



Published in final edited form as:

J R Stat Soc Ser A Stat Soc. 2012 July 1; 175(3): . doi:10.1111/j.1467-985X.2011.01003.x.

Sample size estimation for alternating logistic regressions analysis of multilevel randomized community trials of under-age drinking

Beth A. Reboussin,

Wake Forest University School of Medicine, Winston-Salem, USA

John S. Preisser,

University of North Carolina School of Public Health, Chapel Hill, USA

Eun-Young Song, and

Wake Forest University School of Medicine, Winston-Salem USA

Mark Wolfson

Wake Forest University School of Medicine, Winston-Salem USA

Summary

Under-age drinking is an enormous public health issue in the USA. Evidence that community level structures may impact on under-age drinking has led to a proliferation of efforts to change the environment surrounding the use of alcohol. Although the focus of these efforts is to reduce drinking by individual youths, environmental interventions are typically implemented at the community level with entire communities randomized to the same intervention condition. A distinct feature of these trials is the tendency of the behaviours of individuals residing in the same community to be more alike than that of others residing in different communities, which is herein called ‘clustering’. Statistical analyses and sample size calculations must account for this clustering to avoid type I errors and to ensure an appropriately powered trial. Clustering itself may also be of scientific interest. We consider the alternating logistic regressions procedure within the population-averaged modelling framework to estimate the effect of a law enforcement intervention on the prevalence of under-age drinking behaviours while modelling the clustering at multiple levels, e.g. within communities and within neighbourhoods nested within communities, by using pairwise odds ratios. We then derive sample size formulae for estimating intervention effects when planning a post-test-only or repeated cross-sectional community-randomized trial using the alternating logistic regressions procedure.

Keywords

Alternating logistic regressions; Clustering; Community trial; Pairwise odds ratio; Sample size

1. Introduction

Under-age drinking is a long-standing major public health problem in the USA. Alcohol is the drug that is most commonly used by youths aged 12–20 years, with a higher percentage of youths drinking alcohol than using tobacco or illicit drugs (Substance Use and Mental Health Services Administration, 2010). Although under-age drinking decreased following changes in

the minimum age of purchase in the mid-1980s, prevalence rates remain at relatively high levels. Data suggested a decline in current drinking, binge drinking and heavy drinking between 2002 and 2008, but this trend seems to have ended with the most recent data in 2009 (Substance Use and Mental Health Services Administration, 2010). Current drinking, binge drinking, heavy drinking and drinking to intoxication remain alarmingly high among youths. In 2009, 27.2% of youths aged 12–20 years reported drinking during the past 30 days, 18.1% reported binge drinking and 5.4% reported heavy drinking (Substance Use and Mental Health Services Administration, 2010). These findings are concerning given the widespread individual and social consequences that are associated with under-age drinking. Alcohol use in adolescence is associated with the four leading causes of death within this age group: motor vehicle crashes, other unintentional injuries, homicide and suicide (Centers for Disease Control and Prevention National Center for Injury Prevention and Control, 2004). Numerous studies have found that adolescent drinkers are at increased risk of engaging in violent and aggressive behaviours, increased sexual activity and unsafe sexual practices, and sexual victimization (Durant *et al.*, 1997; Cooper, 2002; Swahn *et al.*, 2004; Champion *et al.*, 2004). Adolescent alcohol use may also have deleterious effects on psychosocial development. Early onset drinking is a risk factor for progression to illicit drug use and alcohol abuse and dependence (Anthony and Petronis, 1995). It may also have long-lasting effects on intellectual capabilities (Brown *et al.*, 2000) and ability to attain educational and occupational goals (Braun *et al.*, 2000).

Evidence that community level structures may have an enabling effect on under-age drinking has led to a proliferation of efforts to change the environment surrounding alcohol use (Chaloupka *et al.*, 2002; Toomey and Wagenaar, 2002; Song *et al.*, 2009). These have included efforts to change public policy, institutional policy, law enforcement practices and availability of alcohol (Hingson *et al.*, 1996; Grube, 1997; Wagenaar *et al.*, 2000a, b). Although the ultimate focus of these efforts is to reduce under-age drinking by individual youths, interventions are typically implemented at the community level. Randomization of individuals to intervention or comparison conditions to study the effect of community level interventions, however, is problematic. In the case of a public policy intervention, for example, it would be difficult to limit the effects of a policy only to those youths in a community who are assigned to the intervention condition. As a result, the standard approach in these types of studies is to randomize entire communities to the intervention or comparison condition while still measuring impact at the level of the individual. This is in contrast with clinical trials where randomization and measurement both occur at the individual level. A distinct feature of community-randomized trials that has implications for both study design and analysis is the tendency of the health behaviours of individuals residing in the same geographical area to be more alike than those of others residing in different geographical areas, which is hereafter called ‘clustering’. Further, although the intervention may be implemented at the level of the community and impact assessed at the level of the individual, a more complex hierarchical structure often exists; neighbourhoods are nested within communities (e.g. cities or towns), households are nested within neighbourhoods and individuals are nested within households. The magnitude of the clustering of behaviours within each of these levels may differ and could alter the sample size that is required for detecting a significant intervention effect (Donner *et al.*, 1981).

There are primarily two families of models for analysing data from community-randomized trials that account for clustering: cluster-specific and population-averaged (PA) models. Although both of these approaches have their merits, the choice between them really depends on the research question being investigated. For covariates that do not vary within a cluster-like intervention condition, PA models are often recommended because of their regression parameter interpretation (Zeger *et al.*, 1988; Neuhaus *et al.*, 1991; Heagerty, 1999; Klar and Donner, 2001; Preisser *et al.*, 2003). In the PA model, the regression parameter describes the average change in response across subsets of the population defined by the intervention

condition. For cluster-specific models, the interpretation of the regression parameter is specific to a given cluster. It describes the difference in response when that particular cluster would have been observed in the comparison condition and the intervention condition; a contrast that is not observed in any single cluster (Heagerty, 1999). When efforts are focused on changing public policy or law enforcement practices, the population information rather than the cluster-specific information regarding an intervention effect is often considered more relevant.

In this paper, we focus on the effect of a law enforcement intervention on the prevalence of a set of binary responses, specifically under-age drinking behaviours, and therefore favour the PA model interpretation. The most common PA modelling approach for clustered binary data is the first-order generalized estimating equations approach of Liang and Zeger (1986). This approach focuses on the marginal prevalences and treats the clustering as a nuisance through specification of a simple 'working correlation matrix'. It has the advantage of providing consistent intervention effect estimates even if the clustering as measured by the intraclass correlation coefficient (ICC) is not correctly specified. However, more accurate modelling of the clustering can result in gains in efficiency for the intervention effect (Fitzmaurice, 1995; Wang and Carey, 2003; Ye and Pan, 2006). Heterogeneity in the clustering at the level of the community and neighbourhood may impact on our power to detect intervention effects if they are not accounted for in the design stage. The cluster structure itself may also be of scientific interest. Knowledge that under-age drinking clusters to a greater or lesser degree within neighbourhoods compared with within communities might influence our interpretation of the intervention impact as well as our design of future interventions (Crespi *et al.*, 2009). To address interests in both the prevalence and the cluster structure requires an approach that simultaneously fits models to the marginal prevalences and the within-cluster association.

Estimating equations approaches that simultaneously estimate the prevalence and within-cluster association parameters were introduced by Prentice (1988), Lipsitz *et al.* (1991) and Liang *et al.* (1992), among others. Whereas the approach of Prentice (1988) permits model building on the pairwise within-cluster correlations the other two approaches use pairwise odds ratios PWOR. Odds ratios have traditionally been used in epidemiology as a measure of association for binary responses and are considered easier to interpret. However, the use of the ICC in community trials, including binary outcomes, is well established. The choice between odds ratios and correlations to model within-cluster associations is a matter of personal preference, as inferences on marginal mean model parameters are not likely to be much affected by the choice (Lipsitz *et al.*, 1991). Yet, estimating equation methods for community trials based on PWOR have not been described to the extent that methods based on the ICC have been described.

For the researcher who prefers the pairwise odds ratio over the intraclass correlation as a measure of within-cluster association, the approach of Lipsitz *et al.* (1991) has the advantage of providing consistent mean parameter estimates even if the within-cluster association is misspecified; the price that is paid is the potential for a substantial loss of efficiency for the association parameters. The approach of Liang *et al.* (1992) lacks this robustness property but gains in efficiency are realized by jointly estimating the mean and association parameters in a single estimating equation. This second-order approach, however, is not computationally feasible for large cluster sizes that are frequently encountered in community trials. As an alternative to the approach of Liang *et al.* (1992), the alternating logistic regression (ALR) procedure (Carey *et al.*, 1993) was developed as a computationally feasible implementation of generalized estimating equations which provides consistent estimates of the mean regression parameters even if the association model is not correctly specified. At the same time, it has been shown to be nearly as efficient for association parameter estimation as the second-order estimating equations of Liang *et al.* (1992). Thus, among a collection of estimating equations approaches for conducting inference on PA models of interest, the ALR approach is well suited

for the analysis of data from randomized community trials. Furthermore, it has been widely applied in settings other than randomized community trials in which interest was in conducting inference for PA models for marginal means and within-cluster pairwise odds ratios (Petronis and Anthony, 2000, 2003; Preisser *et al.*, 2003; Ananth and Kantor, 2004; Metlay *et al.*, 2007; Curtis *et al.*, 2009; Wells *et al.*, 2009) Formulae have been provided for calculating design effects when using ALRs to estimate prevalence rates from a cross-sectional survey with a single level of clustering (Katz *et al.*, 1993). We extend this work to derive sample size formulae when using ALRs to estimate an intervention effect in randomized community trials with multiple levels of clustering. The ‘Enforcing underage drinking laws randomized community trial’ (EUDLRCT) is used to illustrate the sample size formulae proposed.

2. ‘Enforcing underage drinking laws randomized community trial’

The EUDL programme is a national initiative, which is funded by the United States Office of Juvenile Justice and Delinquency Prevention, intended to increase enforcement of under-age drinking laws and to reduce under-age drinking. Each year since 1998, each of the 50 states was awarded a block grant to support and enhance state and local efforts to prohibit the sale and consumption of alcoholic beverages to and by minors. In addition, each year since the programme began, discretionary grants were awarded competitively to a subset of the states to expand the number of communities taking a comprehensive approach to prevention of under-age drinking.

The EUDLRCT was funded under the 2003 fiscal year appropriation to compare communities that had been randomized to receive discretionary grants with matched comparison communities not randomized to such intense interventions. States responding to a solicitation for the EUDLRCT were required to provide a list of 14–28 cities or towns that were interested in, and eligible for, participation in the EUDLRCT should the state be funded. Eligibility requirements included

- a. being an incorporated city or town with a population of between 25000 and 200000 and
- b. not having engaged in high levels of certain programmatic activities to reduce under-age drinking in the two years preceding the date of the solicitation.

On the basis of their proposals, five states were funded to participate in the EUDLRCT: California, Connecticut, Florida, Missouri and New York (Wolfson *et al.*, 2005). Using city or town (which are herein collectively referred to as communities) characteristics of population size, median family income and the percentages of the population that were black, Hispanic, spoke Spanish and were currently in college, Mahalanobis distance measures were computed for all possible pairs of nominated communities within a state. Following creation of pairs with the lowest distance measures, a random-number generator was employed to assign communities to either the intervention or control condition. This process resulted in good balance on a variety of community level characteristics (Wolfson *et al.*, 2005). 34 intervention communities were funded to participate in the EUDLRCT and matched to 34 comparison communities.

Data were collected by using a variety of surveys as part of the evaluation of the EUDLRCT. In this paper, we focus on a telephone survey of 14–20-year-olds in each of the 68 communities. Repeated cross-sectional surveys were administered preintervention (or early in the intervention period) in 2004, 2 years later in 2006 and 3 years later in 2007. We consider five self-reported outcomes of interest: past 30-day use of alcohol, becoming drunk, binge drinking, non-violent consequences due to drinking and attempts at purchasing alcohol. Although randomization and delivery of the intervention occurred at the level of the community, we recognize that youths are further nested within neighbourhoods and the magnitude of clustering

of alcohol behaviours may vary by level. Therefore, beginning in 2006, youths were asked during the telephone survey ‘Could you please tell me what your current address is?’. If the respondent refused to answer, they were asked to provide their zip code. A second source of the youths’ addresses was the commercial firm that provided the telephone numbers for the survey sample. Addresses were then geocoded by the Population Research Institute at Pennsylvania State University. The addition of these data permits estimation of clustering at the level of both the community and the neighbourhood (or census tract).

3. Modelling multilevel geographic clustering with pairwise odds ratios

The ALR approach jointly estimates two models: a model for the marginal prevalence and a model for the within-cluster association. First, we consider the model for the marginal prevalence for the three-level data in the EUDL example. Since we do not have youths’ addresses for the preintervention period, we focus on the comparison of the prevalence of under-age drinking during the first post-intervention assessment in 2006. In randomized community trials with a moderate to large number of communities, it is not necessary to adjust for preintervention, since groups are assumed to be balanced with respect to outcomes and covariates as a result of randomization. The hypothesis that the prevalence of past 30-day drinking during the post-test period, for example, differs between intervention and comparison communities can be tested by fitting the logistic regression model

$$\text{logit}(\pi_{ijk}) = \beta_0 + \beta_1 X_i \quad (1)$$

where $\pi_{ijk} = E[Y_{ijk}] = \text{pr}(Y_{ijk} = 1)$ is the probability that the k th youth in the j th neighbourhood in the i th community receiving the intervention ($X_i = 1$) or comparison ($X_i = 0$) condition reports past 30-day drinking. The intervention effect is given by β_1 , and is the log-odds ratio comparing the odds of past 30-day drinking in the post-test period for youths in intervention communities with the odds of past 30-day drinking for youths in control communities. Model (1) can be expanded to include important individual, neighbourhood and community level covariates that are associated with alcohol use.

The regression parameters β_0 and β_1 in model (1) are estimated while simultaneously accounting for the association between responses from the same geographic area as measured by the pairwise odds ratio PWOR. The pairwise odds ratio between two responses Y_{ijk} and Y_{ilm} from the i th community is given by

$$\alpha_{ijklm} = \frac{\text{pr}(Y_{ijk}=1, Y_{ilm}=1) \text{pr}(Y_{ijk}=0, Y_{ilm}=0)}{\text{pr}(Y_{ijk}=1, Y_{ilm}=0) \text{pr}(Y_{ijk}=0, Y_{ilm}=1)}$$

It is interpreted as the odds of past 30-day drinking for a youth given that a randomly chosen youth residing in the same community is a past 30-day drinker relative to the odds that a randomly chosen youth is not a past 30-day drinker. Consistent with odds ratios from logistic regression analysis, PWOR takes a value of 1.0 when there is no clustering of the outcome of interest. If PWOR is greater than 1.0 then the past 30-day drinking of one youth is statistically dependent on the past 30-day drinking of another randomly chosen youth residing in the same community, over the expectation based on randomly paired selections of youths without respect to communities. In the ALR framework, the pairwise odds ratio for the three-level data in the EUDL example can be modelled by

$$\log(\alpha_{ijklm}) = \lambda_0 + \lambda_1 Z_{ijklm} \quad k \neq m \quad (2)$$

where $Z_{ijklm} = 1$ if $j = l$ (i.e. if youths k and m reside in the same neighbourhood) and $Z_{ijklm} = 0$ otherwise. Then, $\exp(\lambda_0 + \lambda_1)$ is the within-neighbourhood PWOR and $\exp(\lambda_0)$ is the within-community–between-neighbourhood PWOR. The within-neighbourhood PWOR estimates the odds of past 30-day drinking for a youth from a neighbourhood where a randomly chosen youth from the same neighbourhood has the same outcome relative to the odds that a youth does not have the same outcome. The within-community–between-neighbourhood PWOR is the odds of past 30-day drinking for a youth from a community where another randomly chosen youth from the same community but different neighbourhood has the same outcome relative to the odds that the randomly chosen youth does not have the same outcome. Because the within-community–between-neighbourhood PWOR considers pairs of respondents from the same community but different neighbourhoods, it reflects the residual clustering of past 30-day drinking after accounting for within-neighbourhood clustering. Estimation of the ALR model parameters β_0 and β_1 in model (1) and λ_0 and λ_1 in model (2) involves alternating between a logistic regression using a first-order generalized estimating equation for the parameters in model (1) and an offset logistic regression of Y_{ijk} on $Z_{ijklm}Y_{ilm}$ ($k > m$) for the association parameters in model (2). The motivation for estimating the association parameters in model (2) by an offset logistic regression is given by Firth (1992) and Diggle (1992). Details are provided in Carey *et al.* (1993). In this paper, the ALR approach is applied by using the PROC GENMOD procedure in SAS version 9.2 (SAS Institute, 2008) with a REPEATED statement and the LOGOR and SUBCLUST options.

4. Trial results

Differences in self-reported under-age drinking behaviours between intervention and comparison communities at post-intervention were examined on the basis of yes–no responses to questions about past 30-day drinking, binge drinking in the past 2 weeks, becoming drunk during the past month, non-violent consequences as a result of drinking in the past year and making an attempt to purchase alcohol in the past month. If communities differ in terms of their composition with respect to factors that are associated with under-age drinking, it is possible that both the intervention effect and the estimates of clustering could be artefacts of these differences. For this reason, several individual level characteristics of the participants some of which most certainly mark varying levels of vulnerability, predisposition or risk of under-age drinking, were incorporated in model (1) in an effort to examine whether their statistical adjustment might attenuate any intervention effects as well as estimates of clustering. We concentrated our analyses on the sociodemographic characteristics of age, gender, race, mother's education and living in a single-parent household which were previously found in the EUDLRCCT to be associated with under-age drinking (Song *et al.*, 2009; Reboussin *et al.*, 2010). We considered two models for the pairwise odds ratios PWOR. In model 1, the within-neighbourhood and within-community–between-neighbourhood PWORs were assumed to depend on intervention conditions. Model 2 constrained the PWORs to be equal across intervention and comparison conditions as in equation (2).

Results are presented in Table 1. Each model is fitted on the basis of 5760 youths from 68 communities residing in 1283 neighbourhoods in 2006. The average number of youths in each community is 85 (range 28–151), with the average number of neighbourhoods within a community equal to 19 (range 7–48). The average number of youths within a neighbourhood is 4 (range 1–41). Adjustment for potential confounding variables in model (1) had little effect on the intervention odds ratios (unadjusted model results are not shown). There was also little difference in the intervention odds ratios and 95% confidence intervals for the two different PWOR models. Although the PWORs were generally larger in the communities that were

assigned to the comparison condition based on model 1, the Wald tests comparing models 1 and 2 were not statistically significant for any of the drinking outcomes. On the basis of the simpler model 2, the largest estimated intervention effect was for becoming drunk during the past month. Youths in the intervention communities have a 10% decreased odds of reporting becoming drunk in the past month as a youth in a comparison community at post-intervention (odds ratio OR = 0.90; 95% confidence interval CI = 0.74,1.10). This result, however, was not statistically significant. With regard to clustering, past 30-day drinking, becoming drunk, non-violent consequences and making an attempt to purchase alcohol all cluster significantly within neighbourhoods as indicated by the within-neighbourhood PWORs in Table 1 for model 2. The significance of the clustering even after adjustment for sociodemographic characteristics provides evidence that the clustering is not an artefact of the composition of communities with respect to individual level factors associated with under-age drinking. Binge drinking did not cluster significantly within neighbourhoods for unadjusted or adjusted models. The greatest magnitude of clustering within neighbourhoods was for making an attempt to purchase alcohol (PWOR = 1.49; 95% CI = 1.16,1.91). A youth living in a neighbourhood where another randomly chosen youth from the same neighbourhood makes a purchase attempt has 1.49 times the odds of making a purchase attempt relative to the odds that a randomly chosen youth does not make a purchase attempt. Other significant within-neighbourhood PWORs were relatively smaller, ranging from 1.13 to 1.15. Once clustering within neighbourhoods had been taken into account, there remained residual clustering of outcomes within communities. The clustering within neighbourhoods was generally greater in magnitude than the residual clustering within communities and between neighbourhoods. These PWORs were only significantly different for experiencing non-violent consequences ($p = 0.0266$).

We then compared the ALR results with those obtained by using the approach of Prentice (1988). Model (1) for the intervention effect is the same for the ALR and Prentice (1988) approaches. Prentice (1988), however, modelled the within-cluster associations by using ICCs in contrast with the ALR approach, which uses PWORs. Models for the approach of Prentice (1988) were fitted by using the SAS interactive matrix language macro that is available at <http://www.bios.unc.edu/jpreisse>. The results are displayed in Table 2. There was little difference in the estimated intervention effects between the two approaches. Similarly to the ALR approach, the intervention effect odds ratios and 95% confidence intervals were generally the same under the two different PWOR models and the Wald tests comparing models 1 and 2 were not statistically significant. Also consistent with the ALR results, the within-neighbourhood ICCs were statistically significant for all the drinking outcomes except binge drinking and the within-neighbourhood and the within-community-between-neighbourhood ICCs were only significantly different for experiencing non-violent consequences. There was a substantive difference, however, in the results for the clustering of making an attempt to purchase alcohol. Under the ALR approach, the PWORs for this outcome were considerably larger than for the other drinking outcomes. The magnitude of the ICCs for making an attempt to purchase alcohol were relatively similar in magnitude to the other drinking outcomes by using the approach of Prentice (1988). On the basis of the ALR approach, a youth has a 49% increased risk of making an attempt to purchase alcohol if another youth in the same neighbourhood makes an attempt to purchase alcohol whereas they have only a 10% increased risk of becoming drunk in the past month if another youth in their neighbourhood reports becoming drunk. The within-neighbourhood ICC, however, for making an attempt to purchase alcohol is 0.020 and the within-neighbourhood ICC for becoming drunk is 0.021. Not only are these findings qualitatively different from the ALR findings, but also the interpretation of the ICC and statements about strength of associations are less clear for binary responses.

5. Sample size formulae

We follow the approach to power calculations for community-randomized trials which are based on community summary statistics appropriate to the study design (Preisser *et al.*, 2003, 2007). The general approach is to specify the hypothesis of interest as $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$ where $\delta = \mu_1 - \mu_0$ is the intervention effect and $\mu_h = E[S_{hi}]$ is the expected value of the summary statistic S_{hi} for the i th community receiving the intervention ($h = 1$) or comparison ($h = 0$) condition. Since communities are assumed to be statistically independent, deriving an expression for the variance of the community summary statistic in each condition (intervention and comparison) is an important step in determining sample size.

5.1. Post-test-only design

For a post-test-only randomized community trial design where the response is measured only after delivery of the intervention and not at baseline, statistical inference is typically based on the logistic regression model (1). For the purpose of sample size considerations, it is convenient to replace the ijk subscript in π_{ijk} with a single subscript h , because the only covariate in the model is whether the community received the intervention ($h = 1$) or was in the comparison ($h = 0$) condition. The intervention effect under this model is given by $\beta_1 = \delta$ where $\delta = \text{logit}(\pi_1) - \text{logit}(\pi_0)$; the log-odds ratio comparing the odds of reporting the behaviour in the post-test period for youths in the intervention communities with the odds of reporting the behaviour for youths in comparison communities. The relevant summary statistic, S_{hi} for the i th community receiving the h th condition, is the logit of the observed proportion p_{hi} reporting the behaviour at post test. An approximately unbiased estimator of δ is then

$d = \sum_{i=1}^{C_1} S_{1i}/C_1 - \sum_{i=1}^{C_0} S_{0i}/C_0$ with large sample variance $\sigma_d^2 = \sigma_1^2/C_1 + \sigma_0^2/C_0$ where σ_h^2 is the large sample variance of S_{hi} , C_1 is the number of intervention communities and C_0 is the number of comparison communities. For a sufficiently large number of intervention and comparison communities, the sample sizes that are needed for testing the intervention by using an approximate two-sided Wald test with significance level α and power $1 - \beta$ is

$$\begin{aligned} C_1 &= (\sigma_1^2 + \sigma_0^2/r)(z_{1-\alpha/2} + z_{1-\beta})^2 / \delta^2, \\ C_0 &= (r\sigma_1^2 + \sigma_0^2)(z_{1-\alpha/2} + z_{1-\beta})^2 / \delta^2 \end{aligned} \quad (3)$$

where z_c is the $(100c)$ th percentile of the standard normal distribution and $r = C_0/C_1$. These methods can be easily modified for sample size planning based on a one-sided hypothesis test. If Φ is the cumulative normal distribution function, then the power $1 - \beta$ is given by

$$1 - \beta = \Phi \left(\frac{\delta}{\sqrt{\sigma_d^2}} - z_{1-\alpha/2} \right). \quad (4)$$

In the case of a three-level design where neighbourhoods are nested within communities and individuals are nested within neighbourhoods as in the EUDLRCT, the observed proportion reporting the behaviour in the post-test period for a youth in the i th community receiving the h th condition can be expressed as

$$p_{hi} = \frac{\sum_{j=1}^{N_{ih}} \sum_{k=1}^{n_{ijh}} Y_{ijkh}}{\sum_{j=1}^{N_{ih}} n_{ijh}}$$

where $Y_{ijkh} = 1$ if the k th youth in the j th neighbourhood in the i th community receiving the h th condition reports the behaviour and $Y_{ijkh} = 0$ otherwise, N_{ih} is the number of neighbourhoods in community i receiving condition h and n_{ijh} is the number of youths in neighbourhood j in community i receiving condition h . For purposes of sample size calculation, we assume that the number of neighbourhoods within communities and the number of youths within each neighbourhood is the same, i.e. $N_{ih} = N$ and $n_{ijh} = n$. It follows that the large sample variance of the summary statistic $S_{hi} = \text{logit}(p_{hi})$ is

$$\sigma_h^2 = \frac{1 + (n-1)\varphi_h^{(w)} + n(N-1)\varphi_h^{(b)}}{Nn\pi_h(1-\pi_h)} \quad (5)$$

where $\varphi_h^{(w)}$ is the correlation of behaviours between youths residing in the same neighbourhood (within-neighbourhood correlation) and $\varphi_h^{(b)}$ is the correlation of behaviours between two youths residing in the same community but different neighbourhoods (within-community–between-neighbourhood correlation). Details regarding the derivation of σ_h^2 in equation (5) can be found in Appendix A. When compared with a single-level design with $\varphi_h^{(w)} = \varphi_h^{(b)} = 0$, the variance of the summary statistic S_{hi} in our three-level design is inflated by a variance inflation factor or design effect $D = 1 + (n-1)\varphi_h^{(w)} + n(N-1)\varphi_h^{(b)}$ (Teerenstra *et al.*, 2010). This is equivalent to the design effects reported for a three-level design with continuous outcomes (Heo and Leon, 2008).

For a two-level design, the correlation between two binary outcomes can be re-expressed in terms of the marginal prevalence π_h and the joint probability that two youths chosen at random report the behaviour as shown by Katz *et al.* (1993). For a three-level design, there are two sources of correlation as described above: $\varphi_h^{(w)}$ and $\varphi_h^{(b)}$. Extending the approach for a two-level design to a three-level design, we express the within-neighbourhood correlation $\varphi_h^{(w)}$ by using first principles in terms of the marginal prevalence π_h and the joint probability that two youths from the same neighbourhood report the behaviour, $p_{h(11)}^{(w)}$,

$$\varphi_h^{(w)} = \frac{p_{h(11)}^{(w)} - \pi_h^2}{\pi_h(1-\pi_h)},$$

and the within-community–between-neighbourhood correlation $\varphi_h^{(b)}$ in terms of the marginal prevalence π_h and the joint probability that two youths from the same community but different neighbourhoods report the behaviour, $p_{h(11)}^{(b)}$,

$$\varphi_h^{(b)} = \frac{p_{h(11)}^{(b)} - \pi_h^2}{\pi_h(1-\pi_h)}.$$

On the basis of the result of Dale (1986) and provided in Liang *et al.* (1992) for a single-level design, we express the joint probabilities $p_{h(11)}^{(w)}$ and $p_{h(11)}^{(b)}$ as functions of the prevalence π_h

and the within-neighbourhood PWOR, which is herein called $\alpha_h^{(w)}$, and the within-community–between-neighbourhood PWOR, which is herein called $\alpha_h^{(b)}$, respectively. Specifically,

$$p_{h(11)}^{(l)} = \frac{1 - 2\pi_h(1 - \alpha_h^{(l)}) - [\{1 - 2\pi_h(1 - \alpha_h^{(l)})\}^2 - 4\alpha_h^{(l)}\pi_h^2(\alpha_h^{(l)} - 1)]^{1/2}}{2(\alpha_h^{(l)} - 1)}$$

if $\alpha_h^{(l)} \neq 1$; otherwise $p_{h(11)}^{(l)} = \pi_h^2$ if $\alpha_h^{(l)} = 1$ for $l \equiv w, b$. The variance of S_{hi} in equation (5), and therefore the sample size and power computations in equations (3) and (4), can now be expressed as functions of the marginal prevalence π_h , the within-neighbourhood PWORs $\alpha_h^{(w)}$, the within-community–between-neighbourhood PWOR $\alpha_h^{(b)}$, the number of neighbourhoods N and the number of youths within each neighbourhood, n . It is clear from the preceding formulae that $\alpha_1^{(l)} = \alpha_0^{(l)}$ implies that $\varphi_1^{(l)} = \varphi_0^{(l)}$ when $\pi_1 = \pi_0$, but not generally otherwise. The practical implication is that, for a given outcome, the models for the PWORs in Table 1 and the models for the ICCs in Table 2 cannot both be true.

5.2. Repeated cross-sectional design

For a pretest–post-test repeated cross-sectional design, the hypothesis that the intervention and comparison communities differ in the change in prevalence over time is tested by fitting the logistic regression model

$$\text{logit}(\pi_{ijkt}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2t} + \beta_{12} X_{1i} X_{2t} \quad (6)$$

where $\pi_{ijkt} = E[Y_{ijkt}] = P(Y_{ijkt} = 1)$ is the probability that the k th youth in the j th neighbourhood in the i th community receiving the intervention ($X_{1i} = 1$) or comparison condition ($X_{1i} = 0$) reports past 30-day drinking at pretest ($X_{2t} = 0$) or post test ($X_{2t} = 1$). The intervention effect is given by β_{12} and is the difference between the log-odds ratio post test for the intervention and comparison communities and the log-odds ratio pretest. We note that for a repeated cross-sectional design, although the same communities and neighbourhoods are followed over time, different samples of youths are taken at pretest and post test. Similar to the derivation of the post-test-only design effect, we replace the $ijkt$ -subscript in π_{ijkt} with two subscripts h and t , resulting in π_{ht} since the only covariates in the model are whether the community received the intervention ($h = 1$) or comparison condition ($h = 0$) and whether the measurement was taken at pretest ($t = 0$) or post test ($t = 1$). The intervention effect under this model is given by $\beta_{12} = \delta$ where $\delta = \text{logit}(\pi_{11}) - \text{logit}(\pi_{10}) - \text{logit}(\pi_{01}) + \text{logit}(\pi_{00})$. Sample size is based on the summary statistic $S_{hi} = \text{logit}(p_{hi1}) - \text{logit}(p_{hi0})$ where

$$p_{hit} = \sum_{j=1}^N \sum_{k=1}^n Y_{ijkht} / Nn.$$

The approximate variance of S_{hi} is

$$\sigma_h^2 = \frac{1+(n-1)\varphi_{h0}^{(w)}+n(N-1)\varphi_{n0}^{(b)}}{Nn\pi_{h0}(1-\pi_{h0})} + \frac{1+(n-1)\varphi_{h1}^{(w)}+n(N-1)\varphi_{h1}^{(b)}}{Nn\pi_{h1}(1-\pi_{h1})} - \frac{2\{\varphi_{h2}^{(w)}+(N-1)\varphi_{h2}^{(b)}\}}{N\sqrt{\{\pi_{h1}(1-\pi_{h1})\pi_{h0}(1-\pi_{h0})\}}} \quad (7)$$

where there are now six types of intraclass correlations; $\varphi_{h0}^{(w)}$ is the correlation between two youths residing in the same neighbourhood at pretest, $\varphi_{h0}^{(b)}$ is the correlation between two youths residing in the same community but different neighbourhoods at pretest, $\varphi_{h1}^{(w)}$ is the correlation between two youths residing in the same neighbourhood at post test, $\varphi_{h1}^{(b)}$ is the correlation between two youths residing in the same community but different neighbourhoods at post test, $\varphi_{h2}^{(w)}$ is the correlation between two youths residing in the same neighbourhood at different time periods of measurement and $\varphi_{h2}^{(b)}$ is the correlation between two youths residing in the same community but different neighbourhoods at different time periods of measurement. Similar to the post-test-only design in Section 5.1, the correlations can be re-expressed as functions of the marginal prevalences π_{ht} , the within-neighbourhood PWOR at time t , $\alpha_{ht}^{(w)}$, the within-community-between-neighbourhood PWOR at time t , $\alpha_{ht}^{(b)}$, the within-neighbourhood PWOR at different time periods of measurement, $\alpha_{h2}^{(w)}$, and the within-community-between-neighbourhood PWOR at different time periods of measurement, $\alpha_{h2}^{(b)}$. These formulae along with the derivation of σ_h^2 in equation (7) are provided in Appendix A.

5.3. Power analysis for the trial

We evaluated the statistical power for a post-test-only three-level design on the basis of known estimates of clustering from the EUDLRCT and explored sensitivity to the within-neighbourhood PWOR $\alpha_h^{(w)}$ and the within-community-between-neighbourhood PWOR $\alpha_h^{(b)}$. Consistent with the findings from the EUDLRCT in Table 1, we assumed that $\alpha_h^{(w)} = \alpha^{(w)}$ and $\alpha_h^{(b)} = \alpha^{(b)}$, i.e. the PWORs between communities assigned to the intervention and comparison conditions are equal. We considered the power to detect an intervention effect for experiencing non-violent consequences. This was the only outcome in which the within-neighbourhood and within-community-between-neighbourhood PWORs were significantly different. Four sets of PWORs were explored: $(\alpha^{(w)}, \alpha^{(b)}) = (1.00, 1.00)$, corresponding to no clustering within or between neighbourhoods, $(\alpha^{(w)}, \alpha^{(b)}) = (1.14, 1.05)$, corresponding to the magnitude of clustering seen in Table 1 for experiencing non-violent consequences, $(\alpha^{(w)}, \alpha^{(b)}) = (1.50, 1.00)$, corresponding to moderate neighbourhood clustering with no residual clustering within communities and between neighbourhoods, and $(\alpha^{(w)}, \alpha^{(b)}) = (1.50, 1.50)$, corresponding to moderate within-neighbourhood and within-community-between-neighbourhood clustering. We assumed that the prevalence at post test in the comparison communities was equal to 0.27. This is the prevalence of experiencing non-violent consequences in the EUDLRCT comparison communities at post test. The statistical power for detecting a statistically significant intervention odds ratio, $\exp(\beta_1)$, by using a two-sided test with a significance level of 0.05 is presented in Fig. 1.

For an equal allocation design with 34 communities assigned to each condition as in the EUDLRCT, 19 neighbourhoods within each community and four youths per neighbourhood (equivalent to the average number of neighbourhoods and youths within neighbourhoods in the EUDLRCT), an intervention odds ratio of 0.83 representing a 17% decreased odds of experiencing non-violent consequences in the intervention communities is detectable with 80% power in the absence of clustering at both the neighbourhood and the community level, $(\alpha^{(w)},$

$\alpha^{(b)} = (1.00, 1.00)$. In the presence of a small amount of within-neighbourhood and within-community–between-neighbourhood clustering, as seen in the EUDLRCT, $(\alpha^{(w)}, \alpha^{(b)}) = (1.14, 1.05)$, the power to detect the same intervention effect reduces to 60%; an intervention odds ratio of 0.79, or decreased odds of 21% or greater in the intervention communities is detectable with 80% power. However, in the absence of within-community–between neighbourhood clustering but moderate within-neighbourhood clustering, $(\alpha^{(w)}, \alpha^{(b)}) = (1.50, 1.00)$, a smaller intervention effect of 18% decreased odds is detectable with 80% power. In the presence of moderate clustering at both the neighbourhood and the community level, $(\alpha^{(w)}, \alpha^{(b)}) = (1.50, 1.50)$, an intervention odds ratio of 0.62 is detectable with 80% power, representing a much larger intervention effect of 38% decreased odds of experiencing non-violent consequences in the intervention communities.

For planning a future trial, we examined the sensitivity of the sample size determinations to assumptions regarding the PWORs. First we examine the sensitivity of the assumption of equal PWORs across conditions, i.e. $\alpha_h^{(w)} = \alpha^{(w)}$ and $\alpha_h^{(b)} = \alpha^{(b)}$ for $h = 0, 1$. Since the PWORs were not significantly different between intervention and comparison conditions for any of the EUDLRCT outcomes, we considered two scenarios. First, we consider the estimates of clustering for becoming drunk in the past month. Although not significantly different, these PWORs exhibited the largest differences between intervention and comparison conditions among the EUDLRCT outcomes. On the basis of the estimates in Table 1 for model 1, we assume $(\alpha_0^{(w)}, \alpha_0^{(b)}) = (1.10, 1.12)$ for comparison communities and $(\alpha_1^{(w)}, \alpha_1^{(b)}) = (1.18, 1.08)$ for intervention communities. We compare this with estimates of sample size when the PWORs are assumed to be equal across conditions, in which case $(\alpha^{(w)}, \alpha^{(b)}) = (1.13, 1.10)$ on the basis of model 2 in Table 1. In a second scenario, we consider larger differences between PWORs, specifically $(\alpha_0^{(w)}, \alpha_0^{(b)}) = (1.75, 1.50)$ and $(\alpha_1^{(w)}, \alpha_1^{(b)}) = (1.10, 1.05)$. We compare this with the average of the PWORs for the intervention and comparison conditions calculated on the log-odds-ratio scale consistent with model (2). Because there are equal numbers of intervention and comparison communities, the $\log(\text{PWOR})$ s are weighted equally. After transforming back to the odds ratio scale, the average PWORs are $(\alpha^{(w)}, \alpha^{(b)}) = (1.39, 1.26)$. Despite the fact that the EUDLRCT study was underpowered to detect the observed intervention effect for becoming drunk, the effect of the intervention was not clinically meaningful. Therefore, we assume that in planning a future trial the intervention is modified on the basis of what was learned in the initial study and the expected intervention odds ratio will be 0.80 in the new trial. We assume that the prevalence of becoming drunk is the same as in the original EUDLRCT study for the comparison communities at post test, specifically 0.25. We also assume that the number of neighbourhoods in each community and the number of youths in each neighbourhood are the same as in the EUDLRCT, $(N, n) = (19, 4)$. For the first scenario using the estimates of the PWORs from the EUDLRCT, the number of communities required to achieve 80% power to detect an odds ratio of 0.80 is $C = 54$ communities regardless of whether the PWORs depend on the intervention condition. In the second scenario with larger differences in the PWORs, the number of communities required was $C = 98$ when the PWORs varied by condition and $C = 99$ when the PWORs were assumed equal across conditions.

We then examined the sensitivity of the sample size determinations to the assumption that the PWORs are different within neighbourhoods and within communities and between neighbourhoods. In this case, we consider the estimates of clustering for experiencing non-violent consequences: the only outcome in which the PWORs were significantly different. We assume that $(\alpha^{(w)}, \alpha^{(b)}) = (1.14, 1.05)$ from Table 1, model 2. We compare this with a sample size based on estimates of clustering obtained when assuming $(\alpha^{(w)}, \alpha^{(b)}) = \alpha$, or an exchangeable PWOR-structure. For non-violent consequences, the exchangeable PWOR-estimate was 1.06 (the results are not shown). We assume that the prevalence of experiencing

non-violent consequences is 0.27 as in the EUDLRCT at post test for the comparison communities and $(N, n) = (19, 4)$. To detect an intervention odds ratio of 0.80 with 80% power, $C = 39$ communities are required when the PWORs are assumed unequal and $C = 41$ when the exchangeable PWOR structure is assumed. We then considered a larger difference in the PWORs which were not realized in the EUDLRCT example of $(\alpha^{(w)}, \alpha^{(b)}) = (1.50, 1.05)$ and compared this with the average PWOR when assuming $\alpha^{(b)} = \alpha^{(w)} = \alpha$. Because the numbers of within-neighbourhood and within-community-between-neighbourhood pairs are not equal, we calculate the average PWOR as the weighted average on the log-odds-ratio scale of within-neighbourhood and within-community-between-neighbourhood PWORs with weights equal to the inverse of the total sum (in a cluster) of the number of pairs of youths within neighbourhoods and the number of pairs of youths within communities and between neighbourhoods. This results in an average PWOR of 1.07. The number of communities required to achieve 80% power to detect an intervention odds ratio of 0.80 is $C = 42$ communities if the PWORs are assumed unequal and $C = 43$ if they are assumed equal.

Finally, we compared the sample size that is required for an analysis by using the Prentice (1988) approach for experiencing non-violent consequences with the sample size based on an ALR analysis. Similarly to the ALR model results, the within-cluster associations that were estimated by using the ICCs did not differ between intervention and comparison conditions but they were significantly different within neighbourhoods and within communities and between neighbourhoods for experiencing non-violent consequences. Therefore, we compare the sample sizes based on estimates of within-cluster associations from model 2 in Tables 1 and 2. We assume that the prevalence is 0.27 at post test in the comparison communities as above and that $(N, n) = (19, 4)$. We assume that $(\phi^{(w)}, \phi^{(b)}) = (0.024, 0.009)$ from Table 2 for the approach of Prentice (1988). Using formula (5) for the variance of the summary statistic, $C = 38$ communities are required to achieve 80% power to detect an odds ratio of 0.80 by using the approach of Prentice (1988). This is in comparison with the sample size that was obtained above based on the ALR approach of $C = 39$ communities.

6. Discussion

In this paper, we presented a population-averaged approach to the design and analysis of community-randomized trials with multiple levels of nesting when outcomes are binary. The population-averaged interpretation for the law enforcement intervention in our community-randomized trial of under-age drinking was considered more relevant and therefore preferred over a cluster-specific interpretation. In contrast with estimation approaches that estimate within-cluster associations by using ICCs, the ALR approach that was presented in this paper uses pairwise odds ratios. Odds ratios are readily accepted as a measure of association between binary variables in the epidemiologic literature and therefore have a familiar interpretation when extended to estimating clustering within geographic areas. As stated earlier, the choice between odds ratios and correlations to model within-cluster associations is a matter of personal preference. For the EUDLRCT example, the estimated intervention effects were the same regardless of whether the ALR approach or the approach of Prentice (1988) which uses ICCs to measure the within-cluster associations was used. This is consistent with others who have found that inferences on the mean model parameters are not likely to be much affected by the choice of odds ratios or ICCs to model the within-cluster association (Lipsitz *et al.*, 1991).

In general, the findings for the within-cluster associations by using the two approaches were qualitatively similar. The possible exception was for making an attempt to purchase alcohol. Estimates based on the ALR approach suggest that making an attempt to purchase alcohol clusters to a greater degree than the other drinking outcomes both within neighbourhoods and within communities and between neighbourhoods. This finding is consistent with the substantive literature that suggests that, even though youths are influenced by the behaviours

of other youths and often obtain alcohol from social sources, there is a greater concentration of youths attempting to purchase alcohol in neighbourhoods with greater availability of alcohol. Greater availability may be realized through less enforcement of under-age drinking and purchase laws, a higher concentration of alcohol outlets in the neighbourhood and the presence of more outlets that are willing to sell to under-age drinkers because of the presence of similar outlets nearby (Paschall *et al.*, 2007; Chen *et al.*, 2009; Reboussin *et al.*, 2011). Therefore, it is reasonable to expect that the behaviour of making an attempt to purchase alcohol may be more strongly influenced by community or neighbourhood level *structures* than the actual behaviour of drinking (e.g. becoming drunk and past 30-day drinking). The estimates of clustering for making an attempt to purchase alcohol based on Prentice (1988) are not consistent with this hypothesis; they are similar in magnitude to the other drinking outcomes. The metric for the ICCs, however, makes it more difficult to interpret these findings from a substantive standpoint for binary responses. The within-neighbourhood PWOR from the ALR approach for making an attempt to purchase alcohol has the interpretation that youths are 49% more likely to report making an attempt to purchase alcohol if a youth residing in the same neighbourhood also reports making an attempt to purchase alcohol relative to if that youth does not report making an attempt to purchase alcohol. It is less clear how to interpret the within-neighbourhood ICC of 0.020 from the approach of Prentice (1988) for this same outcome. Although both estimates are statistically significant, when the clustering is of interest from a scientific and not just a design standpoint, the pairwise odds ratio interpretation may be more informative.

Application of the proposed sample size formulae for an ALR analysis illustrated that, even with the relatively small within-neighbourhood and within-community–between-neighbourhood clustering that was observed in the EUDLRCT, the power to detect intervention odds ratios of 0.83 detectable in the absence of clustering was reduced to 60%. In the presence of moderate within-neighbourhood clustering but no within-community–between-neighbourhood clustering, the effect on the power was much less. The introduction of moderate within-community–between-neighbourhood clustering had the greatest effect on the power. As seen in formulae (5) for the variance of the summary statistic, the within-community–between-neighbourhood clustering is multiplied by a factor of $n(N - 1)$ compared with the within-neighbourhood clustering which is multiplied by a factor of $n - 1$. In our EUDLRCT example, $n = 4$, which was quite small. For larger neighbourhoods, we might expect to see a greater effect for the within-neighbourhood clustering in the absence of the within-community–between-neighbourhood clustering. This also demonstrates that, in the presence of significant within-neighbourhood clustering and no residual clustering within communities, increasing the number of neighbourhoods N rather than the number of youths n within a neighbourhood will have a greater positive influence on the power.

Sensitivity analyses revealed that for balanced designs ($C_0 = C_1$, and N and n fixed across clusters) the sample size calculations were not sensitive to the simplifying assumptions of equal PWORs across intervention conditions or geographic areas (within and between neighbourhoods) for the EUDLRCT example or in scenarios where we considered larger differences in the PWORs. Even though the magnitude of the differences and the influence on sample size were minimal, the information that is gained by modelling the PWORs may be of scientific interest. For both the ALR and Prentice (1988) approaches, the within-neighbourhood clustering was generally greater than the within-community–between-neighbourhood clustering. The statistically significant difference in these PWORs for experiencing non-violent consequences could inform future interventions. By implementing interventions at the neighbourhood level rather than the community level, a greater effect on non-violent consequences may be realized. It may also lead to future work that could provide insights into why non-violent consequences cluster to a greater degree within neighbourhoods in contrast with other outcomes in which there is less distinction between within-neighbourhood and

between-neighbourhood clustering of outcomes. For example, it might lead to investigations into what neighbourhood level structures (e.g. parties where parents provide alcohol) might explain the clustering of non-violent consequences within neighbourhoods in contrast with community level structures like law enforcement activities that were the focus of the EUDLRCT interventions.

For the outcome that we considered in which the within-neighbourhood and within-community-between-neighbourhood within-cluster associations were significantly different, the sample size estimates were the same for both the ALR and the Prentice (1988) approaches. However, this may not always be so. In fact, although the results are not shown, if we consider the outcome of becoming drunk in the past month and use the estimates of the PWORs and ICCs from model 2 in Tables 1 and 2, the sample size that is required to detect an intervention odds ratio of 0.80 with 80% power is $C = 54$ communities on the basis of the ALR approach and $C = 48$ communities on the basis of the Prentice (1988) approach. In general, we would not expect the results to be equivalent because the models for the within-cluster associations are not equivalent. Specifically, when the marginal mean model (1) has covariates (e.g. condition, gender and age), the within-cluster association model (2), in which the PWOR-structure of homogeneity applies across clusters, does not correspond to a within-cluster association model based on pairwise correlations whose structure homogeneously applies across clusters, and vice versa. Although the results are not widely different, they highlight the importance of using the same model for analysis that was used in the sample size computation.

Finally, we mention that sample size formulae for community-randomized trials based on equal cluster sizes, including those which we have proposed, will overestimate power for samples with unequal cluster sizes, when the total sample size is fixed (Donnar and Klar, 2000). Even when balanced sample sizes are planned, realized sample sizes, and particularly cluster sizes, are frequently unbalanced in ways that cannot be anticipated. When the sizes of all clusters are known, Donner and Klar (2000) have suggested inserting an adjusted mean cluster size into sample size formulae for a simple post-test-only design. The expression for the mean cluster size gives a value that is smaller than the mean cluster size, reflecting power loss due to imbalance. Because actual cluster sizes are commonly not known in advance, simulation methods that consider a wide range of cluster size distributions would be informative in assessing the effect of unbalanced sample sizes on power.

In summary, ALRs are a computationally feasible implementation of the generalized estimating equations procedure for fitting population-average models for the analysis of community trials. In comparison with cluster-specific models for binary data, the user has more control over model specification for the within-cluster association, because models for the prevalence and association of responses within the same cluster are specified separately (Preisser *et al.*, 2004). In our example, we could simultaneously model clustering at the level of the community and the neighbourhood as well as between conditions. For the repeated cross-sectional design, we presented models that allow for the specification of different magnitudes of clustering within and between time at each level. This increased flexibility afforded by population-averaged approaches represents a significant advance for the design of community-randomized trials with binary outcomes and multiple levels of nesting. It provides consistent estimates of the intervention effect even if the association model is not correctly specified. Moreover, we showed in a community trial to reduce under-age drinking that intervention effects were similar under four different within-cluster association models: one pair of models based on PWORs; the other on ICCs. We have demonstrated that there are not meaningful differences between the ALR and Prentice (1988) approaches with regard to the estimation and interpretation of the intervention effects, so that, when the pairwise odds ratio interpretation is preferred for interpretation of the association structure either from a design or scientific standpoint, the ALR approach and the accompanying sample size formulae that were presented

in this paper are well suited for the design and analysis of multilevel community-randomized trials.

References

- Ananth CV, Kantor ML. Modeling multivariate binary responses with multiple levels of nesting based on alternating logistic regressions: an application to caries aggregation. *J Dentl Res*. 2004; 83:776–781.
- Anthony JC, Petronis KR. Early onset drug use and risk of later drug problems. *Drug Alchl Depend*. 1995; 40:9–15.
- Braun BL, Hannan P, Wolfson M, Jones-Webb R, Sidney S. Occupational attainment, smoking, alcohol intake, and marijuana use: ethnic-gender differences in the CARDIA study. *Addict Behav*. 2000; 25:399–414. [PubMed: 10890293]
- Brown SA, Tapert SF, Granholm E, Delis DC. Neurocognitive functioning of adolescents: effects of protracted alcohol use. *Alchl Clin Exptl Res*. 2000; 24:164–171.
- Carey VJ, Zeger S, Diggle P. Modeling multivariate binary data with alternating logistic regressions. *Biometrika*. 1993; 80:517–526.
- Centers for Disease Control and Prevention National Center for Injury Prevention and Control. Web-based injury statistics query and reporting system (WISQARS). Centers for Disease Control and Prevention; Atlanta: 2004.
- Chaloupka FJ, Grossman M, Saffer H. The effects of price on alcohol consumption and alcohol-related problems. *Alchl Res Hlth*. 2002; 26:221–234.
- Champion HL, Foley KL, Durant RH, Hensberry R, Altman D, Wolfson M. Adolescent sexual victimization, use of alcohol and other substances, and other health risk behaviors. *J Adolesc Hlth*. 2004; 35:321–328.
- Chen JJ, Gruenewald PJ, Remer LG. Does alcohol outlet density affect youth access to alcohol? *J Adolesc Hlth*. 2009; 44:582–589.
- Cooper ML. Alcohol use and risky sexual behaviour among college students and youth: evaluating the evidence. *J Stud Alchl Suppl*. 2002; 14:101–117.
- Crespi CM, Wong WK, Mishra SI. Using second-order generalized estimating equations to model heterogeneous intraclass correlation in cluster-randomized trials. *Statist Med*. 2009; 28:814–827.
- Curtis JR, Arora T, Xi J, Silver A, Allison JJ, Chen L, Saag KG, Schenck A, Westfall AO, Colon-Emeric C. Do physicians within the same practice setting manage osteoporosis patients similarly?: implications for implementation research. *Osteoporos Int*. 2009; 20:1921–1927. [PubMed: 19319619]
- Dale J. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*. 1986; 42:909–917. [PubMed: 3814731]
- Diggle PJ. Discussion on ‘Multivariate regression analysis for categorical data’ (by K.-Y. Liang, S. L. Zeger and B. Qaqish). *J R Statist Soc B*. 1992; 54:28–29.
- Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidem*. 1981; 114:906–914.
- Donner, A.; Klar, N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London: Arnold; 2000.
- Durant RH, Kahn J, Beckford PH, Woods ER. The association of weapon carrying and fighting on school property and other health risk and problem behaviors among high school students. *Arch Pediatr Adolesc Med*. 1997; 151:360–366. [PubMed: 9111434]
- Firth D. Discussion on ‘Multivariate regression analysis for categorical data’ (by K.-Y. Liang, S. L. Zeger and B. Qaqish). *J R Statist Soc B*. 1992; 54:24–26.
- Fitzmaurice GM. A caveat concerning independence estimating equations with multivariate binary data. *Biometrics*. 1995; 51:309–317. [PubMed: 7766784]
- Grube JW. Preventing sales of alcohol to minors: results from a community trial. *Addiction*. 1997; 92 (suppl 2):S251–S260. [PubMed: 9231448]

- Heagerty PJ. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*. 1999; 55:688–698. [PubMed: 11314994]
- Heo M, Leon AC. Statistical power and sample size requirements for three level hierarchical cluster randomized trials. *Biometrics*. 2008; 64:1256–1262. [PubMed: 18266889]
- Hingson R, McGovern T, Howland J, Heeren T, Winter M, Azkocs R. Reducing alcohol-impaired driving in Massachusetts: the Saving Lives Program. *Am J Publ Hlth*. 1996; 86:791–797.
- Katz J, Carey VJ, Zeger SL, Sommer A. Estimation of design effects and diarrhea clustering within households and villages. *Am J Epidem*. 1993; 138:994–1006.
- Klar N, Donner A. Current and future challenges in the design and analysis of cluster randomization trials. *Statist Med*. 2001; 20:3729–3740.
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
- Liang KY, Zeger S, Qaqish B. Multivariate regression analysis for categorical data (with discussion). *J R Statist Soc B*. 1992; 54:3–40.
- Lipsitz S, Laird N, Harrington D. Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*. 1991; 78:153–160.
- Metlay JP, Camargo CA Jr, MacKenzie T, McCulloch C, Maselli J, Levin SK, Kersey A, Gonzales R. Cluster-randomized trial to improve antibiotic use for adults with acute respiratory infections treated in emergency departments. *Ann Emerg Med*. 2007; 50:221–230. [PubMed: 17509729]
- Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Statist Rev*. 1991; 59:25–35.
- Paschall MJ, Grube JW, Black C, Ringwalt CL. Is commercial alcohol availability related to adolescent alcohol sources and alcohol use? : findings from a multi-level study. *J Adolesc Hlth*. 2007; 41:168–174.
- Petronis KR, Anthony JC. Perceived risk of cocaine use and experience with cocaine: do they cluster within US neighborhoods and cities? *Drug Alchl Depend*. 2000; 57:183–192.
- Petronis KR, Anthony JC. A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. *J Epidem Commtty Hlth*. 2003; 57:893–900.
- Preisser JS, Arcury TA, Quandt SA. Author reply to “Detecting occupational illness clustering with alternating logistic regressions applied to longitudinal data”. *Am J Epidem*. 2004; 160:505–506.
- Preisser JS, Reboussin BA, Song EY, Wolfson M. The importance and role of intracluster correlations in planning cluster trials. *Epidemiology*. 2007; 18:552–560. [PubMed: 17879427]
- Preisser JS, Young ML, Zaccaro DJ, Wolfson M. An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. *Statist Med*. 2003; 22:1235–1254.
- Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics*. 1988; 44:1033–1048. [PubMed: 3233244]
- Reboussin BA, Preisser JS, Song EY, Wolfson M. Geographic clustering of underage drinking and the influence of community characteristics. *Drug Alchl Depend*. 2010; 106:38–47.
- Reboussin BA, Song EY, Wolfson M. The impact of alcohol outlet density on the geographic clustering of underage drinking behaviors within census tracts. *Alchlism Clin Exptl Res*. 2011; 35:1541–1549.
- SAS Institute. *SAS Version 9.2*. Cary: SAS Institute; 2008.
- Song EY, Reboussin BA, Foley KL, Kaltenbach LA, Wagoner KG, Wolfson M. Selected community characteristics and underage drinking. *Subst Use Misuse*. 2009; 44:179–194. [PubMed: 19142820]
- Substance Use and Mental Health Services Administration. *Results from the 2009 National Survey on Drug Use and Health, vol. I, Summary of National Findings*. Rockville: Substance Use and Mental Health Services Administration; 2010.
- Swahn MH, Simon TR, Hammig BJ, Guerrero JL. Alcohol-consumption behaviors and risk for physical fighting and injuries among adolescent drinkers. *Addictn Behav*. 2004; 29:959–963.
- Teerenstra S, Lu B, Preisser JS, van Achterberg T, Borm GF. Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics*. 2010; 66:1230–1237. [PubMed: 20070297]
- Toomey TL, Wagenaar AC. Environmental policies to reduce college drinking: options and research findings. *J Stud Alchl*. 2002; (suppl):193–205.

Wagenaar AC, Murray DM, Gehan JP, Wolfson M, Forster JL, Toomey TL, Perry CL, Jones–Webb R. Communities mobilizing for change on alcohol: outcomes from a randomized community trial. *J Stud Alchl.* 2000a; 61:85–94.

Wagenaar AC, Murray DM, Toomey TL. Communities mobilizing for change on alcohol (CMCA): effects of a randomized trial on arrests and traffic crashes. *Addiction.* 2000b; 95:209–217. [PubMed: 10723849]

Wang YG, Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalized estimating equations performance. *Biometrika.* 2003; 90:29–41.

Wells JE, Degenhardt L, Bohnert KM, Anthony JC, Scott KM. Geographical clustering of cannabis use: results from the New Zealand Mental Health Survey 2003–2004. *Drug Alchl Depend.* 2009; 99:309–316.

Wolfson, M.; Song, E.; Martin, B.; Wagoner, K.; Brown, V.; Suerken, C.; Reboussin, B.; Foley, K.; Preisser, J.; Garner, G.; Hulme, S.; Miller, C. Report. Wake Forest University School of Medicine; Winston-Salem: 2005. National evaluation of the Enforcing Underage Drinking Laws Randomized Community Trial: year 1 report.

Ye H, Pan J. Modeling of covariance structures in generalized estimating equations for longitudinal data. *Biometrika.* 2006; 93:927–941.

Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equations approach. *Biometrics.* 1988; 44:1049–1060. [PubMed: 3233245]

Appendix A

A.1. Derivation of σ_h^2 for post-test-only design

The summary statistic for the i th community receiving the h th condition based on model (1) is $S_{hi} = \text{logit}(p_{hi})$. For a three-level design, p_{hi} can be expressed as

$$p_{hi} = \sum_{j=1}^N \sum_{k=1}^n Y_{ijkh} / Nn,$$

where N is the number of neighbourhoods in each community and n is the number of youths in each neighbourhood. From first principles, the variance of p_{hi} is given by

$$\begin{aligned} \text{var}(p_{hi}) &= \frac{1}{N^2 n^2} \left\{ \sum_{j=1}^N \text{var} \left(\sum_{k=1}^n Y_{ijkh} \right) + \sum_{j=1}^N \sum_{l < j=1}^N \text{cov} \left(\sum_{k=1}^n Y_{ijkh}, \sum_{k=1}^n Y_{ilkh} \right) \right\} \\ &= \frac{1}{N^2 n^2} \left[\sum_{j=1}^N \left\{ \sum_{k=1}^n \text{var}(Y_{ijkh}) + \sum_{k=1}^n \sum_{m < k=1}^n \text{cov}(Y_{ijkh}, Y_{ijmh}) \right\} + \sum_{j=1}^N \sum_{l < j=1}^N \sum_{k=1}^n \sum_{m=1}^n \text{cov}(Y_{ijkh}, Y_{ilmh}) \right] \\ &= \frac{1}{N^2 n^2} \left[\sum_{j=1}^N \left\{ \sum_{k=1}^n \pi_h (1 - \pi_h) + \sum_{k=1}^n \sum_{m < k=1}^n \varphi_h^{(w)} \pi_h (1 - \pi_h) \right\} + \sum_{j=1}^N \sum_{l < j=1}^N \sum_{k=1}^n \sum_{r=1}^n \varphi_h^{(b)} \pi_h (1 - \pi_h) \right] \\ &= \frac{1}{N^2 n^2} \{ Nn\pi_h(1 - \pi_h) + Nn(n - 1)\varphi_h^{(w)}\pi_h(1 - \pi_h) + N(N - 1)n^2\varphi_h^{(b)}\pi_h(1 - \pi_h) \} \\ &= \frac{\pi_h(1 - \pi_h)}{Nn} \{ 1 + (n - 1)\varphi_h^{(w)} + n(N - 1)\varphi_h^{(b)} \}. \end{aligned}$$

It follows from the delta method that the variance of the summary statistic $S_{hi} = \text{logit}(p_{hi})$ is $S'_{hi}(p_{hi})^2 \text{var}(p_{hi})$ so

$$\sigma_h^2 = \frac{1+(n-1)\varphi_h^{(w)} + n(N-1)\varphi_h^{(b)}}{Nn\pi_h(1-\pi_h)}.$$

A.2. Derivation of σ_h^2 for repeated cross-sectional design

The summary statistic for the i th community receiving the h th condition based on model (6) is $S_{hi} = \text{logit}(p_{hi1}) - \text{logit}(p_{hi0})$. For a three-level design, p_{hit} can be expressed as

$$p_{hit} = \sum_{j=1}^N \sum_{k=1}^n Y_{ijkht} / Nn$$

where $t = 0$ is pretest and $t = 1$ is post test. Using the result for the variance of $\text{logit}(p_{hi})$ above for a post-test-only design, we have that

$$\text{var}\{\text{logit}(p_{hi0})\} = \frac{1+(n-1)\varphi_{h0}^{(w)} + n(N-1)\varphi_{h0}^{(b)}}{Nn\pi_{h0}(1-\pi_{h0})}$$

and

$$\text{var}\{\text{logit}(p_{hi1})\} = \frac{1+(n-1)\varphi_{h1}^{(w)} + n(N-1)\varphi_{h1}^{(b)}}{Nn\pi_{h1}(1-\pi_{h1})}.$$

The variance of the summary statistic S_{hi} is given by

$$\sigma_h^2 = \text{var}\{\text{logit}(p_{hi1})\} + \text{var}\{\text{logit}(p_{hi0})\} - 2\text{cov}\{\text{logit}(p_{hi1}), \text{logit}(p_{hi0})\}.$$

Applying the multivariate delta method to the covariance, we have that

$$\text{cov}\{\text{logit}(p_{hi1}), \text{logit}(p_{hi0})\} = S'_{hi1}(p_{hi1})\text{cov}(p_{hi1}, p_{hi0})S'_{hi0}(p_{hi0})$$

where $S_{hi1} = \text{logit}(p_{hi1})$ and $S_{hi0} = \text{logit}(p_{hi0})$. The covariance between p_{hi0} and p_{hi1} is given by

$$\begin{aligned} \text{cov}(p_{hi0}, p_{hi1}) &= \frac{1}{N^2n^2} \left\{ \sum_{j=1}^n \sum_{k=1}^n \sum_{m=1}^n \text{cov}(Y_{ijkh1}, Y_{ijmh0}) + \sum_{j=1}^N \sum_{l < j=1}^N \sum_{k=1}^n \sum_{m=1}^n \text{cov}(Y_{ijkh1}, Y_{ilmh0}) \right\} \\ &= \frac{1}{N^2n^2} [Nn^2\varphi_{h2}^{(w)} \sqrt{\{\pi_{h1}(1-\pi_{h1})\pi_{h0}(1-\pi_{h0})\}} + N(N-1)n^2\varphi_{h2}^{(b)} \sqrt{\{\pi_{h1}(1-\pi_{h1})\pi_{h0}(1-\pi_{h0})\}}] \\ &= \frac{\sqrt{\{\pi_{h1}(1-\pi_{h1})\pi_{h0}(1-\pi_{h0})\}}\{\varphi_{h2}^{(w)} + (N-1)\varphi_{h2}^{(b)}\}}{N}. \end{aligned}$$

Finally, $S'_{hit}(p_{hit})=1/\pi_{ht}(1-\pi_{ht})$ so that

$$\begin{aligned} \text{cov}\{\text{logit}(p_{hi1}), \text{logit}(p_{hi0})\} &= \frac{\sqrt{\{\pi_{h1}(1-\pi_{h1})\pi_{h0}(1-\pi_{h0})\}}\{\varphi_{h2}^{(w)}+(N-1)\varphi_{h2}^{(b)}\}}{N\pi_{h0}(1-\pi_{h0})\pi_{h1}(1-\pi_{h1})} \\ &= \frac{\varphi_{h2}^{(w)}+(N-1)\varphi_{h2}^{(b)}}{N\sqrt{\{\pi_{h0}(1-\pi_{h0})\pi_{h1}(1-\pi_{h1})\}}}. \end{aligned}$$

This can be substituted along with the derivations above into σ_h^2 to obtain the result in equation (7).

A.3. Formulae for repeated cross-sectional design

The correlations $\varphi_{ht}^{(l)}$ between individuals at the same time period of measurement t can be re-expressed as

$$\varphi_{ht}^{(l)} = \frac{p_{ht(11)}^{(l)} - \pi_{ht}^2}{\pi_{ht}(1-\pi_{ht})}$$

for $l = b, w, h = 0, 1$ and $t = 0, 1$. On the basis of the result in Dale (1986), the joint probabilities $p_{ht(11)}^{(l)}$ can be expressed as functions of the prevalences π_{ht} , the within-neighbourhood PWOR at time t , which is herein called $\alpha_{ht}^{(w)}$, and the within-community-between-neighbourhood PWOR, which is herein called $\alpha_{ht}^{(b)}$. Specifically,

$$p_{ht(11)}^{(l)} = \frac{1-2\pi_{ht}(1-\alpha_{ht}^{(l)}) - [\{1-2\pi_{ht}(1-\alpha_{ht}^{(l)})\}^2 - 4\alpha_{ht}^{(l)}\pi_{ht}^2(\alpha_{ht}^{(l)}-1)]^{1/2}}{2(\alpha_{ht}^{(l)}-1)},$$

for $l \equiv b, w, h = 0, 1$ and $t = 0, 1$. The correlation $\varphi_{h2}^{(l)}$ between individuals at different time periods of measurement is given by

$$\varphi_{h2}^{(l)} = \frac{p_{h2(11)}^{(l)} - \pi_{h0}\pi_{h1}}{\sqrt{\{\pi_{h0}(1-\pi_{h0})\pi_{h1}(1-\pi_{h1})\}}}$$

for $l \equiv b, w$ and $h = 0, 1$. On the basis of the result in Dale (1986), the joint probabilities $p_{h2(11)}^{(l)}$ can be expressed as functions of the prevalences π_{ht} , the within-neighbourhood PWOR at different time periods of measurement, which is herein called $\alpha_{h2}^{(w)}$, and the within-community-between-neighbourhood PWOR at different time periods of measurement, which is herein called $\alpha_{h2}^{(b)}$. Specifically,

$$p_{h2(11)}^{(l)} = \frac{1 - (\pi_{h0} + \pi_{h1})(1 - \alpha_{h2}^{(l)}) - [\{1 - (\pi_{h0} + \pi_{h1})(1 - \alpha_{h2}^{(l)})\}^2 - 4\alpha_{h2}^{(l)}\pi_{h0}\pi_{h1}(\alpha_{h2}^{(l)} - 1)]^{1/2}}{2(\alpha_{h2}^{(l)} - 1)},$$

for $l \equiv b, w$ and $h = 0, 1$.

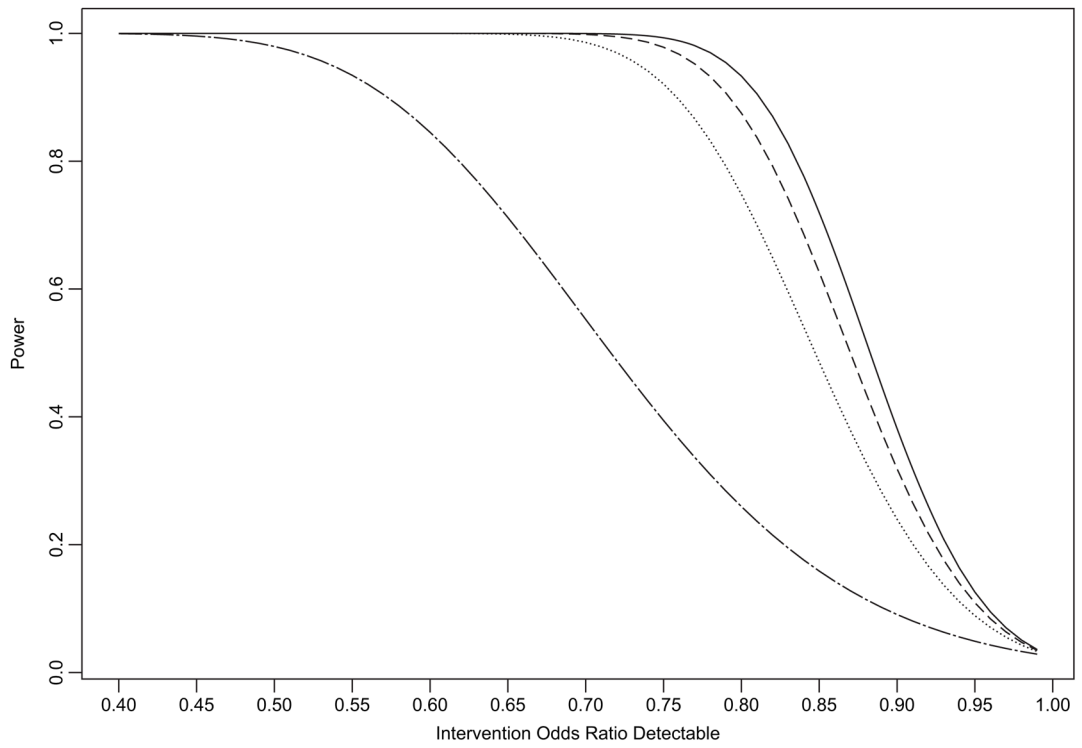


Fig. 1. Statistical power for comparing prevalences at post test between intervention and comparison communities when the prevalence at post test for the comparison communities is 0.27, $C = 34$, $N = 19$ and $n = 4$ for pairs of within-neighbourhood and within-community-between neighbourhood pairwise odds ratios $(a^{(w)}, a^{(b)})$: —, (1.00,1.00); , (1.14,1.05); -- --, (1.50,1.00); — · —, (1.50,1.50)

Table 1
Intervention effects (odds ratios) and estimates of clustering (PWORs) for the EUDLRC estimated by ALRs, USA 2006

Outcome	Model [†]	Intervention effect [‡] , odds ratio (95% confidence interval)	Within-neighbourhood PWOR (95% confidence interval)		Within-community-between-neighbourhood PWOR (95% confidence interval)		p-value [§]
			Control	Intervention	Control	Intervention	
Past 30-day drinking	1	0.96 (0.79,1.17)	1.18 (1.03,1.36)	1.13 (1.01,1.26)	1.15 (1.06,1.26)	1.07 (0.99,1.15)	0.3236
	2	0.97 (0.79,1.18)	1.15 (1.05,1.26)		1.11 (1.05,1.18)		0.3312
Binge drinking	1	0.99 (0.77,1.27)	1.13 (0.92,1.39)	1.06 (0.88,1.27)	1.14 (1.03,1.27)	1.12 (1.00,1.26)	0.8828
	2	0.99 (0.77,1.27)	1.09 (0.95,1.26)		1.13 (1.05,1.22)		0.6116
Becoming drunk	1	0.90 (0.74,1.10)	1.10 (0.92,1.30)	1.18 (1.03,1.34)	1.12 (1.02,1.23)	1.08 (0.97,1.20)	0.2276
	2	0.90 (0.74,1.10)	1.13 (1.01,1.26)		1.10 (1.03,1.18)		0.4587
Non-violent consequences	1	0.95 (0.81,1.13)	1.13 (1.04,1.24)	1.14 (1.00,1.30)	1.07 (1.01,1.14)	1.03 (0.98,1.08)	0.5848
	2	0.96 (0.81,1.13)	1.14 (1.05,1.23)		1.05 (1.01,1.09)		0.0266
Purchase attempt	1	0.94 (0.65,1.36)	1.57 (1.07,2.31)	1.41 (1.05,1.90)	1.43 (1.17,1.74)	1.41 (1.09,1.82)	0.9067
	2	0.94 (0.65,1.36)	1.49 (1.16,1.91)		1.42 (1.21,1.66)		0.6830

[†]Model 1, four-parameter PWOR-model with separate PWORs for intervention and control conditions; model 2, two-parameter PWOR-model with equal PWORs for intervention and control conditions.

[‡]The prevalence model adjusts for individual level covariates: age, gender, race, mother's education and living in a single-parent household.

[§]The p-value for model 1 is from a 2 degrees of freedom Wald test that compares the nested four-parameter (model 1) and two-parameter (model 2) models. The p-value for model 2 is from a 1 degree of freedom Wald test that the within-neighbourhood and within-community-between-neighbourhood PWORs are equal.

Table 2

Intervention effects (odds ratios) and estimates of clustering (ICCs) for the EUDLRTC estimated by the approach of Prentice (1988), USA, 2006

Outcome	Model [†]	Intervention effects [‡] , odds ratio (95% confidence interval)	Within-neighbourhood ICC (95% confidence interval)		Within-community-between-neighbourhood ICC (95% confidence interval)		p-value [§]
			Control	Intervention	Control	Intervention	
Past 30-day drinking	1	0.96 (0.79,1.18)	0.029 (0.001,0.057)	0.025 (0.001,0.050)	0.027 (0.010,0.044)	0.012 (-0.002,0.026)	0.7226
	2	0.97 (0.80,1.18)	0.027 (0.008,0.046)		0.020 (0.009,0.031)		0.3270
Binge drinking	1	1.01 (0.78,1.30)	0.009 (-0.009,0.026)	0.011 (-0.014,0.036)	0.011 (0.0002,0.021)	0.008 (-0.002,0.018)	0.9437
	2	1.00 (0.78,1.29)	0.010 (-0.005,0.025)		0.009 (0.002,0.016)		0.9539
Becoming drunk	1	0.91 (0.74,1.11)	0.015 (-0.015,0.045)	0.027 (0.002,0.053)	0.018 (0.002,0.035)	0.010 (-0.007,0.027)	0.3091
	2	0.91 (0.74,1.11)	0.021 (0.001,0.041)		0.014 (0.002,0.027)		0.3587
Non-violent consequences	1	0.95 (0.80,1.12)	0.023 (0.005,0.040)	0.025 (-0.002,0.052)	0.012 (0.00003,0.024)	0.005 (-0.004,0.014)	0.8311
	2	0.95 (0.80,1.13)	0.024 (0.008,0.040)		0.009 (0.001,0.016)		0.0289
Purchase attempt	1	1.01 (0.70,1.47)	0.022 (-0.011,0.055)	0.017 (0.002,0.031)	0.023 (0.008,0.038)	0.016 (0.005,0.028)	0.9976
	2	1.01 (0.70,1.46)	0.020 (0.001,0.038)		0.020 (0.010,0.029)		0.9928

[†]Model 1, four-parameter ICC model with separate ICCs for intervention and control conditions; model 2, two-parameter ICC model with equal ICCs for intervention and control conditions.

[‡]The prevalence model adjusts for individual level covariates: age, gender, race, mother's education and living in a single-parent household.

[§]The p-value for model 1 is from a 2 degrees of freedom Wald test that compares the nested four-parameter (model 1) and two-parameter (model 2) models. The p-value for model 2 is from a 1 degree of freedom Wald test that the within-neighbourhood and within-community-between-neighbourhood ICCs are equal.