Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Estimation of adjusted rate differences using additive negative binomial regression

Mark W. Donoghoe*and Ian C. Marschner

Rate differences are an important effect measure in biostatistics and provide an alternative perspective to rate ratios. When the data are event counts observed during an exposure period, adjusted rate differences may be estimated using an identity-link Poisson generalised linear model, also known as additive Poisson regression. A problem with this approach is that the assumption of equality of mean and variance rarely holds in real data, which often show overdispersion. An additive negative binomial model is the natural alternative to account for this, however, standard model-fitting methods are often unable to cope with the constrained parameter space arising from the non-negativity restrictions of the additive model. In this paper, we propose a novel solution to this problem using a variant of the ECME algorithm. Our method provides a reliable way to fit an additive negative binomial regression model and also permits flexible generalisations using semi-parametric regression functions. We illustrate the method using a placebo-controlled clinical trial of fenofibrate treatment in patients with type II diabetes, where the outcome is the number of laser therapy courses administered to treat diabetic retinopathy. An R package is available that implements the proposed method.

Keywords: ECME algorithm; Negative binomial regression; Overdispersion; Rate difference; Semi-parametric regression

1. Introduction

In many biostatistical contexts we observe an outcome that counts the number of times an event of interest occurred and often these counts are observed over differing units of exposure, such as differing time periods of follow-up. When a collection of covariates is also available, such data allow us to build a regression model quantifying the effects of the covariates on the rate at which events occur. In this context, the parameters in an additive regression

Department of Statistics, Macquarie University, New South Wales 2109, Australia
NHMRC Clinical Trials Centre, University of Sydney, New South Wales 2006, Australia
** Correspondence to: Mark Donoghoe, Department of Statistics, Macquarie University, New South Wales 2109, Australia*
Email: mark.donoghoe@mq.edu.au

Contract/grant sponsor: The Australian Research Council (grant number DP110101254).

model represent adjusted rate differences, and provide an alternative to the adjusted rate ratios from a multiplicative model.

Rate differences may often be preferable to rate ratios as a measure of effect size in biostatistical applications. In studies evaluating an intervention, a rate difference quantifying the effect of the intervention directly relates to the expected number of events that may be prevented by its use. For example, in vaccine studies the rate difference is called the vaccine-attributable reduction, and can provide better information about the number of infections prevented than the vaccine efficacy, which is a relative effect size [1]. A similar distinction exists in epidemiological cohort studies where the rate difference is called the attributable risk. Rate differences are also useful in health economics, because they allow quantification of the cost of an intervention per event prevented within a given time period. Furthermore, risk factor models used for prediction and stratification are sometimes best presented in terms of rate differences rather than rate ratios, both from an interpretation perspective and because additive models sometimes fit the data better than multiplicative models.

A common approach for estimating adjusted rate differences is to fit a generalised linear model (GLM) with an identity link function, and assume that the observed counts have a Poisson distribution. We will refer to this analysis as additive Poisson regression, and a reliable algorithm for maximum likelihood estimation in this model has been described by Marschner [2]. Provided the mean model is correctly specified, the parameter estimates from such a model will be consistent [3, p. 80], but the Poisson constraint that the variance of the count is equal to its expected value often does not hold true in real data. In particular, it is common for count data to exhibit overdispersion, where the variance exceeds the mean [4]. An important consequence of overdispersion is that the estimated standard errors from a Poisson model will be underestimates, leading to overly narrow confidence intervals and inflated type I error rates [5].

Several model-based alternatives to Poisson GLMs for analysing overdispersed count data have been presented in the literature, and we discuss these in detail in Section B of the supplementary materials available online. Negative binomial regression is the most common of these approaches, and has been reviewed by various authors; see for example the books by Cameron and Trivedi [6] and Hilbe [7]. Some others, such as generalised negative binomial, generalised Poisson and Poisson–inverse Gaussian regression have been implemented in SAS [8], R [9] and Stata [10].

These implementations typically rely on gradient-based algorithms such as Newton–Raphson to maximise the likelihood. With multiplicative models this is usually reliable, because on the log link scale the parameter space is unconstrained. As such, the multiplicative model is the only available option in some of these implementations. While it is possible to estimate adjusted rate differences from a multiplicative model, using bootstrap resampling or the delta method to construct confidence intervals [11], it is important to recognise that this rate difference differs depending on the value of the adjustment covariates.

In situations where an additive model provides a better fit to the data when compared to a multiplicative model, or there is little difference in fit but rate differences are the favoured measure of association, the additive model should be preferred. Goodness of fit may be assessed by the use of model indices such as the AIC [12], AIC_c [13] or BIC [14]. Alternatively, we can compare the observed number of events to the number predicted under each model, using a Chi-squared test or an application of Hosmer and Lemeshow's deciles approach [15].

The existing methods for fitting overdispersed count regression models can be modified in order to apply the same gradient-based algorithms to an additive model, but in practice this presents challenges. For additive models, a restriction must be placed on the possible values of the parameters to ensure that estimated mean counts and rates are non-negative. For this reason, when applied to additive models, these standard algorithms often suffer convergence issues and can be unreliable.

The purpose of this paper is to address these challenges by presenting a reliable and flexible approach for additive negative binomial regression and using it to estimate adjusted rate differences in count data subject to

overdispersion. Section 2 will outline the additive model that we will use based on the negative binomial distribution, and then Section 3 will describe how a variant of the Expectation–Conditional Maximisation–Either (ECME) algorithm can be used to reliably fit this model. We will then extend this in Section 4 to allow additional flexibility through smooth semi-parametric regression functions. A detailed illustrative application of the methods will then be provided in Section 5, using data from a large clinical trial in type II diabetics where the outcome is the number of laser therapy courses administered to treat diabetic retinopathy.

2. Additive negative binomial regression

The negative binomial distribution can be derived in several ways [16], but in this context we motivate its use by introducing multiplicative Gamma-distributed errors into the Poisson model. We begin by assuming that the number of events Y_i for individual i $(i = 1, \ldots, n)$ is distributed as

$$Y_i \sim \text{Poisson}(N_i \lambda_i).$$

Here, N_i is the fixed exposure over which Y_i is observed, such that λ_i represents the event rate per unit of exposure for individual i. Typically, N_i will correspond to a fixed time period of observation for individual i, so that λ_i is the event rate per unit of time, although other types of exposure are also possible. When modelling count data rather than rates, we can set $N_i = 1$ for all i so that λ_i is the expected count.

Under an additive regression model, the event rate is a linear function of the covariate vector $\boldsymbol{x}_i$ and its associated parameter vector $\boldsymbol{\theta}^*$, which we denote by

$$\lambda_i = \Lambda(\boldsymbol{x}_i, \boldsymbol{\theta}^*) = \boldsymbol{x}_i \boldsymbol{\theta}^*. \tag{1}$$

This is an identity-link Poisson GLM in which component j of $\boldsymbol{\theta}^*$ is the adjusted rate difference per unit change in component j of $\boldsymbol{x}_i$. A number of authors have discussed methods for fitting this model to obtain adjusted rate differences; see for example [2] and [7, pp. 152–157]. This model is a natural alternative to the multiplicative Poisson model with log link function, where the (exponentiated) parameters are adjusted rate ratios.

The Poisson model has the restrictive assumption that $\text{Var}(Y_i) = \mathbb{E}(Y_i)$, which is frequently violated in the direction of overdispersion, $\text{Var}(Y_i) > \mathbb{E}(Y_i)$. A natural way to accommodate this overdispersion is to adopt a more general mean–variance relationship, $\text{Var}(Y_i) = \sigma^2 \mathbb{E}(Y_i)$, for $\sigma^2 > 1$. For notational convenience we will rewrite this relationship in the equivalent form $\text{Var}(Y_i) = (1 + \phi)\mathbb{E}(Y_i)$, for $\phi > 0$. Then the desired mean–variance relationship can be achieved by generalising the Poisson model such that the mean is perturbed by multiplicative errors,

$$Y_i \mid \eta_i \sim \text{Poisson}(N_i \lambda_i \eta_i),$$

where η_i is a random variable with $\mathbb{E}(\eta_i) = 1$ and $\text{Var}(\eta_i) = \phi/(N_i \lambda_i)$. This can be confirmed by a straightforward application of the law of total variance.

One possible cause of such overdispersion is unobserved heterogeneity in which some individuals are more event-prone than others due to factors that are not captured by the observed covariates. Whatever the cause of the overdispersion, λ_i retains its interpretation as an event rate, and $\boldsymbol{\theta}^*$ its interpretation as a vector of adjusted rate differences, averaged over the population.

Specification of the model is completed by assuming a distribution for η_i. A commonly used approach is to assume that η_i is Gamma-distributed, which dates back to Greenwood and Yule [17]. This results in a negative binomial marginal distribution for Y_i, equivalent to the model referred to as NegBin I by Cameron and Trivedi [18].

The parameter $\phi > 0$ measures the extent of the overdispersion, and the distribution of Y_i converges to a Poisson distribution as $\phi \to 0$.

It should be noted that other types of negative binomial models are also possible. In particular, if η_i is assumed to be Gamma-distributed with constant variance ϕ, the resulting negative binomial distribution has a quadratic mean–variance relationship, $\text{Var}(Y_i) = \mathbb{E}(Y_i)(1 + \phi\mathbb{E}(Y_i))$. This model is referred to as NegBin II by Cameron and Trivedi [18] and various other comparative discussions of the two models have been provided; see for example [7]. One advantage that is sometimes claimed for the NegBin II model is that, unlike NegBin I, when ϕ is fixed it is in the exponential family and therefore falls within the GLM framework. Nonetheless, this does not invalidate the use of the NegBin I distribution with an additive model for its mean. Indeed, in Section 5 we will present a motivating application in which our additive NegBin I approach is more appropriate, and this model will be our main focus in this paper. However, comparison with the additive NegBin II model will also be considered.

Expressing our model in terms of the usual negative binomial parameters leads to

$$Y_i \sim \text{NegBin}(r_i(\boldsymbol{\theta}), p) \qquad \text{where} \qquad r_i(\boldsymbol{\theta}) = \frac{1}{\phi} N_i \lambda_i = N_i \Lambda(\boldsymbol{x}_i, \boldsymbol{\theta})$$

with $\boldsymbol{\theta} = \boldsymbol{\theta}^*/\phi$ and $p = \phi/(1 + \phi)$. By finding the MLEs $\hat{\boldsymbol{\theta}}$ and $\hat{p}$ of this marginal negative binomial distribution, we can transform them back to obtain the MLEs of the coefficient of overdispersion $\hat{\phi}$ and the adjusted rate differences $\hat{\boldsymbol{\theta}}^*$:

$$\hat{\phi} = \frac{\hat{p}}{1 - \hat{p}} \qquad \text{and} \qquad \hat{\boldsymbol{\theta}}^* = \hat{\phi}\hat{\boldsymbol{\theta}}. \tag{2}$$

However, in practice the additive structure for $r_i(\theta)$ leads to substantial model-fitting challenges stemming from the constraint $r_i(\theta) \geq 0$. Indeed, as discussed in Section 6, there is no standard commercial software package that fits this model. In the next section we will describe how the additive structure combined with the fact that p is constant across individuals allows us to use the convolution properties of the negative binomial distribution to construct an underlying latent outcome model. This will then allow us to implement the ECME algorithm [19], which provides a stable approach to model fitting.

3. Model fitting

In this section we describe how the additive negative binomial model can be reliably fitted using a variant of the ECME algorithm. To do this, we need to first describe the constrained parameter space and a latent outcome model that will be used as the complete-data model in an ECME algorithm. We then describe a variant of the ECME algorithm based on the combinatorial EM algorithm of Marschner [20].

3.1. Parameter space

In describing the model-fitting procedure, it is helpful to keep the categorical and continuous covariates separate. We will examine a model that includes A categorical covariates, $u_{i1}, \ldots, u_{iA}$, and B continuous covariates, $v_{i1}, \ldots, v_{iB}$, for each individual i. Without loss of generality, the possible values of the a^{th} categorical covariate are $\{1, \ldots, k_a\}$, and the b^{th} continuous covariate can take any value in the range $[v_b^{(0)}, v_b^{(1)}]$, where $v_b^{(0)} = \min_i\{v_{ib}\}$ and $v_b^{(1)} = \max_i\{v_{ib}\}$.

The parameter vector $\boldsymbol{\theta}$ associated with this model has $J = 1 + \sum_{a=1}^{A} k_a + B$ components: $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_A, \beta_1, \ldots, \beta_B)^\top$, where α_0 is an intercept term, each $\boldsymbol{\alpha}_a = (\alpha_a(1), \ldots, \alpha_a(k_a))$ is a vector of parameters associated with the a^{th} categorical covariate and β_b is the parameter associated with the b^{th} continuous covariate.

Thus the additive model is

$$\Lambda(\boldsymbol{x}_i, \boldsymbol{\theta}) = \alpha_0 + \sum_{a=1}^{A} \alpha_a(u_{ia}) + \sum_{b=1}^{B} \beta_b v_{ib}, \tag{3}$$

which for simplicity we can express in the form equivalent to (1)

$$\Lambda(\boldsymbol{x}_i, \boldsymbol{\theta}) = \sum_{j=1}^{J} \theta_j x_{ij},$$

where θ_j is the j^{th} component of the parameter vector $\boldsymbol{\theta}$ and we have defined the covariate vector for individual i as $\boldsymbol{x}_i = (1, \boldsymbol{u}'_{i1}, \ldots, \boldsymbol{u}'_{iA}, v_{i1}, \ldots, v_{iB})$, where $\boldsymbol{u}'_{ia}$ is a vector of length k_a with 1 in the u_{ia}^{th} position and 0 elsewhere.

The parameter space for the additive negative binomial model requires that the expected counts are non-negative and the Gamma scale parameter is positive, that is $(\boldsymbol{\theta}, p) \in \Theta \times \Phi$, where

$$\Theta = \{\boldsymbol{\theta} : \Lambda(\boldsymbol{x}, \boldsymbol{\theta}) \geq 0, \, \boldsymbol{x} \in \mathcal{X}\} \qquad \text{and} \qquad \Phi = (0, 1) \tag{4}$$

for a covariate space $\mathcal{X}$, which consists of all possible combinations of the levels of the categorical covariates and any value of the continuous covariates within their observed ranges

$$\mathcal{X} = \prod_{a=1}^{A} \{1, \ldots, k_a\} \times \prod_{b=1}^{B} [v_b^{(0)}, v_b^{(1)}].$$

Note that the parameter space (4) does not place any restrictions on the individual components of the parameter vector $\boldsymbol{\theta}$, but we will introduce such restrictions in defining the latent outcome model.

3.2. Latent outcome model

In order to define a complete-data model for the ECME algorithm, we begin by considering a particular $\boldsymbol{t} = (t_1, \ldots, t_{A+B})$, giving the reference level for each of the $A + B$ covariates. For each categorical covariate $a = 1, \ldots, A$, we choose $t_a \in \{1, \ldots, k_a\}$ and set $\alpha_a(t_a) = 0$. For each of the continuous covariates $b = 1, \ldots, B$, the reference level is either the minimum or maximum observed value. That is, $t_{A+b} = v_b^{(\varsigma_b)}$ for a choice of $\varsigma_b \in \{0, 1\}$. For a particular ς_b, we define the shifted covariate $v'_{ib} = (-1)^{\varsigma_b}(v_{ib} - v_b^{(\varsigma_b)})$, which will be non-negative for all i, and (3) can be equivalently written as

$$\begin{aligned} \Lambda(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Lambda(\boldsymbol{x}'_i, \boldsymbol{\theta}') \\ &= \alpha'_0 + \sum_{a=1}^{A} \alpha_a(u_{ia}) + \sum_{b=1}^{B} \beta'_b v'_{ib} \\ &= \sum_{j=1}^{J} \theta'_j x'_{ij}, \end{aligned} \tag{5}$$

where $\boldsymbol{x}'_i = (1, \boldsymbol{u}'_{i1}, \ldots, \boldsymbol{u}'_{iA}, v'_{i1}, \ldots, v'_{iB})$ and $\boldsymbol{\theta}' = (\alpha'_0, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_A, \beta'_1, \ldots, \beta'_B)^\top$, with

$$\alpha'_0 = \alpha_0 + \sum_{b=1}^{B} \beta_b v_b^{(\varsigma_b)} \qquad \text{and} \qquad \beta'_b = (-1)^{\varsigma_b} \beta_b. \tag{6}$$

For a particular choice of $\boldsymbol{t}$, the complete-data model consists of J independent latent negative binomial random variables with common p underlying the observed Y_i. That is,

$$Y_i = \sum_{j=1}^{J} \mathcal{Y}_i^{(j)},$$

where each $\mathcal{Y}_i^{(j)} \sim \text{NegBin}(r_{ij}(\boldsymbol{\theta}'), p)$ and $r_{ij}(\boldsymbol{\theta}') = N_i \theta'_j x'_{ij}$.

Because the sum of independent negative binomial random variables with common p is itself negative binomial, the complete-data model is equivalent to the observed-data model, with the additional restriction that the $r_{ij}(\boldsymbol{\theta}')$ parameter of each of the latent variable distributions must be non-negative. As each x'_{ij} is defined to be non-negative for all i and j, this natural restriction imposed by the complete-data model for a given $\boldsymbol{t}$ is effectively a non-negativity constraint on each of the components of $\boldsymbol{\theta}'$.

From (5) it is clear that for any choice of the reference level vector $\boldsymbol{t}$, $\Lambda(\boldsymbol{t}, \boldsymbol{\theta}) = \alpha'_0$, and the non-negativity constraints on the transformed parameters ensure that the fitted value for any other transformed covariate vector will be larger than this. Thus, the parameter space related to the complete-data model for a particular choice of $\boldsymbol{t}$ is $\Theta(\boldsymbol{t}) \times \Phi$, where

$$\Theta(\boldsymbol{t}) = \{\boldsymbol{\theta} : \Lambda(\boldsymbol{x}, \boldsymbol{\theta}) \geq \Lambda(\boldsymbol{t}, \boldsymbol{\theta}) \geq 0,\ \boldsymbol{x} \in \mathcal{X}\}.$$

There are $\prod_a k_a \times 2^B$ possible choices for the reference level vector $\boldsymbol{t}$, which define a family of complete-data models. Importantly, for any $\boldsymbol{\theta} \in \Theta$, the smallest fitted value must correspond to one of these possible choices of $\boldsymbol{t}$. Hence, if we can find the constrained MLE associated with each of the complete-data models, the MLE for the observed-data model is simply the constrained MLE that attains the highest likelihood. This means that the MLE can be found by cycling through each of the possible choices of $\boldsymbol{t}$, which constitutes an implementation of the combinatorial EM algorithm described by Marschner [20].

A useful feature of this approach is that it is straightforward to consider a parameter space in which a particular coefficient β_b associated with a continuous covariate is constrained to be non-negative. This can be achieved by considering only the reference level vectors in which $\varsigma_b = 0$. Likewise, a non-positivity constraint can be imposed by considering only the reference level vectors in which $\varsigma_b = 1$. We will take advantage of this feature to include semi-parametric terms in our model, described in Section 4.

3.3. ECME algorithm

We find the MLE for a particular complete-data model by using the ECME algorithm described by Liu and Rubin [19]. In the ECME algorithm, the M-step of the EM algorithm [21] is replaced by a series of conditional maximisation (CM) steps, which act on a subset of the unknown parameters while keeping the others fixed at their current estimates. It differs from the ECM algorithm [22] in that some of the CM-steps can be designed to maximise the observed-data log-likelihood rather than the expected complete-data log-likelihood from the E-step. In order to ensure that the likelihood increases monotonically, Meng and Van Dyk [23] noted that the CM-steps maximising the observed-data log-likelihood must be performed after all steps that maximise the expected complete-data log-likelihood at each iteration.

After removing terms that do not depend on $\boldsymbol{\theta}$ or p, the observed-data log-likelihood for the additive negative binomial model is

$$\ell(\boldsymbol{\theta}, p; \boldsymbol{Y}) = \sum_{i=1}^{n} \log(\Gamma(Y_i + r_i(\boldsymbol{\theta}))) - \log(\Gamma(r_i(\boldsymbol{\theta}))) + r_i(\boldsymbol{\theta}) \log(1-p) + Y_i \log(p),$$

and the complete-data log-likelihood for a chosen $\boldsymbol{t}$ has a similar form:

$$L(\boldsymbol{\theta}', p; \boldsymbol{\mathcal{Y}}) = \sum_{i=1}^{n}\sum_{j=1}^{J} \log(\Gamma(\mathcal{Y}_i^{(j)} + r_{ij}(\boldsymbol{\theta}'))) - \log(\Gamma(r_{ij}(\boldsymbol{\theta}'))) + r_{ij}(\boldsymbol{\theta}')\log(1-p) + \mathcal{Y}_i^{(j)}\log(p).$$

Given a set of starting parameter estimates $\hat{\boldsymbol{\theta}}'_{(0)}$ and $\hat{p}_{(0)}$, the E-step at the $(c+1)^{\text{th}}$ iteration requires calculation of

$$Q(\boldsymbol{\theta}', p \mid \hat{\boldsymbol{\theta}}'_{(c)}, \hat{p}_{(c)}) = \mathbb{E}\left(L(\boldsymbol{\theta}', p; \boldsymbol{\mathcal{Y}}) \mid \boldsymbol{Y}, \hat{\boldsymbol{\theta}}'_{(c)}, \hat{p}_{(c)}\right),$$

for which we use the property that the conditional distribution of negative binomial random variables with common p given their sum is beta-binomial [24]. That is,

$$\begin{aligned}\Pr_{(c)}(\mathcal{Y}_i^{(j)} = y \mid Y_i) &= \Pr(\mathcal{Y}_i^{(j)} = y \mid Y_i, \hat{\boldsymbol{\theta}}'_{(c)}, \hat{p}_{(c)}) \\ &= \binom{Y_i}{y} \frac{\mathrm{B}\left(y + r_{ij}(\hat{\boldsymbol{\theta}}'_{(c)}), Y_i + r_i(\hat{\boldsymbol{\theta}}_{(c)}) - (y + r_{ij}(\hat{\boldsymbol{\theta}}'_{(c)}))\right)}{\mathrm{B}\left(r_{ij}(\hat{\boldsymbol{\theta}}'_{(c)}), r_i(\hat{\boldsymbol{\theta}}_{(c)}) - r_{ij}(\hat{\boldsymbol{\theta}}'_{(c)})\right)},\end{aligned}$$

for $y = 0, \ldots, Y_i$, where $\mathrm{B}(\cdot, \cdot)$ denotes the Beta function.

The first CM-step involves conditional maximisation of Q with respect to $\boldsymbol{\theta}'$, holding p constant at its current estimate $\hat{p}_{(c)}$. To do this, for each θ'_j we must find the non-negative root of the derivative

$$\frac{\partial Q}{\partial \theta'_j} = \sum_{i=1}^{n} N_i x'_{ij} \left\{ \sum_{y=0}^{Y_i} (\psi(y + r_{ij}(\boldsymbol{\theta}')) - \psi(r_{ij}(\boldsymbol{\theta}'))) \Pr_{(c)}(\mathcal{Y}_i^{(j)} = y \mid Y_i) + \log(1 - \hat{p}_{(c)}) \right\}.$$

This does not have an explicit solution, but we can exploit the property of the digamma function

$$\psi(y + r) - \psi(r) = \sum_{k=0}^{y-1} \frac{1}{r+k} \leq \frac{y}{r} \qquad y = 1, 2, \ldots$$

in order to find an upper bound for the root

$$\hat{\theta}'_{j(c+1)} \leq \frac{\sum_{i=1}^{n} \mathbb{E}_{(c)}\left(\mathcal{Y}_i^{(j)} \mid Y_i\right)}{\log(\frac{1}{1-\hat{p}_{(c)}}) \sum_{i=1}^{n} N_i x'_{ij}} = U_{j(c+1)} \qquad \text{where} \qquad \mathbb{E}_{(c)}\left(\mathcal{Y}_i^{(j)} \mid Y_i\right) = Y_i \frac{r_{ij}(\hat{\boldsymbol{\theta}}'_{(c)})}{r_i(\hat{\boldsymbol{\theta}}_{(c)})}.$$

This allows the straightforward use of an omnibus root-finding routine, such as `uniroot` in R, to find the root in the finite interval $[0, U_{j(c+1)}]$.

In the second CM-step, we keep the additive part $\hat{\boldsymbol{\theta}}'_{(c+1)}$ fixed at its current estimate, and update the estimate for p by maximising the observed-data log-likelihood, giving

$$\hat{p}_{(c+1)} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} r_i(\hat{\boldsymbol{\theta}}_{(c+1)})}.$$

If our starting estimates are within the parameter space for a given complete-data model, the updated estimates at each iteration are guaranteed to remain within the parameter space, and the likelihood will increase until it converges to its maximum value within the restricted parameter space $\Theta(\boldsymbol{t})$.

This process is repeated for each possible choice of the reference vector $\boldsymbol{t}$ in order to find the global maximum, which is the constrained MLE with the highest likelihood. However, there are some situations in which we can stop early, having only searched a subset of the restricted parameter spaces [20]. In particular, if we find a stationary

point in the interior of the parameter space $\Theta(\boldsymbol{t})$, we can be sure that it is the overall MLE.

Finally, after finding the overall MLE $(\hat{\boldsymbol{\theta}}', \hat{p})$, we can use the inverse of (6) to obtain the unshifted intercept and linear slope parameters, and then (2) to calculate the estimated coefficient of overdispersion $\hat{\phi}$, and the estimated rate difference parameters $\hat{\boldsymbol{\theta}}^* = (\hat{\alpha}_0^*, \hat{\boldsymbol{\alpha}}_1^*, \ldots, \hat{\boldsymbol{\alpha}}_A^*, \hat{\beta}_1^*, \ldots, \hat{\beta}_B^*)^\top$.

For categorical covariates, $\hat{\alpha}_a^*(u)$ is the rate difference associated with level u compared to the reference level t_a, adjusted for the other $A + B - 1$ covariates. For continuous covariates, $\hat{\beta}_b^*$ is the adjusted rate difference associated with a one-unit increase in the b^{th} continuous covariate.

3.4. Variance estimation

If our distributional assumption and mean model are correct, the MLE resulting from this method will be consistent, with an asymptotic multivariate normal distribution. We can obtain an estimate of the covariance matrix for the parameter estimates $(\hat{\boldsymbol{\theta}}^*, \hat{\phi})$ by using the inverse of the observed information matrix, evaluated at the MLE. The relevant formulae are given in Section A of the supplementary materials available with the online version of this article.

However, asymptotic normality of the MLE may be questionable if the estimate is on or close to the boundary of the parameter space. In this case, confidence intervals for parameter estimates may be obtained by using a resampling method such as the bootstrap. The stability of our fitting method ensures that the MLE can be obtained in every resampled dataset, and so there will be no bias caused by non-convergence in some samples.

4. Semi-parametric model

In some situations it may be desirable to relax the linearity restriction on the effect of continuous covariates on the expected count or rate, and instead adjust for these covariates semi-parametrically. To allow this, we extend (3) to include C additional unspecified functions, so

$$\Lambda(\boldsymbol{x}_i, \boldsymbol{\theta}) = \alpha_0 + \sum_{a=1}^{A} \alpha_a(u_{ia}) + \sum_{b=1}^{B} \beta_b v_{ib} + \sum_{c=1}^{C} f_c(w_{ic}).$$

As demonstrated by Donoghoe and Marschner [25] for the additive Poisson model, it is possible to incorporate the estimation of the unknown functions f_c into the additive negative binomial model by using B-spline basis functions. This allows us to retain the stability of our estimation method, while still ensuring that the fitted means are non-negative. The approach relies on the fact that it is straightforward to impose non-negativity constraints on any of the parameters associated with continuous covariates. This is easy to do for the additive negative binomial model also, as discussed in Section 3.2. We now describe the B-spline model.

4.1. B-spline model

For notational convenience, in this subsection we will first consider the case of $C = 1$ and drop the subscript c from the unspecified function f_c. We parameterise these regression functions using the B-spline model

$$f(w) = \sum_{d=1}^{D} \gamma_d B_d(w),$$

where $\{B_d;\, d = 1, \ldots, D\}$ are B-spline basis functions [26].

Let $\boldsymbol{\tau}$ be a set of $D + \kappa$ knots on which the B-spline basis is defined, where $\kappa - 1$ is the desired degree of the smooth curve and $D - \kappa$ of the knots are distinct turning points in the interior of the range of the continuous

covariate. Then the above model means that f is restricted to belong to $\mathcal{B}(\boldsymbol{\tau})$, the function space defined by the B-spline basis on $\boldsymbol{\tau}$. For simplicity, here we use $\kappa = 3$ and choose the $D - 3$ turning points to be evenly spaced quantiles of the observed $\{w_i\}$, but these methods will work with any choice of κ and knot vector. We allow the level of smoothness to vary by fitting models with different numbers of turning points and choosing the one with the smallest AIC [12]. Other suitable model selection criteria can also be used, such as the small-sample version, AIC_c [13].

Each basis function B_d is positive in (τ_d, τ_{d+3}) and zero elsewhere, meaning that if all of the γ_d coefficients are non-negative, the resulting f will also be strictly non-negative. Furthermore, the B-spline bases are normalised such that $\sum_{d=1}^{D} B_d(w) = 1$ for all w, so we must include an identifiability constraint on the coefficients of each curve.

We begin by choosing a reference level $t_{A+B+1} \in \{1, \ldots, D\}$, and impose an identifiability constraint by setting $\gamma_{t_{A+B+1}} = 0$. The remaining coefficients can be estimated by treating the basis function values $B_d(w_i)$ as continuous covariates in the ECME algorithm, restricting their associated parameters to be non-negative. The resulting f will belong to $\mathcal{B}^+(\boldsymbol{\tau}; t_{A+B+1})$, which denotes the subspace of strictly non-negative curves in $\mathcal{B}(\boldsymbol{\tau})$ that have their shape constrained by the choice of t_{A+B+1}. We repeat this for all D possible choices of t_{A+B+1}, and the estimate with the highest likelihood is the MLE for the semi-parametric model.

Extension to $C > 1$ is straightforward: for each $c = 1, \ldots, C$, we choose a reference level $t_{A+B+c} \in \{1, \ldots, D_c\}$, set $\gamma_{ct_{A+B+c}} = 0$ and use the ECME algorithm to find the MLE, constraining the remaining B-spline coefficients to be non-negative. To find the global maximum, we repeat this for all $\prod_{c=1}^{C} D_c$ possible choices of these reference levels.

4.2. Monotonicity restriction

This approach also allows us to impose a monotonicity restriction on any of the smooth curves. A sufficient condition for the monotonicity of f is that coefficients of the B-spline basis functions are themselves strictly non-decreasing, that is $\gamma_1 \leq \cdots \leq \gamma_D$ [27]. If we set $\gamma_1 = 0$ and introduce parameters $\delta_2, \ldots, \delta_D$ that represent the increments between successive coefficients, so

$$\delta_d = \gamma_d - \gamma_{d-1} \qquad d = 2, \ldots, D,$$

an equivalent condition is to require these increments to be non-negative.

Under this parameterisation,

$$f(w) = \sum_{d=2}^{D} \delta_d \left(\sum_{e=d}^{D} B_e(w) \right), \tag{7}$$

so a monotonicity restriction can be applied simply by treating the partial sums of B-spline bases in (7) as continuous covariates, and constraining their associated parameters to be non-negative in the CEM algorithm. Because we only need to consider one parameterisation for each monotonic regression function, extension to $C > 1$ is trivial, and in fact the inclusion of additional semi-parametric monotonic covariates does not require additional applications of the ECME algorithm.

5. Application

The Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) study was a double-blind randomised clinical trial examining the effect of fenofibrate treatment on cardiovascular risk in 9795 participants with type II diabetes, aged between 50–75 years over a median follow-up time of 5 years [28]. A pre-specified secondary endpoint of the study was laser photocoagulation treatment for diabetic retinopathy, a microvascular complication of diabetes. Laser

therapy is used to slow or prevent vision loss caused by retinopathy, but is associated with visual field reduction and other side-effects [29], so reducing the need for its use would be a positive outcome.

The vast majority of patients (96%) did not undergo laser therapy during the follow-up period, but individuals could have multiple courses of laser therapy, and we wish to estimate the effect of fenofibrate treatment on the number of laser therapy courses administered per 5 patient-years. In the primary analysis of these data [30], a Poisson GLM with a log link function was used to estimate a rate ratio (RR) for the treatment effect. The overdispersion apparent under the Poisson model was accounted for by rescaling the estimated standard errors.

There was no evidence of an interaction between fenofibrate treatment and the presence of known prior retinopathy at baseline on the multiplicative scale, with an estimated overall RR of 0.63 (95% CI 0.49–0.81) from this quasi-Poisson model. Accounting for overdispersion by using a distributional form, a multiplicative NegBin I model gives an estimated RR of 0.69 (95% CI 0.56–0.84), with a linear mean–variance relationship of $\text{Var}(\mu) = 3.9\mu$. A multiplicative NegBin II model gives 0.62 (95% CI 0.48–0.81), with quadratic mean–variance relationship $\text{Var}(\mu) = \mu + 30.8\mu^2$. The rate ratio estimates from these multiplicative models can be applied to the observed mean rates of laser therapy in the placebo group for those with and without known prior retinopathy to obtain estimates of the absolute rate reduction, and hence the expected number of laser therapy courses that may be avoided, by administering fenofibrate in these cohorts.

Alternatively, by using an additive model we are able to directly estimate these absolute rate differences and their confidence intervals. The model will also allow us to adjust for individuals' duration of diabetes prior to study entry, which is strongly associated with the risk of retinopathy, and confounded with prior retinopathy status.

Assuming no interaction between treatment and known prior retinopathy, we estimated the effect of fenofibrate, first unadjusted for any covariates, then adjusted for baseline retinopathy, in an additive NegBin I model. The results are shown in Figure 1 as rate differences per 5 patient-years, along with their 95% confidence intervals.

Diabetes duration up to 20 years was then entered into the model as a flexible smooth term with a maximum of 5 internal knots. The inclusion of diabetes duration substantially improved the model fit as measured by the *AIC* (3853.4 with diabetes duration compared to 4014.3 without), and led to a noticeable reduction in the magnitude of the treatment effect estimate.

In order to obtain an estimate of the rate difference for fenofibrate separately in patients with and without known prior retinopathy, the model must include an interaction between treatment allocation and retinopathy status. The estimates from this model, with and without adjustment for diabetes duration, are also shown in Figure 1. As expected from the analysis that showed no interaction on a multiplicative scale, the common relative effect of treatment is manifested in a smaller absolute rate difference for individuals at low risk, that is those without known prior retinopathy, compared to those at high risk. After adjustment for diabetes duration, this interaction is marginally significant ($p = 0.044$).

Under this model, the effect of diabetes duration on the rate of laser treatment is assumed to be the same for patients with and without known prior retinopathy. The estimated effect is shown in Figure 2, with a 95% confidence interval obtained using the information matrix standard errors. This effect of increasing diabetes duration was allowed to be non-monotonic, and the estimate shows a small reduction in rate just prior to 20 years, which may be due to sampling error. The estimated effect of diabetes duration under a monotonicity constraint is also shown in Figure 2.

The estimated rates under our additive NegBin I model were very similar to those under an equivalent additive Poisson model, and did not exceed 0.75 events per 5 patient-years in either case. The coefficient of overdispersion in the negative binomial model was estimated to be $\hat{\phi} = 2.44$ (with estimated standard error 0.25), such that the linear mean–variance relationship has a gradient of 3.44. The test statistic for a score test of the Poisson versus the NegBin I distribution, which has a standard normal distribution under the null hypothesis [18], is 114.6, suggesting strong evidence of overdispersion. Likewise the likelihood ratio test statistic, with a $0.5\chi^2(1)$ distribution under the

null [31], is 1476.9.

Accounting for overdispersion by using a distributional model allows us to estimate the number of patients in our sample expected to undergo a certain number of laser therapy courses, and these are shown in Figure 3 along with the observed histogram. The additive Poisson model substantially underestimates the number of patients who would not undergo any laser therapy, overestimates the number receiving a single course, and underestimates the number receiving 3 or more courses. In contrast, the negative binomial models capture the large proportion of zeros well: when we considered models that account for zero-inflation, these converged to the uninflated form.

Both additive and multiplicative NegBin I models appear to fit the observed counts well, including in the tail of the distribution. The Chi-squared statistic for the observed versus expected counts shown in Figure 3 was 3.44 for the additive model and 3.08 for the multiplicative model, both with an associated p-value greater than 0.9. The AIC for the multiplicative model was 3837.2, which was also marginally better than the 3851.3 of the additive model, but the goodness-of-fit statistic based on Hosmer and Lemeshow's deciles approach [15] favoured the additive model (66.2 versus 74.7, both $p > 0.95$). Overall, both models appear to have good fit to the data. Thus, if rate differences were preferred then the additive model would be appropriate.

In the tail of the distribution, the additive NegBin I model appears to provide better fit than an additive NegBin II model, which has an estimated mean–variance relationship of $\text{Var}(\mu) = \mu + 9.9\mu^2$. The NegBin II model underestimates the number of patients receiving between 2–6 courses of laser therapy, suggesting that the quadratic mean–variance relationship has inferior fit to the linear overdispersion of the NegBin I model for this dataset. This is supported by a comparison of AIC values (4132.8 for NegBin II versus 3851.3 for NegBin I) and Chi-squared goodness of fit statistics (48.3 for NegBin II versus 3.44 for NegBin I) for the fitted models. Empirically, a plot of the observed means and variances of the event rate within groups of patients defined by their assigned treatment, prior retinopathy status and diabetes duration (Figure 4) also suggests that a linear mean–variance relationship is more appropriate for these data.

As an extension, we considered the possibility that the effect of increasing diabetes duration may differ between patients with and without known prior retinopathy. We fitted a similar additive NegBin I model, with a monotonicity constraint but without the assumption of a common effect, and the expected rates under each model are shown in Figure 5. The model that allowed different diabetes duration curves had superior fit in terms of AIC (3842.7 versus 3852.8), and shows a dramatic increase in the expected rate between 5 and 15 years of diabetes duration for patients with known prior retinopathy. As can be seen by examining the vertical distance between the pairs of lines in each case however, the estimated treatment effect is virtually identical under both models.

6. Discussion

We have described a stable method for finding the MLE in additive negative binomial regression, which allows for estimation of adjusted rate differences in the presence of overdispersion. It also allows for smooth semi-parametric regression using B-splines, with optional monotonicity constraints on these curves. The method respects the natural non-negativity restriction on the fitted means, which can cause issues for approaches that employ standard gradient-based algorithms such as Newton–Raphson.

Section B of the supplementary materials available online gives a detailed summary of existing software that can potentially fit models for estimating adjusted rate differences in the presence of overdispersion. At present, the `gamlss` package in R [9] is the only other available implementation of a method for fitting additive NegBin I models. However, it employs a variant of the Newton–Raphson algorithm, which may perform poorly when the MLE is on or near the boundary of the parameter space. In the current version (4.3), the routine stops with an error if the estimates move outside the parameter space at any iteration, and although this may be improved by

step-size reduction, it is not straightforward to guarantee convergence in all cases. Together with the superiority of the additive NegBin I model in the analysis of Section 5, these considerations illustrate the practical usefulness of our method.

Our method has been implemented in the `addreg` package in R [32], which is available from the Comprehensive R Archive Network (CRAN). Currently the package implements a basic version of the method, searching every possible constrained parameter space consecutively. We plan that future releases will focus on optimising the computational efficiency using techniques to reduce the number of ECME algorithms that need to be run, such as those discussed by Marschner [20], and by taking advantage of the independent nature of the multiple ECME algorithms using parallel implementation on a multi-core processor.

The additive NegBin I model that we consider in this paper has a linear relationship between the conditional mean and variance, distinguishing it from the additive NegBin II model, which has a quadratic relationship. Both of these are nested in the NegBin-p model [18], which could be used to test which is more appropriate, but there are no existing methods for fitting this model with an additive mean. Alternatively, a scatterplot similar to that used by Armitage [33] or Figure 4 can help to distinguish between NegBin I and NegBin II. Other forms of overdispersion could be investigated using models such as the generalised Poisson [34] and Poisson–inverse Gaussian [35].

Both overdispersion and zero-inflation may be observed in the same data as a result of the data generating process. A possible extension of our method is the inclusion of zero-inflation by introducing a latent Bernoulli random variable into the complete-data model associated with the CEM algorithm. By utilising the methods of Marschner and Gillett [36] and Donoghoe and Marschner [37], stability of the algorithm could be maintained for any choice of logit, log or identity link in the regression model for the binary component. Such extensions are the subject of ongoing research.

Acknowledgement

The authors thank Tony Keech and the FIELD study investigators for providing the laser therapy dataset for analysis.

References

1. Greenwood B. Interpreting vaccine efficacy. *Clinical Infectious Diseases* 2005; **40**(10):1519–1520, doi:10.1086/429833.
2. Marschner IC. Stable computation of maximum likelihood estimates in identity link Poisson regression. *Journal of Computational and Graphical Statistics* 2010; **19**(3):666–683, doi:10.1198/jcgs.2010.09127.
3. Winkelmann R. *Econometric Analysis of Count Data*. 5th edn., Springer-Verlag: Berlin, 2008.
4. Tang W, He H, Tu XM. *Applied Categorical and Count Data Analysis*. CRC Press: Boca Raton, FL, 2012.
5. Gourieroux C, Monfort A, Trognon A. Pseudo maximum likelihood methods: Theory. *Econometrica* 1984; **52**(3):681–700, doi:10.2307/1913471.
6. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*. Cambridge University Press: Cambridge, UK, 1998.
7. Hilbe JM. *Negative Binomial Regression*. 2nd edn., Cambridge University Press: Cambridge, UK, 2011.
8. Chou NT, Steenhard D. A flexible count data regression model using SAS PROC NLMIXED. *SAS Global Forum* 2009; :Paper 250–2009.
9. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2005; **54**(3):507–554, doi:10.1111/j.1467-9876.2005.00510.x.
10. Hardin JW, Hilbe JM. *Generalized Linear Models and Extensions*. 3rd edn., Stata Press: College Station, TX, 2012.
11. Zhu H, Wu X. Confidence interval estimation for number of patient-years needed to treat. *Pharmaceutical Statistics* 2014; **13**(6):403–409, doi:10.1002/pst.1650.

12. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **19**(6):716–723, doi:10.1109/tac.1974.1100705.
13. Hurvich CM, Simonoff JS, Tsai CL. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1998; **60**(2):271–293, doi:10.1111/1467-9868.00125.
14. Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**(2):461–464, doi:10.1214/aos/1176344136.
15. Hosmer DW, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods* 1980; **9**(10):1043–1069, doi:10.1080/03610928008827941.
16. Johnson NL, Kotz S, Kemp AW. *Univariate Discrete Distributions*. 2nd edn., Wiley-Interscience: New York, 1993.
17. Greenwood M, Yule GU. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society* 1920; **83**(2):255–279, doi:10.1086/429833.
18. Cameron AC, Trivedi PK. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1986; **1**(1):29–53, doi:10.1002/jae.3950010104.
19. Liu C, Rubin DB. The ECME algorithm - a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 1994; **81**(4):633–648, doi:10.2307/2337067.
20. Marschner IC. Combinatorial EM algorithms. *Statistics and Computing* 2014; **24**(6):921–940, doi:10.1007/s11222-013-9411-7.
21. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 1977; **39**(1):1–38.
22. Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 1993; **80**(2):267–278, doi:10.2307/2337198.
23. Meng XL, Van Dyk D. The EM algorithm—an old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1997; **59**(3):511–567, doi:10.1111/1467-9868.00082.
24. Wisniewski TKM. Another statistical solution of a combinatorial problem. *The American Statistician* 1966; **20**(3):25, doi:10.1080/00031305.1966.10480403.
25. Donoghoe MW, Marschner IC. Flexible regression models for rate differences, risk differences and relative risks. *International Journal of Biostatistics* 2015; **11**(1):91–108, doi:10.1515/ijb-2014-0044.
26. Ramsay JO. Monotone regression splines in action. *Statistical Science* 1988; **3**(4):425–441, doi:10.1214/ss/1177012761.
27. Leitenstorfer F, Tutz G. Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics* 2007; **8**(3):654–673, doi:10.1093/biostatistics/kxl036.
28. FIELD Study Investigators. Effects of long-term fenofibrate therapy on cardiovascular events in 9795 people with type 2 diabetes mellitus (the FIELD study): randomised controlled trial. *The Lancet* 2005; **366**(9500):1849–1861, doi:10.1016/S0140-6736(05)67667-2.
29. Aiello LM. Perspectives on diabetic retinopathy. *American Journal of Ophthalmology* 2003; **136**(1):122–135, doi:10.1016/S0002-9394(03)00219-8.
30. Keech AC, Mitchell P, Summanen PA, O'Day J, Davis TME, Moffitt MS, Taskinen MR, Simes RJ, Tse D, Williamson E, *et al.*. Effect of fenofibrate on the need for laser treatment for diabetic retinopathy (FIELD study): a randomised controlled trial. *The Lancet* 2007; **370**(9600):1687–1697, doi:10.1016/S0140-6736(07)61607-9.
31. Lawless JF. Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 1987; **15**(3):209–225, doi:10.2307/3314912.
32. Donoghoe MW. *addreg: Additive Regression for Discrete Data* 2015. R package version 2.0. http://CRAN.R-project.org/package=addreg.
33. Armitage P. Studies in the variability of pock counts. *The Journal of Hygiene* 1957; **55**(4):564–581, doi:10.1017/S0022172400037451.
34. Consul PC. *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker Inc.: New York, 1989.
35. Dean C, Lawless JF, Willmot GE. A mixed Poisson-inverse-Gaussian regression model. *The Canadian Journal of Statistics* 1989; **17**(2):171–181, doi:10.2307/3314846.
36. Marschner IC, Gillett AC. Relative risk regression: reliable and flexible methods for log-binomial models. *Biostatistics* 2012; **13**(1):179–192, doi:10.1093/biostatistics/kxr030.
37. Donoghoe MW, Marschner IC. Stable computational methods for additive binomial models with application to adjusted risk differences. *Computational Statistics and Data Analysis* 2014; **80**:184–196, doi:10.1016/j.csda.2014.06.019.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

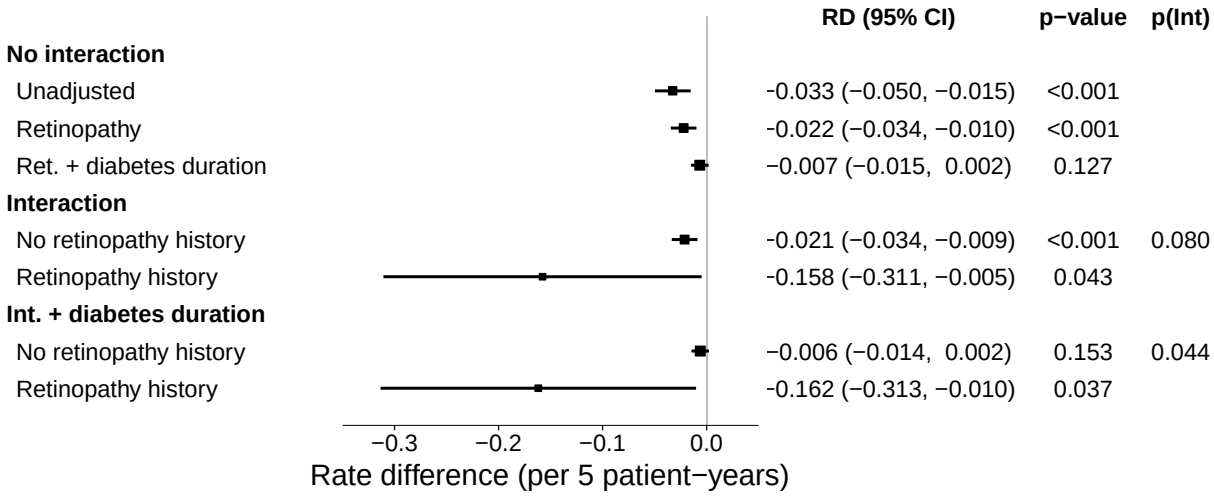


Figure 1. Estimated effect of fenofibrate on the rate of laser therapy in the FIELD study from additive negative binomial models, unadjusted and adjusted for prior retinopathy status and diabetes duration. p(Int) denotes the p-value for testing an interaction between treatment and retinopathy history.

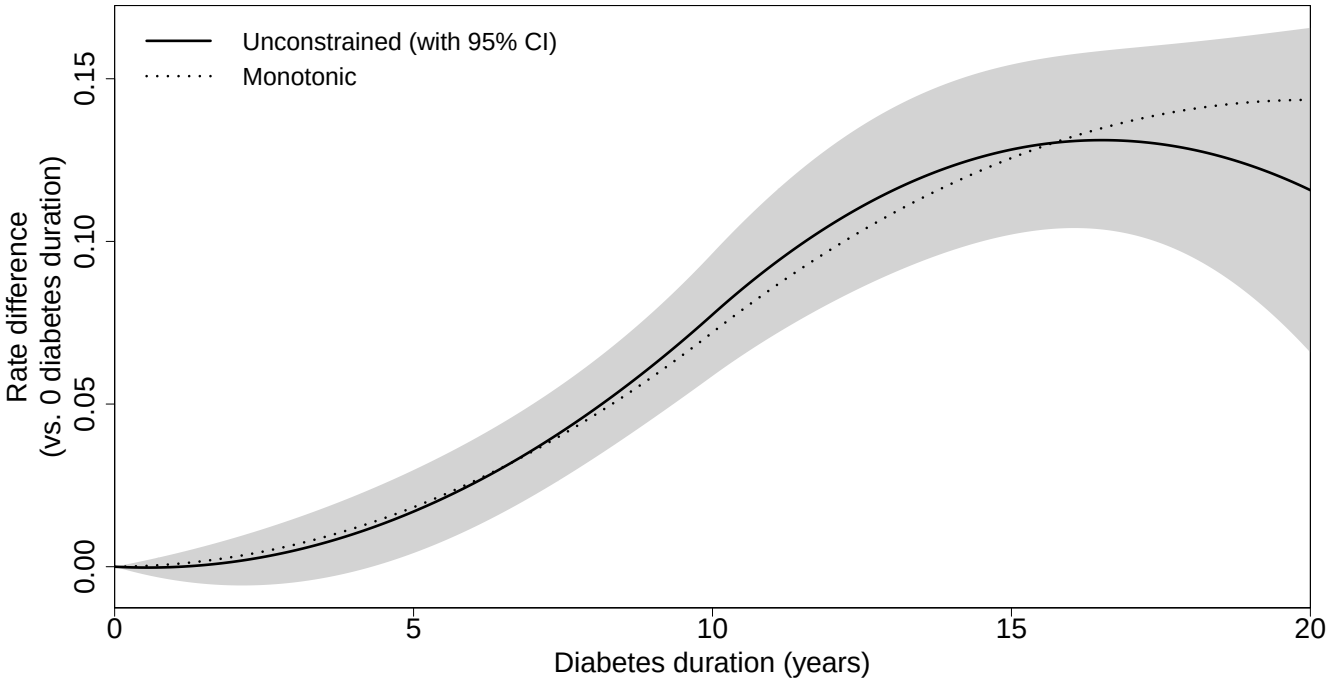
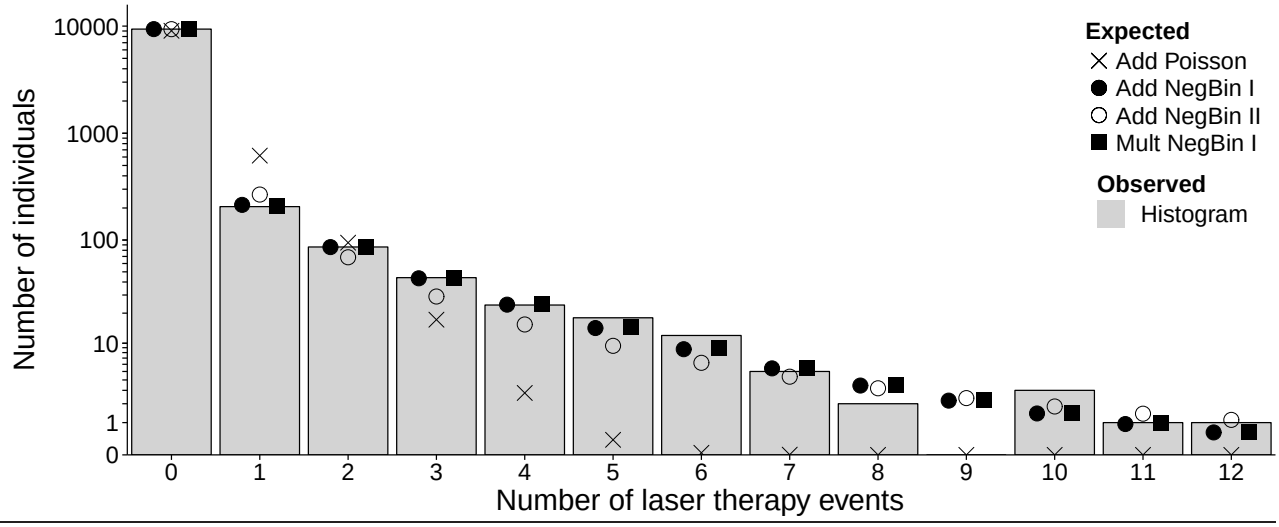


Figure 2. Laser therapy rate difference associated with increasing diabetes duration in the FIELD study under an additive negative binomial model.



	0	1	2	3	4	5	6	7	8	9	10	11	12
Observed	9376	206	86	44	24	18	12	5	2	0	3	1	1
Expected													
Add Poisson	9045.1	618.1	94.4	17.2	2.8	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Add NegBin I	9372.5	213.9	85.7	43.4	24.2	14.2	8.7	5.4	3.4	2.2	1.4	0.9	0.6
Add NegBin II	9362.7	267.3	68.7	29.0	15.5	9.4	6.2	4.4	3.2	2.4	1.8	1.4	1.1
Mult NegBin I	9376.9	209.1	85.1	43.5	24.4	14.5	8.8	5.5	3.5	2.3	1.5	1.0	0.6

Figure 3. Observed number of individuals with each number of laser therapy courses in the FIELD study, compared to the expected number under additive Poisson, NegBin I and NegBin II models, and a multiplicative NegBin I model. The y-axis has been log-scaled so that values in the upper tail can be compared despite the heavy skew (96% of individuals had zero events).

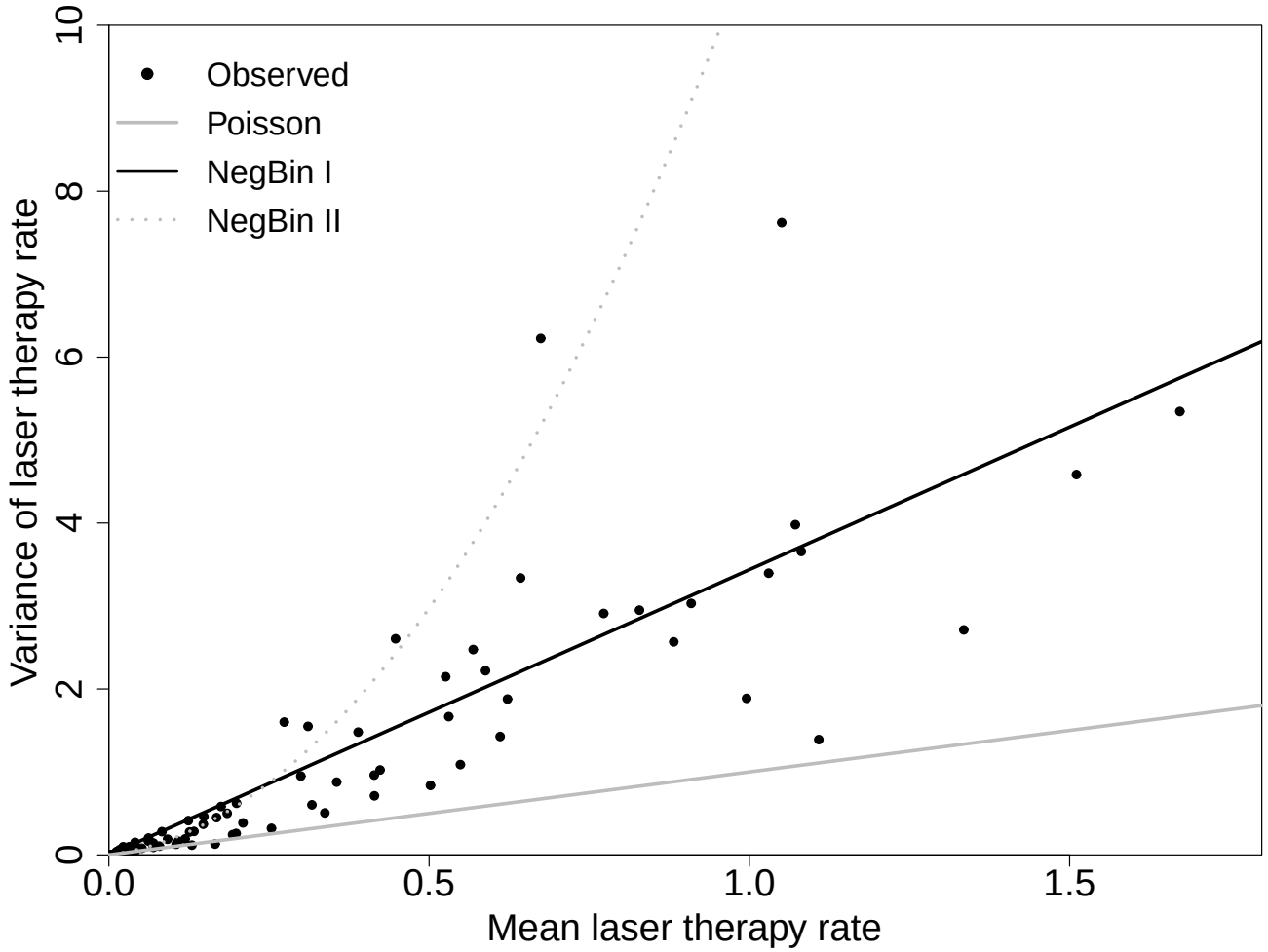


Figure 4. Observed mean and variance of laser therapy rates in the FIELD study, for groups defined by assigned treatment, prior retinopathy status and diabetes duration. Lines show the estimated mean–variance relationship from fitted Poisson, NegBin I and NegBin II models.

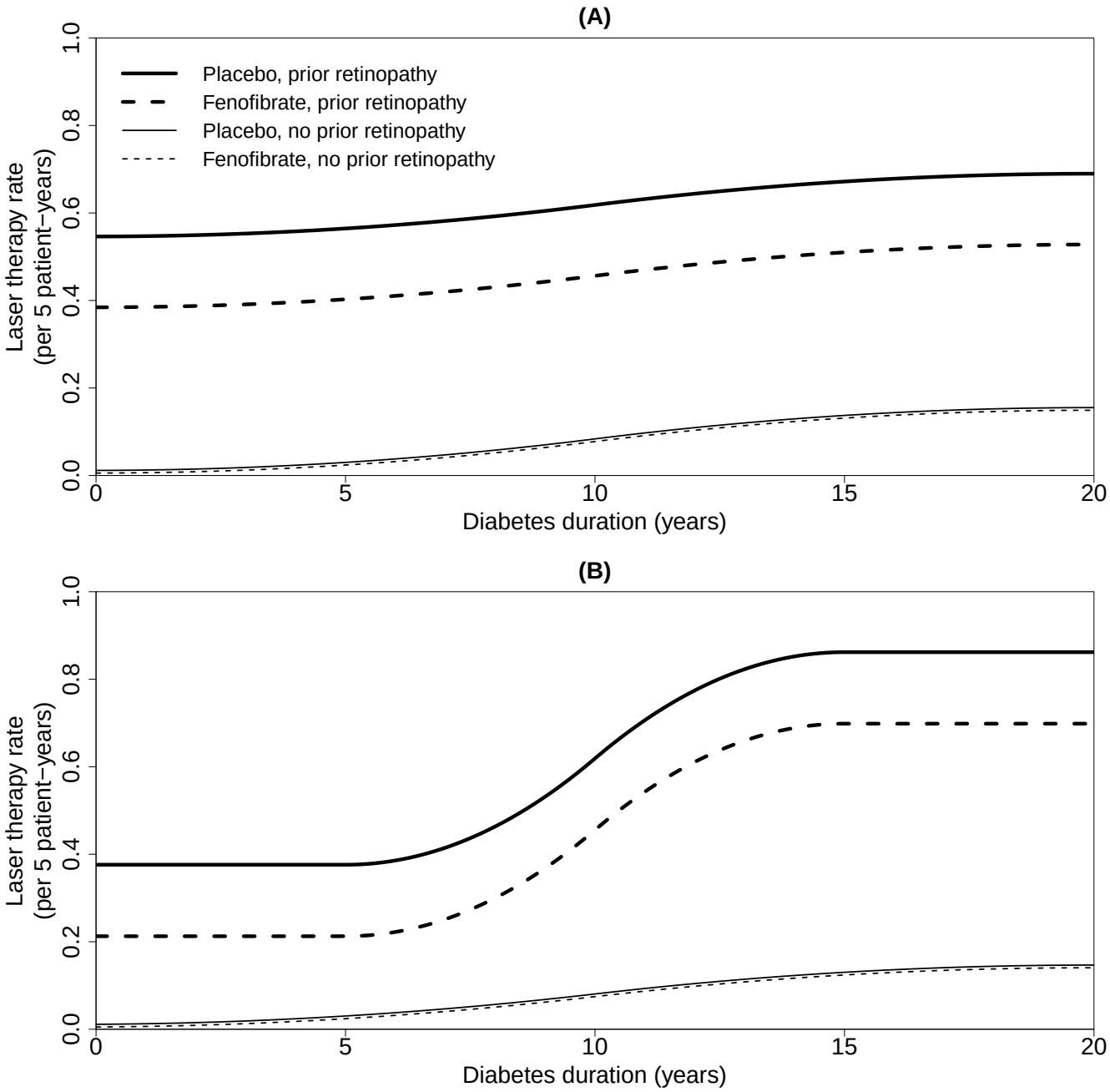


Figure 5. Fitted rates of laser therapy per 5 patient-years by baseline diabetes duration in the FIELD study under an additive negative binomial model. In (A), the effect of diabetes duration was restricted to be the same for all patients, whereas in (B) it was allowed to differ for those with and without known prior retinopathy.