C. Breen, A. Shakeshaft, T. Slade,
C. D'Este and R.P. Mattick

Alcohol-related crime: Finding a suitable measure for community-level analyses using routinely collected data

NDARC Technical Report No. 317

ALCOHOL-RELATED CRIME: FINDING A SUITABLE MEASURE FOR COMMUNITY-LEVEL ANALYSES USING ROUTINELY COLLECTED DATA

Courtney Breen, Anthony Shakeshaft, Tim Slade, Catherine D'Este and Richard P Mattick

Technical Report Number 317

Funded by Alcohol Education and Rehabilitation Foundation as part of the Alcohol Action in Rural Communities (AARC) Project

ISBN: 978-0-7334-3030-5

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Aims Alcohol misuse causes substantial public health harm. Strategies have been proposed to reduce alcohol-related harm at the community-level, which requires suitable community-level measures to monitor changes over time and between communities. For alcohol-related crime, certain offences occurring at certain times that often involve alcohol have been used as a proxy measure. There is currently no adequate empirical rationale for identifying the most reliable proxy measure of alcohol-related crime. This report examines the suitability of three measures of alcohol-related crime. **Methods** Police records of reported incidents from twenty communities in NSW, Australia, that were involved in a community-wide randomised controlled trial to reduce alcohol-related harm were examined. Three measures were derived; i) serious assaults only, ii) a broader range of assaults and iii) assaults and public nuisance offences. Hierarchical linear models (HLM) account for various sources of variability and correlation of longitudinal data and were used to determine reliability estimates for model parameters and in the calculation of the intraclass correlations (ICC). **Results** The broadest measure of alcohol-related crime (assaults and public nuisance offences) was found to have the highest reliability estimates between communities at a given time point and over time. This measure also had the highest ICC, indicating relatively more variability in the measure can be attributed to differences between towns rather than changes over time. **Conclusions** The HLM approach gives more accurate reliability estimates than could be assessed using a repeated measures ANOVA. For the communities from where these data derive, the broadest measure is the most reliable for comparing rates of alcohol-related crime between them, and for assessing intervention effects over time.

ACKNOWLEDGEMENTS

This work was supported by the Alcohol Education and Rehabilitation Foundation as part of the Alcohol Action in Rural Communities (AARC) project. AARC is registered with the Australian Clinical Trials Registry (ACTRNO12607000123448). Thanks to Laura Vogl for her assistance using HLM.

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Globally, alcohol accounts for 3.7% of all deaths and 4.4% of the burden of disease (Begg et al., 2007; World Health Organization, 2007). In Australia, 15% of this burden is directly attributable to alcohol-related crime at an estimated annual cost of $1.6 billion (Collins and Lapsley, 2008). Recent attempts to reduce these harms and costs have focused on strategically mobilising and co-ordinating resources at the community level rather than attempting to modify individuals' behaviours (Giesbrecht, 2007; Holder, 2000; Holder et al., 2000; Stafström and Larsson, 2007; Wallin, 2007). This has prompted the development of new evaluation designs appropriate for whole communities (Hawkins et al., 2007; Sanson-Fisher et al., 2007) and the use of routinely collected data suitable for longitudinal analyses.

Routinely collected data have a number of advantages over survey data: they are less expensive; they can be applied to defined communities by using post-code or local government area as the unit of analysis; they are not biased by non-consent (generally their use does not require individual consent); and they can be used retrospectively (Treno and Holder, 1997). The primary disadvantage of routinely collected data is there is little evidence for their validity (that they measure what they purport to measure) and reliability (their consistency over time or between groups/administrators), both of which are critical, especially for establishing the cost-benefit of government policies and public health interventions (Gruenewald et al., 1997; Shakeshaft et al., 1997; World Health Organization, 2000).

In Australia, data routinely collected by police provide a measure of alcohol involvement in crime with alcohol-related offences flagged as such in police databases. Although evidence suggests that routinely collected data underestimate the extent of alcohol involvement in crime, the difficulty of quantifying their validity is clearly evidenced by the lack of published validition studies (Ireland and Thommeny, 1993; McClelland and Teplin, 2001; Palk et al., 2007). The reliability of routinely collected data is also problematic: the decision to flag an incident as alcohol-related in the police database is a subjective judgement by an individual officer at a particular time and is influenced by policing practices (Brinkman et al., 2001). Establishing the reliability of measures of alcohol-related crime, and thus the true extent of differences between locations or over time, is challenging.

Surrogate or proxy measures are less reliant on individual judgement and have been widely used internationally to improve the reliability of measures of alcohol- related harm (Chikritzhs et al., 2000; Holder and Wagenaar, 1994; Matthews et al., 2002a; Wagenaar and Holder, 1991). These measures apply a consistent formula to routinely collected data, based on knowledge of harms. Night-time serious assaults, for example, have been used as a measure of alcohol-related crime because the majority of these are alcohol-related and they are reported more consistently by police, compared to less serious crimes, such as malicious damage (Matthews et al., 2002a). Despite the wide-spread use of proxy measures of alcohol-related crime, there do not appear to by any published assessments of their reliability, making it difficult to separate real, as opposed to artefactual, changes in alcohol-related crime over time or between locations.

Given that current proxy measures of alcohol-related crime use data are collected over time, their reliability could be quantified using repeated measures analysis of variance. This approach relies on assumptions that may not be applicable to longitudinal data: it does not model all possible sources of variability over time; it assumes observations conform to a specific correlation structure; and it assumes complete data for all observations (Laenen et al., 2009a). This approach also assumes the intercept and slope are fixed values that do not vary across units (eg. communities). Hierarchical linear modelling (HLM) regression analyses (also called random effects regression or multi-level modelling) provide viable alternatives that have the potential to more efficiently account for various sources of variability and correlation in repeated measurements (Laenen et al., 2009a; Mujahid et al., 2007).

Accounting for the variability in intercept and slope allows for more accurate estimates of change over time, or between groups. A greater understanding of where the variability in measurement occurs can indicate the utility of a measure appropriate for a particular purpose. For example, a measure with little variability between communities may be useful for measuring an intervention effect, while measures with relatively more variation across communities might be useful for identifying the factors that contribute to differences between communities.

HLM analyses are increasingly being used for longitudinal data (Cheng et al., 2009) and can provide valuable information on the suitability of measures including an assessment of their reliability.

1.1 Study Aims

This study assesses three alcohol-related crime measures within a linear growth model framework, to demonstrate the utility of the approach and ascertain which is the most suitable for identifying differences between communities and differences within communities over time.

2. METHODS

2.1 Communities and data source

This study uses routinely recorded police incident data from 20 rural communities in NSW involved in a randomised controlled trial (RCT) aimed at reducing alcohol harm, the Alcohol Action in Rural Communities (AARC) project (Breen et al., 2010). Communities were invited to participate if they had an urban-centre locality (UCL) population between 5,000 and 20,000 (n=27) (Australian Bureau of Statistics) were at least 100km away from a major urban centre (population ≥ 100,000) (n=24) and were not currently involved in another alcohol-related study (n=20). De-identified unit records of criminal incidents recorded routinely by police for January 2001 to December 2005 (prior to the commencement of any intervention) were provided by the NSW Bureau of Crime Statistics and Research (Bureau of Crime Statistics and Research New South Wales). Incidents were selected on the basis of the postcode in, and the date on, which they occurred. Given criminal incidents were selected on the basis of the postcode where the incident occurred, per capita rates were standardised using Postal Area Population (see Table 1), even though communities were selected for the RCT based on their UCL population.

2.2 Definition of alcohol-related crime measures

In order to find the optimal balance between including a range of crimes identified as often involving alcohol and maximising reliability, three measures were assessed. Crimes known to often involve alcohol were included. For example, previous research showed that 77% of street incidents (assault, offensive behaviour, offensive language, malicious damage, domestic violence, noise complaint and drink driving) involved alcohol (Ireland and Thommeny, 1993).

Although driving offences involving alcohol that are recorded in police incident data have been included in previous research (Matthews et al., 2002a) they were excluded from this study due to changes in recording practices in NSW in 2003 and because similar data are available from NSW Road Traffic Authority (Czech et al., 2010).

Measure one: serious assault

Assaults are used in proxy measures of alcohol-related crime because they are relatively common and often involve alcohol (Briscoe and Donnelly, 2003; English et al., 1995; Ireland and Thommeny, 1993). Limiting the measure to serious assaults reduces the number of incidents that are included but optimises inter-rater reliability, because the greater severity of serious assaults means they are recorded more consistently by police (Matthews et al., 2002a). Serious assaults included actual bodily harm and grievous bodily harm (including malicious wounding).

Measure two: assaults

Given serious assaults alone under-estimate the extent of alcohol involvement, this measure includes common assault and sexual assaults (including the sub-categories sexual offences; sexual assault, aggravated sexual assault, indecent assault or act of indecency, aggravated indecent assault or act of indecency). Sexual assaults have been excluded from previous national-level analyses because data were not available for all jurisdictions (Matthews et al., 2002a) but as data are available in NSW they were included in this measure.

Measure three: public nuisance

This measure was devised to capture a wider range of crimes that have been identified as often involving alcohol (Donnelly et al., 2007; Ireland and Thommeny, 1993). In addition to offences in the assaults and serious assaults measures, it includes malicious damage to property and street offences (offensive conduct, offensive language and wilful and obscene exposure).

2.3 Definition of alcohol and non alcohol-related times

Since proxy measures involve defined types of crimes occurring at defined times, times identified previously as likely to be alcohol-related in NSW were used in this study (Chikritzhs et al., 2000; Matthews et al., 2002a; Matthews et al., 2002b). Alcohol-related times are: Sun 10pm - Mon 6am, Mon 10pm - Tues 2am, Wed 10pm - Thurs 2am, Fri 10pm - Sat 6am and Sat 6pm - Sun 6am. Non alcohol-related times are: Mon 6am - Mon 6pm, Tues 6am - Tues 2pm, Wed 10am - Wed 2pm, Thurs 6am - Thurs 2pm, Fri 6am - Fri 10am. The alcohol and non alcohol time periods do not account for all time because incidents in some time periods are neither clearly alcohol, nor non-alcohol, related. Equal hours are covered, however, in alcohol and non-alcohol times.

2.4 Statistical methods

All data manipulation was conducted in SPSS and analysis performed in HLM 6.02 (Raudenbush et al., 2005).

Ratio method of analyses

Alcohol-related crime is estimated as the ratio of incidents occurring at alcohol times relative to the same incident type occurring at non-alcohol-related times. This ratio method minimises the impact of extraneous variables (such as police activity and population sizes) (Matthews et al., 2002b). The ratio is treated as a continuous variable with a normal distribution. An inspection of normal probability plots and the Shapiro-Wilk test (Shapiro and Wilkinson, 1965) found that for the majority of time points the ratios were normally distributed for all three measures. The ratios were calculated for each six-monthly interval (January-June, July – December) for 2001-2005. The ratios were plotted by community and examined for seasonal variation.

Hierarchical linear models and reliability estimates

The reliability of the crime ratios between communities and over time (as opposed to the within-community reliability of individuals) were calculated using hierarchical linear models. These models take into account the hierarchical nature of the data (multiple observations over time are nested within each community), account for the effect of time (data collected at temporally contiguous points are more likely to be similar than data at temporally divergent points) and allow for the explicit modelling of between-community heterogeneity (i.e. they account for the likelihood that alcohol-related crime will differ between communities at the initial time point and/or in their rate of change over time). The modelling of between community heterogeneity is achieved by including both fixed and random effects, representing community-specific intercepts and community-specific trajectories in the models. Parameter estimation was based on restricted maximum likelihood.

Each community was represented, at level 1 in the HLM analysis, by an individual trajectory. These trajectories become outcome variables in a level-2 model, where they may depend on community-level characteristics. The current analysis is focussed on the unconditional growth model (i.e. the model that is not conditioned on any measured characteristics of the communities). For each measure a random intercept only model and a random intercept and random slope model were examined.

For each of the three measures, the variation in crime ratios is partitioned into variation between communities at a given time point and variation between communities over time (expressed as correlation coefficients). In addition, an overall reliability estimate was calculated using formula outlined in Raudenbush and Bryk (Raudenbush and Bryk, 2002) (pg. 166). For the random intercept only model, an intraclass correlation (ICC) was calculated. The ICC quantifies how much of the total variability in each measure is accounted for by variability between communities.

The models used notation and methods employed by Raudenbush and colleagues (Raudenbush and Sampson, 1999; Raudenbush and Bryk, 2002). The level 1 model is:

$$(RATIO)_{ti} = \pi_{0i} + \pi_{1i}(TIME)_{ti} + e_{ti}$$

This model indicates that each community's crime ratio is a function of crime at the first time point (this is the intercept and is specified by π_{0i}), the linear effect of time on crime (this is the coefficient associated with TIME and specified by π_{1i}), plus an error term representing the unique contribution of each community at each time point (this is specified by e_{ti} and is assumed to be normally distributed with mean 0 and variance σ^2).

The level 2 model incorporates the random effects for the intercept and the slope:

$\pi_{0i} = \beta_{00} + r_{0i}$ for the intercept and

$\pi_{1i} = \beta_{10} + r_{1i}$ for the slope

This model contains terms representing the average crime ratio intercept for all communities (β_{00}) and the average crime ratio slope for all communities (β_{10}). These parameters are associated with a certain amount of error. The variability in any given community's intercept around the average community intercept (r_{0i}) is assumed to be normally distributed with mean zero and variance τ 00. The variability in any given community's slope around the average community slope (r_{1i}) is also assumed to be normally distributed with mean zero and variance τ 11. For the random intercept only model the variability in the slope (r_{1i}) is assumed to be fixed.

The level 1 and level 2 models are combined:

$$(RATIO)_{ti} = [\beta_{00} + \beta_{10}(TIME)_{ti}] + [r_{0i} + r_{1i}(TIME)_{ti} + e_{ti}]$$

The set of parameters within the first square brackets represent the fixed effects and those within the second represent the random effects. The fixed effects ("mean initial crime ratio") represent the mean crime ratio for all 20 communities at Time 0 (Jan-Jun

2001). The "mean growth rate" represents the expected change in crime ratios for a fixed unit of time (six months). The random effects represent the amount of variation across all 20 communities in the initial crime ratio and the amount of variation across all 20 communities in the growth rates.

Reliability of the estimated random effect parameters was calculated. Statistically, reliability was interpreted in the language of classical test theory as "the ratio of true score or parameter variance, relative to the observed score or total variance" (Raudenbush & Bryk, 2002, p.46). More conceptually, reliability is a function of two aspects: variability between communities; and variability over time within each community. Variability between communities is a function of the degree to which the parameter estimates (i.e. the intercept or growth parameters) vary from community to community. The more the estimates vary (holding constant sample size) the greater the reliability. Variability over time within each community is a function of the precision with which each community's regression equation is estimated where precision relates to the residual variance (the amount of deviation of each community's observed scores around a fitted regression line). Precision increases with increasing sample size. These components of reliability are related and cannot be assessed in isolation (Laenen et al., 2009b).

Reliability is estimated separately for the intercept and growth random effects. For the intercept random effect reliability is estimated as the proportion of total variance in estimated intercept parameter values that is explained by differences between communities and is estimated from the regression models that contain both a random intercept and random slope.

There are ten separate time points and ten separate reliabilities were calculated for the intercept random effect parameter by centring the TIME variable at each of the ten time points, and re-estimating the intercept random effect parameters included in the model for each TIME variable. The reliability of the growth random effect parameter does not change with the use of different TIME variables.

3. RESULTS

3.1 Community characteristics

The population size for the twenty communities ranged from 6571 to 29005. The communities included in the analysis had SEIFA decile scores ranging from 1 to 6, indicating scores in the bottom 60% of NSW. Descriptive information on the communities is presented in Table 1.

Table 1: Descriptive statistics for community characteristics of the 20 communities

Measures	Minimum	Maximum	Mean	Standard deviation
Population[1]	6571	29005	14106	6397
Proportion young males[1]	0.05	0.07	0.06	0.01
Proportion indigenous[1]	0.01	0.17	0.05	0.04
Remoteness indicator (ARIA score)	0.98	7.72	2.90	1.47
Socioeconomic indicator[1,2] (SIEFA disadvantage decile)	1	6	3.40	1.31
Number of pubs /clubs [3]	5.41	19.04	10.58	3.88
Number of wholesalers/retailers [3]	0.89	9.37	3.37	2.12
Number of other licensed [3]	7.05	26.62	14.03	5.43
Number of police officers [4]	2.70	44.09	17.94	10.02
Number of highway patrol officers [4]	0	6.94	2.96	1.80
Number of General Practitioners [5]	6.17	36.7	11.13	6.73

1. From 2001 Census Postal Area population (ABS).
2. Low scores for SEIFA indicate high levels of disadvantage.
3. Number per 10 000 population. Information published by the Office of Gaming and Racing (Gaming and Racing A NSW Government Department, 2004).
4. Number per 10 000 population. Compiled from information obtained through Local Area Commands.
5. Number per 10 000 population. Compiled from information obtained through Divisions of General Practice.

3.2 Crime ratio

The alcohol-related crime ratio for each community at each six monthly time period is presented for each measure in Figures 1-3. A ratio of 2.2, for example, indicates that 2.2 crimes occurred in an alcohol-related time for every crime in a non alcohol-related time.

Figure 1: Serious assaults ratio for each community, 2001-2005

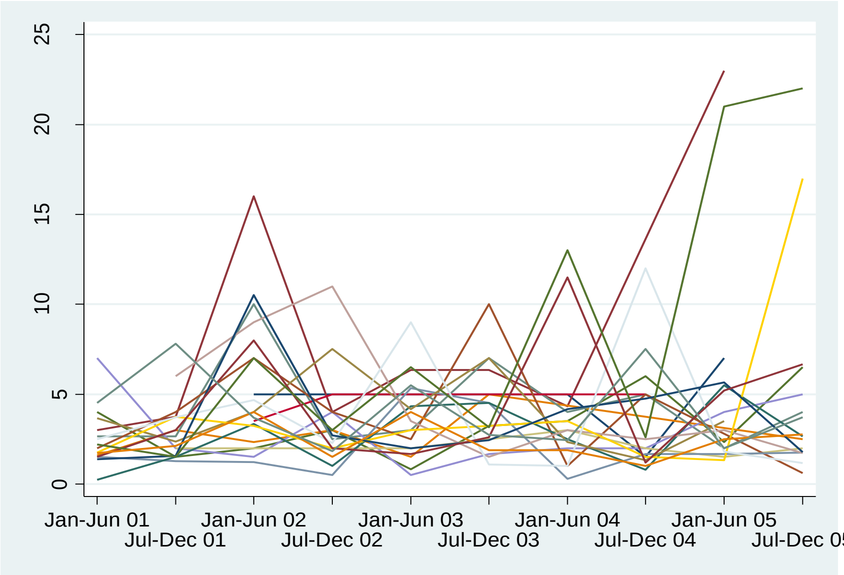


Figure 1 shows that there is more variability within and between communities in alcohol-related crime ratios using the serious assaults measure, compared to the assaults (Figure 2) and public nuisance (Figure 3) measures.

Figure 2: Assaults ratio for each community, 2001-2005

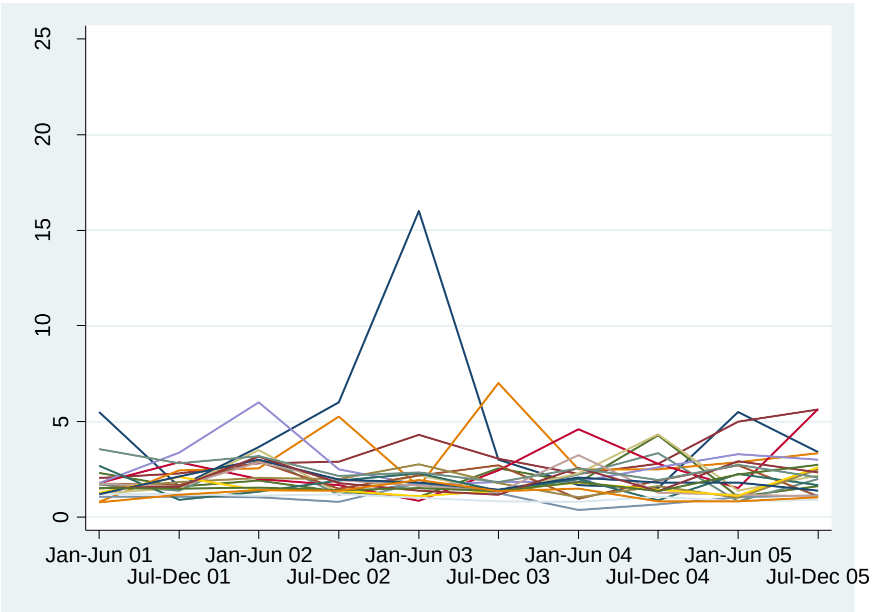
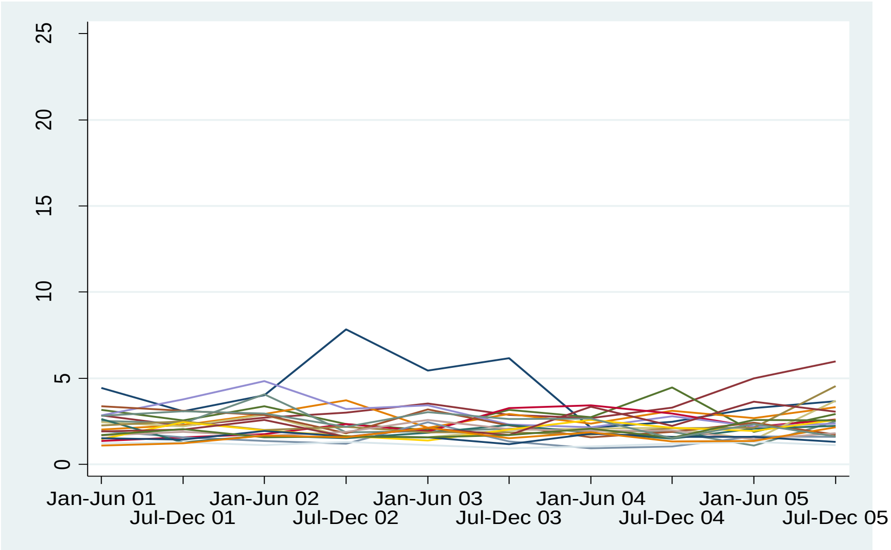


Figure 3: Public nuisance ratio for each community, 2001-2005



3.3 Regression models

Table 2 summarises estimates of fixed and random effects for the random intercept and random slope models three measures.

Table 2: Hierarchical linear regression models for three different measures of alcohol-related crime at Time 0 (January – June 2001).

	Serious Assault	**Assault**	**Public Nuisance**
Fixed Effects	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**
Mean initial crime ratio	3.13 (0.46)**	2.04 (0.24)**	2.26 (0.21)**
Mean growth rate	0.21 (0.13)	0.02 (0.03)	0.01 (0.03)
Random Effects	**Variance component, df, chisq**	**Variance component, df, chisq**	**Variance component, df, chisq**
Initial crime ratio	0.36, 19, 17.80	0.56, 19, 31.97*	0.74, 19, 110.05**
Growth rate	0.22, 19, 49.48**	0.00008, 19, 11.72	0.01, 19, 53.14**
Residual error	9.59	1.66	0.45

* p<0.05, ** p<0.001
The covariance structure was modelled as unstructured (modelling the structure as compound symmetry or autoregressive did not substantially improve the fit)
Quadratic time effects were also tested but were non-significant for all three measures.
The significant mean initial crime ratio indicates significantly different from zero.

The serious assault measure has the highest mean initial alcohol-related crime ratio (i.e. at time 0) (3.13) but also the largest standard error (0.46), followed by the public nuisance (2.26) and assault measures (2.04). The public nuisance measure has the lowest standard error of the three measures (0.21).

The mean growth estimates show the communities' alcohol-related crime ratios increased every six months by an average of 0.21 (serious assaults), 0.02 (assaults) and 0.01 (public nuisance). The increases were not statistically significant for all three measures, indicating that the estimates of alcohol-related crime produced by each measure remain relatively stable over time.

The random effects results (Table .2) demonstrate that the communities vary significantly in their initial crime ratio for both the assault and the public nuisance measures, indicating that for these measures, there is significant variation between communities in these measures relative to the total amount of variation in these measures. The communities also vary significantly in how the crime ratios change over time (growth rates) according to both the serious assaults and the public nuisance measures of alcohol-related crime. These results suggest that while, on average, there is no significant growth in alcohol-related crime over time the individual community crime trajectories (i.e. the growth in crime over time for each community) are substantially different from one another.

Table 3 provides the reliability estimates for both the estimated initial crime ratio and growth rate parameters for the random intercept and random slope model. The range of reliabilities for the estimated initial crime ratio parameter for the ten time points is also presented.

Table 3: Random intercept and random slope model: Reliability estimates for three different measures of alcohol-related crime at Time 0 (January -June 2001).

	Serious Assault	Assault	Public Nuisance
Reliability			
Initial crime ratio parameter	0.08	0.49	0.83
(range [1])	(0.01-0.77)	(0.49-0.78)	(0.79-0.91)
Growth rate parameter	0.59	0.004	0.64

[1] The range represents the range in reliability estimates obtained for the ten time points

At Time 0, the initial crime ratio is highest for the public nuisance measure (0.83) and lowest for the serious assault measure (0.08). The growth rate estimate is poorest for the assault measure (0.004) and highest for the public nuisance measure (0.64)

In addition to Table 3, reliability estimates were obtained for each six monthly period. The initial crime ratio reliability estimates vary for all three measures at every time point. The reliability estimates range from 0.01 -0.77 for the serious assault measure, 0.49-0.78 for the assault measure and 0.79-0.91 for the public nuisance measure. The lowest reliability estimate for the public nuisance measure is greater than the highest reliability estimate of the other two measures.

When considering both the initial crime ratio and growth rate estimates for reliability, the results suggest the public nuisance measure is most reliable at measuring initial crime estimates (i.e. the ratio at any time point - reliability estimate 0.79-0.91) and over time (0.64).

Table 4 provides the reliability estimates and the intraclass correlation coefficients for the random intercept only model, where the variability of the slope is assumed to be fixed.

Table 4: Random intercept only model: Reliability estimates and ICC for three different measures of alcohol-related crime for all time points (January -June 2001).

	Serious Assault	Assault	Public Nuisance
Reliability	0.54	0.78	0.90
ICC	0.13	0.27	0.46

The reliability estimates are highest for the public nuisance measure (0.90) and lowest for the serious assault measure (0.54). The ICC is highest for the public nuisance measure (0.46) indicating that more of the total variability is accounted for by community variability.

4. DISCUSSION

This is the first study to assess the suitability of proxy measures of alcohol-related crime, both at any given time between groups and between groups over time. The broadest measure, the public nuisance measure, was the most reliable measure of alcohol-related crime at the community level and is considered to be the most useful to examine differences between location and over time. The serious assaults measure had the greatest residual variance while the public nuisance measure had the least. The public nuisance measure also had the highest reliability for initial crime ratios, that is, the ratio at any time point (0.79-0.91) and growth rates (0.64). The public nuisance measure also has the largest ICC, indicating it may be useful to examine difference between communities.

Using routinely collected police data to measure alcohol-related crime was feasible for these communities, and is likely to be applicable to larger communities. Previous research at the state and national level focussed on proxy measures using only serious assaults (Matthews et al., 2002a). The current study indicates that, at the local community level, there are likely to be too few serious assaults for this measure to be reliable over time and location. The assault measure, which incorporates a wider range of assault offences, does not provide a more reliable estimate than serious assaults. Conversely, the public nuisance measure was the most reliable and demonstrated substantial signal in the data (indicated by community differences in initial crime ratios and growth ratios), meaning that modelling community differences is an appropriate method of identifying community-level characteristics associated with different rates of alcohol-related crime and evaluating community-level intervention and public policy effects.

Reliability is not a fixed property of a measure but is population dependent, with more homogenous populations leading to lower reliabilities. Laenen and colleagues (Laenen et al., 2009a; Laenen et al., 2009b), found reliability was lower with more complex scenarios (such as in longitudinal research) than previously reported and stress the importance of studying reliability within a longitudinal setting using longitudinal data. Research suggests that for data with additional interaction terms, such as intervention or treatment effects, linear mixed models could be used to ascertain reliability taking into account the more complex data structure (Laenen et al., 2009b). The current study has focussed on the unconditional growth model to ascertain the most reliable measure, however, future analyses could include an intervention effect.

A reliable measure is population dependent (Laenen et al., 2009b) and may not be reliable for all purposes. Different aspects of HLM analyses can be examined to assess the utility of measures, depending on the purpose of which the measure is intended. In this study a measure with larger reliability estimates and larger ICC is considered more useful than a measure with smaller reliability and ICC estimates, because the primary purpose was to identify a measure to examine differences between communities. If a measure was being sought to examine intervention effects, a measure with a lower ICC may be considered more useful, as a measure with little variability provides greater statistical power to detect a difference.

4.1 Methodological considerations

HLM analyses, which account for variability over time, are appropriate for assessing measures designed to examine trends over time. In this study, HLM were used to decompose variance in the models to facilitate comparative analysis of the three measures, rather than the traditional purpose of examining hierarchical data. Although the generalisability of these results is restricted by the sample size of 20 communities, data for ten time points were analysed allowing for detailed examination of variance over time. In any case, the reliability of a measure should be checked when applied to different populations (Laenen et al., 2009b).

This study used incidents that were reported to, or detected by, police but arrest data could be used if available. Arrest data have been recommended over police calls or convictions as calls reflect the subjective view of a criminal act and convictions reflect legal manoeuvring (Gruenewald et al., 1997). There may be insufficient arrest data at the community level to ensure reliable estimates over time, given arrest data only account for a small proportion of crime.

4.2 Surrogates or proxy measures

Distinctions must be made between the use of data to estimate the magnitude of a problem at one point and change over time (World Health Organization, 2000). A surrogate, although not intended for estimating prevalence, is useful in examining trends and intervention effects over time and location, as it has the ability to reduce measurement error and can be based on data collected routinely at relatively low cost.

Issues of detection bias with the proxy measure approach remain. The inclusion of crimes that are influenced by police activity can result in detection bias. The selection of offence type for inclusion in the proxy may impact the reliability of the measure. In this study the inclusion of additional offences improved the reliability of the measure, although the extent of detection bias is unknown. Applying the same approach to different communities or over time minimises relative differences.

The recent introduction of the automated flag for alcohol-related police incidents in NSW may provide more reliable community level data for examination over time (Wiggers, 2007). It is not known if the automated approach or the proxy measure is more reliable, but the proxy approach is useful in other states and internationally where alcohol incidents are not automatically flagged. In addition, when comparisons between states or countries are required the proxy approach may be necessary.

4.3 Conclusion

The need for indicators of the health and the social impact of alcohol use has been specifically identified in the National Preventative Health Strategy in Australia (Australian Government Department of Health and Ageing, 2009). This study provides a method for developing and testing the reliability of estimates of population measures for alcohol-related crime. HLM is an innovative and rigorous method of measuring the reliability of measures used in longitudinal research (Laenen et al., 2009b). The HLM approach used in this study gives more accurate reliability estimates than could be assessed using a repeated measures ANOVA, as HLM models take into account the variability in the intercept and slope.

For the communities from which the data used in this analysis derive, the public nuisance measure is the most appropriate for determining whether community characteristics account for different rates of alcohol-related crime in communities, such as population levels of consumption or numbers of police, and for use as an outcome measure in determining the effectiveness of interventions aimed at reducing alcohol-related crime. Applying HLM to assess the utility of alcohol harm data from complementary sources, such as emergency departments and traffic accidents, would enable more comprehensive analyses of the deleterious impact of alcohol on communities and more precise measurement of the impact of community-level interventions and policy changes.

5. REFERENCES

Australian Bureau of Statistics Census Data by location 2001. http://www.abs.gov.au/ Accessed 2005. http://www.abs.gov.au/ accessed on 2005>.

Australian Government Department of Health and Ageing (2009). Australia: The Healthiest Country by 2020. National Preventative Health Strategy – the roadmap for action. Australian Government Department of Health and Ageing, Canberra.

Begg, S., Vos, T., Barker, B., Stevenson, C., Stanley, L., Lopez, A. (2007). The burden of disease and injury in Australia, 2003. In: AIHW (Ed.), PHE82. Australian Institute of Health and Welfare, Canberra.

Breen, C., Shakeshaft, A., Doran, C., Sanson-Fisher, R., Mattick, R. (2010). The cost-effectiveness of postal surveys in improving survey response rates: a randomised controlled trial. *Australia and New Zealand Journal of Public Health;* 34, 508-512.

Brinkman, S., Chikritzhs, T., Stockwell, T., Mathewson, P. (2001). An indicator approach to the measurement of alcohol-related violence. In: Williams, P. (Ed.), Alcohol, young persons and violence. Australian Institute of Criminology, Canberra.

Briscoe, S., Donnelly, N. (2003). Problematic licenses premises for assault in inner Sydney, Newcastle and Wollongong. *The Australian and New Zealand Journal of Criminology;* 36, 18-33.

Bureau of Crime Statistics and Research New South Wales Definitions and explanatory notes for recorded crime statistics. http://www.bocsar.nsw.gov.au.,accessed on Jan 2009.

Cheng, J., Edwards, L., Maldonado-Molina, M., Komro, K., Muller, K. (2009). Real longitudinal data analysis for real people: Building a good enough mixed model. *Statistics in Medicine;* 29, 504-520.

Chikritzhs, T., Stockwell, T., Heale, P., Dietze, P., Webb, M. (2000). National Indicators Project Technical Report No. 2: Trends in alcohol-related road injury in Australia, 1990-1997. National Drug Research Institute, Perth.

Collins, D.J., Lapsley, H.M. (2008). The costs of tobacco, alcohol and illicit drug abuse to Australian society 2004/05. National Drug Strategy Monograph Series No 64. Department of Health and Ageing, Canberra.

Czech, S., Shakeshaft, A., Byrnes, J., Doran, C. (2010). Comparing the cost of alcohol-related traffic crashes in rural and urban environments. *Accident Analysis and Prevention;* 42, 1195-1198.

Donnelly, N., Scott, L., S., P., Weatherburn, D., Shanahan, M., Hansen, F. (2007). Estimating the short-term cost of police time spent dealing with alcohol-related crime in NSW. Monograph Series No 25. National Drug Law Enforcement Research Fund, Hobart.

English, D., Holman, C., Milne, E., Winter, M., Hulse, G., Codde, J., Bower, C., de Klerk, N., Knuiman, M., Kurinczuk, J., Lewin, G., Ryan, G. (1995). The quantification of drug caused morbidity and mortality in Australia, 1995 edition. Commonwealth Department of Human Services and Health, Canberra. pp. 227-230.

Gaming and Racing A NSW Government Department (2004). Index of New South Wales licensed premises and registered clubs 2004. NSW Government.

Giesbrecht, N. (2007). Community-based prevention of alcohol problems: Addressing the challenges of increasing deregulation of alcohol. *Substance Use & Misuse;* 42, 1813-1834.

Gruenewald, P., Treno, A., Taff, G., Klitzner, M. (1997). Measuring community indicators: a systems approach to drug and alcohol problems. Sage Publishers, London.

Hawkins, N., Sanson-Fisher, R., Shakeshaft, A., D'Este, C., Green, L. (2007). The multiple baseline design for evaluating population-based research. *American Journal of Preventive Medicine;* 33, 162-168.

Holder, H. (2000). Community prevention of alcohol problems. *Addictive Behaviours;* 25, 843-859.

Holder, H., Gruenewald, P., Ponickie, W., Treno, A., Grube, J., Salz, R., Voas, R., Reynolds, R., Davis, J., Sanchez, L., Gaumont, G., Roeper, P. (2000). Effect of community-based interventions on high risk drinking and alcohol-related injuries. *Journal of the American Medical Association;* 284, 2341-2347.

Holder, H., Wagenaar, A.C. (1994). Mandated server training and reduced alcohol-involved traffic crashes: A time series analysis of the Oregon experience. *Accident Analysis and Prevention;* 26, 89-97.

Ireland, C., Thommeny, J. (1993). The crime cocktail: Licensed premises, alcohol and street offences. *Drug and Alcohol Review;* 12, 143-150.

Laenen, A., Alonso, A., Molenberghs, G., Vangeneudden, T. (2009a). Reliability of a longitudinal sequence of scale ratings. *Psychometrika;* 74, 49-64.

Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T., Mallinckrodt, C.H. (2009b). Using longitudinal data from a clinical trial in depression to assess the reliability of its outcome scales. *Journal of Psychiatric Research;* 43, 730-738.

Matthews, S., Chikritzhs, T., Catalano, P., Stockwell, T., Donath, S. (2002a). Trends in alcohol-related violence in Australia, 1991/92-1999/00. National Drug Research Institute, Perth.

Matthews, S., Donath, S., Chikritzhs, T., Catalano, P., Stockwell, T. (2002b). National Alcohol Indicators Project Technical Report No. 5. Trends in alcohol-related violence in Australia, 1991/92-1999/00 National Drug Research Institute, Perth.

McClelland, G., Teplin, L. (2001). Alcohol intoxication and violent crime: Implications for public health policy. *The American Journal on Addictions;* 10, 70-85.

Mujahid, M., Diez Roux, A., Morenoff, J., Raghunathan, T. (2007). Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *American Journal of Epidemiology;* 165, 858-867.

Palk, G., Davey, J., Freeman, J. (2007). Prevalence and characteristics of alcohol-related incidents requiring police attendance. *Journal of Studies on Alcohol and Drugs;* 68, 575-581.

Raudenbush, S., Bryk, A., Congdon, R. (2005). Hierarchical Linear and Nonlinear Modelling. Scientific Software International.

Raudenbush, S., Sampson, R. (1999). Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observations of neighborhoods. *Sociological Methodology;* 29, 1-41.

Raudenbush, S.W., Bryk, A.S. (Eds.), (2002). Hierarchical Linear Models: Applications and data analysis methods Sage, Thousand Oakes.

Sanson-Fisher, R., Boneveski, B., Green, L., D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine;* 33, 155-161.

Shakeshaft, A., Bowman, J., Sanson-Fisher, R. (1997). Behavioural alcohol research: new directions or more of the same? *Addiction;* 92, 1411-1422.

Shapiro, S.S., Wilkinson, M.B. (1965). An anlaysis of variance test for normality (complete samples). *Biometrika;* 52, 591-611.

Stafström, M., Larsson, S. (2007). The Trelleborg Project: A process evaluation of a multi-sector community intervention to reduce alcohol consumption and related harm. *Substance Use & Misuse;* 42, 2041-2051.

Treno, A., Holder, H. (1997). Measurement of alcohol-involved injury in community prevention: The search for a surrogate III. *Alcoholism: Clinical and Experimental Research;* 21, 1695-1703.

Wagenaar, A., Holder, H. (1991). Effects of alcoholic beverage server liability on traffic crash injuries. *Alcoholism: Clinical and Experimental Research;* 15, 942-947.

Wallin, E. (2007). Dissemination of prevention: Community action targeting alcohol use-related problems at licensed premises. *Substance Use and Misuse;* 42, 2085-2097.

Wiggers, J. (2007). Reducing alcohol-related violence and improving community safety: The alcohol linkage program. *NSW Public Health Bulletin;* 18, 83-85.

World Health Organization (2000). International guide for monitoring alcohol consumption and related harm. World Health Organization, Department of Mental Health and Substance Dependence, Noncommunicable Diseases and Mental Health Cluster, Geneva.

World Health Organization (2007). WHO Expert Committee on Problems Related to Alcohol Consumption. Technical Report series no 944. WHO, Geneva.