VERSITA
EMERGING SCIENCE PUBLISHERS

NAD

Research report

Was the STAD programme really that successful?

THOR NORSTRÖM & BJÖRN TROLLDAL

ABSTRACT

AIM – A community intervention programme STAD was launched in Stockholm in January 1998, which included training in responsible beverage service and stricter enforcement of existing alcohol laws. An evaluation suggested that during the first 33 months of the programme, the level of police-recorded violence dropped by a striking 29%. We propose to probe the robustness of this estimate, which is often cited as evidence of the effectiveness of these kinds of intervention. In this paper, we reanalyse the underlying data by applying alternative model specifications. DATA AND METHODS – We reanalysed the original data on police-recorded violence from January 1994 to September 2000 by autoregressive integrated moving average (ARIMA) modelling. We estimated models based on raw data and seasonally differenced data; we also varied the definition of control area and applied the statistical technique of difference-in-differences modelling. RESULTS – The estimated intervention effects from these model specifications were all strongly significant statistically, ranging between 21% and 32%. CONCLUSION – Estimates based on a variety of model specifications were generally somewhat lower than those previously reported. However, the new estimates were all strongly statistically significant and fairly uniform with regard to effect size, which suggests that the findings of a substantial impact of the STAD programme are indeed quite robust.

KEY WORDS – alcohol, on-licensed premises, assaults, violence, RBS programme, community intervention, ARIMA, difference-in-differences

Submitted 27.09.2012 Final version accepted 19.12.2012

Background

In the late 1990s, the STAD project in the Swedish capital – Stockholm Prevents Alcohol and Drug Problems – developed a multi-component Responsible Beverage Service (RBS) programme to reduce violence in and around bars. The evaluation of the STAD programme suggested that the programme had been relatively effective (Wallin, Norström, & Andreasson, 2003). This has prompted its subsequent adoption elsewhere in Sweden. In 2008, 253 or 90% of the Swedish municipalities claimed to have implemented the programme in full or in part (Trolldal, Brännström, Paschall, & Leifman, 2013). As this spread of the programme should rest on robust evi-

Acknowledgements
This study was funded by the Swedish Council for Working Life and Social Research (FAS, grant number 2009-1705).

dence, our paper undertakes a reanalysis of the data underlying the evaluation. As an introduction, we shall first briefly describe the STAD programme.

The STAD programme: a multi-component RBS scheme

The notion underlying multi-component RBS programmes is that alcohol consumption and alcohol-related problems are affected by a multitude of factors, such as social norms, law enforcement, alcohol availability (serving practices) and consumption patterns. The hypothesis is that the more factors are targeted in an intervention programme, the larger the potential effect (Holder, 1998). The STAD programme therefore comprised several components:

1) An RBS action group with representatives of various stakeholders (municipality; police; County Council; National Institute of Public Health; organisation for owners of on-licensed premises; union of employees at on-licensed premises; and selected owners of popular nightclubs). This group served as a forum for information and discussion.
2) A two-day RBS training session for employees at on-licensed premises, which covered the following issues: medical and social effects (including violence) of alcohol; Alcohol Act; drug problems at on-licensed premises; and conflict management.
3) Stricter enforcement of existing alcohol laws.

The programme was run in 1998–2001 (see Wallin et al., 2003 for a more detailed description of the STAD programme).

The review by Jones, Hughes, Atkinson and Bellis (2011) indicates that multi-component programmes may be effective, although the small number of well-designed evaluations renders this conclusion tentative. One example of an effective multi-component programme are the Safety Action Projects in Queensland, Australia (Homel, Carvolth, Hauritz, Mcilwain, & Teague, 2004), which led to marked reductions in aggression and violence. Another example is the Community Trial Project in the USA (Holder et al., 2000), where the findings indicate significant reductions in incidents of drinking, drunk driving and assaults.

Evaluation of the STAD programme

The intervention was evaluated by applying a time-series quasi-experimental design to assess the effect of the intervention on police-recorded violence occurring between 10pm and 6am. This was accomplished by comparing the violence trajectory in the intervention area during the pre- and post-intervention period with the development of violence in a control area. The intervention area comprised the Central City of Stockholm while a neighbouring district (Södermalm) constituted the control area. The data was analysed by autoregressive integrated moving average (ARIMA) modelling (Box & Jenkins, 2008). The result suggested that the programme yielded a statistically significant reduction in police-recorded violence of 29% (Wallin et al., 2003). A cost-effectiveness analysis (Månsdotter, Rydberg, Wallin, Lindholm, & Andréasson, 2007) of the programme concluded that “the monetary and human benefits have been considerable”.

However, a recent evaluation (Trolldal et al., 2013) of the sequels of the STAD programmes in Sweden shows markedly weaker effects of about 9%. This triggers the question whether the estimate reported by Wallin et al. (2003) was biased upwards. The kind of design that was used for evaluating the STAD programme is usually considered to be methodologically strong, which is indicated by the highest quality assessment rating (on a 3-level scale) of the evaluation in the review by Jones et al. (2011). However, even though the use of a control area is a sound device to control for extraneous factors, it is far from fool-proof. An obvious potential threat to the validity of the results is that the outcome (the violence rate in this case) in either the intervention or the control area is affected by some factor that is not common to both areas. Take, for example, the experiences reported by Norström and Skog (2005). In their evaluation of the Saturday opening of alcohol retail shops in Sweden in 2000, it appeared that one of the outcomes (drunk driving) was affected by an intensification of police surveillance in the intervention area during the post-intervention period. Further, there was a marked, and unexpected, increase in alcohol sales in part of the experiment area. More detailed analyses revealed that this was due to an extraordinary surge in sales in six outlets along the border to Norway. The increased alcohol sales were thus attributable to increased cross-border trade. Had these two mishaps not been detected, and corrected for, the outcome would have been much distorted. Turning to the evaluation of the STAD programme, it may be noted that Södermalm (the control area) exhibited an upward trend in violence during the post-intervention area. This may well reflect the influence of some factor common both to the intervention area and the control area, in which case it makes sense to include Södermalm as control. However, one cannot exclude the possibility that the increase of violence in the control area was due to some local factor. This raises the question to what degree the estimated intervention effect hinges on the violence trend in the control area.

A related concern pertains to the robustness of the estimated intervention effect. As most analysts of time-series data have experienced, estimates tend to be affected by model specifications; e.g., it may make a big difference whether the estimation is based on the raw data or the differenced data. Different statistical techniques may also yield different outcomes. For instance, Wicki and Gmel (2011), in their evaluation of an alcohol policy intervention in Geneva, reported marked differences in outcomes when they compared the type of modelling applied by Wallin et al. (2003) with the difference-in-differences modelling (Ashenfelter & Card, 1985). However, Wallin et al. (2003) do not discuss to what degree their estimate is sensitive to the choice of statistical techniques or to differences in model specifications. It thus seems warranted to probe the robustness of this estimate, as it is often cited as evidence of the effectiveness of these kinds of intervention (Månsdotter et al., 2007; Jones et al., 2011).

Aims

The specific aims of this article are to elucidate the following questions: 1) To what degree is the estimated intervention effect contingent on the trend in the control area? 2) How sensitive is the estimated intervention effect to changes in model specification, and to the choice of statistical techniques?

Data and methods

We analysed the same data set as Wallin et al. (2003), which consisted of police-recorded violence including the following offences: assaults, illegal threats and harassment, violence and threats targeted at officials (including police officers and doormen). The indicator comprised all such reported offences committed between 10pm and 6am in the intervention area, i.e., Stockholm Central City, and in the control area, i.e., Södermalm. The data spanned the period from January 1994 to September 2000. As the intervention began in January 1998, the pre-intervention period included 48 months and the intervention period covered 33 months.

Wallin et al. assessed the intervention effect by estimating the following model:

$$LnVE_t = a+b_1 lnVC_t+b_2 I_t+N_t$$

where VE and VC denote the violence indicator in the experimental and control area, respectively. Wallin et al. showed that the intervention intensity increased gradually over the intervention period (January 1998 to September 2000). We adopted the same coding of the intervention variable (I), which thus took the value 0 before the intervention, and then gradually increased during the intervention period to attain the value 1 in the last intervention month. In August 1994, there was an annual Water Festival that was discontinued during the intervention period (the last festival was in August 1998). We controlled for this event in the same way as did Wallin et al., by including a dummy variable that was coded 1 for August 1999 and August 2000, and 0 otherwise. N is the noise term which captures other factors affecting the outcome. The parameter b2 expresses the intervention effect. The percentage effect is obtained from the expression:
(exp [b2] - 1) * 100.

As noted above, Wallin et al. used Södermalm as a control area. Figure 1 shows that Södermalm had an upward trend in violence during the post-intervention period (estimated at 0.8% (SE=0.3; p=.028) per month). If this resulted from a factor that only affected the control area but not the intervention area, the result may well be an over-estimation of the intervention effect. In our reanalyses, we estimated one model with no control area, and one model with an alternative neighbouring control area (Kungsholmen), which displayed a stationary trajectory in violence during the intervention period (estimated at 0.1% (SE=0.6; p=.92) per month). Further, Wallin et al. analysed the raw data, while our reanalyses also estimated the models on the basis of seasonally differenced data.

We applied the same technique for time-series analysis as Wallin et al., i.e., the method developed by Box and Jenkins (Box & Jenkins, 2008), often referred to as ARIMA-modelling (autoregressive integrated moving average). A distinctive feature of this method is that the noise (error) term, which includes explanatory variables not considered in the model, is allowed to have a temporal structure that is modelled and estimated in terms of autoregressive or moving average parameters. The model residuals should not differ from white noise; this was tested using the Box-Ljung Q statistics. (SPSS 17.0 was used for this analysis.)

We also applied an alternative modelling technique usually referred to as difference-in-differences estimation (DiD) (Ashenfelter & Card 1985). This technique is often used by economists when evaluating policy reforms that have been implemented in a quasi-experimental fashion such as in only a geographically delimited part of a country (Angrist & Krueger, 1999). The DiD model is specified as follows:

$$V_{it} = a + b_1 E_i + b_2 T + b_3 I_i + N_{it}$$

where V is the violence indicator, and E is a dummy variable that captures possible differences between the experiment area (City) and the control area (Södermalm) before the intervention. It takes the value 1 in the experiment area, and 0 in the control area. T is a time period dummy that captures factors that would affect violence even in the absence of the intervention. It takes the value 1 in the intervention period and 0 in the pre-intervention period. I is the intervention variable; it takes positive values only in the experiment area during the intervention period (gradually increasing as described above), and 0 otherwise. In addition, we included a dummy variable for the August Water Festivals (coded as above). The parameter of interest is thus b_3, which captures the intervention effect. Bertrand, Duflo and Mullainathan (2004) observe that most applications of DiD do not consider that the residuals are likely to be autocorrelated with the concomitant risk of underestimating the standard error of the effect estimate. Thus, we consider the residual structure by including AR- and/or MA-parameters.

Results

Table 1 (see next page) displays the outcomes from the various model estimations. The first model accords with the model applied by Wallin et al., reproducing their estimated intervention effect and implying a reduction in violence of 29%. If the alternative control area (Kungsholmen) is used, the estimated intervention effect drops somewhat (to 24%), and if no control area is included (model 3), the effect is down to 21%. Seasonal differencing of the data (model 4-6) also tended to lower the estimated intervention effect. It is also seen that the estimate based on the DiD model was compatible with remaining estimates. Generally, none of the estimated intervention effects was significantly different from one another; the t-value for the most significant difference (the one between model 1 and model 3) was equal to 1.28.

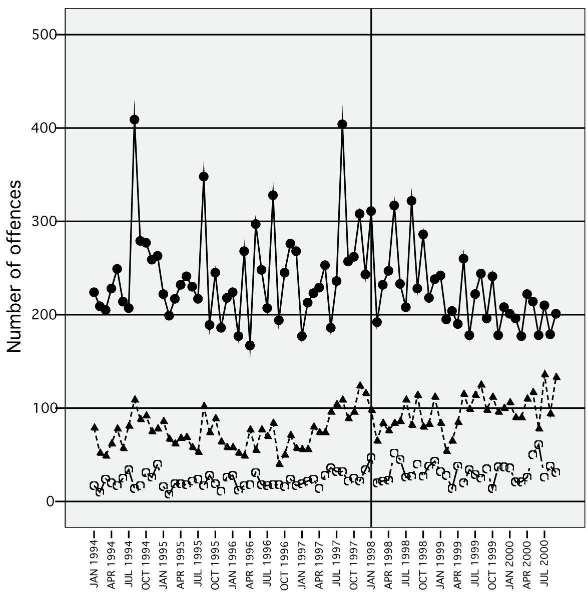

Figure 1. Violence in Stockholm Central City (filled circles), Södermalm (triangles) and Kungsholmen (open circles).

Table 1. Estimated intervention effect on police-recorded violence in experiment area for different model specifications

Model	Control area	Filter	Model	Intervention		Effect %	Control		Q
				Est	SE		Est	SE	
1	Södermalm	None	ARIMA (1,0,0)(2,0,0)	-0.344***	0.046	-29	0.313***	0.054	14.99, p>.24
2	Kungsholmen	None	ARIMA (1,0,0)(2,0,0)	-0.280***	0.059	-24	0.099*	0.047	15.03, p>.24
3	None	None	ARIMA (0,0,0)(2,0,0)	-0.242***	0.065	-21			18.82, p>.09
4	Södermalm	SD	ARIMA (0,0,0)(2,1,0)	-0.296***	0.065	-26	0.196**	0.065	10.04, p>.61
5	Kungsholmen	SD	ARIMA (0,0,0)(2,1,0)	-0.242***	0.068	-21	0.094(*)	0.047	9.26, p>.68
6	None	SD	ARIMA (0,0,0)(2,1,0)	-0.248**	0.069	-22			13.06, p>.36
7	Södermalm	None	DiD (1,0,0)(1,0,0)	-0.390**	0.131	-32			11.25, p>.51

SD = seasonal differencing
ARIMA=autoregressive integrated moving average
DiD=difference-in-differences
Q=Box-Ljung test for residual autocorrelation (lag 12)
*** p<.001; **p<.01; *p<.05

Discussion

The original evaluation of the STAD programme (Wallin et al. 2003) suggested that this multi-component intervention was relatively effective in reducing violence in and around bars. Considering that this result triggered the dissemination of the programme to most other Swedish communities (Trolldal et al., 2013), it seemed feasible to probe how robust it was. In this paper we thus tested how the choice of control area and model specification affected the outcome. Although these alternative approaches tended to yield somewhat lower effect estimates, we conclude that the assessment reported by Wallin et al. has been corroborated by our findings. However, this conclusion should be accompanied with a couple of caveats. First, it cannot be precluded that the decrease in violence in the experiment area is due to some extraneous factor that is not captured by any of the alternative control areas. For instance, a decrease in the bar density in the experiment area during the intervention period would distort the outcome. Although there is little reason to believe that this was the case, there is no data available to test this assumption. Second, as noted at the beginning of the paper, the sequels of the STAD programmes in Sweden seem to have smaller effects than the original programme. As our analysis suggests that this discrepancy is not due to an overestimation of the effect of the original programme, it makes sense to probe some alternative explanation to this. An important element in this context is that the original STAD programme was a single intervention, which was ambitiously designed and meticulously implemented. There is thus growing evidence that the pioneering version of prevention programmes yield larger effects than the widely disseminated sequel programmes that operate on a more ongoing basis (Hallfors et al. 2006; Ringwalt, Clark, Hanley, Shamblen, & Flewelling, 2010). A plausible explanation is that wider dissemination tends to be associated with lower programme fidelity (Flay et al., 2005).

Declaration of interest None.

Thor Norström, Professor
Swedish Institute of Social Research (SOFI)
Stockholm University
E-mail: thor.norstrom@sofi.su.se

Björn Trolldal, PhD
STAD (Stockholm Prevents Alcohol and Drug Problems)
Stockholm Centre for Psychiatric Research and Education
Stockholm County Council Health Care Provision, and Karolinska Institute
E-mail: bjorn.trolldal@sll.se

REFERENCES

Angrist, J.D., & Krueger, A.B. (1999). Empirical strategies in labor economics. In O.C. Ashenfelter & D. Card (Eds.), *Handbook of labor economics, Vol. 3A* (pp. 1278–1357). Amsterdam: North-Holland.

Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics, 67*(4), 648–660.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics, 119*(1), 249–275.

Box, G.E.P., & Jenkins, G.M. (2008). *Time Series Analysis: Forecasting and control*. New York: Wiley.

Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., Moscicki, E.K., Schinke, S., Valentine, J.C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*(3), 151–175.

Hallfors, D. & Cho, H. & Sanchez, V., Khatapoush, S., Kim, H.M., & Bauer, D. (2006): Efficacy vs effectiveness trial results of an indicated "model" substance abuse program: Implications for public health. *American Journal of Public Health, 96*(12), 2254–2259.

Holder, H. (1998). *Alcohol and the Community: A systems approach to prevention*. New York: Cambridge University Press.

Holder, H., Gruenewald, P.J., Ponicki, W.R., Treno, A.J., Grube, J.W., Saltz, R.F.,Voas, R.B., Reynolds, R., Davis, J., Sanchez, L., Gaumont, G., & Roeper, P. (2000). Effect of community-based interventions on high-risk drinking and alcohol-related injuries. *JAMA: The Journal of the American Medical Association, 284*(18), 2341–2347.

Homel, R., Carvolth, R., Hauritz, M., Mcilwain, G., & Teague, R. (2004). Making licensed venues safer for patrons: What environmental factors should be the focus of interventions? *Drug and Alcohol Review, 23*(1), 19–29.

Jones, L., Hughes, K., Atkinson, A.M., & Bellis, M.A. (2011). Reducing harm in drinking environments: A systematic review of effective approaches. *Health & Place, 17*(2), 508–518.

Månsdotter, A.M., Rydberg, M.K., Wallin, E., Lindholm, L.A., & Andréasson, S. (2007). A cost-effectiveness analysis of alcohol prevention targeting licensed premises. *The European Journal of Public Health, 17*(6), 618–623.

Norström, T., & Skog, O.J. (2005). Saturday opening of alcohol retail shops in Sweden:

An experiment in two phases. *Addiction, 100*(6), 767–776.
Ringwalt, C.R., Clark, H.K., Hanley, S., Shamblen, S.R., & Flewelling, R.L. (2010). The effects of project ALERT one year past curriculum completion. *Prevention Science, 11*(2), 172–184.
Trolldal, B., Brännström, L., Paschall, M.J., & Leifman, H. (2013). Effects of a multi-component responsible beverage service program on violent assaults in Sweden. *Addiction, 108*(1), 89–96.
Wallin, E., Norström, T., & Andreasson, S. (2003). Alcohol prevention targeting licensed premises: A study of effects on violence. *Journal of Studies on Alcohol, 64*(2), 270–277.
Wicki, M., & Gmel, G. (2011). Hospital admission rates for alcoholic intoxication after policy changes in the canton of Geneva, Switzerland. *Drug and Alcohol Dependence, 118*(2–3), 209–215.