Harvard University
Harvard University Biostatistics Working Paper Series

Year 2013 *Paper* 157

Efficient Estimation of Risk Ratios From Clustered Binary Data

Matthew Cefalu* Eric Tchetgen Tchetgen†

*Harvard School of Public Health, m.s.cefalu@gmail.com

†Harvard School of Public Health, etchetge@hsph.harvard.edu

http://biostats.bepress.com/harvardbiostat/paper157

Efficient estimation of risk ratios from clustered binary data

Matthew Cefalu

Department of Biostatistics, Harvard School of Public Health

Eric Tchetgen Tchetgen

Department of Epidemiology and Biostatistics, Harvard School of Public Health

March 18, 2013

Abstract

Risk ratios are often the target of inference in epidemiologic studies. The log-binomial model is a natural choice that readily returns risk ratios, but suffers from well known convergence issues. Alternate methods have been proposed to estimate risk ratios for a common binary outcome; however, there has been little work in estimating risk ratios for clustered binary data. The modified Poisson regression approach can be used to take clustering into account through the use of generalized estimating equations, but leads to a potentially inefficient estimator due to the incorrect distributional assumption. In this article, we derive an estimate of the risk ratio that accounts for clustering in the outcome, does not rely on an estimate of the baseline risk for consistency, and delivers asymptotically efficient estimates of the risk ratio parameter. An alternative efficient estimator is provided that bounds the predicted probability by 1, thus guaranteeing stable performance of the estimator. A simulation study is provided verifying that the proposed estimator outperforms the modified Poisson approach as well as estimators that assume no clustering. We apply our method to the *Young Citizens* study, a cluster randomized trial involving a behavioral intervention deigned to train children aged 10-14 years to educate their communities about HIV.

1 Introduction

Risk ratios are often the target of inference in epidemiologic studies. They allow a researcher to easily evaluate the multiplicative association between risk factors and binary outcomes. The log binomial model (Wacholder, 1986) is a natural choice that readily returns risk ratios, but suffers from well known convergence issues (Zou, 2004). The traditional approach to avoid convergence issues is to report odds ratios by using logistic regression as the odds ratio provides a good approximation of the risk ratio when the outcome is rare. However, it is often the case that the outcome is not rare within all levels of risk factors, and using logistic regression will lead to overestimation of the

risk ratio. Further, the odds ratio effect measure may be misinterpreted by non-experts (Knol et al., 2011).

Several methods have been proposed to estimate risk ratios for a common binary outcome (Wacholder, 1986; Lee, 1994; Skove et al., 1998; Greenland, 2004; Zou, 2004; Spiegelman and Hertzmark, 2005; Chu and Cole, 2010; Tchetgen Tchetgen, 2012). Each of these methods, except for Lee (1994) and Tchetgen Tchetgen (2012), share the requirement that the log-baseline risk must be estimated in order to obtain a consistent estimate of the risk ratios. This requirement is not easily satisfied, and may lead to a violation of the model restriction that all predicted probabilities are less than 1. Worse, failure to satisfy the model conditions often results in a lack of convergence of the estimation procedures.

Recently, methods have been proposed to address these issues. Chu and Cole (2010) developed a Bayesian approach that incorporates the model restriction in the estimation procedure, while Tchetgen Tchetgen (2012) presents a frequentist approach that allows for consistent and efficient estimation of the risk ratios that does not rely on obtaining an estimate for the baseline risk. It was shown that a simple plug-in estimate of the baseline risk may be used without altering the large sample efficiency of the estimated risk ratios. Another, the modified Poisson regression approach, has been widely cited and adopted as a simple method of risk ratio estimation for both observational and intervention studies (Zou, 2004). This method uses a Poisson distribution for the data in place of the Bernoulli distribution.

However, there has been little work in estimating risk ratios for clustered binary data. Such data could arise from a cluster randomized trial or from a study with repeated measures on an individual (e.g. longitudinal data). Yelland et al. (2011) provide evidence that the modified Poisson regression approach can be used to take clustering into account through the use of generalized estimating equations (GEE) (Liang and Zeger, 1986). They showed that for both observational and intervention

studies, the modified Poisson regression approach using GEEs to account for clustering results in small relative bias and near nominal confidence interval coverage. A major drawback of this approach is that the covariance structure is guaranteed to be misspecified because of the incorrect distributional assumption, leading to a potentially inefficient estimator. Note that the misspecified covariance structure is by choice and is chosen to improve numerical convergence.

In this article, we generalize the work of Tchetgen Tchetgen (2012) to allow for clustered outcomes in the estimation of risk ratios. We show that our method does not rely on an estimate of the baseline risk for consistency and delivers asymptotically efficient estimates of the risk ratios. A slight modification to the approach is described that guarantees the estimated probabilities are bounded by 1. Therefore, the method guarantees stable performance of the estimated risk ratios. We provide a simulation study under both correct and incorrect specification of the working correlation structure that verifies the proposed estimator outperforms the modified Poisson approach as well as estimators that assume no clustering.

We apply our method to the *Young Citizens* study (Kamo et al., 2008), a cluster randomized trial involving a behavioral intervention deigned to train children aged 10-14 years to educate their communities about HIV.

2 Methods

2.1 Independent outcomes

To begin, we give a brief review of the work of Tchetgen Tchetgen (2012). Consider independent binary outcomes Y_i and a set of q covariates X_i with:

$$log(P(Y_i = 1|X_i)) = log(\mathrm{E}[Y_i|X_i]) = \alpha_0 + X_i\beta_0$$

where the parameter of interest is the q-dimensional vector of log relative risks, β_0.

Tchetgen Tchetgen (2012) provided a simple estimator of β_0 that is asymptotically efficient, in the sense that it has the minimal variance of any regular and asymptotically linear (Bickel et al., 1998) estimator of β_0. Specifically, a large class of estimators was derived that contains many common estimators of the risk ratio as well as the semiparametric efficient estimator. First, an initial consistent estimate of β_0 is provided that is free of the intercept and can be constructed by solving the equation $0 = \sum_{i:Y_i=1}(Z_i - \exp\{\widehat{\beta}W_i\})W_i$, where $W_i = -(X_i - \bar{X})$ and $Z_i = 0$ for all i. This corresponds to an artificial case only model in which the pseudo-outcome Z_i is assumed to follow a Poisson distribution with mean given by the intercept-free multiplicative model $\exp(\beta W_i)$, which facilitates its use with standard regression software. Then, the class of one-step update estimators is given by:

$$\widehat{\beta}(w) = \widehat{\beta} + \left[\sum_i Y_i \widehat{T}_i(w) X_i^T\right]^{-1} \left[\sum_i Y_i \widehat{T}_i(w)\right]$$

where $\widehat{\beta}$ is an initial consistent estimate of β_0 and

$$\widehat{T}_i(w) = \left\{ w_i - \frac{\sum_i w_i \exp(\widehat{\beta}^T X_i)}{\sum_i \exp(\widehat{\beta}^T X_i)} \right\}$$

It was shown that $w_i = X_i$ is asymptotically equivalent to the Breslow-Lee estimator, $w_i = \exp(-\widehat{\beta}^T X_i)(X_i - \mathbf{X})$ returns $\widehat{\beta}$ exactly, and $\widehat{\beta}(w_{opt})$ is asymptotically efficient, with

$$w_{opt,i} = (1 - \widehat{p}_i)^{-1} \left[X_i - \frac{\sum_i X_i (1 - \widehat{p}_i)^{-1} \widehat{p}_i}{\sum_i (1 - \widehat{p}_i)^{-1} \widehat{p}_i}\right]$$

and

$$\widehat{p}_i = \exp(\widehat{\beta}^T X_i) \sum_j Y_j \exp(-\widehat{\beta}^T X_j)/n$$

In general, the difficulty in estimating β_0 lies in the fact that an estimate of the predicted risk $\widehat{p}_i$ must be provided and must be such that predicted probability is bounded by 1 on the support of X. The estimator $\widehat{\beta}(w_{opt})$ (and hence $\widehat{p}_i$) uses a simple plug-in estimate for the log-baseline risk, but any consistent estimate of α_0 could be used without affecting the large sample efficiency of $\widehat{\beta}(w_{opt})$. However, this does not guarantee the predicted probability is bounded by 1 on the support of X. Tchetgen Tchetgen (2012) provides a solution that bounds the predicted probability without requiring an estimate of the baseline risk and will be discussed in detail in Section 3.1

2.2 Correlated outcomes

We generalize the approach of Tchetgen Tchetgen (2012) to allow for correlation among the outcomes. Let $\mathbf{Y}_i$ be a k-dimensional response vector and $\mathbf{X}_i$ be a (kxq) matrix of covariates for $i = 1, \ldots, n$. Consider the semiparametric model with the only restriction

$$\mathrm{E}\left[\mathbf{Y}|\mathbf{X}\right] = \mu(\mathbf{X}|\alpha_0, \beta_0) = \exp\left(\alpha_0 \mathbf{1}_k + \mathbf{X}\beta_0\right)$$

where β_0 is a q-dimensional parameter of interest. Note that all observations share a common intercept, but this assumption can easily be relaxed as discussed in Section 3.2 below. The key in the derivation of our estimator is that our model is semiparametric in the sense that we allow the intercept and the dependence between outcomes to remain unrestricted by treating them as nuisance parameters. As a result, our inferences are robust to misspecification of the baseline risk and working covariance structure.

We briefly review the principles of semiparametric theory. Consider a model $\mathcal{M}$ with param-

eters (ϕ, η), where ϕ is a finite dimensional parameter of interest and η is a potentially infinite dimensional nuisance parameter. Define the nuisance tangent space Λ for the semiparametric model $\mathcal{M}$ as the mean-square closure of scores for the nuisance parameter η along all regular parametric submodels. The efficient score s_ϕ^{eff} for the parameter ϕ in the model $\mathcal{M}$ is the orthogonal projection of the score s_ϕ for ϕ onto the ortho-complement $\Lambda^\perp$ to the nuisance tangent space Λ in the Hilbert space $\mathcal{L}_2 \equiv \mathcal{L}_2(\mathcal{F}_0)$ of mean zero functions with inner product $\mathrm{E}_{F_0}(T_1^T T_2)$, where F_0 is the distribution function that generated the data (Bickel et al., 1998).

Define the restricted mean model as $\mathcal{M}_{RM} = \{F_0 : \mathrm{E}[Y|\mathbf{X}] = \exp(\alpha_0 \mathbf{1}_k + \mathbf{X}\beta_0)\}$, $\theta_0 = (\alpha_0, \beta_0)$ and let $D_\beta(\mathbf{X}) = \frac{\partial \mu(\mathbf{X};\theta_0)}{\partial \beta^T}$. Bickel et al. (1998) gives the set of all influence functions for β_0 in the restricted mean model $\mathcal{M}_{RM}$ is given by:

$$\Lambda_{RM}^\perp = \left\{ \varphi(\mathbf{X}) = \mathrm{E}\left[A(\mathbf{X}) D_\beta(\mathbf{X})\right]^{-1} A(\mathbf{X})\epsilon : A(\mathbf{X}) \text{ arbitrary} \right\}$$

As stated before, we treat the baseline risk as a nuisance parameter in our semiparametric model. Therefore, the nuisance tangent space Λ_{RM} needs to additionally span the space of scores for α_0. In other words, $\Lambda = \Lambda_{RM} + \Lambda_\alpha$, where Λ_α is the closed linear space spanned by scores for α_0 along all regular parameteric submodels, or $\Lambda^\perp = \Lambda_{RM}^\perp \cap \Lambda_\alpha^\perp$, where Λ is the nuisance tangent space of the semiparametric model in which the baseline risk is a nuisance parameter. Using this result, one can characterize the set of influence functions for any regular and asymptotically linear estimator of β_0 in the semiparametric model that treats α_0 as a nuisance parameter. Proofs of all the following results are provided in the Appendix.

Result 1: The set of all influence functions of β_0 can be characterized by the set:

$$\Lambda^{\perp} = \left\{ \varphi(\mathbf{X}) = E\left[A(\mathbf{X})D_{\beta}(\mathbf{X})\right]^{-1} A(\mathbf{X})\epsilon : A(\mathbf{X}) = h(\mathbf{X}) - \frac{E\left[h(\mathbf{X})\mu(\mathbf{X};\theta_0)\right]}{E\left[\mu^T(\mathbf{X};\theta_0)\mu(\mathbf{X};\theta_0)\right]}\mu^T(\mathbf{X};\theta_0), h(\mathbf{X}) \text{ arbitrary} \right\}$$

This implies that for any choice of $h(\mathbf{X})$, $U(h;\mathbf{X}) = A(\mathbf{X})\epsilon$ can be used as an estimating equation and the resulting estimator has influence function belonging to $\Lambda^{\perp}$.

Given that we have characterized the set of all influence functions, a result due to Bickel et al. (1998) states that, under certain regularity conditions, any regular and asymptotically linear estimator of β_0 that can be obtained by solving an estimating equation has an influence function belonging to $\Lambda^{\perp}$ and asymptotic distribution given by:

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(\mathbf{X}) + o_p(1)$$

Standard application of the central limit theorem implies:

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{E}[\varphi^{\otimes 2}]) \tag{1}$$

As we now show, the benefit of treating the log-baseline risk as a nuisance parameter in a semiparametric model is that solving an estimating equation for β_0 whose influence function belongs to $\Lambda^{\perp}$ is robust to misspecification of the baseline risk $exp(\alpha_0)$.

Result 2: Consider any $U(h;\mathbf{X},\alpha_0,\beta_0)$ as defined in Result 1, and replace the log-baseline risk α_0 with any arbitrary value α. Then,

$$\mathrm{E}\left[U(h;\mathbf{X},\alpha,\beta_0)\right] = 0$$

Result 2 implies that we have a set of unbiased estimating equations for β_0 that are robust to misspecification of α_0; therefore, a working estimate of the baseline risk can be used in place of the true baseline risk, and the resulting estimators are regular and asymptotically linear with influence functions belonging to $\Lambda^{\perp}$. The estimator provided for independent outcomes in Section 2.1 has influence function belonging to $\Lambda^{\perp}$ by taking $h(\mathbf{X}) = D_{\beta}^{T}(\mathbf{X})V_{ind}^{-1}(\mathbf{X}) - \frac{\mathrm{E}\left[D_{\beta}^{T}(\mathbf{X})V_{ind}^{-1}(\mathbf{X})\mu(\mathbf{X}|\theta_0)\right]}{\mathrm{E}\left[\mu^{T}(\mathbf{X}|\theta_0)V_{ind}^{-1}(\mathbf{X})\mu(\mathbf{X}|\theta_0)\right]}\mu^{T}(\mathbf{X}|\theta_0)V_{ind}^{-1}(\mathbf{X})$, where $V_{ind}(\mathbf{X}) = diag\{\mu(\mathbf{X}|\theta_0)(1-\mu(\mathbf{X}|\theta_0))\}$ and remains robust to misspecification of the baseline risk for clustered outcomes. However, the estimator provided for independent outcomes is inefficient in the setting of clustered outcomes because it fails to consider the covariance structure between the clustered outcomes.

Result 3: The efficient score for β_0 in $\mathcal{M}$ is given by $U(h^{eff};\mathbf{X})$ with

$$h^{eff} = D_{\beta}^{T}(\mathbf{X})V^{-1}(\mathbf{X}) - \frac{E\left[D_{\beta}^{T}(\mathbf{X})V^{-1}(\mathbf{X})\mu(\mathbf{X}|\theta_0)\right]}{E\left[\mu^{T}(\mathbf{X}|\theta_0)V^{-1}(\mathbf{X})\mu(\mathbf{X}|\theta_0)\right]}\mu^{T}(\mathbf{X}|\theta_0)V^{-1}(\mathbf{X})$$

where $V(\mathbf{X}) = E[\epsilon\epsilon^{T}|\mathbf{X}]$.

The efficient score $U(h^{eff};\mathbf{X})$ given in Result 3 can be used as an estimating equation. The resulting estimator $\widehat{\beta}^{eff}$ is efficient in large samples and has asymptotic distribution given by Equation 1. In practice, estimation of the nuisance parameters (α_0 and $V^{-1}(\mathbf{X})$) is needed. We have already shown in Result 2 that any estimating equation for β_0 whose influence function belongs to $\Lambda^{\perp}$ is robust to misspecification of the log-baseline risk; as a direct result, the efficient score $U(h^{eff};\mathbf{X})$ is robust to misspecification of the log-baseline risk. Further, estimating equations for β_0 given by $\Lambda^{\perp}$ do not depend on the covariance structure $V(\mathbf{X})$ for unbiasedness. Therefore, any estimate of $V(\mathbf{X})$ can be used in $U(h^{eff};\mathbf{X})$ and the resulting estimator still has influence function belonging to $\Lambda^{\perp}$.

To construct the efficient estimate of the log risk ratio β_0, we will use the efficient score in an estimating equation. Specifically, let $\widehat{\beta}^{eff}$ be the solution to:

$$\sum_{i=1}^{n} U(h^{eff}; \mathbf{X}_i, Y_i) = 0 \tag{2}$$

A theorem due to Bickel et al. (1998) states that for any initial $n^{1/2}$-consistent estimator of β_0, an efficient estimator can be constructed by a one-step update in the direction of the estimated efficient score using:

$$\widehat{\beta}^{eff} = \widehat{\beta} - \left[\sum_i \dot{\widehat{s}}_\beta^{eff}\right]^{-1} \sum_i \widehat{s}_\beta^{eff}$$

where $\widehat{s}_\beta^{eff}$ is an empirical version of s_β^{eff} (and $\sum_i \dot{\widehat{s}}_\beta^{eff}$ is an empirical estimator of the expected derivative of the efficient score) obtained by replacing all expectations by their empirical counterpart, with β_0 estimated by $\widehat{\beta}$ and $\exp(\alpha_0)$ estimated by the plug-in estimator $\sum_i \mathbf{1}_k^T \mathbf{Y}_i \exp(-\mathbf{X}_i\widehat{\beta})$. Bickel et al. (1998) also states under standard regularity conditions, $n^{1/2}(\widehat{\beta}^{eff} - \beta_0)$ is asymptotically normal with mean zero and variance given as before.

In practice, each expectation is replaced with its empirical counterpart, so that $\widehat{\beta}^{eff}$ is simple to calculate. One can use the estimate provided for independent outcomes as an initial $\widehat{\beta}$; however, based on our simulations in Section 3.3, a better choice is to use the modified Poisson estimator. Note that the efficient estimator $\widehat{\beta}^{eff}$ is only feasible if $V(\mathbf{X})$ is known. Since this covariance function is unknown, it must be modeled.

A major contribution of this method is that it allows a researcher to capture the correlation among the clustered outcomes by modeling of $V^{-1}(\mathbf{X})$, which in turn may be used to increase

the efficiency if correctly specified. Modeling the covariance structure for binary outcomes can be a challenging task. Consider the parameterization in terms of correlations proposed by Bahadur (1961). If we let $R_j = \frac{Y_j - \mu_j}{\{\mu_j(1-\mu_j)\}^{1/2}}$, $\rho_{jk} = \text{corr}(Y_j Y_k) = \text{E}(R_j R_k)$, $\rho_{jkl} = \text{E}(R_j R_k R_l)$ and so on. Then,

$$Pr(\mathbf{Y} = \mathbf{y}) = \prod_{j=1}^{k} \mu_j^{y_j}(1-\mu_j)^{(1-y_j)} \left(1 + \sum_{j<k} \rho_{ik} r_j r_k + \sum_{j<k<l} \rho_{ikl} r_j r_k r_l + ... + \rho_{1...k} r_1 r_2 \cdots r_k\right)$$

We proceed under the common assumption that all 3^{rd} order or higher correlations are zero, so that all that must be specified to estimate $V^{-1}(\mathbf{X})$ is a working correlation structure, $\mathbf{R}(\boldsymbol{\rho})$. Since the model does not put any restriction on $V^{-1}(\mathbf{X})$, we additionally allow for a dispersion parameter ϕ, and $\widehat{V}(\mathbf{X}_i) = \phi \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\rho}) \mathbf{A}_i^{1/2}$, where $\mathbf{A}_i = \text{diag}[\widehat{\mu}_{\text{i}}(1 - \widehat{\mu}_{\text{i}})]$. Common choices of correlation structures include exchangeable, autoregressive, and unstructured and details of the choices and estimation of correlation parameters can be found in Liang and Zeger (1986). As a note, in theory $\phi = 1$, but we have found that allowing it be estimated from the data improves finite sample variance estimation.

3 Additional results and simulation

3.1 An alternate efficient estimator

Estimation of $\widehat{\beta}^{eff}$ depends on $\widehat{\mathbf{A}}_{ij}^{1/2} = [\widehat{\mu}_{ij}(1 - \widehat{\mu}_{ij})]^{1/2}$ through the covariance function, which is only defined for $0 \leq \widehat{\mu}_{ij} \leq 1$. As such, the efficient estimator may run into convergence issues if the estimated risks are not bounded by 1. To get around such a problem, we adopt the method proposed by Tchetgen Tchetgen (2012). Specifically, let

$$logit(\mu_{ij}) = logit(\exp\left(\alpha + \mathbf{X}_{i(j)}\beta_0\right))$$

Then, ignoring knowledge about the functional form of the predicted risk, fit the model:

$$logit(\mu_{ij}) = \xi(\mathbf{X}_{i(j)}\beta_0)$$

where $\xi(\cdot)$ is an unrestricted function, and $\mathbf{X}_{i(j)}\beta_0$ is replaced with the initial estimate $\mathbf{X}_{i(j)}\widehat{\beta}$. Any nonparametric technique can be used to approximate $\xi(\cdot)$ including polynomial series, kernel smoothing, wavelet regression, or spline regression (Wasserman, 2005; Friedman et al., 2008). Let $\widehat{\xi}_{ij} = \widehat{\xi}(\mathbf{X}_{i(j)}\widehat{\beta})$ denote such an estimator, and the resulting $\widetilde{\mu}_{ij} = expit\left\{\widehat{\xi}_{ij}\right\}$ is used in the place of μ_{ij} in the updating of $\widehat{\beta}^{eff}$.

Here, we briefly illustrate that polynomial series regression does not change the efficiency of the resulting estimator. Let $\phi_k(M_i) = M_i^k$ for $k = 1, ..., K$. Then, for fixed K, let $\widetilde{p}_i$ denote the predicted probabilities obtained by standard logistic regression of Y_i on $\{\phi_k(M_i) : k \leq K\}$ using the data $\{(M_i, Y_i) : i = 1, ..., n\}$. A result due to Hirano et al. (2003) implies that since $\xi(\cdot)$ has at least four bounded derivatives, setting $K = Cn^{1/6}$ for some constant C is sufficient for the resulting estimator $\widetilde{\mu}_i$ to converge to μ_i at rates no slower than $n^{1/4}$, and the resulting estimator $\widetilde{\beta}^{eff}$ of β_0 is semiparametric efficient.

3.2 A more general model

All previous results were derived for the model that assumes a common baseline risk for observations within a cluster, but easily extend to a model that allows for different baseline risks. Such models are useful in the context of repeated measures over time (i.e. longitudinal data), and allow for the model to capture the risk changing over time.

As before, let $\mathbf{Y}_i$ be a k-dimensional response vector and $\mathbf{X}_i$ be a (kxq) matrix of covariates for $i = 1, \ldots, n$. Consider the semiparametric model where the only restriction is

$$\mathrm{E}\left[\mathbf{Y}|\mathbf{X}\right] = \mu(\mathbf{X}|\alpha_0, \beta_0) = \exp\left(\alpha_0 + \mathbf{X}\beta_0\right)$$

where β_0 is a q-dimensional parameter of interest and α_0 is a k-dimensional vector of log-baseline risks. Following the same development as before, it can be shown that the set of influence functions for β_0 treating the vector of baseline risks α_0 as a nuisance parameter are of the form:

$$\Lambda^{\perp} = \Big\{\varphi(\mathbf{X}) = \mathrm{E}\left[A(\mathbf{X})D_{\beta}(\mathbf{X})\right]^{-1} A(\mathbf{X})\epsilon : A(\mathbf{X}) = h(\mathbf{X}) - \mathrm{E}\left[h(\mathbf{X})M(\mathbf{X};\theta_0)\right]\mathrm{E}\left[M^T(\mathbf{X};\theta_0)M(\mathbf{X};\theta_0)\right]^{-1} M^T(\mathbf{X};\theta_0)$$
$$, h(\mathbf{X}) \text{ arbitrary}\}$$

where $D_{\beta}(\mathbf{X}) = \frac{\partial \mu(\mathbf{X};\theta_0)}{\partial \beta^T}$ and $M(\mathbf{X};\theta_0) = diag(\mu(\mathbf{X};\theta_0))$.

This set contains influence functions of all regular and asymptotically linear estimators of β_0 when the baseline risk is arbitrarily flexible. As such, this set is contained in the set of influence functions derived in Result 1 because assuming a common baseline risk is a more restrictive model. Similarly (but not exclusively), this set could also be used to construct regular and asymptotically linear estimators of β_0 in the context of longitudinal data where the baseline risk is indexed by time, $\alpha(t)$.

3.3 Simulations

In this section, we empirically verify the efficiency of the proposed estimator, and its robustness to misspecification of the covariance structure. We compare three estimators: (1) the estimator of Tchetgen Tchetgen (2012) which ignores possible correlation of the clustered outcomes; (2) the modified Poisson approach assuming an exchangeable correlation structure; and (3) our proposed

estimator $\widehat{\beta}^{eff}$ assuming an exchangeable correlation structure.

The data is generated in a manner to reflect a cluster randomized trial for a binary treatment, and is generated as follows: (1) for each independent cluster i, generate $\mathbf{X}_i$ as $q-1$ normal random vectors and a vector of treatment indicator variables; and (2) generate the $k-$dimensional response $\mathbf{Y}_i$ such that $log(\mathrm{E}[Y_i|\mathbf{X}]) = \alpha_0 + \mathbf{X}_i\beta_0$ with correlation structure given by $\mathbf{R}$. The baseline risk was chosen to be 0.37. Various relative risks and two correlation structures were considered. First, the exchangeable correlation structure assumes all pairwise correlations between observations within a cluster are equal to ρ. This structure is widely used in practice and is useful in capturing the overall correlation within a cluster. The second correlation structure we consider mimics what might be expected if the clusters are households where the first two observations in each cluster are the parents and the remaining observations are the children. This household correlation structure is given by:

$$\begin{pmatrix} 1 & 0.05 & 0.1 & 0.1 & 0.1 \\ 0.05 & 1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1 & 0.3 & 0.3 \\ 0.1 & 0.1 & 0.3 & 1 & 0.3 \\ 0.1 & 0.1 & 0.3 & 0.3 & 1 \end{pmatrix} \tag{3}$$

Table 1 provides the absolute bias and mean squared error of each estimator for estimating the relative risk of the binary treatment when there are 1000 clusters of size 5 and the true correlation structure is either exchangeable with $\rho = 0.3$ or the household structure given in Equation 3. Recall that the working correlation structure for the modified Poisson and the efficient estimator is assumed to be exchangeable. The estimator that assumes independent observations has the

highest mean squared error under each value of the relative risk, and the efficient estimator has the smallest mean squared error. These results are as anticipated; accounting for the correlation in the outcome improves the efficiency of both the modified Poisson and the efficient estimator. Although the modified Poisson approach accounts for correlation, it is inefficient due to misspecification of the covariance structure (due to the misspecification of the distribution). The efficient estimator correctly models this covariance structure, and as a result has the smallest mean squared error.

Consider the results when the relative risk of the binary treatment is 1.05 in Table 1 under the exchangeable correlation structure; we note that the three estimators have approximately the same absolute bias ($2.98\text{x}10^{-3}$, $2.67\text{x}10^{-3}$, and $2.89\text{x}10^{-3}$), but that the efficient estimator has the smallest mean squared error of $1.93\text{x}10^{-3}$ compared to $2.61\text{x}10^{-3}$ and $2.00\text{x}10^{-3}$. Moving to the case where the relative risk of the binary treatment is 2, accounting for the correlation in the outcome dramatically reduces the bias, with the bias of the estimator that assumes independence equal to $6.18\text{x}10^{-3}$ and that of the efficient estimator equal to $0.12\text{x}10^{-3}$.

Consider the situations in Table 1 where the true correlation structure is the household structure given in Equation 3. Here, the modified Poisson and efficient estimator incorrectly assume that the working correlation structure is exchangeable, but still show a reduction in mean squared error when compared to the estimator that assumes independence. The same patterns are observed under the misspecification of the covariance structure as were observed under the correct specification, with the estimator that assumes independent observations having the highest mean squared error under each value of the relative risk. In each case, the efficient estimator has smaller mean squared error than the estimator that assumes independent observations. Further, the bias of the efficient estimator remains small under the misspecification of the correlation structure. Under the case when the relative risk of the binary treatment is 2, the efficient estimator has a bias and mean squared error of $1.35\text{x}10^{-3}$ and $3.58\text{x}10^{-3}$, respectively, while the estimator assuming independence has a larger

bias and mean squared error at $10.48\text{x}10^{-3}$ and $3.89\text{x}10^{-3}$, respectively.

Table 2 is a reproduction of Table 1 but for a continuous covariate in place of the binary treatment. The results follow a similar pattern.

The results of these simulations verify that the proposed efficient estimator reduces mean squared error of the estimated risk ratios across a variety of simulated scenarios. All estimators considered in this simulation study are consistent and provide asymptotically valid inference. However, it appears that accounting for clustering in the outcomes reduces finite sample bias.

4 Application: Young Citizens Data

We applied our proposed estimator for the risk ratio to data from the *Young Citizens* study (Kamo et al., 2008). The trial involved a behavioral intervention designed to train children aged 10-14 years to educate their communities about HIV. The study involved 30 communities that were paired based on a clustering algorithm incorporating demographics, and one community in each pair randomly assigned treatment group with the other assigned to the control group. Residents within each community were surveyed post-intervention to determine their beliefs about the ability to children to teach the community about HIV. The primary outcome of this study was a composite scored reflecting the strength of this belief. However, to illustrate our estimator, we chose to consider a secondary outcome of the study, specifically the residents' beliefs regarding whether or not the AIDS problem was getting worse in their community (Stephens et al., 2012). This outcome was derived by collapsing a 4-point scale with values "strongly agree", "agree", "disagree", or "strongly disagree" into two values, "agree" or "disagree".

We estimated the risk ratio of the intervention using the efficient estimator given in Section 2.2 assuming an exchangeable correlation structure, the modified Poisson approach assuming an exchangeable correlation structure, and the estimator that assumes independence given in Section

2.1. Additionally, we estimate the odds ratio of the intervention using a GEE with a logit link and assuming an exchangeable correlation structure. In all of the estimators, we control for the baseline covariates residential or urban community, religion, ethnic group, and indicators of wealth by including the covariates into the linear predictor of the mean.

Table 3 provides the estimated risk ratio of the intervention, the standard error, and the 95% confidence interval for each of the estimators considered. We would like to note that standard GEE for the log-binomial model with correlated data failed to converge, and as such, a different approach must be taken to estimate the risk ratios. The outcome is not rare ($\sim$82% responded "agree"); therefore, using odds ratios to estimate the risk ratio is not valid.

The efficient estimator and that of the modified Poisson approach provide similar estimates of the log-risk ratio, -0.0188 and -0.0206, respectively, with the efficient estimator slightly smaller in magnitude. The standard error of the efficient estimator is 0.0375, compared to 0.0406 for the modified Poisson approach. This corresponds to an empirical asymptotic relative efficiency of 0.85 for the modified Poisson compared to the efficient estimator, and is reflected in by a narrowing of the confidence intervals. Neither approach leads to significant effects at the $\alpha = 0.05$, but the results do illustrate the efficient estimator has tighter confidence intervals than that of the modified Poisson approach. Also provided in Table 3 is the log-odds ratio estimated using a GEE with a logit link and assuming an exchangeable correlation structure. The estimated log-odds ratio is -0.1222, illustrating that the odds ratio is not a good approximation of the risk ratio in the trial and likely overestimates the relative risk of the intervention.

5 Discussion

In this paper, we have proposed an efficient estimator of the risk ratio that accounts for clustering among binary outcomes. We prove that this estimator is robust to misspecification of the baseline

risk, in the sense that the estimator does not directly rely on an estimate of the baseline risk for consistency, and showed that it has the smallest asymptotic variance of any regular and asymptotically linear estimator. Further, a modification of the estimator is provided that guarantees the predicted probability is bounded by 1 (a model restriction), and as a result, guarantees stable performance of the estimator.

Simulations confirm that the proposed estimator has smaller variance than estimators that assume independence and the modified Poisson approach both under correct and incorrect specification of the correlation structure. Additionally, the simulations suggest that the proposed estimator may have smaller finite sample bias in the estimation of the risk ratios when compared to estimators that assume independence. Therefore, it is important to account for correlation among clustered outcomes both to improve efficiency and to remove finite sample bias.

The gains in efficiency of the proposed estimator when compared to the modified Poisson approach are due to allowing for correct specification of the underlying data distribution. A priori, the modified Poisson approach incorrectly models the data as a Poisson distribution, leading to a misspecification of the covariance structure and ruling out the possibility of an efficient estimator. The estimator proposed in this paper allows for correct distributional assumptions, and avoids the common drawbacks of this assumption by being robust to misspecification of the baseline risk.

6 Acknowledgements

We would like to thank Felton Earls and Mary Carlson for providing the *Young Citizens* data, and Lisa Yelland for the sharing of her code. Support for this research was provided by National Institute of Environmental Health Sciences grant 5T32ES007142 and NIH grants R01ES020337-01, R21ES019712, U54GM088558, and R0151164

7 Appendix

Proof of Result 1: Recall that the nuisance tangent space is characterized by $\Lambda = \Lambda_{RM} + \Lambda_\alpha$, where Λ_{RM} is the nuisance tangent space from the restricted mean model and Λ_α is the closed linear space spanned by scores for α_0 along all regular parametric submodels. For any $A(X)\epsilon \in \Lambda_{RM}^{\perp}$, then

$$\begin{aligned}
\Pi\left[A(\mathbf{X})\epsilon|(\Lambda_{RM}+\Lambda_\alpha)^{\perp}\right] &= A(\mathbf{X})\epsilon - \Pi\left[A(\mathbf{X})\epsilon|\Lambda_{RM}+\Lambda_\alpha\right] \\
&= A(\mathbf{X})\epsilon - \Pi\left[A(\mathbf{X})\epsilon|\left\{\Lambda_\alpha - \Pi\left[\Lambda_\alpha|\Lambda_{RM}\right]\right\}\right] \\
&= A(\mathbf{X})\epsilon - \Pi\left[A(\mathbf{X})\epsilon|\Lambda_\alpha^*\right] \\
&= A(\mathbf{X})\epsilon - \frac{\mathrm{E}\left[A(\mathbf{X})\epsilon\epsilon^T V^{-1}(\mathbf{X})M(\mathbf{X})\mathbf{1}_k\right]}{\mathrm{E}\left[\mu^T(\mathbf{X})V^{-1}(\mathbf{X})\mu(\mathbf{X})\right]}\mu^T(\mathbf{X})V^{-1}(\mathbf{X})\epsilon \\
&= A(\mathbf{X})\epsilon - \frac{\mathrm{E}\left[A(\mathbf{X})\mu(\mathbf{X})\right]}{\mathrm{E}\left[\mu^T(\mathbf{X})V^{-1}(\mathbf{X})\mu(\mathbf{X})\right]}\mu^T(\mathbf{X})V^{-1}(\mathbf{X})\epsilon
\end{aligned}$$

where Λ_α^* is the closed linear space spanned by the efficient score for α_0 in $\mathcal{M}_{RM}$. Therefore, we have characterized the set of all influence functions for β_0 in the model $\mathcal{M}_{RM}$ that treats the baseline risk as a nuisance parameter as:

$$\Lambda_1^{\perp} = \left\{\varphi(\mathbf{X}) = \mathrm{E}\left[A(\mathbf{X})D_\beta(\mathbf{X})\right]^{-1}A(\mathbf{X})\epsilon : A(\mathbf{X}) = h(\mathbf{X}) - \frac{\mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X};\theta_0)\right]}{\mathrm{E}\left[\mu^T(\mathbf{X};\theta_0)V^{-1}(\mathbf{X})\mu(\mathbf{X};\theta_0)\right]}\mu^T(\mathbf{X};\theta_0)V^{-1}(\mathbf{X}), h(\mathbf{X}) \text{ arbitrary}\right\}$$

All that is left is to show $\Lambda^{\perp} = \Lambda_1^{\perp}$. For any $h(\mathbf{X}) \in \Lambda_1^{\perp}$, let $S(\mathbf{X}) = \left[h(\mathbf{X}) - \frac{\mathrm{E}[h(\mathbf{X})\mu^T(\mathbf{X})\mu(\mathbf{X})]}{\mathrm{E}[\mu^T(\mathbf{X})\mu(\mathbf{X})]}\right]\mu^T(\mathbf{X})$. Then,

$$\mathrm{E}\left[S(\mathbf{X})\mu(\mathbf{X})\right] = 0$$

so that $\Lambda_1^{\perp} \subset \Lambda^{\perp}$. Alternately, for any $S(\mathbf{X}) \in \Lambda^{\perp}$, let $h(\mathbf{X}) = S(\mathbf{X}) - \frac{\mathrm{E}[S(\mathbf{X})\mu(\mathbf{X})]}{\mathrm{E}[\mu^T(\mathbf{X})V^{-1}(\mathbf{X})\mu(\mathbf{X})]}\mu^T(\mathbf{X})V^{-1}(\mathbf{X})$. Then,

$$\mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X})\right] = 0$$

implying that $\Lambda^{\perp} \subset \Lambda_1^{\perp}$, and we are done.

Proof of Result 2: Let $U(h; \mathbf{X}, \alpha_0, \beta_0)$ be as defined in Result 1. Replace the log-baseline risk α_0 with an arbitrary value α. Then, for all h,

$$\begin{aligned}
\mathrm{E}[U(h;\mathbf{X},\alpha,\beta_0)] &= \mathrm{E}\left[h(\mathbf{X})\epsilon(\mathbf{X};\alpha,\beta_0) - \frac{\mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X};\alpha,\beta_0)\right]}{\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha,\beta_0)\right]}\mu^T(\mathbf{X};\alpha,\beta_0)\epsilon(\mathbf{X};\alpha,\beta_0)\right] \\
&= \mathrm{E}\left[h(\mathbf{X})(Y-\mu(\mathbf{X};\alpha,\beta_0))\right] - \frac{\mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X};\alpha,\beta_0)\right]}{\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha,\beta_0)\right]}\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)(Y-\mu(\mathbf{X};\alpha,\beta_0))\right] \\
&= \mathrm{E}\left[h(\mathbf{X})\mathrm{E}[Y|\mathbf{X}]\right] - \mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X};\alpha,\beta_0))\right] - \frac{\mathrm{E}\left[h(\mathbf{X})e^{\mathbf{X}\beta_0}e^{\alpha}\right]}{\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha,\beta_0)\right]}\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mathrm{E}[Y|\mathbf{X}]\right] \\
&\quad + \frac{\mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X};\alpha,\beta_0)\right]}{\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha,\beta_0)\right]}\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha,\beta_0)\right] \\
&= \mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X};\alpha_0,\beta_0)\right] - \frac{\mathrm{E}\left[h(\mathbf{X})e^{\mathbf{X}\beta_0}e^{\alpha}\right]}{\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha,\beta_0)\right]}\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha_0,\beta_0)\right] \\
&= \mathrm{E}\left[h(\mathbf{X})\mu(\mathbf{X};\alpha_0,\beta_0)\right] - \frac{\mathrm{E}\left[h(\mathbf{X})e^{\mathbf{X}\beta_0}e^{\alpha_0}\right]}{\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)\mu(\mathbf{X};\alpha,\beta_0)\right]}\mathrm{E}\left[\mu^T(\mathbf{X};\alpha,\beta_0)e^{\mathbf{X}\beta_0}e^{\alpha}\right] \\
&= 0
\end{aligned}$$

Proof of Result 3: Recall the efficient score is defined by $s_{\beta}^{eff} = \Pi[s_{\beta}|\Lambda^{\perp}]$, where s_{β} is the score for β_0. Under the restricted moment model, the efficient score (Bickel et al., 1998) for

$\theta_0 = (\alpha_0, \beta_0)^T$ is given by:

$$s_\theta^{eff,RM} = (s_\alpha^{RM}, s_\beta^{RM})^T = \Pi\left[s_\theta | \Lambda_{RM}^{\perp}\right] = D^T(\mathbf{X})V^{-1}(X)\epsilon = (\mathbf{1}_k, \mathbf{X})^T M(\mathbf{X}|\theta_0)V^{-1}(\mathbf{X})\epsilon$$

where $D(\mathbf{X}) = \frac{\partial \mu(\mathbf{X}|\theta)}{\partial \theta^T}$, $M(\mathbf{X}|\theta) = \text{diag}\left\{\mu(\mathbf{X}|\theta)\right\}$ is the (kxk) diagonal matrix made up of the elements of μ, and $V^{-1}(\mathbf{X}) = \text{E}\left[\epsilon\epsilon^T\right]^{-1}$. Then, by definition of the efficient score and using arguments similar to Result 1:

$$s_\beta^{eff} = s_\beta^{RM} - \Pi\left[s_\beta^{RM} | \Lambda_\alpha^*\right]$$

where Λ_α^* is the closed linear space spanned by the efficient score for α_0 in $\mathcal{M}_{RM}$. Thus,

$$\begin{aligned}
s_\beta^{eff} &= s_\beta^* - \Pi\left[s_\beta^* \mid \Lambda_\alpha^*\right] \\
&= s_\beta^* - \text{E}\left[s_\beta^* s_\alpha^{*T}\right] \text{E}\left[s_\alpha^* s_\alpha^{*T}\right]^{-1} s_\alpha^* \\
&= \mathbf{X}^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\epsilon - \text{E}\left[\mathbf{X}^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(X)\epsilon\epsilon^T V^{-1}(\mathbf{X})M^T(\mathbf{X}|\alpha_0, \beta_0)\mathbf{1}_k\right] \\
&\quad \text{E}\left[\mathbf{1}_k^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\epsilon\epsilon^T V^{-1}(\mathbf{X})M^T(\mathbf{X}|\alpha_0, \beta_0)\mathbf{1}_k\right]^{-1} \mathbf{1}_k^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\epsilon \\
&= \mathbf{X}^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\epsilon - \text{E}\left[\mathbf{X}^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})M^T(\mathbf{X}|\alpha_0, \beta_0)\mathbf{1}_k\right] \\
&\quad \text{E}\left[\mathbf{1}_k^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})M^T(\mathbf{X}|\alpha_0, \beta_0)\mathbf{1}_k\right]^{-1} \mathbf{1}_k^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\epsilon \\
&= \mathbf{X}^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\epsilon - \text{E}\left[\mathbf{X}^T M(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\mu(\mathbf{X}|\alpha_0, \beta_0)\right] \\
&\quad \text{E}\left[\mu^T(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\mu(\mathbf{X}|\alpha_0, \beta_0)\right]^{-1} \mu^T(\mathbf{X}|\alpha_0, \beta_0)V^{-1}(\mathbf{X})\epsilon
\end{aligned}$$

References

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1998), *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins series in the mathematical sciences, Springer.

Chu, H. and Cole, S. (2010), "Estimation of Risk Ratios in Cohort Studies With Common Outcomes. A Bayesian Approach," *Epidemiology*, 21, 855–862.

Friedman, J., Hastie, T., and Tibshirani, R. (2008), *The elements of statistical learning*, vol. 2, Springer.

Greenland, S. (2004), "Model-based Estimation of Relative Risks and Other Epidemiologic Measures in Studies of Common Outcomes and in Case-Control Studies," *American Journal of Epidemiology*, 160, 301–305.

Hirano, K., Imbens, G., and Ridder, G. (2003), "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161–1189.

Kamo, N., Carlson, M., Brennan, R., and Earls, F. (2008), "Young citizens as health agents: Use of drama in promoting community efficacy for HIV/AIDS," *Journal Information*, 98.

Knol, M. J., Duijnhoven, R. G., Grobbee, D. E., Moons, K. G., and Groenwold, R. H. (2011), "Potential misinterpretation of treatment effects due to use of odds ratios and logistic regression in randomized controlled trials," *PLoS One*, 6, e21248.

Lee, J. (1994), "Odds Ratio or Relative Risk for Cross-Sectional Data?" *International Journal of Epidemiology*, 23, 201–203.

Liang, K.-Y. and Zeger, S. L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.

Skove, T., Deddens, J., Petersen, M. R., and Endahl, L. (1998), "Prevalence proportion ratios: estimation and hypothesis testing," *International Journal of Epidemiology*, 27, 91–95.

Spiegelman, D. and Hertzmark, E. (2005), "Easy SAS Calculations for Risk or Prevalence Ratios and Differences," *American Journal of Epidemiology*, 162, 199–200.

Stephens, A., Tchetgen Tchetgen, E., and Gruttola, V. (2012), "Augmented generalized estimating equations for improving efficiency and validity of estimation in cluster randomized trials by leveraging cluster-level and individual-level covariates," *Statistics in Medicine*.

Tchetgen Tchetgen, E. (2012), "Estimation of Risk Ratios in Cohort Studies With Common Outcomes: A Simple and Efficient Two-stage Approach," *International Journal of Biostatistics*, In press.

Wacholder, S. (1986), "Binomial Regression in GLIM: Estimating Risk Ratios and Risk Differences," *American Journal of Epidemiology*, 123, 174–184.

Wasserman, L. (2005), *All of nonparametric statistics*, Springer.

Yelland, L. N., Salter, A. B., and Ryan, P. (2011), "Performance of the Modified Poisson Regression Approach for Estimating Relative Risks From Clustered Prospective Data," *American Journal of Epidemiology*, 174, 984–992.

Zou, G. (2004), "A Modified Poisson Regression Approach to Prospective Studies with Binary Data," *American Journal of Epidemiology*, 159, 702–706.

		Independent		Modified Poisson		Efficient		
True CS	Relative Risk	MSE	Bias	MSE	Bias	MSE	Bias	Coverage
Exchangeable	1	2.81	1.52	2.14	2.08	2.09	1.82	94.5
	1.05	2.61	2.98	2.00	2.67	1.93	2.89	95.1
	1.5	3.60	2.58	2.88	0.43	2.83	0.37	94.7
	2	4.57	6.18	3.74	0.41	3.67	0.12	96.2
Household	1	2.05	0.75	1.99	1.18	1.91	0.93	94.5
	1.05	2.16	3.68	2.09	2.56	1.96	2.57	95.6
	1.5	2.77	5.46	2.68	0.07	2.53	1.27	95.8
	2	3.89	10.48	3.53	3.29	3.58	1.35	95.0

Table 1: Bias (10^{-3}) and mean square error (10^{-3}) of the modified Poisson approach and the efficient approach for estimating the relative risk of a binary covariate when there are 1000 clusters of size 5 under an exchangeable working correlation structure. The true correlation structure is either exchangeable with $\rho = 0.3$ or the household structure given in Equation 3.

		Independent		Modified Poisson		Efficient		
True CS	Relative Risk	MSE	Bias	MSE	Bias	MSE	Bias	Coverage
Exchangeable	1	0.31	0.06	0.25	0.08	0.23	0.23	94.7
	1.05	0.33	1.49	0.26	0.20	0.24	0.02	94.5
	1.5	0.74	6.45	0.55	0.68	0.50	1.27	94.5
	2	1.66	12.47	1.23	0.069	1.05	1.68	95.4
Household	1	0.286	0.01	0.284	0.09	0.275	0.15	93.9
	1.05	0.27	1.76	0.28	0.87	0.24	1.13	94.8
	1.5	0.66	13.56	0.44	0.095	0.43	0.50	94.1
	2	2.01	25.66	0.818	1.69	0.816	0.57	93.0

Table 2: Bias (10^{-3}) and mean square error (10^{-3}) of the modified Poisson approach and the efficient approach for estimating the relative risk of a continuous covariate when there are 1000 clusters of size 5 under an exchangeable working correlation structure. The true correlation structure is either exchangeable with $\rho = 0.3$ or the household structure given in Equation 3.

Estimator	log(Risk ratio)	Std. Error	95% Confidence Interval
$\widehat{\beta}^{eff}$	-0.0188	0.0375	(-0.0922 , 0.0547)
$\widehat{\beta}^{MP}$	-0.0206	0.0406	(-0.1002, 0.0590)
$\widehat{\beta}^{OR}$	-0.1222	0.2529	(-0.6179 , 0.3736)

Table 3: Estimated log-risk ratio (or log-odds ratio) of the intervention, the standard error, and corresponding 95% confidence interval. $\widehat{\beta}^{eff}$ is the efficient estimator provided in Section 2.2 assuming an exchangeable correlation structure, $\widehat{\beta}^{MP}$ is the modified Poisson estimator assuming an exchangeable correlation structure, and $\widehat{\beta}^{OR}$ is the log-odds ratio estimated using the GEE with a logit link and assuming an exchangeable correlation structure.