This is a repository copy of *Evaluating the causal impact of individual alcohol licensing decisions on local health and crime using natural experiments with synthetic controls*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/158369/

Version: Published Version

# Evaluating the causal impact of individual alcohol licensing decisions on local health and crime using natural experiments with synthetic controls

Frank de Vocht[1,2,3] (iD), Cheryl McQuire[1] (iD), Alan Brennan[2,4], Matt Egan[2,5], Colin Angus[2,4] (iD), Eileen Kaner[2,6], Emma Beard[2,7] (iD), Jamie Brown[2,7] (iD), Daniela De Angelis[2,8,9], Nick Carter[10], Barbara Murray[11], Rachel Dukes[12], Elizabeth Greenwood[12], Susan Holden[12], Russell Jago[2,13] & Matthew Hickman[1,2] (iD)

## ABSTRACT

**Background and Aims** Given the costs of alcohol to society, it is important to evaluate whether local alcohol licensing decisions can mitigate the effects of alcohol misuse. Robust natural experiment evaluations of the impact of individual licensing decisions could potentially inform and improve local decision-making. We aimed to assess whether alcohol licensing decisions could be evaluated at small spatial scale by using a causal inference framework. **Design** Three natural experiments. **Setting and participants** Three English local areas of 1000–15 000 people each. **Intervention and comparator** The case study interventions were (i) the closure of a nightclub following reviews; (ii) closure of a restaurant/nightclub following reviews and (iii) implementation of new local licensing guidance (LLG). Trends in outcomes were compared with synthetic counterfactuals created using Bayesian structural time–series. **Measurements** Time–series data were obtained on emergency department admissions, ambulance call-outs and alcohol-related crime at the Lower or Middle Super Output geographical aggregation level. **Findings** Closure of the nightclub led to temporary 4-month reductions in antisocial behaviour (−18%; 95% credible interval − 37%, −4%), with no change in other outcomes. Closure of the restaurant/nightclub did not lead to measurable changes in outcomes. The new licensing guidance led to small reductions in drunk and disorderly behaviour (nine of a predicted 21 events averted), and the unplanned end of the LLG coincided with an increase in domestic violence of two incidents per month. **Conclusions** The impact of local alcohol policy, even at the level of individual premises, can be evaluated using a causal inference framework. Local government actions such as closure or restriction of alcohol venues and alcohol licensing may have a positive impact on health and crime in the immediate surrounding area.

**Keywords** Alcohol licensing, Bayesian analysis, counterfactuals, crime, natural experiments, public health, synthetic controls.

## INTRODUCTION

There is extensive literature that describes a relationship between location-based interventions in the alcohol environment, such as changes in alcohol outlet density or in the hours that alcohol can be sold and alcohol-related consequences on crime and health [1,2].

The costs of alcohol to the United Kingdom have been estimated to be somewhere between 21 and 52 billion pounds in 2016 [3]. Local authorities in England have powers to influence local alcohol environment by contributing to the licensing process and control the enforcement of existing licenses and can thereby help in mitigating the impact of alcohol on society, although they cannot, for example, set levels of tax [4]. Licensing legislation requires that all licensing decisions examine evidence about specific outlets or local authorities, yet most published or usable evidence exists for larger spatial aggregates which typically cannot demonstrate causal links between individual outlets and harmful outcomes [3]. This gap in the availability of evidence of effectiveness around individual premises is an important limitation to local licensing decisions (e.g. revoking licenses or reductions of trading hours). Legal advice to practitioners in licensing departments with respect

to cumulative impact zones and other licensing activities indicated that formal evaluations of the actual impact of changes to policy could be important to supplement case law and other sources of evidence in licensing hearings. The availability of quantitative evaluations at small temporal–spatial scales, ideally linked to single premises, would thus be beneficial.

The implementation of local licensing policies or licensing actions such as the closure of specific venues generally results from political processes. The timing of these decisions makes it difficult for research teams to influence when and where these are implemented (or in some situations it would be unethical to do so) and as such evaluations using randomized controlled trial (RCT) designs are not possible. Natural experiments (and other quasi-experimental designs) offer the opportunity to evaluate such interventions [5]. The UK Medical Research Council define natural experiments broadly as studies that divide population into exposed and unexposed groups, but where the allocation to these groups was not under the control of the research team [6]. The strength of the inferences that can be made from natural experiments rely largely on the plausibility that the intervention could be considered to be allocated 'as-if randomized'; i.e. although not controlled by the research team they can be considered random, for example due to 'the forces of nature' [7,8]. In such circumstances, natural experiments can generate estimates of effects comparable in causal interpretation to those from controlled trials, while they have the added benefits of avoiding the artificial setting common to trials which can improve their external validity [8]. Another benefit is that they can be conducted using routinely collected data, which (i) enables relatively quick evaluations at low cost as the relevant data are already being collected when an intervention is introduced (and hence pre-intervention data is already available) and (ii) in addition to the evaluation of planned public health interventions also allows for the evaluation of (unplanned) withdrawal of such interventions [9].

We aimed to utilize the benefits from natural experiments to assess the possibility of evaluating the impact of local alcohol licensing interventions on health and crime at small geospatial levels up to single premises and their immediate surroundings. To do so, here we compare temporal trends to their counterfactuals which, because these are by definition unmeasurable, are estimated by using synthetic controls [10]. The synthetic controls are calculated using Bayesian structural time–series, a methodology that was previously shown to be beneficial in the evaluations of alcohol policies at the Local Authority level [4]. Specifically, we developed three case studies together with local practitioners to evaluate the impact of two licensing decisions and the introduction and subsequent defunding of new guideline documentation on health and crime in their

immediate vicinity, with the aim to directly support local council practice.

## METHODS

### Case studies

Based on the outcomes of a work-shop with practice partners from Local Councils, Public Health England and academic colleagues, three cases from different regions in England were selected for inclusion in the study and developed. These included (i) the closure of a venue following a licensing committee review, (ii) the closure of two colocated venues following review and (iii) where new local licensing guidance and increased inspections were established (iiia) and subsequently defunded and restructured (iiib), providing two intervention points for analyses. Outcomes were selected prior to analyses for each of the case studies through discussions between local government officials and the research team and included time–series of reported numbers of specific crimes and health outcomes, covering a pre- and post-intervention period. Data were obtained by the practice partners from existing collaborations, including multi-agency groups, and included recorded numbers of calls for service from police records classified based on national criteria based on locations of the incidents, hospital admissions data based on the home address of the individuals and ambulance data based on the location of the incidents. Counts were aggregated to lower layer super output geographical areas (LSOA; 1000–3000 people) for the first two case studies and middle-layer super output areas (5000–15 000 people) for case study 3 [11]. More details of each case study are provided in the Supporting information.

### Counterfactuals

The measured time–series were compared to the counterfactual time–series, which are the potential outcomes in the case the intervention had not been introduced [12] and the difference interpreted as the effect of the intervention. Counterfactual series were constructed by synthetic controls methods [10] using the weighted average of the outcome variable from a group of local control areas (described in the Statistical methodology). Control areas were areas at the same geospatial level selected from the same Local Authority and were selected by the practice partners as comparable to the case area. We limited the selection of control areas to those from the same Local Authority in this study because (a) at such small geographical-level local practitioners would be best placed to select comparable controls and (b) these areas would be subject to similar trends resulting from policies and priorities within the Local Authority. In addition, by limiting the selection of areas to the same Local Authority this would facilitate future

local use of the methodology. A possible limitation of selection by local practitioners may be the risk of implicit or explicit bias in the selection of the controls, but the complexity of how these are combined to create synthetic counterfactuals mitigates the impact of this. Each synthetic control was based on different combinations of the selected control areas (ranging from two to six), depending on the case study.

### Statistical methodology

Bayesian structural time–series have been described in detail elsewhere [13,14], but in brief are state-space models that can incorporate trends, seasonality and regression components. Two linked equations describe (a) how the outcomes are related to an underlying latent state, and (b) how the latent state changes over time. The errors of different state components are assumed to be normally distributed with a mean of zero and independent of each other. The regression component of the model enables the inclusion of covariates. To develop the synthetic controls, a spike-and-slab prior was placed on the regression coefficients, where the 'spike' determines the probability that a covariate (here, the time–series of the outcome in a control area) has a non-zero coefficient based on independent Bernoulli distributions and the 'slab' is a weakly informative Gaussian prior with a large variance. We used Bayesian model averaging across all combinations of predictors to construct an outcome time–series from the covariate time–series [15]. The posterior predictive density is a joint distribution over all data points [14], and the posterior distribution of the model parameters was estimated using a sequence of Gibbs sampling steps from a Markov chain with a stationary distribution; i.e. stochastic search variable selection [16]. The posterior distribution of the time–series can then be projected forward using the regression component of the model. The difference between the measured and projected (or forecasted) time–series can be interpreted as the effect of the intervention. The 95% Bayesian credible interval (CrI) provides an estimate of the precision of average and cumulative effects over the forecasting time horizon and posterior tail area probabilities, calculated as the samples where the CrI of the cumulative effect excluded zero, are interpreted as classical *P*-values [13]. One minus the posterior tail area probability can also be interpreted as the probability of a causal effect [17].

Individual model specifications were required for each case study and outcome combination because of differences in the data with respect to length of the time–series, number of control sets, temporal variability and other temporal patterns (such as seasonality). Each initial model was based on a random walk plus noise, or local-level, model and sequentially other structures—autoregressive of order

1 [AR(1)] model, local-linear time trends and additional seasonal components [18] were assessed. Common to all analyses were the following specifications: expected explained variance was set to 90%, and priors for the mean of each outcome and its standard deviation were defined as the arithmetic mean of the outcome in the pre-intervention period and 10% of its standard deviation, respectively, and the upper limit for the standard deviation was set to 200% of that of the pre-intervention time–series. Model evaluation was conducted graphically based on plots, and statistically using root mean square errors for the pre-intervention (training) time-period (partial) autocorrelation plots and Ljung–Box tests to assess independent distribution of residuals and precision of counterfactual estimates (standard deviation of forecasts). Markov Chain Monte Carlo samples (10 000) were used and convergence evaluated using Geweke and Raftery & Lewis diagnostics.

Falsification tests in which different implementation dates (temporal falsification) and replacement of case areas by control areas (spatial falsification) were also conducted to strengthen causal inference [19].

Bayesian structural time–series models were implemented using the *bsts* [18] and CausalImpact [17] packages in R (version 3.2.4).

The analysis plan was not pre-registered and the results should therefore be considered exploratory in nature.

## RESULTS

### Case study 1: closure of a nightclub

This nightclub was opened in September 2011 but was put under review by local government following concerns about crime and antisocial behaviour at the premises, and eventually forced to close in September 2013 (see OSM). Time–series of reported incidents of crime (all offenses), specifically antisocial behaviour, and ambulance call-outs and emergency admissions to hospital for alcohol (i.e. the outcome data) from 2010 to September 2014 (12 months post-closure) were used to evaluate the intervention, but incident counts from one to four were not provided because of identifiability concerns, and these were imputed using random imputation with equal inclusion probabilities. Evaluation of 12-month follow-up was modelled based on a semi-local linear trend model with two seasonal components (12 months and four seasons) and 4-month follow-up (see below) based on a local level with seasonal components [18]. The results for all outcomes are shown in Fig. 1. There was little evidence of a longer-term impact at 12 months on any of the selected outcomes as a result of the closure of the venue (Table 1). Graphical observation of the temporal patterns was indicative of an initial, but temporary, reduction in the incidence of reported incidents of antisocial
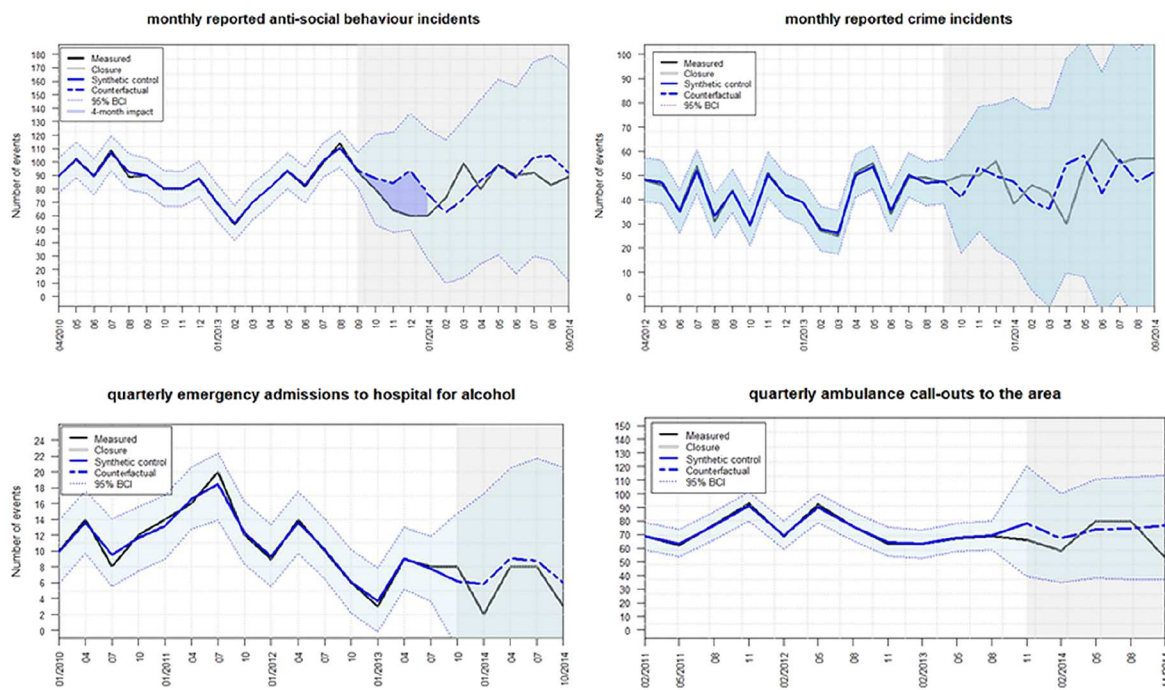
**Figure 1** Case study 1: evaluation of closure of a venue following a licensing committee review. Impact of *post-hoc* 4-month evaluation highlighted at top left [Colour figure can be viewed at wileyonlinelibrary.com]

**Table 1** Case study 1 evaluation of the impact of closure of a venue following review on selected outcomes, and results of temporal and spatial falsification tests for observed effects.

*Evaluation of the 12-month impact of the closure of the venue*

| Outcome | 12-month average impact | 95% CrI | Posterior tail-area probability |
|---|---|---|---|
| Anti-social behaviour[a] | +8% | −63%, +49% | 0.36 |
| Crime (all offences) [1] | +4% | −69%, +77% | 0.45 |
| Ambulance call-outs[b] | −9% | −36%, +20% | 0.22 |
| Emergency Admission to hospital for alcohol[b] | −19% | −155%, +113% | 0.39 |
| *Post-hoc* evaluation of 4-month impact | | | |
| Outcome | 4-month average impact | 95% credible interval | Posterior tail-area probability |
| Anti-social behaviour | −18% | −37%, −4% | 0.01 |

*Temporal falsification tests[c]*

| Antisocial behaviour | 4-month average impact | 95% CrI | Posterior tail-area probability |
|---|---|---|---|
| 6 months earlier | −7% | −22%, +16% | 0.15 |
| 6 month later | −9% | −46%, +17% | 0.31 |

*Spatial falsification tests[d]*

| | | | |
|---|---|---|---|
| Area 2 | +11% | −44%, +69% | 0.32 |
| Area 3 | −39% | −109%, +47% | 0.10 |
| Area 4 | −20% | −158%, +77% | 0.43 |

[a]Monthly number of reported events; [b]quarterly number of events; [c]the time-point of the intervention was artificially moved forwards or backwards; [d]each control area was artificially assigned as the intervention area. CrI = credible interval.

behaviour. *Post-hoc* analyses of this pulse impact (i.e. the shortest period of impact [20]) indicated a 4-month impact of −18% (95% CrI = −37%, −4%), equating to 60 averted incidents. Falsification tests indicated that this was specific to the time-point and the area of closure of the venue (as summarized in Table 1).

**Case study 2: closure of a restaurant and cocktail bar**

Both co-located premises were closed in November 2016 following a licence review brought by the police on the grounds of crime and disorder related to the premises. Time–series of reported incidents of crime (all offences) and antisocial behaviour specifically were available for 2015–17. AR(1) models with linear local trend and monthly seasonal components were used to model the counterfactuals, and the results for the closure of the venues are shown in Fig. 2. Uncertainty is large as a result of the low numbers of incidents in these LSOAs, and based on point estimates and precision there was little indication of a measurable effect on the incidence of reported antisocial behaviour or crime incidents in the subsequent 12-month period.

**Case study 3a and b: local licensing guidance**

In 2012 the introduction of new statutory guidance facilitated the development of Local Licensing Guidance (LLG) for specific areas which was taken up for a designated area within an urban conurbation. The LLG was in place during 2013/14, during which there was active engagement, including additional inspections of venues, by Public Health locality teams. However, as a result of budget cuts and changes to resource allocation there were no more active follow-ups after the 12-month period.

Time–series of monthly reported incidents of drunk and disorderly behaviour, sexual offences, antisocial behaviour and domestic violence were available for the period 2008–14. Different model specifications were required to best model the temporal patterns of each outcome: drunk and disorderly behaviour was modelled on an AR(1) autoregressive model with a monthly seasonal component, sexual offences on a local level model with a monthly seasonal component and antisocial behaviour and domestic violence both on a semi-local linear time trend and a monthly seasonal component. The results of the introduction of the LLG (3a) are shown in Table 2 and Fig. 3, and suggest weak evidence for a reduction in the incidence of reported drunk and disorderly behaviour incidents [12-month impact of −42% (95% CrI = −109%, +23%) compared to the counterfactual]. This, however, corresponds to fewer than one averted incident per month (95% CrI = −2, +0.3) on average. Falsification tests indicated that this weak evidence of effect was specific to the time-point as well as to the area (with the exception of one neighbouring area, where a comparable effect was observed).

Results of the evaluation of the impact of defunding and restructuring of the LLG initiative (3b) are shown in Table 3 and Fig. 4. Strong evidence of a reduction in antisocial behaviour events compared to the counterfactual was observed, which was thought to be related change of

the role of the Community Safety Manager at the same time, which contributed to a loss of engagement with the local community and businesses and a reduction in the number of incidents reported. Indeed, falsification tests indicated this change was not specific to the time-point or area in which the LLG initiative was defunded. There was also moderate evidence of an increase in reported incidents of domestic violence of +11% (−10%, +35%), which corresponds to two additional incidents per month. Falsification tests indicated that the increase happened at approximately the same time as the withdrawal of public health, and was also not observed in any of the other areas.

## DISCUSSION

The analyses presented in this paper highlight that it is possible to quantitatively evaluate the small-area impact of local changes to the local alcohol environment up to the level of single premises and their immediate surroundings in a robust manner through synthetic control methodology. Our single case studies did not (and would not be expected to) result in large effects, and they are implemented in complex local systems [21]. Hence, the quantitative results we report in this study should be appraised in conjunction with local information about implementation and context from local practice partners. As such, the evidence from these natural experiments contribute to the evidence of effects of local licensing policies at the Local Authority level in England [4,22,23] and is also in agreement with results from natural experiments looking at effects of choices in the alcohol environment more broadly [24–26].

In case study 1 there was strong evidence for an immediate, but relatively short-term, impact of closure of the venue on the incidence of reported antisocial behaviour, but not for local crime rates, emergency admissions to hospital for alcohol or for ambulance call-outs. For that immediate 4-month period, the impact equated to 15 incidents (95% CrI = −3 to − 30 incidents) averted per month, or 60 in total. However, this 4-month pulse impact was defined *post hoc* and there was no evidence of impact after this period. Falsification tests provided reassurance that the observed impact could be ascribed to the closure of the venue, although weak evidence of an observed effect in one of the control areas highlight the complex setting, with many additional exogenous factors affecting these local systems. Feedback from the practice partner indicated that a reduction in antisocial behaviour after closure of the venue conformed to their expectations, given that this was one of the main reasons that resulted in its closure. However, there is no clear explanation why this would have returned to the normal level after the initial pulse.

In case study 2 there was little evidence that closure of the venues had resulted in measurable effects on reported incidents of crimes or antisocial behaviour in the area
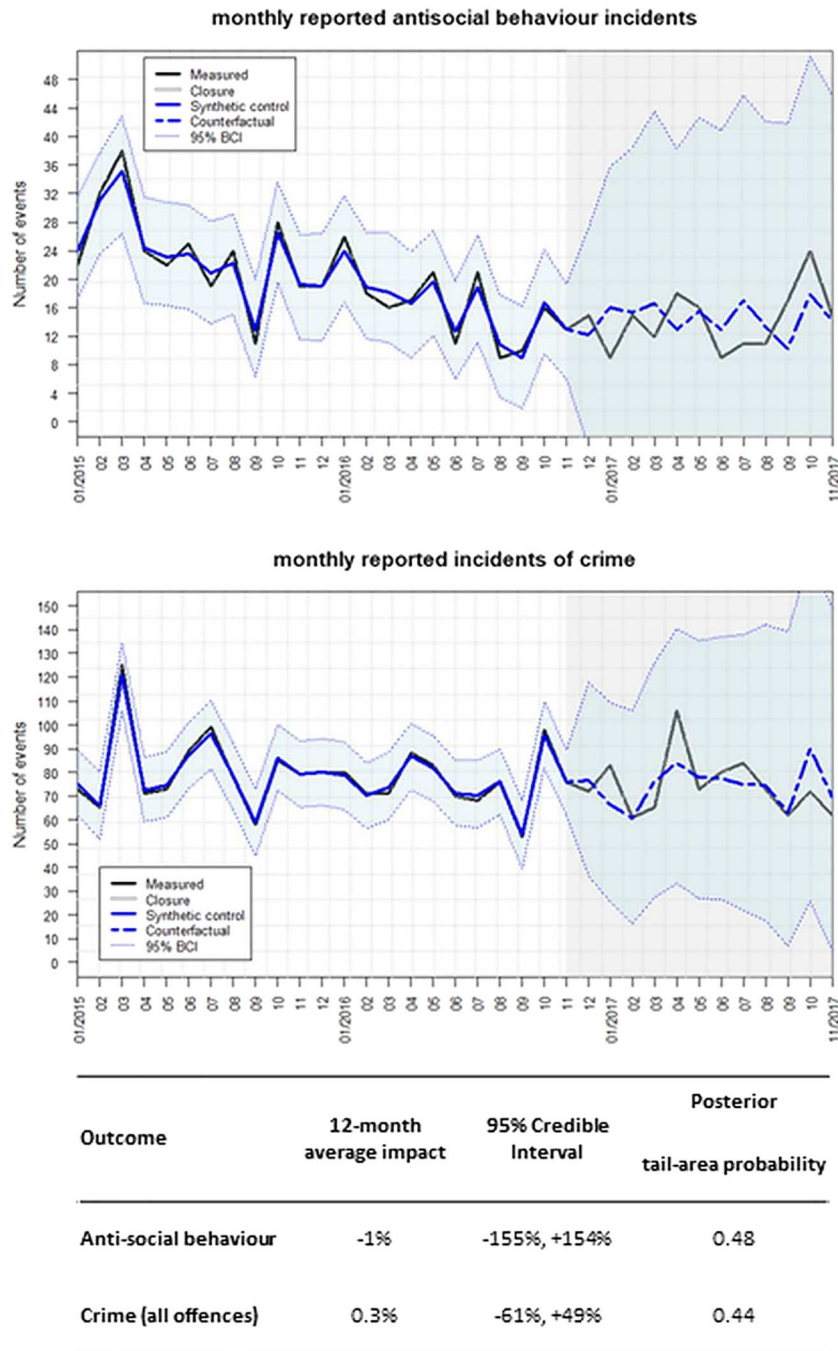
**monthly reported antisocial behaviour incidents**



**monthly reported incidents of crime**



| Outcome | 12-month average impact | 95% Credible Interval | Posterior tail-area probability |
|---|---|---|---|
| Anti-social behaviour | -1% | -155%, +154% | 0.48 |
| Crime (all offences) | 0.3% | -61%, +49% | 0.44 |

**Figure 2** Case study 2: evaluation of the impact of closure of two venue on the monthly incidence of reported antisocial behaviour and crimes in the immediate local area [Colour figure can be viewed at wileyonlinelibrary.com]

immediately surrounding the venues. The influence of high variability and exogenous factors that affect the time–series (here the sudden reduction in recorded antisocial behaviour in the year prior to the closure of the venues) complicated statistical inferences. There may not have been direct effects on the immediate vicinity of the venues, because these were located in the connecting area between two night-life focal points in the conurbation and thus any effects (if these occurred) would have been observed elsewhere.

Case study 3 described weak to moderate evidence of a small impact on the incidence of reported drunk and disorderly behaviour in the area following the introduction of LLG, corresponding to approximately one to two averted events per month compared to the counterfactual. Comparable effects were observed in some adjacent areas as well which, the local practitioners indicated, also received additional inspections by the Public Health team in that period, providing additional weight to the evidence for a small effect as a result of the LLG. The subsequent defunding and

**Table 2** Case study 3a: evaluation of the impact of new LLG and increased inspections on selected outcomes, as well as results of temporal and spatial falsification tests for observed effects.

| Monthly number of reported incidents | 12-month average impact | 95% Credible Interval | Posterior tail-area probability |
|---|---|---|---|
| Drunk and disorderly behaviour | −42% | −109%, +23% | 0.10 |
| Sexual offences | +5% | −95%, +90% | 0.44 |
| Antisocial behaviour | −12% | −95%, +36% | 0.40 |
| Domestic violence | +0.7% | −28%, +30% | 0.48 |
| Temporal falsification for drunk & disorderly behaviour[a] | 12-month average impact | 95% Credible Interval | Posterior tail-area probability |
| 6 months earlier | −1% | −95%, +91% | 0.49 |
| 6 months later | −27% | −115%, +61% | 0.27 |
| Spatial falsification for drunk & disorderly behaviour[b] | 12-month average impact | 95% CrI | Posterior tail-area probability |
| Control area 1 | −9% | −64%, +43% | 0.36 |
| Control area 2 | −53% | −119%, +14% | 0.06 |
| Control area 3 | −46% | −183%, +89% | 0.23 |
| Control area 4 | +27% | −106%, +156% | 0.33 |
| Control area 5 | +329%[c] | −148%, +821% | 0.08 |
| Control Area 6 | −64% | −196%, +60% | 0.15 |

[a]The time-point of the intervention was artificially moved forwards or backwards. [b]Each control area was artificially assigned as the intervention area. [c]This percentage is misleading because it is based on very small incidence. CrI = credible interval; LLG = local licensing guidance.
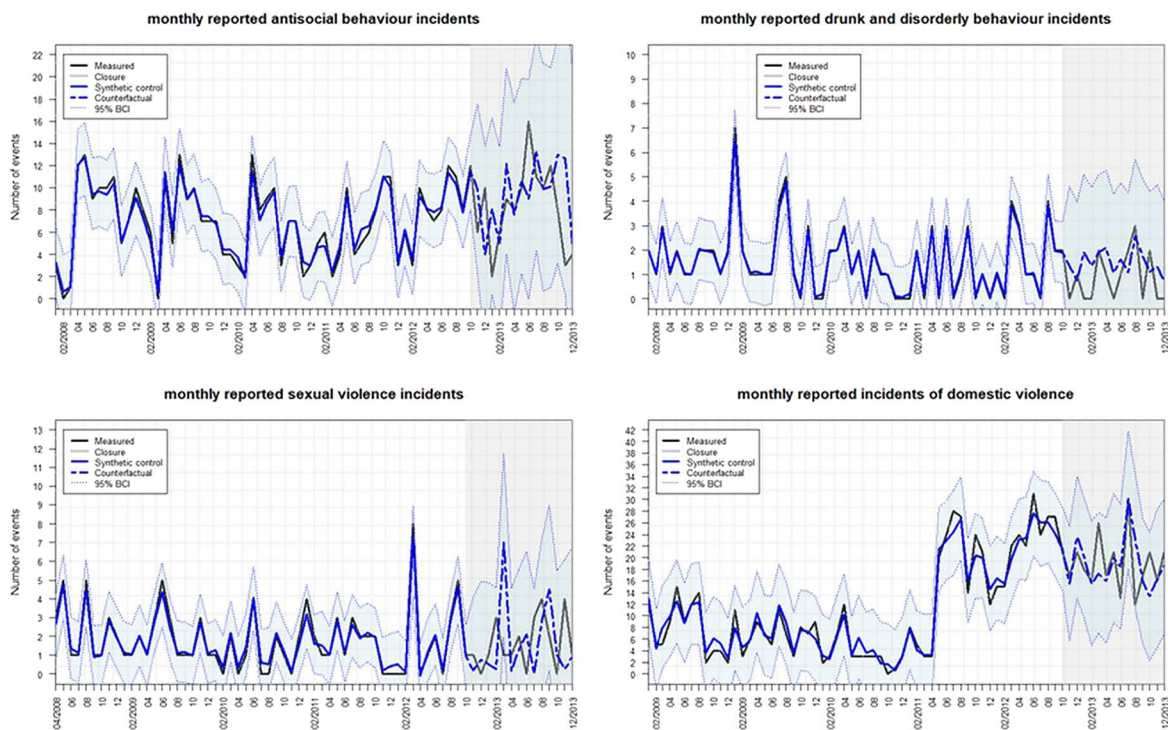


**Figure 3** Case study 3a: evaluation of the introduction of the Local Licensing Guidance and increased inspections [Colour figure can be viewed at wileyonlinelibrary.com]

restructuring of the LLG provided moderate evidence that this had resulted in a small increase in incidents of domestic violence of approximately two additional incidents monthly compared to the counterfactual. In interpreting these effects, it is important to note that these small-scale interventions do not occur in a vacuum and, as a consequence of the restructuring process other work being carried out at the time, including work on domestic violence, also changed, which may have contributed to the observed excess incidents of domestic violence.

**Table 3** Case Study 3b: evaluation of the impact of defunding and restructuring of LLG on selected outcomes, as well as results of temporal and spatial falsification tests for observed effects.

| Outcome | 12-month average impact | 95% CrI | Posterior tail-area probability |
|---|---|---|---|
| Drunk and disorderly behaviour | −37% | −213, +123% | 0.31 |
| Sexual offences | −8% | −152%, +138% | 0.46 |
| Antisocial behaviour | −61% | −109%, −19% | 0.01 |
| Domestic violence | 11% | −10%, +35% | 0.14 |

*Falsification tests for antisocial behaviour*

| Temporal falsification[a,b] | 4 month average impact | 95% CrI | Posterior tail-area probability |
|---|---|---|---|
| 3 months earlier | −3% | −33%, +24% | 0.43 |
| 3 months later | −41% | −79%, −4% | 0.01 |
| Spatial falsification[a,c] | 4 month average impact | 95% CrI | Posterior tail-area probability |
| Control area 1 | −6% | −79%, +62% | 0.44 |
| Control area 2 | −24% | −119%, +71% | 0.31 |
| Control area 3 | −49% | −112%, +19% | 0.07 |
| Control area 4 | −38% | −124%, +42% | 0.2 |
| Control area 5 | −42% | −131%, +46% | 0.17 |
| Control area 6 | −43% | −95%, +9% | 0.05 |

*Falsification tests for domestic violence*

| Temporal falsification test[a,b] | 4-month average impact | 95% CrI | Posterior tail-area probability |
|---|---|---|---|
| 3 months earlier | 4% | −21%, 36% | 0.42 |
| 3 month later | −0.20% | −22%, 23% | 0.48 |
| Spatial case area[a,c] | 4 month average impact | 95% CrI | Posterior tail-area probability |
| Control area 1 | 7% | −14%, 28% | 0.25 |
| Control area 2 | 6% | −18%, 34% | 0.32 |
| Control area 3 | −5% | −19%, 12% | 0.27 |
| Control area 4 | −6% | −32%, 22% | 0.32 |
| Control area 5 | −3% | −38%, 31% | 0.42 |
| Control area 6 | 25% | −36%, 72% | 0.15 |

[a]3 months instead of 6 months temporal falsification was a priori selected to minimize the effects of the introduction of the LLG and additional inspections, which overlaps with the earlier temporal falsification test. This allowed for 4-month impact periods which, for consistency, were also used for the spatial falsification tests; [b]the time-point of the intervention was artificially moved forwards or backwards; [c]each control area was artificially assigned as the intervention area.
CrI = credible interval; LLG = local licensing guidance.

These analyses illustrate the difficulties of the statistical modelling in these highly localized settings, with low numbers of events, high temporal variability and relatively small expected effect sizes (if any). The Bayesian structural time–series approach allows, in theory, for an accurate modelling of counterfactual time–series but may have benefited from longer time–series to increase statistical power, and also suffers when no events occur on many time-points. Moreover, these natural experiments occur in a complex systems setting, but these analyses are based on traditional linear evaluations of cause and effect [21]. For example, the pulse effect observed in case study 1 may be the result of an unknown feedback loop not taken into account by the method. Comparable control areas were selected from the same local authorities in all cases and had varying distances to the case areas. The main assumption of this approach is that the correlations between outcomes prior to the evaluated intervention remain the same in the post-intervention period. Although there was little evidence of spillover effects, which have been reported in geographical studies of alcohol [27], but not in England in previous analyses [22], some evidence of spatial clustering was observed, such as in study 3 where the falsification test highlighted similar effects of the introduction of LLG in a neighbouring area. Most unobserved confounding factors, such as macro-economic changes and specific local policies or priorities, are in theory accounted for because the synthetic controls were based on weighted combinations of comparable control areas from the same Local Authority and/or are unlikely to change significantly over these time–periods such as, for example, area-level demographics or deprivation. However, residual confounding cannot be excluded because the researchers were not in control of the randomization process. Generally, the use of synthetic controls has advantages over other analytical methods for natural experiments [10], but regardless of the analytical method the stronger the argument is that the allocation of the intervention and the selection of
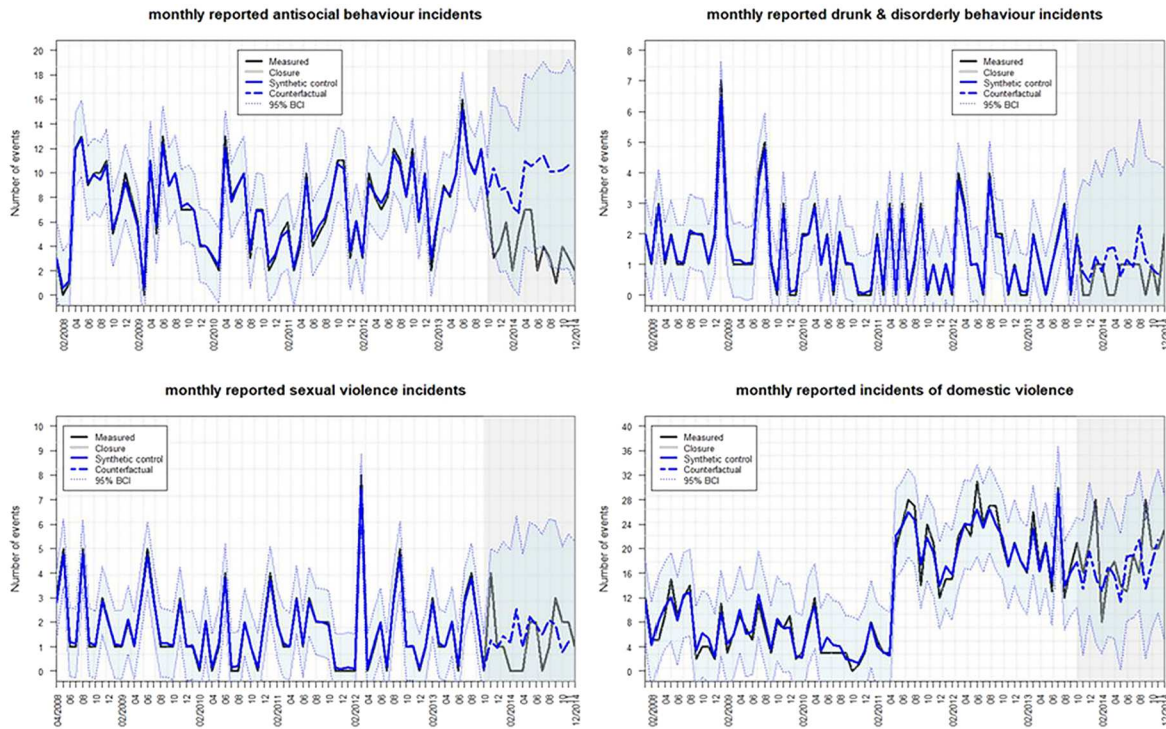
**Figure 4** Case study 3b: evaluation of the defunding and restructuring of the Local Licensing Guidance [Colour figure can be viewed at wileyonlinelibrary.com]

control areas may mimic a random process (as-if randomization), the less likely that the results may be significantly confounded [7].

The main strengths of this study were that these analyses demonstrate that in local policy interventions where randomized controlled trials (RCTs) are not possible, natural experiments in which as-if randomization enables assignment on observed and unobserved factors [28] can be a valuable alternative for robust evaluation if sufficient data are available [6]. Moreover, this study indicates that comparisons to counterfactuals can be performed at small area-levels with low numbers of incidents, although with the expected effect sizes in these case studies, moderate strength of the evidence was the maximum achievable [29]. The comparison to the counterfactual was an important strength of the current methodology, with advantages over other alternative methods for the evaluation of natural experiments such as pre–post designs or interrupted time–series [6,10] (note that there are alternative statistical methods to the Bayesian structural time–series that enable comparison to controls). Specific strengths related to the use of Bayesian structural time–series include the inclusion of subjective knowledge to defining the priors, the inclusion of posterior uncertainty about the covariates in the model and their coefficients, the accommodation of serial autocorrelation and cross-sectional and longitudinal heterogeneity and the inclusion of Bayesian model averaging to mitigate issues around selection of covariates and overfitting [14].

To our knowledge, these were the first studies in which the effect of small-scale interventions, at the levels of individual premises in two of the case studies, were evaluated using a quantitative method that enabled robust comparison to counterfactuals. The use of this methodology to conduct small-scale evaluation has the potential to bridge the gap between the predominantly data-orientated approach by public health teams and the need for contextualized evidence for presentation to councillors [30], at least in principle. However, the small effect sizes and complex modelling provide moderate evidence of effects, at best. Although in this study we used this analytical framework in the context of evaluation of the impact of local alcohol (licensing) policies, it is a general framework for evaluation of natural experiments, and as such is likely to be also applicable to the evaluation of natural experiments and other quasi-experimental in the broader public health setting.

### Declaration of interests

### Acknowledgements

## Author's affiliations

Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK,[1] The National Institute for Health Research School for Public Health Research (NIHR SPHR), London, UK,[2] The National Institute for Health Research Applied Research Collaboration West (NIHR ARC West), Bristol, UK,[3] SCHaRR, University of Sheffield, Sheffield, UK,[4] London School of Hygiene and Tropical Medicine, London, UK,[5] Newcastle University, Newcastle upon Tyne, UK,[6] University College London, London, UK,[7] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK,[8] Statistics, Modelling and Economics Department, National Infection Service, Public Health England, London, UK,[9] Bristol City Council, Bristol, UK,[10] Medway Council, Chatham, Kent, UK,[11] Leeds City Council, Leeds, UK[12] and Centre for Exercise, Nutrition and Health Sciences, School for Policy Studies, University of Bristol, Bristol, UK[13]

## References

1. Campbell C. A., Hahn R. A., Elder R., Brewer R., Chattopadhyay S., Fielding J. *et al.* The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. *Am J Prev Med* 2009; **37**: 556–69.

2. Hahn R. A., Kuzara J. L., Elder R., Brewer R., Chattopadhyay S., Fielding J. *et al.* Effectiveness of policies restricting hours of alcohol sales in preventing excessive alcohol consumption and related harms. *Am J Prev Med* 2010; **39**: 590–604.

3. Burton R., Henn C., Lavoie D., O'Connor R., Perkins C., Sweeney K. *et al.* A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective. *Lancet* 2017; **389**: 1558–80.

4. de Vocht F., Tilling K., Pliakas T., Angus C., Egan M., Brennan A. *et al.* The intervention effect of local alcohol licensing policies on hospital admission and crime: a natural experiment using a novel Bayesian synthetic time-series method. *J Epidemiol Community Health* 2017; **71**: 912–8.

5. Shadish W. R., Cook T. D., Campbell D. T. *Experimental and Quasi-Experimental Designs, 2nd edn.* Belmont, CA: Wadsworth, Cengage Learning; 2002, p. 623.

6. Craig P., Cooper C., Gunnell D., Haw S., Lawson K., Macintyre S. *et al.* Using natural experiments to evaluate population health interventions: new medical research council guidance. *J Epidemiol Community Health* 2012; **66**: 1182–6.

7. Dunning T. Improving causal inference: strengths and limitations of natural experiments. *Polit Res Q* 2008; **61**: 282–93.

8. Bärnighausen T., Tugwell P., Røttingen J. A., Shemilt I., Rockers P., Geldsetzer P. *et al.* Quasi-experimental study designs series—paper 4: uses and value. *J Clin Epidemiol* 2017; **89**: 21–9.

9. Craig P., Gibson M., Campbell M., Popham F., Katikireddi S. V. Making the most of natural experiments: what can studies of the withdrawal of public health interventions offer? *Prev Med (Balt)* 2018; **108**: 17–22.

10. Bouttell J., Craig P., Lewsey J., Robinson M., Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Community Health* 2018; **72**: 673–8.

11. Office for National Statistics. Census geography [internet]. 2019 [cited 2019 Jun 11]. Available at: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography (accessed 28 February 2020).

12. Glass T., Goodman S., Hernán M. A., Samet J. M. Causal inference in public health. *Annu Rev Public Health* 2013; **34**: 61–75.

13. Brodersen K. H., Gallusser F., Koehler J., Remy N., Scott S. L. Inferring causal impact using Bayesian structural time-series models. *Ann Appl Stat* 2015; **9**: 247–74.

14. Scott S. L., Varian H. R. Predicting the present with Bayesian structural time series. *Int J Math Model Numer Optim* 2014; **5**: 4.

15. Hoeting J. A., Madigan D., Raftery A. E., Volinsky C. T. Bayesian model averaging: a tutorial. *Stat Sci* 1999; **14**: 382–417.

16. George E. I., McCulloch R. E. Stochastic search variable selection. In: Gilks W. R., Richardson S., Spiegelhalter D. J., editors. *Markov Chain Monte Carlo in Practice*. London, UK: Chapman and Hall; 1996, pp. 203–14.

17. CausalImpact BK. An R package for causal inference using Bayesian structural time-series models [internet]. 2015. Available at: https://google.github.io/CausalImpact/CausalImpact.html (accessed 28 February 2020).

18. Scott SL. Package 'bsts'. Bayesian Structural Time Series [internet]. 2017. Available at: https://cran.r-project.org/web/packages/bsts/bsts.pdf (accessed 28 February 2020).

19. Craig P., Katikireddi S. V., Leyland A., Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annu Rev Public Health* 2017; **38**: 39–56.

20. Sherrieb K., Norris F. H. Public health consequences of terrorism on maternal–child health in New York City and Madrid. *J Urban Health* 2013; **90**: 369–87.

21. Rutter H., Savona N., Glonti K., Bibby J., Cummins S., Finegood D. T. *et al.* The need for a complex systems model of evidence for public health. *Lancet* 2017; **390**: 2602–4.

22. De Vocht F., Heron J., Angus C., Brennan A., Mooney J., Lock K. *et al.* Measurable effects of local alcohol licensing policies on population health in England. *J Epidemiol Community Health* 2016; **70**: 231–7.

23. de Vocht F., Tilling K., Campbell R., Hickman M. Inferring the intervention effect of local alcohol licensing policies on hospital admission and violent crime: a natural experiment with Bayesian synthetic controls. *Lancet* 2016; **388**: 43.

24. Humphreys D. K., Eisner M. P., Wiebe D. J. Evaluating the impact of flexible alcohol trading hours on violence: an interrupted time series analysis. *PLOS ONE* 2013; **8**: e55581.

25. Callan C. M., Boyle A. A. Has the licensing act 2003 affected violence rates in England and Wales? A systematic review of hospital and police studies. *Eur J Emerg Med* 2018; **25**: 304–11.

26. Pliakas T., Egan M., Gibbons J., Ashton C., Hart J., Lock K. Increasing powers to reject licences to sell alcohol: impacts on availability, sales and behavioural outcomes from a novel natural experiment evaluation. *Prev Med (Balt)* 2018; **116**: 87–93.

27. Brännström L., Trolldal B., Menke M. Spatial spillover effects of a community action programme targeting on-licensed premises on violent assaults: evidence from a natural experiment. *J Epidemiol Community Health* 2016; **70**: 226–30.

28. Waddington H., Aloe A. M., Becker B. J., Djimeu E. W., Hombrados J. G., Tugwell P., *et al.* Quasi-experimental study designs series—paper 6: risk of bias assessment. *J Clin Epidemiol* 2017; **89**: 43–52.

29. Samartsidis P, Seaman SR, Presanis AM, Hickman M, De Angelis D. Review of methods for assessing the causal effect of binary interventions from aggregate time-series observational data [internet]. arXiv.org. 2018. Available at: https://arxiv.org/abs/1804.07683 (accessed 28 February 2020).

30. Mooney J. D., Sattar Z., de Vocht F., Smolar M., Nicholls J., Ling J. Assessing the feasibility of using health information in alcohol licensing decisions: a case study of seven English local authorities. *Lancet* 2016; **388**: 79.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Data S1** Supporting Information.