LETTER TO THE EDITOR

Assessing Reliability of Measures Using Routinely Collected Data

Courtney Breen1,*, Anthony Shakeshaft1, Tim Slade1, Catherine D’Este2 and Richard P Mattick1

1National Drug and Alcohol Research Centre, School of Public Health and Community Medicine, University of New South Wales, Sydney, NSW 2052, Australia and 2Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, King St, University of Newcastle, Callaghan, NSW 2308, Australia

*Corresponding author. Tel.: +61-0-2-9385-0282; Fax: +61-0-2-9385-0222; E-mail: courtney.breen@unsw.edu.au

(Received 14 March 2011; in revised form 30 March 2011; accepted 30 March 2011)

Recent attempts to reduce alcohol-related harms have focused on strategically mobilizing and co-ordinating resources at the community level, to complement attempts to modify individuals’ behaviours (Holder, 2000; Holder et al., 2000; Giesbrecht, 2007; Stafström and Larsson, 2007; Wallin, 2007). This has prompted the development of evaluation designs appropriate for whole communities and renewed interest in the use of routinely collected data to measure community intervention effects (Hawkins et al., 2007; Sanson-Fisher et al., 2007).

Routinely collected data are advantageous for assessing community-level intervention effects and monitoring changes over time, or between communities, as they: are low cost; can be defined by post-code or local government area; are not biased by non-consent (generally their use does not require individual consent); and can be used retrospectively. Routinely collected data is that 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 in separating real from artefactual changes over time, which allows more accurate quantification of the cost–benefit of government policies and public health interventions (Gruenewald et al., 1997; Shakeshaft et al., 1997; World Health Organization, 2000).

In attempting to minimize problems associated with using routinely collected data, surrogate or proxy measures have been developed and used internationally to examine alcohol-related harms (Wagenaar and Holder, 1991; Holder and Wagenaar, 1994; Chikritzhs et al., 2000; Matthews et al., 2002). 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 with less serious crimes, such as malicious damage (Matthews et al., 2002).

The reliability of routinely collected data needs to be assessed separately for different populations because reliability is not a fixed property of a measure but is population-dependent: more homogeneous populations engender less reliable measures (Laenen et al., 2009b). Further, where routinely collected data are used longitudinally, reliability needs to take into account that contiguous data points are correlated, in part because they are not independent and in part because a stable measure should be consistent over time (analogous to test/re-test reliability in psychometric testing). Accurate estimates of community-level intervention effects clearly need to account for this correlation.

Although repeated measures analysis of variance could be used to account for correlated data, it is limited: it 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). An alternative method that is increasingly used for longitudinal data (Cheng et al., 2009) is hierarchical linear modelling (HLM) regression analyses (or random effects regression or multi-level modelling), which accounts for a number of different sources of variability and correlation in repeated measurements (Mujahid et al., 2007; Laenen et al., 2009a).

The particular advantage of HLM is that it models a number of sources of variability over time. Specifically, rather than assuming that the intercept and slope in a regression equation are fixed values, HLM models the observed variability in these regression parameters. In addition, HLM provides information on the precision with which each unit’s regression equation is estimated (i.e. the variability over time for each unit). The variability of regression parameter estimates and the precision are all related components of reliability and cannot be assessed in isolation (Laenen et al., 2009b).

A recent community-wide randomized controlled trial aimed at reducing alcohol harm in 20 communities in NSW, Australia, the Alcohol Action in Rural Communities (AARC) project, provided an opportunity for an applied demonstration of the value of assessing the reliability of different measures of alcohol-related crime derived from routinely collected data using HLM. Although a proxy measure of alcohol-related crime based on national data was identified (i.e. serious assaults; Matthews et al., 2002), such a narrow definition was unlikely to provide a sufficient number of incidents in each community. Consequently, two alternative possible measures were derived: (i) a broader range of assaults and (ii) assaults and public nuisance offences. HLM was used to identify which measure produced the most precise set of co-efficients between the AARC communities, the detail of which is specified elsewhere (Breen et al., 2011a). The measures were assessed using an unconditional linear growth model, although future analyses could include an intervention effect. 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 intra class correlation, indicating relatively more variability in the measure can be attributed to differences between towns rather than changes over time. For the communities from where these data derive, the broadest measure is the most reliable for comparing alcohol-related crime between them (Breen et al., 2011b), and for assessing intervention effects over time.

As the use of proxy measures derived from routinely collected data increase in popularity, it is critical that their reliability be assessed to ensure that the most reliable and statistically efficient measures are used, which will vary depending on the specific purpose of the study. Using the example of routinely collected alcohol-related crime data relevant to whole communities, these analyses show that HLM is an innovative and rigorous method of measuring the relative reliability of different surrogate measures based on routinely collected data, providing an objective basis on which to choose between different measures that are otherwise equally acceptable.

Similar analyses could be undertaken for a range of data sources, including hospital admissions and traffic accident data.

**Funding** — 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 (ACTRN012607000123448).

**REFERENCES**


