

Stable computational methods for additive binomial models with application to adjusted risk differences[☆]

Mark W. Donoghoe^{a,b,*}, Ian C. Marschner^{a,b}

^a*Department of Statistics, Macquarie University, NSW 2109, Australia*

^b*NHMRC Clinical Trials Centre, University of Sydney, NSW 2006, Australia*

Abstract

Risk difference is an important measure of effect size in biostatistics, for both randomised and observational studies. The natural way to adjust risk differences for potential confounders is to use an additive binomial model, which is a binomial generalised linear model with an identity link function. However, implementations of the additive binomial model in commonly used statistical packages can fail to converge to the maximum likelihood estimate (MLE), necessitating the use of approximate methods involving misspecified or inflexible models. A novel computational method is proposed, which retains the additive binomial model but uses the multinomial-Poisson transformation to convert the problem into an equivalent additive Poisson fit. The method allows reliable computation of the MLE, as well as allowing for semi-parametric monotonic regression functions. The performance of the method is examined in simulations and it is used to analyse two datasets from clinical trials in acute myocardial infarction. Source code for implementing the method in R is provided as supplementary material.

Keywords: Additive binomial model, Multinomial-Poisson transformation, Risk difference, Semi-parametric regression

[☆]R code for implementing the methods in this paper can be found as online supplementary material.

*Correspondence to: Department of Statistics, Macquarie University, NSW 2109, Australia. Tel: +61 2 9850 4770. Fax: +61 2 9850 7669

Email address: `mark.donoghoe@mq.edu.au` (Mark W. Donoghoe)

1. Introduction

In biostatistical applications, the risk of an event is the probability of the event occurring within a specific time-frame. Risk difference is then the absolute difference in risk between two groups and is an important measure of effect size. For example, in randomised clinical trials, risk difference can be used to measure the magnitude of the treatment effect, while in observational studies it can be used to quantify the association between a risk factor and a disease event. Risk difference is important in practice because it is easier to interpret than the odds ratio and can present an alternative perspective to the relative risk.

As an important measure of effect size, the risk difference often needs to be adjusted for covariates. Analogous to logistic regression for estimating adjusted odds ratios, the natural model for estimating adjusted risk differences is a binomial generalised linear model (GLM) with identity link function, which we refer to as the additive binomial model.

The purpose of this paper is to address some common computational difficulties that arise with the additive binomial model for adjusted risk difference estimation. These difficulties arise from the requirement that the parameter space is constrained so that the linear probability model only produces probabilities in $[0, 1]$. This means that the model fitting is a constrained optimisation problem, and implementations of Fisher scoring or related procedures in popular statistics packages may fail to converge. Such numerical instability can occur even when the maximum likelihood estimate (MLE) is in the interior of the parameter space.

In light of these problems, there have been many proposals for estimating adjusted risk differences without using the additive binomial model. These include regression-based methods such as ordinary least squares or Poisson GLMs. However, with these methods the model is misspecified and fitted probabilities are not restricted to the $[0, 1]$ range. Alternative approaches can only provide the adjusted risk difference for a single binary comparison, and are essentially approximations to estimates from the additive binomial model. As demonstrated later in the paper, approximate methods for adjusted risk differences can have some undesirable properties, including loss of efficiency and violation of the parameter constraints.

In this paper we show that it is possible retain the natural additive binomial model for adjusted risk difference estimation, without introducing numerical instability into the model fitting process. We propose a compu-

tational method that uses a novel combination of two existing tools, the multinomial-Poisson transformation and a stable method for fitting additive Poisson models. A useful property of our approach is that it can be extended to allow semi-parametric adjustment, which is not available in other approaches.

We begin by specifying the additive binomial model that can be used to estimate adjusted risk differences, along with specification of the constrained parameter space and likelihood function. We then discuss how this can be recast into an equivalent additive Poisson estimation problem, using the multinomial-Poisson transformation. This allows application of stable computational methods for the additive Poisson model in order to fit the additive binomial model. Subsequently we present a range of simulation studies and analyses of two clinical trial datasets which demonstrate the advantages of our approach over competing methods for adjusted risk difference estimation. To facilitate practical implementation of this approach we have provided R code in the supplementary materials for this paper.

2. Method outline

2.1. Model definition

We assume that there are n independent observations $\mathbf{Y} = (Y_1, \dots, Y_n)$, where each observation Y_i is associated with a vector of A categorical and B continuous covariates. The covariate vector for observation i is therefore $\mathbf{x}_i = \{\mathbf{u}_i, \mathbf{v}_i\} = \{u_{i1}, \dots, u_{iA}, v_{i1}, \dots, v_{iB}\}$. Without loss of generality, we assume that $u_{ia} \in \{1, \dots, k_a\}$ and $v_{ib} \in \mathbb{R}$, where k_a is the number of levels of categorical covariate a .

In a binomial GLM, Y_i is the number of events observed in a fixed number N_i of independent Bernoulli trials, where each trial has an event probability $p(\mathbf{x}_i, \boldsymbol{\theta})$ for some parameter vector $\boldsymbol{\theta}$. This event probability is referred to as the risk. With an identity link function, the risk is assumed to have an additive structure

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

with $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_A, \boldsymbol{\beta})$, where $\boldsymbol{\alpha}_a = (\alpha_a(1), \dots, \alpha_a(k_a))$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_B)$. Model (1) requires A identifiability constraints $\alpha_a(r_a) = 0$ for

$a = 1, \dots, A$, where r_a is the chosen reference level for categorical covariate a .

The risk difference for comparing two covariate combinations \mathbf{x}_1 and \mathbf{x}_2 is the difference in risks $p(\mathbf{x}_1, \boldsymbol{\theta}) - p(\mathbf{x}_2, \boldsymbol{\theta})$. Thus, the parameter $\alpha_a(u)$ represents the risk difference for the u^{th} level of the a^{th} categorical covariate versus the reference level r_a , adjusted for the $A - 1$ other categorical covariates and B continuous covariates in the model. Likewise, β_b represents the adjusted risk difference associated with a one-unit increase in the b^{th} continuous covariate.

2.2. Parameter space and likelihood function

Since the linear functions $p(\mathbf{x}, \boldsymbol{\theta})$ are probabilities, they must lie in the interval $[0, 1]$ for all \mathbf{x} in the $(A + B)$ -dimensional covariate space \mathcal{X} . We will define \mathcal{X} as the space containing all possible combinations of the observed values of the covariates, that is $\mathcal{X} = \mathcal{U} \times \mathcal{V}$, where

$$\mathcal{U} = \prod_{a=1}^A \{1, \dots, k_a\},$$

represents all possible combinations of the categorical covariates, and

$$\mathcal{V} = \prod_{b=1}^B [v_b^{(0)}, v_b^{(1)}],$$

is the B -dimensional Cartesian product of the observed ranges of the continuous covariates, with $v_b^{(0)} = \min_i \{v_{ib}\}$ and $v_b^{(1)} = \max_i \{v_{ib}\}$.

We wish to find the MLE $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$, subject to the constraint that $\hat{\boldsymbol{\theta}}$ lies in the parameter space

$$\Theta = \{\boldsymbol{\theta} : 0 \leq p(\mathbf{x}, \boldsymbol{\theta}) \leq 1, \mathbf{x} \in \mathcal{X}\}. \quad (2)$$

The likelihood function for the additive binomial model, excluding a constant term, is

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{Y}) = \prod_{i=1}^n p(\mathbf{x}_i, \boldsymbol{\theta})^{Y_i} (1 - p(\mathbf{x}_i, \boldsymbol{\theta}))^{N_i - Y_i}.$$

This model can, in principle, be fitted by any GLM software that fits identity link binomial models. However, this approach is often numerically

unstable due to the box constraints specified by (2), which can be difficult to handle with standard computational methods such as Fisher scoring. We therefore consider a more reliable approach that involves the novel combination of two existing tools described in the next two subsections.

2.3. Multinomial-Poisson transformation

The multinomial-Poisson (MP) transformation, described by Baker (1994), relates the likelihood for a multinomial model to that of a Poisson model. In general, the MP transformation applies to observations $\mathbf{Z}_i = \{Z_{i1}, \dots, Z_{ij}, \dots\}$ from a multinomial distribution with $j \in J_i$, where J_i is any set of outcome categories for individual i . Here we describe and apply the MP transformation for the special case of $J_i = \{1, 2\}$ for all i , that is, the binomial distribution.

If Y_i are observations from a binomial distribution with N_i trials and event probability $p_i(\boldsymbol{\theta})$ for some parameter vector $\boldsymbol{\theta}$, we define the functions g_{i1} and g_{i2} such that

$$\frac{g_{i1}(\boldsymbol{\theta})}{G_i(\boldsymbol{\theta})} = p_i(\boldsymbol{\theta}) \quad \text{and} \quad \frac{g_{i2}(\boldsymbol{\theta})}{G_i(\boldsymbol{\theta})} = 1 - p_i(\boldsymbol{\theta}),$$

where $G_i(\boldsymbol{\theta}) = g_{i1}(\boldsymbol{\theta}) + g_{i2}(\boldsymbol{\theta})$.

The likelihood function for $\boldsymbol{\theta}$, excluding a multiplicative constant, is therefore

$$\mathcal{L}_B(\boldsymbol{\theta}; \mathbf{Y}) = \prod_{i=1}^n \left(\frac{g_{i1}(\boldsymbol{\theta})}{G_i(\boldsymbol{\theta})} \right)^{Y_i} \left(\frac{g_{i2}(\boldsymbol{\theta})}{G_i(\boldsymbol{\theta})} \right)^{N_i - Y_i}.$$

Using dummy parameters $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_n\}$, with all $\varphi_i > 0$, the MP transformation of \mathcal{L}_B is

$$\mathcal{L}_P(\boldsymbol{\theta}, \boldsymbol{\varphi}; \mathbf{Z}) = \prod_{i=1}^n \prod_{j=1}^2 (\varphi_i g_{ij}(\boldsymbol{\theta}))^{Z_{ij}} \exp(-\varphi_i g_{ij}(\boldsymbol{\theta})), \quad (3)$$

where $Z_{i1} = Y_i$ and $Z_{i2} = N_i - Y_i$.

The MLE of $\boldsymbol{\varphi}$ for fixed $\boldsymbol{\theta}$ is $\hat{\varphi}_i(\boldsymbol{\theta}) = N_i/G_i(\boldsymbol{\theta})$, and substituting this back into (3) gives

$$\mathcal{L}_P(\boldsymbol{\theta}, \hat{\boldsymbol{\varphi}}(\boldsymbol{\theta}); \mathbf{Z}) \propto \mathcal{L}_B(\boldsymbol{\theta}; \mathbf{Y}).$$

Thus, following the work of Richards (1961) on profile likelihoods, the MLE of $\boldsymbol{\theta}$ and the information matrix are identical for $\mathcal{L}_B(\boldsymbol{\theta})$ and $\mathcal{L}_P(\boldsymbol{\theta}, \boldsymbol{\varphi})$.

This means that the MLE for the binomial model may be found by maximising \mathcal{L}_P , which takes the same form as the likelihood for a Poisson model with

$$Z_{ij} \sim \text{Poisson}(\varphi_i g_{ij}(\boldsymbol{\theta})). \quad (4)$$

The problem of finding the MLE for an additive binomial model (1) can thus be transformed into one of finding the MLE of a Poisson model (4) which involves both multiplicative ($\boldsymbol{\varphi}$) and additive ($\boldsymbol{\theta}$) components.

For model (1), the multiplicative component of (4) can be eliminated by defining

$$g_{i1}(\boldsymbol{\theta}) = N_i p(\mathbf{x}_i, \boldsymbol{\theta}) \quad \text{and} \quad g_{i2}(\boldsymbol{\theta}) = N_i(1 - p(\mathbf{x}_i, \boldsymbol{\theta})). \quad (5)$$

Then $G_i(\boldsymbol{\theta}) = N_i$, and the MLEs of the dummy parameters are $\hat{\varphi}_i = 1$ for all i , meaning that the problem reduces to one of finding the MLE of an additive Poisson model.

Note also that the parameter space for $\boldsymbol{\theta}$ which restricts the probabilities $p_i(\boldsymbol{\theta})$ to lie within $[0, 1]$ is the same as that which requires both $g_{i1}(\boldsymbol{\theta}) \geq 0$ and $g_{i2}(\boldsymbol{\theta}) \geq 0$ for all i . That is, the parameter constraints on the binomial probabilities are equivalent to non-negativity constraints on the Poisson means in (4).

2.4. Additive Poisson regression

The MP transformation converts an additive binomial fit into an additive Poisson fit. However, although fitting an additive Poisson model tends to be more numerically stable than fitting an additive binomial model, it can still be subject to instability in standard software. We therefore make use of the method presented by Marschner (2010) for additive Poisson models, which always provides reliable convergence to the MLE. As well as numerical stability, this method also has a number of other advantages.

The approach described by Marschner (2010) is a stable variant of the Expectation-Maximisation (EM) algorithm, and applies to any identity link Poisson GLM. The computational method is an example of a combinatorial EM algorithm, which was presented in general terms by Marschner (2014). The main advantage of this approach is that it reliably accommodates the required non-negativity constraints on the Poisson means $g_{ij}(\boldsymbol{\theta})$ in (5). In addition, the method has some flexible features that enhance its usefulness. Firstly, while always accommodating the non-negativity constraints on the Poisson means, the method allows the model fitting to be conducted either

with or without non-negativity constraints on the individual regression parameters $\boldsymbol{\theta}$. This is a useful feature that we make use of in implementing our method in Section 3.2. Secondly, the method can accommodate semi-parametric monotone regression functions, which allows semi-parametric adjustment of risk differences.

Next we describe in detail how the combination of these two basic methods, the MP transformation and stable additive Poisson regression, yields a reliable method for the additive binomial model that can be used for adjusted risk difference estimation.

3. Additive binomial regression

3.1. Linear covariates

We will begin by examining the case of a single continuous covariate v_i , with no other covariates in the model, so (1) reduces to

$$p(v_i, \boldsymbol{\theta}) = \alpha_0 + \beta v_i.$$

Without loss of generality we use a rescaled version of the continuous covariate

$$v_i^* = \frac{2v_i - (v^{(0)} + v^{(1)})}{v^{(1)} - v^{(0)}},$$

where $v^{(0)} = \min_i\{v_i\}$ and $v^{(1)} = \max_i\{v_i\}$, so that $v_i^* \in [-1, 1]$. Accordingly, we have a rescaled parameter vector $\boldsymbol{\theta}^* = (\alpha_0^*, \beta^*)$, such that $p(v, \boldsymbol{\theta}) = p(v^*, \boldsymbol{\theta}^*)$, using

$$\alpha_0^* = \alpha_0 + \frac{v^{(0)} + v^{(1)}}{2}\beta \quad \text{and} \quad \beta^* = \frac{v^{(1)} - v^{(0)}}{2}\beta. \quad (6)$$

The MP transformation is useful for additive binomial models because an additive model for $p(\cdot, \cdot)$ implies an additive model for $1 - p(\cdot, \cdot)$. Thus, as in (5), we can define

$$g_{ij}(\boldsymbol{\theta}^*) = N_i g_j(v_i^*, \boldsymbol{\theta}^*) \quad j = 1, 2,$$

where

$$\begin{aligned} g_1(v_i^*, \boldsymbol{\theta}^*) &= p(v_i^*, \boldsymbol{\theta}^*) = \alpha_0^* + \beta^* v_i^* \\ g_2(v_i^*, \boldsymbol{\theta}^*) &= 1 - p(v_i^*, \boldsymbol{\theta}^*) = (1 - \alpha_0^*) + \beta^* (-v_i^*). \end{aligned}$$

This leads to a unified additive model

$$g_j(V_{ij}, \boldsymbol{\theta}^*) = \delta_j + \beta^* V_{ij}, \quad (7)$$

where $V_{ij} = (-1)^{j-1} v_i^*$ and $(\delta_1, \delta_2) = (\alpha_0^*, 1 - \alpha_0^*)$. Note that $V_{ij} \in [-1, 1]$ for all i, j , ensuring that the covariate space is preserved in the unified model.

It follows from the MP transformation discussed in Section 2.3 that our problem of finding the MLE for the additive binomial model is equivalent to finding the MLE for an additive Poisson model with $2n$ observations, $\{(Z_{i1}, Z_{i2}), i = 1, \dots, n\}$, where

$$Z_{ij} \sim \text{Poisson}(N_i g_j(V_{ij}, \boldsymbol{\theta}^*)), \quad (8)$$

with $Z_{i1} = Y_i$ and $Z_{i2} = N_i - Y_i$.

Model (8) requires the non-negativity constraints $g_j(V, \boldsymbol{\theta}^*) \geq 0$ for all $V \in [-1, 1]$, which ensures that $p(v^*, \boldsymbol{\theta}^*) \in [0, 1]$ for all $v^* \in [-1, 1]$. Fitting (8) subject to these constraints is achieved using the additive Poisson method of Marschner (2010) with one categorical covariate and one continuous covariate as specified by (7). The final step is then to transform $\boldsymbol{\theta}^*$ back onto its original scale using the relationships in (6).

Extension to $B > 1$ continuous covariates is straightforward. Each covariate is rescaled onto $[-1, 1]$, and the MLE for the rescaled additive binomial model is the same as the MLE for an additive Poisson model with one categorical covariate, B continuous covariates and $2n$ observations. This approach allows multiple linear regression models, which include non-linear polynomial models.

3.2. Categorical covariates

The approach used for continuous covariates does not apply directly to categorical covariates in an additive binomial model. However, a modification of this approach, again using the MP transformation, does allow incorporation of categorical covariates.

We begin by considering the model with a single categorical covariate $u_i \in \{1, 2, \dots, k\}$, so model (1) reduces to

$$p(u_i, \boldsymbol{\theta}) = \alpha_0 + \alpha_1(u_i). \quad (9)$$

Using the identifiability constraint $\alpha_1(1) = 0$, model (9) can be rewritten as a linear model

$$p(u_i, \boldsymbol{\theta}) = \alpha_0 + \sum_{b=2}^k \beta_b v_{ib}, \quad (10)$$

for an appropriately chosen parameterisation, β_b and v_{ib} . There are many possible parameterisations and a natural choice is

$$\beta_b = \alpha_1(b) \quad \text{and} \quad v_{ib} = \mathbf{1}\{u_i = b\}, \quad (11)$$

so that β_b is the contrast between level b and the reference level 1. The representation (10) would then seem to allow the categorical covariate model to be fitted using the methods described in Section 3.1 for linear covariates. In particular, note that the MP transformation described in Section 3.1 can again be applied, so that the additive binomial model (10) can be fitted using the equivalent additive Poisson model (8), with $k - 1$ linear covariates. However, there is a problem in that the procedure described in Section 3.1 will maximize the likelihood function over the parameter space that restricts the fitted event probabilities to be in $[0, 1]$ for all possible covariate combinations (v_{i2}, \dots, v_{ik}) in which each v_{ib} is in $[0, 1]$. This is overly restrictive, because (11) does not allow more than one of the v_{ib} to equal to 1 for each i . Thus, the method of Section 3.1 applied to the linear covariate model (10) would impose additional constraints that would cause the likelihood function to be maximised over a smaller parameter space than is desired.

An alternative parameterisation is

$$\beta_b = \alpha_1(b) - \alpha_1(b - 1) \quad \text{and} \quad v_{ib} = \mathbf{1}\{u_i \geq b\}, \quad (12)$$

so that the parameters β_b represent the increments between successive levels of the categorical covariate. This parameterisation has an analogous problem to that described above for parameterisation (11), so the method of Section 3.1 cannot be applied directly. However, the advantage of (12) is that it allows a simple modification that rectifies the problem.

As described in Section 2.4, Marschner (2010) presented a method for fitting the additive Poisson model, which can be applied subject to non-negativity constraints on the regression coefficients. When applied in the present context this method is an EM algorithm that imposes the constraints $\beta_b \geq 0$ for all $b = 2, \dots, k$, or equivalently, $\alpha_1(1) \leq \alpha_1(2) \leq \dots \leq \alpha_1(k)$. Although this imposes an undesired order restriction on the parameters, this constraint can be removed by repeatedly applying the order-restricted method after permuting the levels of the categorical covariate.

To see this, we first define the set \mathcal{T} , which consists of the $k!$ possible permutations of the levels of u_i . For each permutation $t \in \mathcal{T}$, there is a corresponding vector of permuted parameters $(\alpha_1^{(t)}(1), \alpha_1^{(t)}(2), \dots, \alpha_1^{(t)}(k))$.

Application of the additive Poisson method with non-negativity constraints leads to maximisation of the likelihood over the space

$$\Theta^{(t)} \subset \Theta = \{\boldsymbol{\theta} : 0 \leq p(u, \boldsymbol{\theta}) \leq 1, \text{ for all } u = 1, \dots, k\},$$

where $\Theta^{(t)}$ is the subset of the parameter space Θ that has $\alpha_1^{(t)}(1) \leq \alpha_1^{(t)}(2) \leq \dots \leq \alpha_1^{(t)}(k)$. Since the parameter space Θ may be partitioned into $k!$ such subsets corresponding to the $k!$ orderings $t \in \mathcal{T}$, it follows that Θ is the union of these subsets, $\Theta = \bigcup_t \Theta^{(t)}$. Thus, having found the constrained maximum within each restricted parameter space $\Theta^{(t)}$, the global maximum over Θ will simply be the constrained maximum that achieves the highest likelihood. This procedure of cycling through all possible permutations of the categorical covariate levels, and applying an EM algorithm for each permutation, is an example of a combinatorial EM algorithm (Marschner, 2014).

For the model with $A > 1$ categorical covariates, where covariate a has k_a distinct levels, the same procedure is applicable except that we must consider the Cartesian product of all possible permutations of each covariate. This leads to $K = \prod_{a=1}^A k_a!$ restricted parameter spaces that have to be searched for the MLE. In practice, if one of these spaces is found to have a stationary maximum, then it is the MLE and the algorithm may be halted. The same approach can be combined with the method described in Section 3.1 for linear covariates, to fit the model with A categorical and B continuous covariates specified in (1). In this case the EM algorithm would be applied a maximum of $K = 2^{B+1} \prod_{a=1}^A k_a!$ times.

3.3. Data analysis example 1

The ASSENT-3 study (ASSENT-3 Investigators, 2001) was a clinical trial of 6095 patients with acute myocardial infarction (heart attack), randomly allocated to treatment regimens containing antithrombotic therapies. The primary treatment comparison of interest was between the group allocated to unfractionated heparin (UFH; $n = 2038$) and the group allocated to receive enoxaparin ($n = 2040$). The trial was designed as a non-inferiority study, with the non-inferiority margin being a 1% risk difference in favor of UFH for the composite endpoint of 30-day mortality and in-hospital reinfarction or ischemia. As a brief numerical illustration we consider estimation of the risk difference between UFH and enoxaparin, adjusted for age.

For comparative purposes, we begin by investigating the form of the relationship between age, treatment and risk by fitting binomial GLMs with

Table 1: Comparison of adjusted effect measures in Example 1, based on GLMs with logit, log or identity link functions. Adjusted estimates are displayed for the odds ratio (OR), relative risk (RR) and risk difference (RD). Standard errors were estimated using the information matrix (SE_I) and bootstrap resampling (SE_B), and are shown on the log scale for the logit and log link models.

Parameter	Logit link			Log link			Identity link		
	OR	SE_I	SE_B	RR	SE_I	SE_B	RD	SE_I	SE_B
Treatment	0.70	0.093	0.093	0.74	0.080	0.080	-0.041	0.010	0.010
Age (per year)	1.03	0.004	0.004	1.03	0.003	0.003	0.003	0.0004	0.0003
Intercept	0.06	0.170	0.171	0.06	0.147	0.151	0.044	0.015	0.018
Scaled deviance	1.068			1.065			1.091		

three different link functions: the logit link (adjusted odds ratio), the log link (adjusted relative risk) and the identity link (adjusted risk difference). We estimated the adjusted risk difference using the method presented above, as implemented in an R function called `addbin`. The resulting parameter estimates and standard errors (derived from the observed information matrix) were identical to those found using the `glm` function in R (R Core Team, 2013) and `PROC GENMOD` in SAS (SAS Institute Inc., 2008). For all three link functions, likelihood ratio tests for the inclusion of either a quadratic age relationship or an interaction between age and treatment were non-significant, and the parameter estimates and their approximate standard errors are shown in Table 1.

The deviances for the three alternative link functions are comparable, suggesting that any of these may be appropriate for modelling the risk of an event. This is an example of a scenario in which risk differences may be useful because they may be considered more interpretable than other measures, particularly odds ratios.

We will focus on the additive binomial model in order to obtain the treatment effect as a risk difference, adjusted for age. Whilst asymptotic normality would allow construction of approximate confidence intervals, this assumption is questionable when the MLE is close to the boundary of the parameter space. We will demonstrate the stability of our algorithm by constructing non-parametric confidence intervals based on 1000 bootstrap samples.

The proposed method converged to the MLE in all 1000 bootstrap samples. In contrast, `PROC GENMOD` failed to converge in 214 samples. Figure 1

compares the distribution of the MLEs of each parameter, separated by whether or not PROC GENMOD reached convergence, demonstrating that confidence intervals obtained from only the converged samples would be biased.

The `glm` function available in R failed to converge in only one sample, however, this non-convergence was concerning because both `addbin` and PROC GENMOD converged to a point in the interior of the parameter space. Figure 2 shows the deviance achieved at each iteration of `glm`, demonstrating the potential instability of the algorithm implemented in R. This type of periodic non-convergence in R has been observed in other related contexts; see for example Marschner and Gillett (2012). In contrast, Figure 2 shows that `addbin` exhibited stable, albeit slow, convergence.

The additive model yields an estimate of 4.11% for the adjusted risk difference favouring the enoxaparin arm, with a one-sided 95% confidence interval from bootstrap resampling that extends to a 2.37% risk difference, still in favour of enoxaparin. This is well below the pre-specified 1% margin in favour of UFH, and so we can conclude that enoxaparin is not inferior (and in fact is superior) to UFH after adjusting for age. These adjusted results are consistent with the unadjusted results, which is not unexpected because age was balanced by randomisation. Nonetheless, the example does provide an initial numerical illustration of the method's performance.

4. Simulation study

For a more detailed evaluation of the performance of the MLE from an additive binomial GLM as an estimator of adjusted risk difference, we performed a number of simulation studies. In these simulations, the MLE was computed using the proposed method described in Section 3, as well as using the `glm` routine in R and PROC GENMOD in SAS. We empirically assessed the statistical properties of the proposed method, and compared it with various alternative non-MLE methods for calculating adjusted risk differences.

4.1. Summary of alternative methods

4.1.1. Misspecified regression models

In cases where the additive binomial model fails to converge, alternative models have been suggested in which the distribution of the outcome variable is misspecified in order to estimate the model parameters. Using a Poisson GLM with an identity link was discussed by Spiegelman and Hertzmark (2005), and Cheung (2007) proposed modified least-squares (MLS), where

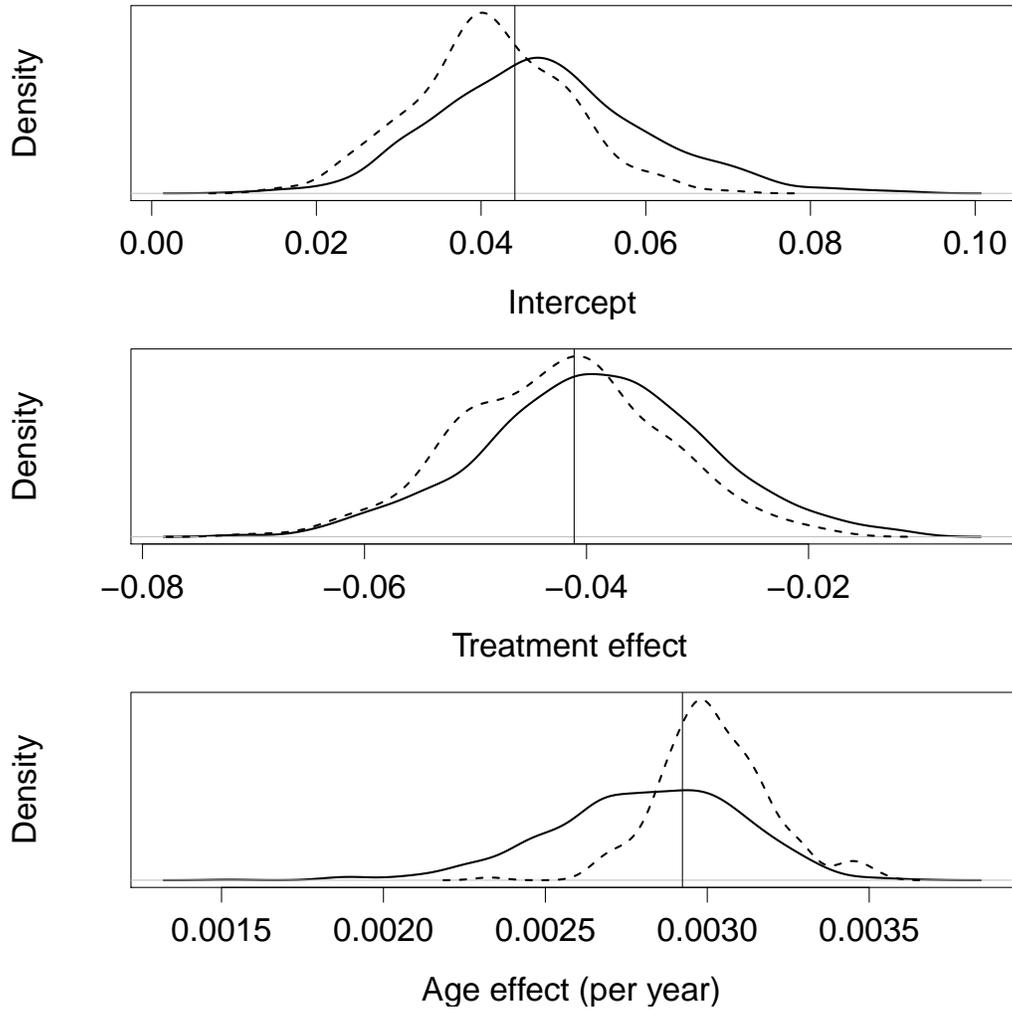


Figure 1: Smoothed density estimates of the MLE from 1000 bootstrap samples in Example 1, using `addbin` to compute the intercept (top), treatment (middle) and age (bottom) parameter estimates. Results are separated by whether `PROC GENMOD` converged (79%, solid line) or not (21%, dashed line). The vertical line shows the parameter estimate in the original data.

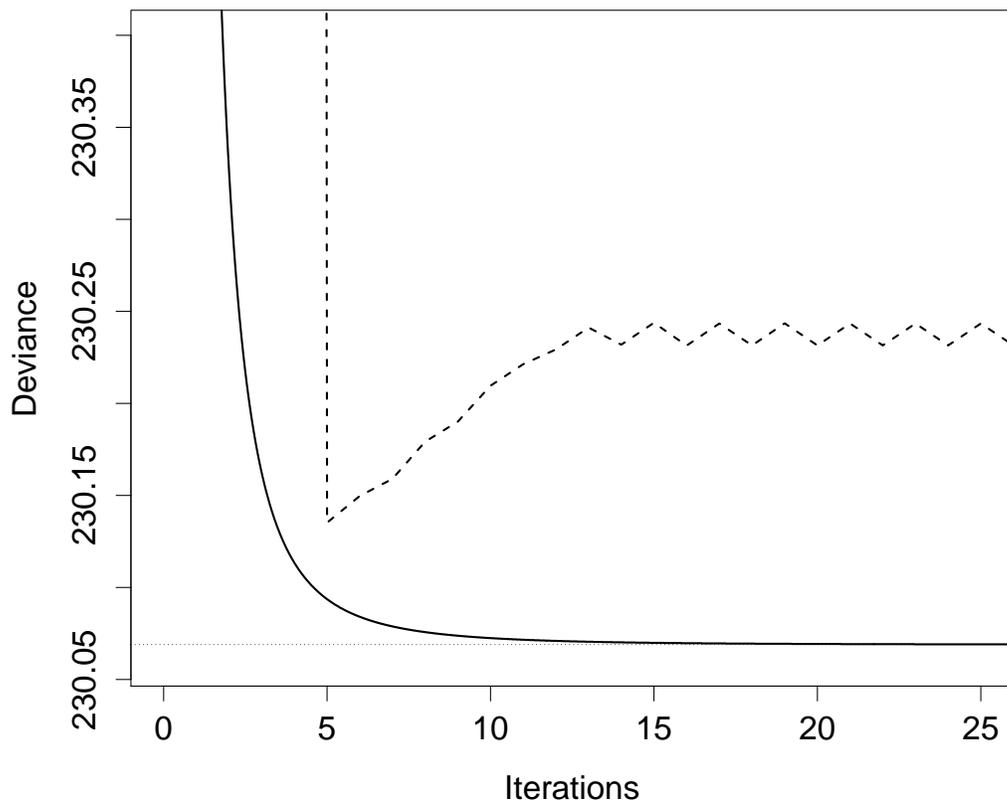


Figure 2: Iteration history for `glm` (dashed line) and `addbin` (solid line) in a bootstrap sample from Example 1. For the `addbin` iteration history, the number of iterations has been divided by 100. The dotted line denotes the optimal deviance.

the binomial risk is represented as the expected value of a binary dependent variable, and ordinary least-squares is used to find parameter estimates. In both cases, a robust variance estimator is used in calculating confidence intervals. The fitted risks from MLS are unrestricted, and the fitted Poisson means are only constrained to be non-negative, so both approaches can produce models with fitted risks outside $[0, 1]$.

4.1.2. Weighted mean methods

Other alternative methods only estimate the adjusted risk difference for a single binary comparison, rather than for a multivariable regression model. The first such methods were based on data in the form of stratified 2×2 contingency tables, with the risk difference estimator being a weighted average of the unadjusted risk differences observed in each stratum. We examine weighting schemes defined by Cochran (1954) (Cochran-Mantel-Haenszel), Kleinbaum et al. (1982) (inverse variance), Rothman and Boice (1982) (null-weighted), Böhning and Sarol (2000), Greenland and Holland (1991), and Mehrotra and Railkar (2000) (minimum-risk). To avoid problems with zero cells, we follow Greenland and Robins (1985) and add $c = 0.5$ to each cell in calculating the inverse variance weights.

4.1.3. Other approximations

There exist other methods that are also restricted to a single binary comparison. Lee (1981) suggested fitting a logistic GLM and finding the average of the difference between the hypothetical fitted risks calculated as if all individuals had been in ‘group 0’ and those calculated as if all individuals had been in ‘group 1’. Stijnen and Van Houwelingen (1993) proposed a pseudo-likelihood approach for sparse stratified data, where the distribution of the response variable is misspecified as a standard normal distribution, such that nuisance parameters are removed from the likelihood and a consistent estimate for the adjusted risk difference can be found. Finally, Lunceford and Davidian (2004) proposed a number of estimators based on propensity scores, where the probability of group assignment must be modelled with respect to the adjustment variables. We examine the IPW2 estimator, later also derived by Ukoumunne et al. (2010), and the double-robustness estimator, which remains consistent if the model for the propensity scores is misspecified.

4.2. Simulation assumptions

We simulated samples of five different sizes, $n = 100, 500$ and 5000 . Motivated by Example 1, the risk for individual i was determined by an

additive model

$$p_i(\mathbf{x}_i, \boldsymbol{\theta}) = \alpha_0 + \alpha_1(u_i) + \beta_1 v_i,$$

where $u_i \in \{0, 1\}$ is the indicator for randomly-allocated treatment group (0 = control, 1 = intervention) and v_i is a continuous covariate for age, generated from a normal distribution with mean 62.5 and variance 10^2 , truncated to lie in the range [40, 85]. The parameter of interest is the adjusted risk difference between the treatment groups, $\alpha_1(1) - \alpha_1(0)$.

With the adjusted treatment effect taking values 0.05 and 0.15, and the gradient of age being 0.0015, 0.0030 or 0.0060 per year, we changed the value of α_0 to provide three different scenarios in which the properties of our method could be tested: an average risk of 0.5; a minimum risk of 0; and a maximum risk of 1.

For each sample size and set of parameter values, we produced 1000 simulations and estimated the parameters in an additive binomial model using the method described in Section 3, as implemented in the `addbin` routine. We estimated the bias of the risk difference parameter estimate, and calculated its sample variance.

We also calculated adjusted risk differences using each of the methods described in Section 4.1, and compared them to the MLE using the estimated mean squared error (MSE). For the misspecified regression models, which provide estimated risks for each individual, we counted the number of simulations in which all fitted risks were valid (within $[0, 1]$).

4.3. Results

The `addbin` routine found the MLE in all 1000 simulated samples for all 18 parameter combinations and all sample sizes, demonstrating its stability. The `glm` routine in R performed almost as well, but failed to converge in a small number of samples ($< 1\%$) with low or high risk ranges. `PROC GENMOD` in SAS converged to the MLE in over 99% of samples where the average risk was 0.5, but consistently failed to converge in around 50% of samples with risks close to 0 or 1, even with $n = 5000$.

4.3.1. Misspecified regression models

The MLE and the misspecified regression models all performed well for an average risk of 0.5, with the relative efficiency of the MLE being in the range 100–106%. Table 2 shows results for scenarios with 5% treatment effect and 0.6% age effect per year.

The additive Poisson model was slightly less efficient than the binomial MLE in terms of both the variance and MSE of its treatment effect estimate, but the difference was usually less than 5%. The estimate from the least-squares method (MLS) generally had a lower variance, though the gain in efficiency was less than 1%.

At the lower and upper ends of the risk range, MLS mostly produced estimates with slightly lower bias but much higher variance than the binomial MLE, and around half of these MLS models had fitted risks outside $[0, 1]$. This led to the MLE being more efficient, in the range 115–125% for this parameter combination. The Poisson model performed similarly to the MLE at low risks, where its non-negativity constraint on the fitted means was imposed. At high risks, the additive Poisson model often produced invalid fitted risks, and was less efficient than the MLE from the additive binomial, with relative efficiency around 125%.

The full range of results are presented in Web Tables 1–6, where it is shown that the differences between methods may be smaller for other parameter combinations; albeit almost always in favour of the binomial MLE. Nonetheless, this example shows the potential for large differences in efficiency between the estimates.

4.3.2. Weighted mean methods

The weighted mean approaches generally performed best when the age covariate was split into 5 categories, and the results for $n = 100$ and $n = 5000$ are shown in Web Tables 7–12. When the average risk was 0.5, the performance of the alternative methods was generally similar to that of the MLE, with the exception of the inverse variance and Böhning-Sarol methods, which had inferior performance. Greenland and Holland’s estimator performed particularly well in the interior of the parameter space for small n , having a small efficiency advantage (5%) compared to the MLE, but this was not consistently true close to the boundaries, where relative efficiency ranged from a 15% advantage to a 20% disadvantage in the parameter combinations we tested. Other estimators also generally suffered at least some loss of efficiency when risks were close to 0 or 1, with the exception of the null-weighted method in small samples, which often had a large bias but small variance, giving efficiency gains of up to 16% in some scenarios.

Table 2: Simulation results for adjusted treatment effect estimates from additive binomial, Poisson and modified least-squares methods with a true treatment risk difference of 5% and age effect of 0.6% per year. Relative MSE is mean squared error relative to the binomial method, and “Valid” refers to the percentage of simulations with all fitted risks in $[0, 1]$.

Risk range	n	Method	Bias (%)	Standard deviation	Relative MSE	Valid (%)
0.34–0.66	100	Binomial	7.25	0.0989	1	100
		Poisson	9.51	0.1017	1.058	99.4
		Least squares	6.74	0.0987	0.996	99.5
	500	Binomial	-1.72	0.0445	1	100
		Poisson	-1.15	0.0449	1.021	100
		Least squares	-1.57	0.0444	0.998	100
	5000	Binomial	-1.15	0.0141	1	100
		Poisson	-1.12	0.0143	1.022	100
		Least squares	-1.16	0.0141	0.999	100
0–0.32	100	Binomial	-6.95	0.0678	1	100
		Poisson	-6.68	0.0678	1.002	100
		Least squares	2.07	0.0745	1.205	47.8
	500	Binomial	-6.43	0.0285	1	100
		Poisson	-6.64	0.0287	1.009	100
		Least squares	-0.36	0.0307	1.140	49.5
	5000	Binomial	-1.51	0.0089	1	100
		Poisson	-1.42	0.0089	0.999	100
		Least squares	-0.48	0.0100	1.252	47.9
0.68–1	100	Binomial	-5.31	0.0669	1	100
		Poisson	6.32	0.0745	1.242	47.7
		Least squares	5.04	0.0719	1.156	48.7
	500	Binomial	-7.74	0.0300	1	100
		Poisson	-1.48	0.0338	1.248	49.4
		Least squares	-1.61	0.0329	1.188	48.4
	5000	Binomial	-1.12	0.0093	1	100
		Poisson	-0.22	0.0104	1.252	52.3
		Least squares	-0.13	0.0102	1.195	51.0

4.3.3. Other approximations

For the approaches that allow only a binary risk difference comparison to be made, we show the empirical properties relative to the binomial MLE in Web Tables 13–18. The estimates from the fitted logistic model and pseudo likelihood approach have almost identical properties, which are very similar to those of the binomial MLE when risks average 0.5. When risks were closer to 0 or 1, these estimators tended to be less efficient, resulting in 10–20% greater efficiency for the MLE in some scenarios. Both propensity score-based methods produced estimates with similar performance to the MLE in the interior of the parameter space. At the upper and lower boundaries, the propensity score methods generally had slightly lower bias but larger variance than the MLE, resulting in efficiency losses of up to 25%.

4.4. Conclusions

Although there were some isolated scenarios in which alternative methods outperformed the additive binomial MLE for estimating adjusted risk difference, when viewed across the full range of scenarios the MLE was the most efficient approach. Those estimators from misspecified regression models were generally less efficient than the correctly-specified binomial model, and also were not constrained to produce fitted risks inside $[0, 1]$. Other approximate methods required additional assumptions, and whilst they performed similarly to the MLE in the interior of the parameter space, they were substantially less efficient near the boundaries. The various weighted methods require that adjustment covariates be categorised, and only challenged the efficiency of the MLE in isolated scenarios.

5. Flexible monotonic regression

In some situations we may be confident of the direction of the effect of a continuous or ordered categorical covariate, but we may not want to restrict the relationship to be linear. To provide for more flexible modelling, we can include unspecified monotonic regression functions in our proposed method. This allows semi-parametric adjustment of risk differences, as well as exploration of an appropriate parametric form for the regression function.

We now include C monotonic covariates $\mathbf{w}_i = \{w_{i1}, \dots, w_{iC}\}$ in the model, where the contribution of w_{ic} to risk is determined by an unspecified non-decreasing function f_c . This leads to a semi-parametric extension of

model (1):

$$p(\mathbf{x}_i, \boldsymbol{\theta}) = \alpha_0 + \sum_{a=1}^A \alpha_a(u_{ia}) + \sum_{b=1}^B \beta_b v_{ib} + \sum_{c=1}^C f_c(w_{ic}). \quad (13)$$

The function f_c is only estimable at the unique observed values of w_{ic} , $z_c(0) < \dots < z_c(l_c)$, and so for each monotonic covariate we introduce l_c parameters

$$\gamma_c(d) = f_c(z_c(d)) - f_c(z_c(d-1)) \text{ for } d = 1, \dots, l_c,$$

with $\gamma_c(0) = f_c(z_c(0)) = 0$. These parameters represent the non-negative increments in risk between the observed covariate values. Model (13) can then be rewritten in the linear form

$$p(\mathbf{x}_i, \boldsymbol{\theta}) = \alpha_0 + \sum_{a=1}^A \alpha_a(u_{ia}) + \sum_{b=1}^B \beta_b v_{ib} + \sum_{c=1}^C \sum_{d=1}^{l_c} \gamma_c(d) h_{icd},$$

where $h_{icd} = \mathbf{1}\{z_c(d) \leq w_{ic}\}$.

This form of the model means that for each monotonic covariate c we have l_c dummy linear covariates $\{h_{icd}; d = 1, \dots, l_c\}$. These covariates can therefore be handled in the same manner as the linear covariates discussed in Section 3.1 after transformation to an additive Poisson model. The only difference with Section 3.1 is that the constraints $\gamma_c(d) \geq 0$ must be imposed to retain monotonicity of the regression function f_c . This can be handled straightforwardly by the additive Poisson method of Marschner (2010), which as discussed previously, allows such non-negativity constraints on the parameters.

5.1. Data analysis example 2

In Section 3.3, we showed an example in which different link functions produced similar fit, and our method allowed us to estimate an adjusted risk difference with bootstrapped confidence intervals. Here we demonstrate an example in which an additive binomial model provides a superior fit, but this is only apparent after identifying an appropriate functional form for a continuous covariate by first including it in the model as a semi-parametric monotonic covariate.

The ASSENT-2 study was a double-blind clinical trial to assess the safety and efficacy of tenecteplase versus alteplase in 16,949 patients with acute myocardial infarction (MI) treated within six hours (ASSENT-2 Investigators, 1999). The primary outcome was 30-day mortality after randomisation.

Marschner and Gillett (2012) analysed the ASSENT-2 data using a binomial GLM with a log link function, focusing on the age-specific relative risk of mortality, adjusting for MI severity, treatment delay and geographic region. Since mortality in the treatment arms was virtually identical, treatment was not included in the model.

We repeated the same analysis of the ASSENT-2 data, but this time with an additive binomial model, such that the parameters represent adjusted risk differences. With age as a 3-level categorical variable (40–59, 60–75, 76–85), and adjusted for MI severity (Killip class I, II or III/IV), treatment delay (< 2 , $2-4$, > 4 hours) and region (Western countries, Latin America, or Eastern Europe), the residual deviance of the additive risk model fitted using `addbin` was 91.92 on 65 degrees of freedom compared to 149.32 for the relative risk model, indicating a superior fit. The fitted risks lie within $[0, 1]$, and both the `glm` function in R and `PROC GENMOD` in SAS successfully converged to the MLE in the main analysis. However, in 1000 bootstrap samples taken in order to estimate 95% confidence intervals, whilst both our method and the `glm` function converged in 100% of replications, `PROC GENMOD` failed to converge in 2.4% of samples.

Since a scaled deviance of $91.92/65 = 1.41$ is not adequate, we further investigated the effect of increasing age on risk by entering it into the model as a 46-level covariate using a flexible monotonic function, and adjusting for the same categorical covariates as above. The adjusted age-specific risk from this model is plotted in Figure 3, shown for the following covariate pattern: low severity event, < 2 hour delay and Western region. The monotonic model is compared to a model in which age is assumed to have a linear effect on risk, as well as a model with a piecewise linear effect of age, where the risk gradient changes at 65 years, as suggested by the shape of the monotonic function. The linear model is clearly inadequate, having a deviance of 926.13 on 744 degrees of freedom. The piecewise linear model, however, compares favourably to the monotonic model and has an adequate fit to the data with a deviance of 722.36 on 743 degrees of freedom.

The parameter values and their 95% confidence intervals (estimated using bootstrap resampling) for the piecewise linear model are shown in Table 3, compared to those from the model with age as a simple linear covariate. Un-

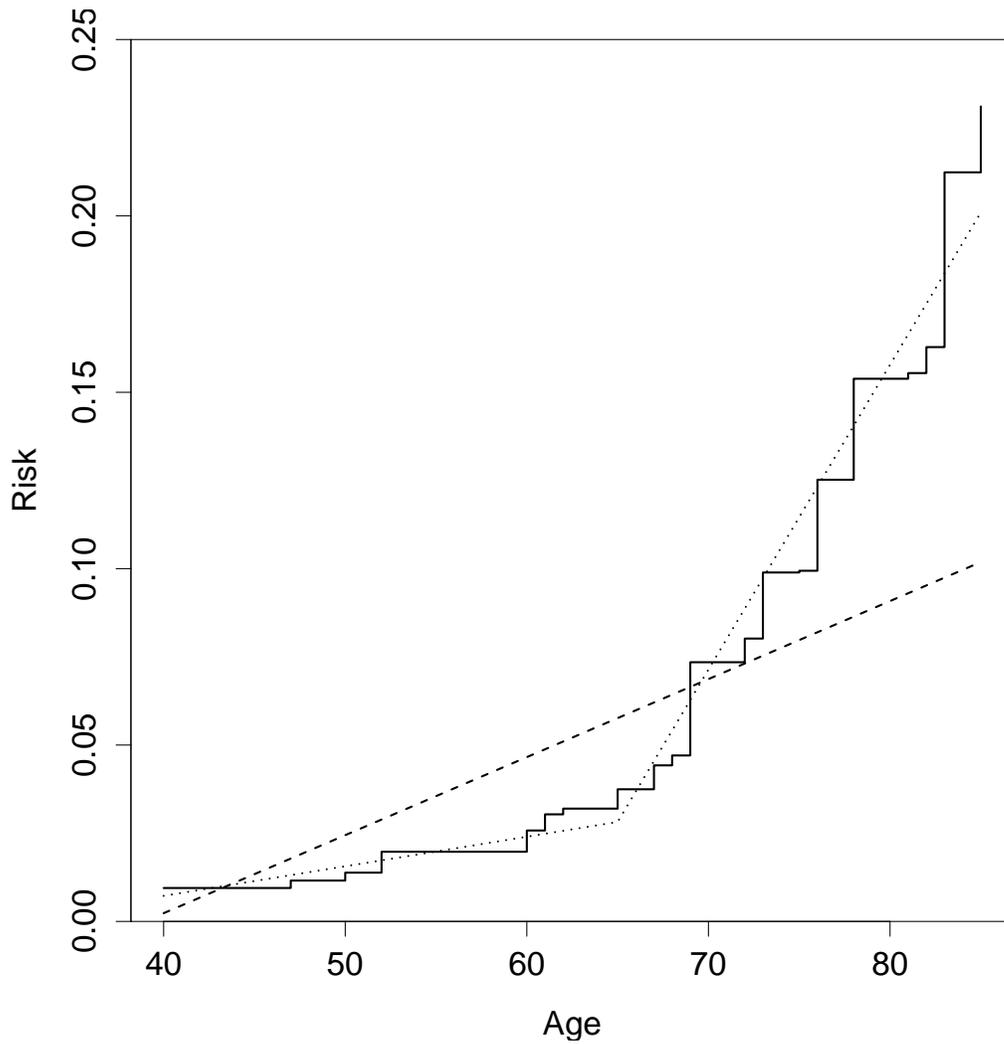


Figure 3: Age-specific risk of heart attack mortality in Example 2, using an additive risk model fitted with `addbin` and adjusted for event severity, treatment delay and region. The effect of age is specified using linear (dashed), piecewise linear (dotted) and semi-parametric monotonic (solid) regression functions. Fitted risks are presented for the Western region with low severity event and treatment delay < 2 hours.

Table 3: Risk differences (RD) and 95% confidence intervals (CI) from an additive risk model with a linear age term, and those from an additive risk model with a piecewise linear age term on the Example 2 data.

	Linear model		Piecewise linear model	
	RD	95% CI	RD	95% CI
Age (per year):				
40–85	0.0022	(0.0020, 0.0023)	—	—
40–65	—	—	0.0008	(0.0006, 0.0011)
65–85	—	—	0.0086	(0.0077, 0.0096)
Severity (Killip class):				
I	0	—	0	—
II	0.066	(0.050, 0.082)	0.061	(0.045, 0.076)
III/IV	0.281	(0.226, 0.335)	0.269	(0.216, 0.322)
Delay:				
< 2 h	0	—	0	—
2–4 h	−0.000	(−0.003, 0.003)	−0.001	(−0.006, 0.004)
> 4 h	0.003	(−0.002, 0.010)	0.002	(−0.005, 0.009)
Region:				
Western	0	—	0	—
Latin America	−0.002	(−0.004, 0.001)	−0.002	(−0.008, 0.016)
Eastern Europe	0.032	(0.013, 0.052)	0.037	(0.018, 0.056)
Deviance	926.13 (744 df)		722.36 (743 df)	

der the piecewise linear model, each year of age up to 65 leads to a mortality risk difference of 0.08%, adjusted for MI severity, treatment delay and region. After 65 years of age, this rises to a risk difference of 0.9% per year. This illustrates that flexible monotonic regression, which is not available in standard implementations of the additive binomial GLM such as in R and SAS, can suggest a simpler parametric form for modelling risk. In constructing confidence intervals for the piecewise linear model, our method converged in all 1000 bootstrap samples, whereas the `glm` function in R failed to converge in one sample and `PROC GENMOD` in SAS failed to converge in over 20% of samples.

6. Discussion

We have described a computational method for estimating adjusted risk differences using the additive binomial model. The proposed approach is a novel combination of two existing methods, the MP transformation and a combinatorial EM algorithm for additive Poisson regression. This leads to a reliable procedure for computing the additive binomial MLE, which avoids the convergence problems inherent in standard GLM algorithms such as Fisher scoring.

Since the proposed method retains the natural additive binomial model, it avoids the need for other approximate methods for adjusted risk differences, which require us to misspecify the outcome distribution, make additional assumptions, or classify covariates into a one-dimensional list of strata. Furthermore, even when a standard GLM algorithm does converge for an additive binomial model, the proposed method may still be advantageous for auxiliary analyses such as the bootstrap, which require convergence in many samples.

Standard algorithms such as Fisher scoring can sometimes be modified to increase their stability. One such approach is offered by the `glm2` package in R (Marschner, 2011), which uses a modified step-halving algorithm to ensure that the deviance will decrease at each iteration. This method converged to the MLE for the model in Example 1, and had a greater percentage of convergence in our simulations than the standard `glm` function, but still failed to converge in some samples. An alternative to standard GLM methods is a generic constrained optimisation algorithm applied to the additive binomial model. For example, Kovalchik et al. (2013) recently developed a method based on an adaptive barrier approach, which also includes the more general LEXPIT model. However, this too failed to converge to the MLE in some of our simulations. Importantly, our method has an advantage over all others in that it allows for the additional flexibility of unspecified monotonic regression functions. This allows semi-parametric adjustment of risk differences, and can also assist in identifying the functional form of continuous covariates.

Adjusted risk differences have wide applicability in biostatistics, and this has led to the use of the additive binomial model in real applications (e.g. Grotvedt et al., 2008; Adelstein et al., 2011). From an individual perspective, risk difference is often a better effect measure than relative risk or odds ratio. Furthermore, from a population health perspective, risk difference is often more relevant than relative measures for assessing the benefit of a population

intervention policy. One reason for this is that the reciprocal of the risk difference can be interpreted as the average number of individuals from the population that need to be treated with the intervention for a given time period to observe one fewer event within that time compared to the control, commonly referred to as the number needed to treat (NNT) (Laupacis et al., 1988).

In some datasets the additive binomial model may better characterise the simultaneous contribution of risk factors to an absolute change in risk, compared to a multiplicative model such as logistic regression. For example, the presence of an interaction between covariates on a multiplicative scale may disappear when their effects are considered on an additive scale, leading to a more parsimonious model for risk.

Our method for additive binomial models has been developed with adjusted risk differences in mind, but this model is also appropriate in many other situations. In epidemiology, adjusted prevalence differences from cross-sectional studies can be estimated using the additive binomial model. Linear probability models are also used in econometrics (Gujarati, 2003) and psychometrics (Maydeu-Olivares, 2005). This suggests that the proposed method may have broad applicability beyond our primary motivation of adjusted risk difference estimation.

Acknowledgements

The authors thank the ASSENT-2 and ASSENT-3 investigators for providing data to the National Health and Medical Research Council Clinical Trials Centre. This research was supported by the Australian Research Council (DP110101254).

Supplementary materials

Web Tables 1–18, referenced in Section 4.3, and R code for implementing this method are available online with this paper.

References

Adelstein, B.-A., Dobbins, T. A., Harris, C. A., Marschner, I. C., Ward, R. L., 2011. A systematic review and meta-analysis of KRAS status as the determinant of response to anti-EGFR antibodies and the impact of

- partner chemotherapy in metastatic colorectal cancer. *European Journal of Cancer* 47 (9), 1343–1354.
- ASSENT-2 Investigators, 1999. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. *The Lancet* 354 (9180), 716–722.
- ASSENT-3 Investigators, 2001. Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: the ASSENT-3 randomised trial in acute myocardial infarction. *The Lancet* 358 (9282), 605–613.
- Baker, S. G., 1994. The multinomial-Poisson transformation. *The Statistician* 43 (4), 495–504.
- Böhning, D., Sarol, J., 2000. Estimating risk difference in multicenter studies under baseline-risk heterogeneity. *Biometrics* 56 (1), 304–308.
- Cheung, Y. B., 2007. A modified least-squares regression approach to the estimation of risk difference. *American Journal of Epidemiology* 166 (11), 1337–1344.
- Cochran, W. G., 1954. Some methods for strengthening the common Chi-squared tests. *Biometrics* 10 (4), 417–451.
- Greenland, S., Holland, P., 1991. Estimating standardized risk differences from odds ratios. *Biometrics* 47 (1), 319–322.
- Greenland, S., Robins, J. M., 1985. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 41 (1), 55–68.
- Grotvedt, L., Stigum, H., Hovengen, R., Graff-Iversen, S., 2008. Social differences in smoking and snuff use among Norwegian adolescents: A population based survey. *BMC Public Health* 8, 322.
- Gujarati, D. N., 2003. *Basic Econometrics*, 4th Edition. McGraw-Hill, New York, p. 582.
- Kleinbaum, D. G., Kupper, L. L., Morgenstern, H., 1982. *Epidemiologic Research: Principles and Quantitative Methods*. Van Nostrand Reinhold, New York, p. 359.

- Kovalchik, S. A., Varadhan, R., Fetterman, B., Poitras, N. E., Wacholder, S., Katki, H. A., 2013. A general binomial regression model to estimate standardized risk differences from binary response data. *Statistics in Medicine* 32 (5), 808–821.
- Laupacis, A., Sackett, D. L., Roberts, R. S., 1988. An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine* 318 (26), 1728–1733.
- Lee, J., 1981. Covariance adjustment of rates based on the multiple logistic regression model. *Journal of Chronic Diseases* 34 (8), 415–426.
- Lunceford, J. K., Davidian, M., 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine* 23 (19), 2937–2960.
- Marschner, I. C., 2010. Stable computation of maximum likelihood estimates in identity link Poisson regression. *Journal of Computational and Graphical Statistics* 19 (3), 666–683.
- Marschner, I. C., 2011. glm2: Fitting generalized linear models with convergence problems. *The R Journal* 3 (2), 12–15.
- Marschner, I. C., 2014. Combinatorial EM algorithms. *Statistics and Computing*. In press. DOI: 10.1007/s11222-013-9411-7.
- Marschner, I. C., Gillett, A. C., 2012. Relative risk regression: reliable and flexible methods for log-binomial models. *Biostatistics* 13 (1), 179–192.
- Maydeu-Olivares, A., 2005. Linear item response theory, nonlinear item response theory, and factor analysis: A unified framework. In: Maydeu-Olivares, A., McArdle, J. J. (Eds.), *Contemporary Psychometrics*. Lawrence Erlbaum Associates, New Jersey, pp. 73–100.
- Mehrotra, D. V., Railkar, R., 2000. Minimum risk weights for comparing treatments in stratified binomial trials. *Statistics in Medicine* 19 (6), 811–825.
- R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>

- Richards, F. S. G., 1961. A method of maximum-likelihood estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 23 (2), 469–475.
- Rothman, K. J., Boice, J. D., 1982. *Epidemiologic Analysis with a Programmable Calculator*, 2nd Edition. Epidemiology Resources Inc., Chestnut Hill, MA, p. 12.
- SAS Institute Inc., 2008. SAS/STAT Software, Version 9.2. Cary, NC.
URL <http://www.sas.com/>
- Spiegelman, D., Hertzmark, E., 2005. Easy SAS calculations for risk or prevalence ratios and differences. *American Journal of Epidemiology* 162 (3), 199–200.
- Stijnen, T., Van Houwelingen, H. C., 1993. Relative risk, risk difference and rate difference models for sparse stratified data: A pseudo likelihood approach. *Statistics in Medicine* 12 (24), 2285–2303.
- Ukoumunne, O. C., Williamson, E., Forbes, A. B., Gulliford, M. C., Carlin, J. B., 2010. Confounder-adjusted estimates of the risk difference using propensity score-based weighting. *Statistics in Medicine* 29 (30), 3126–3136.