

Relative Risks, Odds Ratios and Cluster Randomised Trials

Michael J Campbell, Suzanne Mason, Jon Nicholl

Abstract It is well known in cluster randomised trials with a binary outcome and a logistic link that the population average model and cluster specific model estimate different population parameters for the odds ratio. However, it is less well appreciated that for a log link, the population parameters are the same and a log link leads to a relative risk. This suggests that for a prospective cluster randomised trial the relative risk is easier to interpret. Commonly the odds ratio and the relative risk have similar values and are interpreted similarly. However, when the incidence of events is high they can differ quite markedly, and it is unclear which is the better parameter. We explore these issues in a cluster randomised trial, the Paramedic Practitioner Older People's Support Trial³.

Key words: Cluster randomized trials, odds ratio, relative risk, population average models, random effects models

1 Introduction

Given two groups, with probabilities of binary outcomes of π_0 and π_1 , the Odds Ratio (OR) of outcome in group 1 relative to group 0 is related to Relative Risk (RR) by:

$$OR = RR \frac{(1-\pi_0)}{(1-RR\pi_0)}.$$

If there are covariates in the model, one has to estimate π_0 at some covariate combination. One can see from this relationship that if the relative risk is small, or the probability of outcome π_0 is small then the odds ratio and the relative risk will be close.

One can also show, using the delta method, that approximately

$$SE(\log(OR)) = \frac{SE(\log(RR))}{RR(1-RR\pi_0)} \quad (1).$$

Note that this result is derived for non-clustered data. Since $(1-RR\pi_0)$ is less than one, if the RR is also less than one then the standard error of the OR will be greater than that of the RR.

Relative Risks are a natural outcome from a prospective study and they are easily understood. One approach would be to use logistic regression to estimate the odds ratio and then transform it. However it has been shown that simply transforming the estimated OR to the RR is poor, particularly when covariates are in model². One suggestion is to use Poisson regression to obtain RRs directly². However the standard errors from Poisson regression are not correct because the outcome is 0/1 and not a count. However Zou has shown how to get correct SEs using Robust Standard Errors⁸.

The problem is compounded in the analysis of cluster randomised trials because we have two type of model.

Population average (PA) models take the form

$$h(\mu_{ij}) = \alpha + \beta x_{ij}$$

where $E(Y_{ij}) = \mu_{ij}$ is the expected response for subject j in week i and $h(\cdot)$ is the link function.
 $\text{Var}(Y_{ij}) = v(\mu_{ij})\phi$ and $\text{Corr}(Y_{ij}Y_{ik}) = \rho(\mu_{ij}\mu_{ik})$

For a population average model the β term is interpreted as the average effect over the clusters and the SE is adjusted for clustering using robust standard errors. To fit these models we use Generalised Estimating Equations (GEEs). This is inefficient with few clusters, but there are small sample corrections available.⁴

The alternative is to use cluster specific or random effects (RE) models

$$h(\mu_{ij}) = \alpha^* + \beta^* x_i + \tau_j + \varepsilon_{ij}$$

where τ_j is the random effect in week j , $V(\tau_j) = \sigma_z^2$ and $V(\varepsilon_{ij}) = \sigma^2$. The intra-cluster correlation (ICC) is defined as $\sigma_z^2 / (\sigma_z^2 + \sigma^2)$.

For a cluster specific model, β is estimated by integrating out the random effect and the interpretation is the effect of the intervention *conditional* on being in a cluster

Neuhaus and Jewell⁵ showed that for binary data and a logistic link these two models contain different population parameters. For the logit model, with Gaussian random effect and binomial error we have that

$$\beta \approx (0.35\sigma_z^2 + 1)^{-0.5} \beta^*.$$

Thus for a logistic link, the population parameter in the population averaged model will be shrunk toward the null. However, if the between cluster variance σ_z^2 is small the two parameters will be close. This inequality arises because of the nature of the logit link, and Neuhaus and Jewell showed that for a linear and also a log-linear model that $\beta = \beta^*$.

2. Motivating example

Our motivating example is a study described by Mason et al.³ which is a cluster randomised trial of paramedic practitioners. The intervention was seven paramedics who were trained to provide community assessment and treatment of patients aged over 60 who contacted the emergency ambulance services between 8am and 8pm. The outcome variables were emergency department attendance or hospital admission within 28 days of call, and satisfaction with service. The randomization was to weeks when paramedic practitioners are either operative or non-operative with a block size of 6 and the trial was run over 56 consecutive weeks. The basic results are given in Table 1. One can see that for all outcomes the relative risk and the odds ratio differ. This is particularly apparent for the Emergency department attendance and for the proportion of subjects very satisfied, where the proportions are very high.

Table 1 Results from Paramedic Practitioner Older People's Support Trial³ (outcomes per week)

	Intervention N=1549, 26 weeks	Control N=1469, 30 weeks	RR	OR
Hospital admissions (%)	40.4 (SD 7.2)	46.5 (SD 5.1)	0.87	0.70
Emergency Department attendance (%)	62.6 (SD 5.0)	87.5 (SD 7.2)	0.72	0.24
Very satisfied	85.5	73.8	1.16	2.09

3 Results of fitting different models

Table 2 gives the results of fitting different models to the data using the package Stata⁶, without including covariates. The value β is the $\log_e$ of the OR or RR. For the random effects model we used a Normal error and adaptive Gaussian quadrature to integrate over the random effect. In view of the fact that the Poisson model is not strictly correct⁸, we used a cluster specific non-parametric bootstrap to give a second estimate of the standard error. The coefficient estimates are unchanged by the bootstrap and so not shown. For the population averaged model we used generalized estimating equations with an exchangeable correlation. There is little evidence of clustering since the ICC is small and so one would expect the logistic model estimates of the random effect and population averaged models would be close and this can be seen to be the case. One can see that for both the logistic model and the Poisson model the OR and RR agree with the empirical values in Table 1. As expected the relative risks agree from the two models. However, the standard errors for the random effects Poisson model and the population average Poisson model differ markedly. This is because the random effects model does not use robust standard errors. Using a cluster non-parametric bootstrap the standard errors for the random effects and population averaged models then became similar.

Table 2 Effect of intervention on ED Attendance (allowing for clustering)

Model	OR/RR	β	SE of β
RE logistic	0.238 (OR)	-1.436	0.1011 (0.0958 Bootstrap)
PA logistic	0.239 (OR)	-1.431	0.1068 (Robust)
RE Poisson	0.715 (RR)	-0.335	0.0425 (0.0221 Bootstrap)
PA Poisson	0.715 (RR)	-0.335	0.0234 (Robust)

ICC=0.0049

Note that using the approximate equation (1) we would expect the standard error of the RR to be about 0.0269, which is somewhat larger than that obtained from the models

To simulate the effect of a high ICC we induced a trend in the proportion attending the ED over the time period. This increased the between weeks variance and so increased the ICC. These results can be seen in Table 3.

Table 3 Simulated data with trend –Effect of intervention (not allowing for trend in model)

Model	OR/RR	β	SE of β
RE logistic	0.825 (OR)	-0.1923	0.407 (0.406 Bootstrap)
PA logistic	0.898 (OR)	-0.1078	0.268 (Robust)
RE Poisson	0.975 (RR)	-0.0245	0.067 (0.066 Bootstrap)
PA Poisson	0.974 (RR)	-0.0263	0.065(Robust)

ICC=0.38

We now see the expected result that the OR for the population averaged model is shrunk toward the null relative to that for the logistic. The standard error is also smaller so the p values for the coefficients of the two models are similar (P= 0.64 and 0.69 respectively) .As expected the estimates of the RR are unaffected by the high ICC. The standard errors for the two models of the log RR are close, and the bootstrap estimate does not change the RE estimate.

Of course in reality, if there is an obvious covariate which induces a between cluster variance one would include it in the model. The effect of including a trend term in the model is shown in Table 4, where we can see that the ICC is now much reduced and although the population averaged OR is still attenuated, the effect is much less marked and the standard errors now agree. Here the RE Poisson standard error of the RR is overestimated and the bootstrap reduces it to that from the PA Poisson model.

Table 4 Simulated data with trend –Effect of intervention (trend included in model)

Model	OR /RR	β	SE of β
RE logistic	0.791 (OR)	-0.234	0.118 (0.117 Bootstrap)
PA logistic	0.795 (OR)	-0.229	0.114 (Robust)
RE Poisson	0.967 (RR)	-0.033	0.042 (0.022 Bootstrap)
PA Poisson	0.967 (RR)	-0.033	0.022 (Robust)

ICC=0.017

4. Discussion

There is a saying that an *epidemiologist* thinks an odds ratio is an approximation to a relative risk whereas a *statistician* thinks a relative risk is an approximation to an odds ratio! Both are valid estimates and in a clinical trial there is usually a single summary measure of the outcome and so one can transform between them. However there is a difficulty in transforming the standard error. Because the population average estimate is shrunk toward the null for the OR, the standard error is smaller, thus preserving the P-values. However the simple transformation does not work in the presence of covariates or for cluster trials. The random effects model using Poisson regression gave an inflated standard error. Zou⁸ gave a method of which used a method robust to heteroscedacity in the individual variances to correct the standard error when using Poisson regression. This is difficult to apply for clustered data and so we used a bootstrap method instead which appeared to give reasonable results. The GEE method uses a standard error robust to heteroscedacity in the cluster variances, not the individual variances, so is not exactly Zou's method. However the bootstrap random effect standard error and the robust standard error from the population averaged model appear to agree, but further work is need in this area. Ukoumunne et al⁷ have shown that the GEE method has generally acceptable properties provided the number of clusters per arm is greater than 10, which it is in the examples we are considering.

All estimates from the models for the Paramedic Practitioner Trial are close to the crude estimates, because we did not include covariates. The population average and random effects estimates and SEs are almost the same for logistic regression (because the between weeks variance is small). As expected the population average and random effects estimates are the same for log-linear regression. However, as shown discussed earlier the standard error from the random effects model is not correct and so needs to be adjusted.

As expected, when the data were constructed to have a high intra-cluster correlation the odds ratio was attenuated, and so was the standard error. If there are known covariates which induce a intra-class correlation, then it is a sensible strategy to correct for them, and we have shown that including trend in the model reduces the ICC and so produces results where the random effects and population averaged models agree.

The question remains – should one use odds ratios or relative risks for binary outcomes of cluster randomised trials? We have shown that if the ICC is low, or one can account for the correlation using covariates, then inference using a logistic-linear model and odds ratios is relatively straightforward. In practice we have found there are far fewer convergence problems using a logistic-linear model, and this is particularly true when the incidence of the outcome is high, and so a log-linear model may predict incidences great than one. Thus we would prefer to use odds ratios. However, most non-statisticians find relative risks easier to interpret and there are problems associated with the interpretation of odds ratios when the ICC is large. Thus in the presence of a large ICC, but a relatively low incidence of events, we would suggest the use of Poisson regression to estimate relative risks, but with careful attention to the calculation of the standard errors.

References

1. Campbell MJ, Donner A and Klar N (2007) Developments in cluster randomized trials and Statistics in Medicine. *Statistics in Medicine*, **26**, 2-19
2. Carter, R.E., Lipsitz, S.R., Tilley, B.C.: Quasi-likelihood estimation for relative risk regression models. *Biostatistics*, **6**, 39-44 (2006)
3. Mason, S., Knowles, E., Colwell, B., Dixon, S., Wardrope, J., Gorringer R, et al. Effectiveness of paramedic practitioners in attending 999 calls from elderly people in the community: cluster randomised controlled trial. *BMJ* 2007 doi: 10.1136/bmj.39343.649097.55.
4. Mancl, L. DeRouen, T.A.: A covariance estimator for GEE with improved small sample properties *Biometrics* 57: 126-134 (2001)
5. Neuhaus, J.M., Jewell, N.P.: A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika* **80**, 807-815 (1993)
6. StataCorp Stata: Release 11. Statistical Software. College Station TX: Statacorp LP. (2009)
7. Ukoumunne, O.C., Carlin, J.B., Gulliford, M.C.: A simulation study of odds ratio estimation for binary outcomes from cluster randomized trials. *Statistics in Medicine*, **26**, 3415-3428 (2007)
8. Zou, G.Y.: A modified Poisson regression approach to prospective studies with binary data *Am J Epidemiol* , **159**, 702-706 (2004)