naive use leads to rare or meaningless basis for generalization. For example, setting categorical dummy categorical variables in the model matrix to their means, which, by default, sets them to their sample means or observed proportions. For example, a dummy variable for christian household heads may be set to 0.2, translating to prediction for an household head who is 20% christian. This problem may extend to models with other complex interaction terms [5].

Whole-sample-based approach is not entirely foolproof. For instance, similar to the mean-based approach, in the case of continuous focal predictors, choosing the representative values of the focal predictors can be very challenging especially if the cases are not evenly distributed around the minimum and the maximum values or within some subgroups defined in the population. In addition, whole-sample-based approach can be very computationally intensive for large datasets.

Simulation examples

We now illustrate the construction of predictor effects with isolated confidence intervals and also demonstrate that the mean-based and population approaches produce different results in models with nonlinear link functions and/or additional source of potential bias. In addition, we demonstrate that whole-sample-based approach can be used to correct bias in the predictions.

Isolated confidence intervals

As mentioned above, in some applications, one may be interested in uncertainties associated with the focal predictor only. In that case, it is important to exclude all the uncertainties as a result of the non-focal predictors and only show confidence intervals describing the predictor of interest. To illustrate this, we consider the simulation model described in S1 Appendix.

We start by fitting two models from the simulated data, i.e., consider a cubic polynomial predictor, age, as a focal predictor and also, consider the non-polynomial predictor, wealth index as the focal predictor. After fitting the model, we constructed predictor effects and then compared isolated vs non-isolated confidence intervals associated with these predictions (see Fig 4).

In the absence of complex higher order interaction terms or transformations on the focal predictor in simple linear models, we would expect, at the model center, the observed average, $\overline{hh\ size}$, and the average of the predictions, $\overline{\widehat{hh\ size}}$, to be identical and cross, in the case of isolated confidence bands, at the model center (see Fig 4B). On the other hand, if the focal predictor is characterized by interactions or any other form of transformation, the model center is not necessarily a point in the predictor space. In this case, the isolated confidence bands will be narrower than the non-isolated but not necessarily crossing at the model center (see Fig 4A).

Bias correction

Mean-based and whole-sample-based approaches can produce very different results in models with nonlinear link functions with additional sources of potential bias such as additional non-focal predictor(s), random effect terms or complex interactions. To demonstrate this, we considered a two predictor binary outcome simulation described in S2 Appendix, such that the age effect is way smaller than that of the wealth index, and compared their predictor effect on the predicted probability of improved water quality as shown in Fig 5.

If there was no effect of nonlinear averaging, then we would expect the observed proportion and the average of the predicted probabilities to cross at the model center as we see in Fig 4B. However, the differences we see in Fig 5 are due to nonlinear averaging since both observed status and predicted probabilities are averaged on the response scale as opposed to link scale. In particular, if the range of values are bigger than 0.5

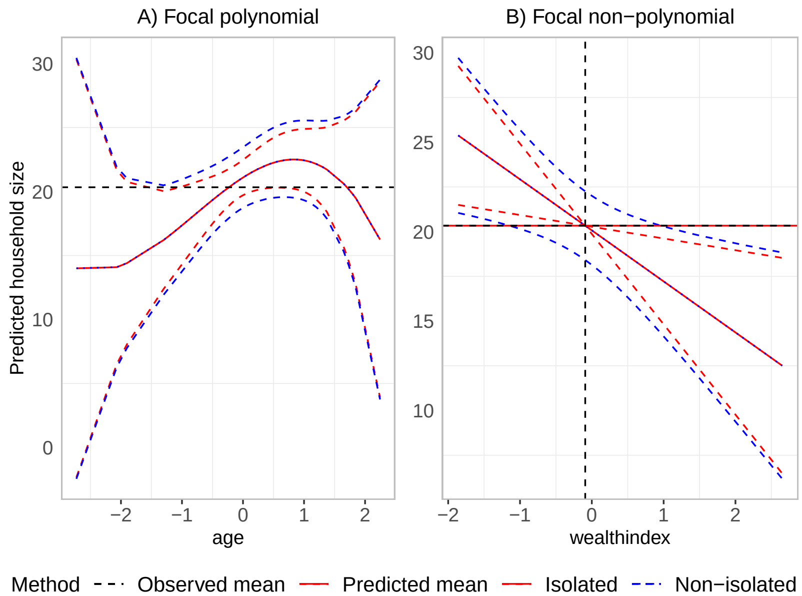

Fig 4. Isolated and non-isolated predictor effects 95% **confidence bands.** These figures compare the predictor effects together with their corresponding 95% confidence bands. A: The focal predictor is a cubic polynomial. B: The focal predictor is not a polynomial function, implying that the complex interaction is in the non-focal predictor. In both cases, the confidence bands are narrower in the isolated case, and crosses at the mean of the focal predictor (model center) in the case of simple (non-polynomial) focal predictor – B. In the case of cubic polynomial focal predictor, the model center is not a point in the predictor space, i.e., mean of the focal predictor, but a combination of all the terms in the cubic polynomial, consequently, the isolated bands do not cross at a point – A. The horizontal black and red dotted lines are the observed and predicted average household size, i.e., $\overline{hh\ size}$ and $\overline{\widehat{hh\ size}}$, respectively; the inner red dotted curves and crossing line are the isolated confidence bands for A and B, respectively, while the outer blue curves are the non-isolated confidence bands in both cases. The central curve and trend line, in A and B, respectively, are the predictor effect of age and wealth index on the predicted household size, respectively.

(seemingly the case here), then we would expected the averages to be slightly higher than what we would expect at the model center, and vice versa.

Mediated effect

Consider the simple indirect mediation simulation described in S3 Appendix. We then fitted two models – *non-mediated* which models z as a function of x and the *mediated* which models z as a function of both x and y. In both cases, we compared mean-based and the whole-sample-based predictions, as shown in Fig 6.

From Fig 6A, we see what we would expect in the absence of additional sources of bias, even though in the simulation, the effect of x on z is mediated through y. By ignoring y in the model, we are still able to capture the effect of x and closely match the observed values using both approaches. However, if there were additional non-focal predictors, we would expect to see the differences similar to those in Fig 5. Including both y and x "dilutes" the direct effect of x on y and as a result, our prediction do not necessarily match the observed binned observations (see Fig 6B).

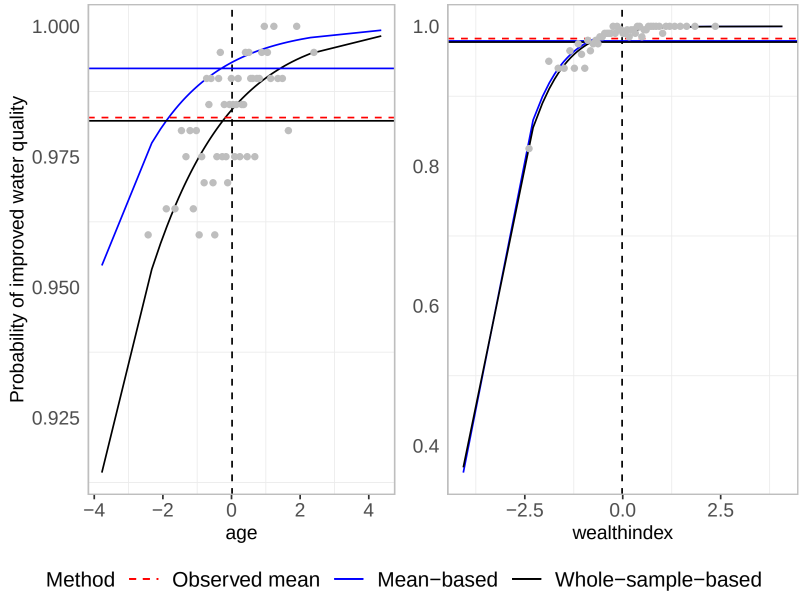

Fig 5. Mean-based and whole-sample-based predictor effects. The two approaches produce relatively different predictor effects for each of the predictors but the difference is more pronounced when the effect of the particular focal predictor has no strong effect – *on the left*. In both cases, age and wealth index effect, the average predicted probability of improved water based on whole-sample-based approach is very close to the observed proportion in the simulated data. The horizontal blue, black and the red dotted lines are the average mean-based, whole-sample-based predicted probability and observed proportion of improved water quality, respectively. The vertical dotted black line represents the mean of the focal predictor, and the point at which it crosses the red dotted line represents the expected "perfect" prediction at the model center. The grey points are binned observations – observed proportions of improved water quality in each bin. We would expect all the curves, horizontal, and vertical lines to cross at the model center but due the nonlinear averaging (at least in the bias corrected whole-sample-based approach) of the predictions on the response scale, this may not always be the case.

Discussion

Our simulations examples majorly focused on simple linear and logistic models due their wide range of usage and application. In addition, these models act as a starting point for building other complex models, including mixed effect models and models with categorical predictors. However, the logic for extension of approaches to more complex models, including other forms of nonlinear link functions is very straightforward. In addition, our R package implementation already extends to and supports most of the nonlinear link functions and mixed model framework, including multivariate binary outcome models.

Although commonly used R software packages, by default, implements mean-based approach, our simulation results demonstrated that the whole-sample-based approach has a potential of yielding results which are more consistent with the observed data. We would, therefore, argue that the use of mean-based approach should have some theoretical justification, especially in complex models.

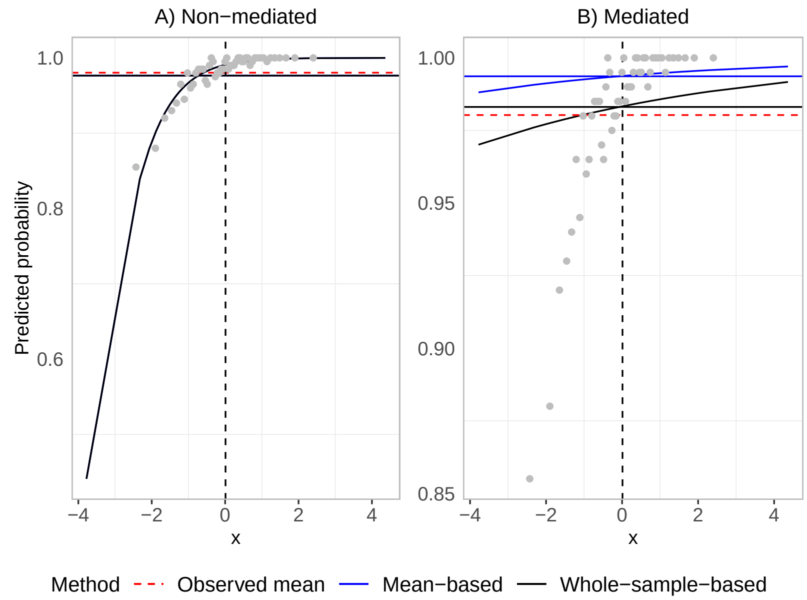

Fig 6. Mean-based and whole-sample-based predictor effects for mediated effect. These figures compare the mean-based and whole-sample-based predictor effect of x from a mediated, $z \sim x + y$, and non-mediated, $z \sim x$, models. A: In the absence of mediator variable, both mean-based and whole-sample-based approaches predictors are identical since there is no any other additional (non-focal predictors) sources of bias. Consequently, in both approaches, the average predicted probability is very close to the observed proportion. B: When the mediator variable is included, there is no direct effect of x on z and as a result the predictor effect curve does not align with the observations. However, the whole-sample-based approach still closely approximates the average proportion in the simulated data. The horizontal blue, black and the red dotted lines are the average mean-based, whole-sample-based predicted probability and observed proportion, respectively. The vertical dotted black line represents the mean of the focal predictor, and the point at which it crosses the red dotted line represents the expected "perfect" prediction at the model center. The grey points are the binned observations.

Conclusion

Generating outcome predictions or predicted probabilities from simple and generalized linear (mixed) models is not only important but also, generating quantities which are consistent with the observed values should be of interest. However, many studies still report coefficient estimates from generalized models like probit, logistic, etc., [5], which are subject of less clarity due to lack of direct link to the outcome of interest.

The argument and results we present in this paper supports a greater need for a shift on focus on what and how to present predictions from generalized models. For example, we believe that isolated confidence intervals could provide more clarity concerning the uncertainty due to the predictor of interests as opposed to the traditional way of incorporating everything.

From our theoretical, methodological and simulation results, researchers using these kind of models should, in the absence of theoretical justification, report predictions based on whole-sample-based approach or at least attempt to do a comparison of the two approaches before settling on the most appropriate in answering their research question. Moreover, we provide R package, **vareffects**, which implements these methods and is available on github (https://github.com/mac-theobio/effects).

Supporting information

S1 Appendix. Cubic polynomial interaction simulation. Consider an hypothetical simulation which simulates household size as a function of household wealth index and cubic function of the age of the household head, specified as follows:

$$\begin{aligned}
\text{hh size}_i &= \beta_0 + \beta_{\mathrm{A}_1}\text{Age}_i + \beta_{\mathrm{A}_2}\text{Age}_i^2 + \beta_{\mathrm{A}_3}\text{Age}_i^3 + \beta_{\mathrm{W}}\text{Wealthindex}_i + \epsilon_i \\
\text{Age}_i &\sim \text{Normal}(0, 1) \\
\text{Wealthindex}_i &\sim \text{Normal}(0, 1) \\
\epsilon_i &\sim \text{Normal}(0, 10) \\
\beta_0 &= 20 \\
\beta_{\mathrm{A}_1} &= 0.1 \\
\beta_{\mathrm{A}_2} &= 0.8 \\
\beta_{\mathrm{A}_3} &= 0.3 \\
\beta_{\mathrm{W}} &= -0.5 \\
i &= 1, \cdots, 100
\end{aligned} \tag{12}$$

S2 Appendix. Binary outcome simulation. Consider a simple simulation for improved water quality in Nairobi slums, such that the status is 1 for improved and 0 for unimproved water quality. In additional to the focal predictor, age of the household head, we add wealth index. In particular:

$$\begin{aligned}
\text{status}_i &\sim \text{Binomial}(1, \text{P}(\text{status}_i = 1)) \\
\text{logit}(\text{P}(\text{status}_i = 1)) &= \eta_i \\
\eta_i &= \beta_0 + \beta_{\mathrm{A}}\text{Age}_i + \beta_{\mathrm{W}}\text{Wealthindex}_i \\
\text{Age}_i &\sim \text{Normal}(0, 1) \\
\text{Wealthindex}_i &\sim \text{Normal}(0, 1) \\
\beta_0 &= 5 \\
\beta_{\mathrm{A}} &= 0.5 \\
\beta_{\mathrm{W}} &= 1.5 \\
i &= 1, \cdots, 10000
\end{aligned} \tag{13}$$

S3 Appendix. Mediated effect simulation. Next, we consider a simple indirect mediation simulation such that x has direct effect on y which in turn has effect on z but x has no direct effect on y, i.e., $x \to y \to z$. In particular:

$$
\begin{aligned}
z_i &\sim \mathrm{Binomial}(1, \mathrm{P}(\mathrm{z_i} = 1)) \\
\mathrm{logit}(\mathrm{P}(z_i = 1)) &= \eta_i \\
\eta_i &= \beta_0 + \beta_{xz} x_i + \beta_{yz} y_i \\
y_i &= \rho x_i + \sqrt{1 - \rho^2} y_y \\
x_i &\sim \mathrm{Normal}(0, 1) \\
y_y &\sim \mathrm{Normal}(0, 1) \\
\rho &= 0.8 \\
\beta_0 &= 5 \\
\beta_{xz} &= 0.2 \\
\beta_{yz} &= 1.5 \\
i &= 1, \cdots, 10000 \qquad (14)
\end{aligned}
$$

Acknowledgments

This work was supported by a grant to Jonathan Dushoff from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery.

Author Contributions

Conceptualization: Jonathan Dushoff
Software: Steve Cygu, Benjamin M. Bolker
Supervision: Jonathan Dushoff, Benjamin M. Bolker
Writing – original draft: Steve Cygu

References

1. Leeper TJ. Interpreting regression results using average marginal effects with R's margins. Reference manual. 2017;32.

2. Fox J, Hong J. Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. Journal of Statistical Software. 2009;32(1):1–24.

3. Leeper TJ, Arnold J, Arel-Bundock V, Leeper MTJ. Package 'margins'. accessed December. 2017;5:2019.

4. Lenth R, Lenth MR. Package 'lsmeans'. The American Statistician. 2018;34(4):216–221.

5. Hanmer MJ, Ozan Kalkan K. Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models. American Journal of Political Science. 2013;57(1):263–277.

6. Schielzeth H. Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution. 2010;1(2):103–113.

7. Duursma R, Robinson A. Bias in the mean tree model as a consequence of Jensen's inequality. Forest Ecology and Management. 2003;186(1-3):373–380.