

Christian Bottomley<sup>1,2</sup> / J. Anthony G. Scott<sup>2,3</sup> / Valerie Isham<sup>4</sup>

# Analysing Interrupted Time Series with a Control

<sup>1</sup> MRC Tropical Epidemiology Group, London School of Hygiene & Tropical Medicine, London, UK, E-mail: christian.bottomley@lshtm.ac.uk

<sup>2</sup> Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK, E-mail: christian.bottomley@lshtm.ac.uk

<sup>3</sup> KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

<sup>4</sup> Department of Statistical Science, University College London, London, UK

## Abstract:

Interrupted time series are increasingly being used to evaluate the population-wide implementation of public health interventions. However, the resulting estimates of intervention impact can be severely biased if underlying disease trends are not adequately accounted for. Control series offer a potential solution to this problem, but there is little guidance on how to use them to produce trend-adjusted estimates. To address this lack of guidance, we show how interrupted time series can be analysed when the control and intervention series share confounders, i. e. when they share a common trend. We show that the intervention effect can be estimated by subtracting the control series from the intervention series and analysing the difference using linear regression or, if a log-linear model is assumed, by including the control series as an offset in a Poisson regression with robust standard errors. The methods are illustrated with two examples.

**Keywords:** interrupted time series, segmented regression, common trend model

**DOI:** 10.1515/em-2018-0010

**Received:** May 29, 2018; **Revised:** January 24, 2019; **Accepted:** January 24, 2019

## 1 Introduction

Interrupted time series (ITS) analysis is an increasingly popular method for evaluating public health interventions (Jandoc et al. 2015). An important feature of the analysis is that it quantifies the population-level impact of the intervention, which often includes herd effects (Armah et al. 2016; Bruhn et al. 2017). ITS effect estimates are therefore different from, and complementary to, the direct effect estimates obtained from individually-randomised studies.

Although useful, a major disadvantage of the method is that estimates of intervention impact are vulnerable to confounding due to trends unrelated to the intervention. Say, for example, a vaccine is introduced in a country that is undergoing rapid economic development. If there is a reduction in disease after vaccine introduction, then it is difficult to know whether this is due to the impact of the vaccine itself or to other factors related to economic development such as improved nutrition.

One potential solution to this problem is to adjust for linear trend using regression: the so-called segmented regression model (Wagner et al. 2002). The model is popular since it is easy to implement and to interpret. However, the model is often difficult to justify since in order to estimate the intervention effect it is necessary to assume that the pre-intervention linear trend would have persisted post intervention had the intervention not been introduced.

Another solution is to use a control series to model the trend in the absence of intervention. The control series should share the same confounders as the intervention series and be unaffected by the intervention. The control might be the *same* outcome in an untreated population—e. g. Clancy et al. (Clancy et al. 2002) used mortality in Ireland outside Dublin as the control to assess the impact of a coal sales ban on mortality in Dublin. Or the control might be a *different* outcome in the same population—e. g. Morgan et al. used aspirin-related deaths as the control to assess the impact of reduced paracetamol pack size on paracetamol-related deaths (Morgan, Griffiths, and Majeed 2007).

Control series are often used descriptively to show that there is no underlying trend, and hence to demonstrate the validity of an intervention effect estimate (Mackenzie et al. 2016; Fowler et al. 2007). But when there is a trend this needs to be accounted for in the analysis. We address this issue here by presenting a simple framework for analysing interrupted time series with a control. First, we review the use of segmented regression to

analyse ITS without a control. Then we show how a control series can be used as an alternative means of adjusting for trend. We illustrate the methodology using data from two ITS studies: an evaluation of a hospital-based infection control programme, and an evaluation of the impact of rotavirus vaccine introduction.

## 2 Interrupted time series model

Consider a series of  $n$  observations on disease incidence,  $y_t$  ( $t = 0, \dots, n - 1$ ), made before and after the introduction of an intervention. For example, the observations might be monthly case counts before and after vaccine introduction. A simple model for these observations is

$$y_t = \alpha + \beta x_t + \epsilon_t. \quad (1)$$

In this model,  $x_t$  denotes the presence/absence of the intervention (i. e.  $x_t = 0$  before the intervention and  $x_t = 1$  after),  $\beta$  is the intervention effect,  $\alpha$  is the expected incidence in the absence of intervention and  $\epsilon_t$  is an error term. The latter is assumed to be *independent* of  $x_t$ , and represents the effects of disease determinants other than  $x_t$ .

This model is misspecified if some determinants of disease incidence are correlated with  $x_t$ . The model would, for example, not be applicable if there was a change in surveillance intensity during the study period or if a second intervention was implemented. In this case, the model should include these confounders such that

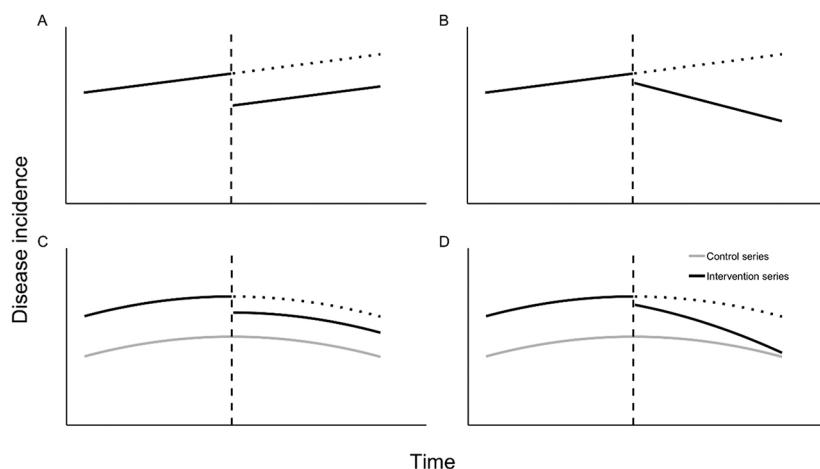
$$y_t = \alpha + \beta x_t + \sum_i \rho_i (z_{it} - z_{i0}) + \epsilon_t, \quad (2)$$

where  $z_{it}$  denotes the  $i$ th confounder,  $\rho_i$  represents its effect on disease incidence, and  $\alpha$  represents the expected incidence at  $t = 0$  (since  $x_t = 0$  and  $z_{it} = z_{i0}$  when  $t = 0$ ). If the confounders are known and measured, then the impact of the intervention can be estimated by fitting a regression model that includes these factors. But what can we do if some or all of the confounders are unobserved? One possible solution is to assume that the combined effect of the confounders is a linear function of time, as is the case if each  $z_{it}$  increases or decreases linearly over time. This is the assumption behind the so-called segmented regression model (Wagner et al. 2002):

$$y_t = \alpha + \beta x_t + \gamma t + \epsilon_t. \quad (3)$$

In the absence of autocorrelation (see below), the model can be fitted by regressing disease incidence on time,  $t$ , and the intervention indicator,  $x_t$ . Often the model is extended to include an interaction between  $t$  and  $x_t$  to allow for a time-varying treatment effect (Figure 1 A & B), i. e.

$$y_t = \alpha + \beta_1 x_t + \beta_2 x_t (t - T) + \gamma t + \epsilon_t. \quad (4)$$



**Figure 1:** Four models for estimating intervention impact from an interrupted time series. The models are based on different assumptions about the nature of the intervention impact and about disease incidence in the counterfactual no intervention scenario (dotted line). In the segmented regression model (A and B) the counterfactual scenario is determined by the pre-intervention trend, and in the common trend model (C and D) it is determined by the control series. The intervention impact may be constant (A and C) or time-dependent (B and D).

In this specification of the model in which we have centred  $t$  about  $T$  in the interaction term,  $\beta_1$  represents the intervention effect at time  $t = T$  and  $\beta_2$  represents the change in the intervention effect for each unit increase in time.

The model assumes the intervention effect increases linearly over time after intervention introduction. But other models of intervention impact, including models that incorporate lag effects, can also be fitted. If, for example, there is a roll out period during which the intervention does not reach its full effect, this could be modelled by including several post-intervention indicators:

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \gamma t + \epsilon_t, \quad (5)$$

where  $x_{1t}$  is an indicator for the first post-intervention period, i. e. the roll out period, and  $x_{2t}$  is an indicator for the second post-intervention period. Faced with many possible models of intervention impact, to avoid overfitting we follow Lopez Bernal, Cummins, and Gasparrini (2017) in recommending that the model be chosen primarily based on biological plausibility rather than statistical criteria.

In addition to the intervention effect, we may also want to include a seasonal component in the model (e. g. by including calendar month as a categorical covariate). Often seasonality will have limited confounding effect, and it will have no confounding effect if both the pre- and post-intervention periods are whole numbers of years. Nevertheless, accounting for seasonality can be useful since doing so will often reduce autocorrelation and error variance.

### 3 Using a control series to adjust for trend

An alternative solution to the problem of confounding is to use the trend observed in a control series to adjust the intervention effect estimate. Usually the control is an age group or population that did not receive the intervention, or those with a related disease. The approach is particularly useful when the trend in incidence is non-linear, and a linear segmented regression analysis is therefore not applicable. But even when the pre-intervention trend is linear, a control can be useful if it provides information on how the intervention series might have behaved in the absence of intervention. As such an ITS analysis that uses a control series may be more convincing than an analysis that uses segmented regression alone.

The key features of the control series are that it should (1) be unaffected by the intervention, and (2) share confounders with the intervention series. If, in addition, we assume the shared confounders have the same effect on both series, then we arrive at the following model:

$$\begin{aligned} y_{It} &= \alpha_I + \beta x_t + g(t) + \epsilon_{It} \\ y_{Ct} &= \alpha_C + g(t) + \epsilon_{Ct}, \end{aligned} \quad (6)$$

where  $y_{It}$  is the intervention series,  $y_{Ct}$  is the control series,  $g(t) = \sum_i \rho_i (z_{it} - z_{i0})$  is the combined effect of the unobserved confounders, and  $\epsilon_{It}$  and  $\epsilon_{Ct}$  are error terms that may be correlated in time (autocorrelated) but are independent of  $x_t$ .

We will refer to the model in eq. (6) as the common trend model since both series share a trend,  $g(t)$ , in the pre-intervention period and would share a trend in the post-intervention period were it not for the impact of the intervention (Figure 1C & D).

Assuming the common trend model, we can eliminate the effect of the unobserved confounders (the trend) by subtracting the control series from the intervention series

$$d_t = y_{It} - y_{Ct} = \alpha_d + \beta x_t + \epsilon_{dt}, \quad (7)$$

where  $\alpha_d = \alpha_I - \alpha_C$  and  $\epsilon_{dt} = \epsilon_{It} - \epsilon_{Ct}$  and both series are of equal length ( $n$ ). Thus, the intervention effect,  $\beta$ , can be estimated by performing a regression where the difference,  $d_t$ , is the outcome and  $x_t$  is an explanatory variable. Assuming a constant intervention effect, as in eq. (7), this amounts to a “difference-in-differences” estimate where the mean of the pre-intervention differences is subtracted from the mean of the post-intervention differences (Angrist and Pischke 2009). In the simple case where the  $\epsilon_{dt}$  are uncorrelated, the means can be compared using a t-test. However, if either the control or intervention series is autocorrelated then the  $\epsilon_{dt}$  will be autocorrelated and it will be necessary to account for this in the regression (see below for further discussion of autocorrelation).

As with the segmented regression model, the common trend model can easily be extended to incorporate more complex intervention effects. If, for example, we expect the treatment effect to increase linearly then we could model this by including a treatment time interaction in the regression, as we did for the segmented regression model eq. (4). Or, if we expect the intervention effect to saturate then we could include several post indicators, as in eq. (5).

We may also include a seasonal component in the model for  $d_t$ , (e. g. by including calendar month as a categorical covariate), if seasonality has not been removed by the process of differencing.

## 4 Testing the common trend assumption

The approach outlined above is only useful if an appropriate control can be identified. Specifically, we must ensure the control is unaffected by the intervention and shares confounders with the intervention series. The latter requirement means that determinants of the intervention series that have changed between the pre- and post-intervention period, apart from the intervention itself, should also be determinants of the control series and vice versa. In particular, a potential control is not suitable if it is affected by an intervention or change in surveillance that does not also affect the intervention series.

As well as using background information to assess the quality of the control, we can test the model by comparing the pre-intervention trend lines: according to the model these should be parallel.

The trend lines can be estimated using one of a number of non-parametric regression techniques. The simplest is the moving average, which is estimated by fixing a time window of width  $2h + 1$  and then for each  $t$  calculating the mean of values between  $t - h$  and  $t + h$ . The method can be improved upon by using a weighted mean, where the weights are determined by proximity to  $t$ , or by using a weighted local linear regression. The latter is known as locally weighted scatter point smoothing (LOESS) and is recommended by Wasserman as a default non-parametric regression method (Wasserman 2006). The procedure can be implemented in R using the `loess` function (R Core Team 2018) or in Stata using the `lowess` command (StataCorp 2017).

In addition to graphically testing the common trend assumption, a statistical test can be conducted by regressing the pre-intervention differences on time, accounting for autocorrelation if necessary. If the data are consistent with the common trend model then there should be no evidence of a trend.

Unfortunately, neither the graphical nor the statistical test guarantees that a common trend model is appropriate. This is because the model requires not only that trend lines are parallel in the pre-intervention period, but also that they would be parallel in the post-intervention period were it not for the effect of the intervention. Because the latter requirement is untestable, a strong biological rationale for the control is necessary even if it passes graphical and statistical tests.

## 5 Autocorrelation

Even after accounting for the trend (either by segmented regression or by fitting a common trend model) and, where relevant, seasonality, the residuals often remain correlated in time. Typically, the correlation is greatest when residuals are close in time. This correlation, which is also known as autocorrelation, needs to be accounted for to obtain correct standard errors and hence correct p-values and confidence intervals. Two approaches are commonly used to account for autocorrelation, both of which can be used to account for many different patterns of autocorrelation.

One approach is to model the error as an autoregressive moving average (ARMA) process and estimate the resulting regression model by maximum likelihood (Box and Tiao 1975; Hyndman and Athanasopoulos 2017). A commonly used ARMA model is the first order autoregressive model where  $\epsilon_t = \phi\epsilon_{t-1} + \eta_t$  and it is assumed that the  $\eta_t$  are independent and normally distributed with zero mean. The model can be fitted using the `arima` function in R or `arima` command in Stata.

Another approach is to fit the regression model ignoring the autocorrelation and adjust the standard errors using the Newey-West method (Newey and West 1987). The method is an extension of the methodology used to obtain robust standard errors (Huber 1967; White 1980) and as such produces standard errors that are valid even when there is heteroscedasticity (heterogeneous error variance). As with the robust standard error methodology, the method is widely applicable, and can be used both in the context of Poisson regression and linear regression. The key assumption is that autocorrelation is zero beyond a certain lag (e. g. zero autocorrelation beyond lag two implies observations that are separated by more than two time units are uncorrelated). To obtain consistent standard error estimates, the cut-off should be small relative to the length of the time series. For example, Wooldridge suggests using the integer part of  $n^{1/4}$  (Wooldridge 2009). Newey-West standard er-

rors can be obtained using the `NeweyWest` function in R (package = `sandwich`) (Zeileis 2004) or `glm` command in Stata (see supplementary information for example Stata code and R code).

## 6 Poisson regression

When disease incidence is expressed as a count, one may wish to use Poisson regression instead of linear regression, especially when the counts are small.

Since Poisson regression is log-linear, we define a log-linear version of the common trend model as follows:

$$\begin{aligned}\log(\mu_{It}) &= \alpha_I + \beta x_t + g(t) \\ \log(\mu_{Ct}) &= \alpha_C + g(t),\end{aligned}\tag{8}$$

where  $\mu_{It}$  and  $\mu_{Ct}$  are the expected counts in the intervention and control series at time  $t$ .

In the appendix, we show that the intervention effect, in the above model can be estimated using Poisson regression with the control series included as an offset. Although the Poisson regression model is misspecified (it is not equivalent to eq. (8)), it gives a consistent estimate of the intervention effect and the confidence intervals are valid provided that robust standard errors (e. g. Newey-West standard errors) are used. A potential constraint is that the offset in a Poisson regression cannot include zero values because the Poisson distribution must have a positive mean. However, this problem can easily be circumvented by replacing zero values of  $y_{Ct}$  with a positive number that is close to zero.

## 7 Multiple control series

In some studies, several plausible controls may be available. This situation might arise, for example, if an intervention is introduced in one state and data are available from other states where the intervention was not introduced. We can combine information from the controls by fitting a model that assumes a common trend across all series.

Mathematically, the (linear) common trend model with multiple controls can be written as

$$y_{it} = \alpha_i + \beta x_{it} + g(t) + \epsilon_{it},\tag{9}$$

where  $i = 1, \dots, K$  denotes the series, one of which is the intervention series and  $K - 1$  of which are controls. The model can be fitted by combining the series into a single outcome vector  $y = (y_{10}, \dots, y_{1n-1}, \dots, y_{K0}, \dots, y_{Kn-1})$ , and conducting a linear regression that includes terms for series, intervention and trend. The series effect,  $\alpha_i$ , is modelled as a fixed effect and the common trend is modelled either by assuming a functional form, e. g. linear, or by using indicators for each time period.

Note that this approach is not limited to linear regression. If the series consist of count data a log-linear version of the common trend model can be fitted using Poisson regression.

Autocorrelation remains an issue whatever type of regression is used. Fortunately, however, it can be accounted for either by using an extension of the Newey-West methodology for multiple times series (panel data), or by treating the series as clusters and using cluster-robust standard errors (Bertrand, Duflo, and Mullainathan 2004). The latter approach works best when the number of clusters (series) is large but simulations suggest that it may also work well when there are few clusters (e. g. <10) provided that standard corrections are implemented (Cameron and Miller 2015; Bottomley et al. 2016).

Recently an alternative method has been proposed that involves using a synthetic control as the comparator (Abadie, Diamond, and Hainmueller 2010). The synthetic control is generated by comparing all linear combinations of the controls and finding the one that most closely resembles the intervention series in the pre-intervention period. Statistical inference is design based and the null distribution is derived by treating each control as the intervention series and recalculating the intervention effect under this assumption. The method therefore requires a reasonably large number of controls. In particular, at least 20 controls are required to have a possibility of achieving a p-value less than 0.05.

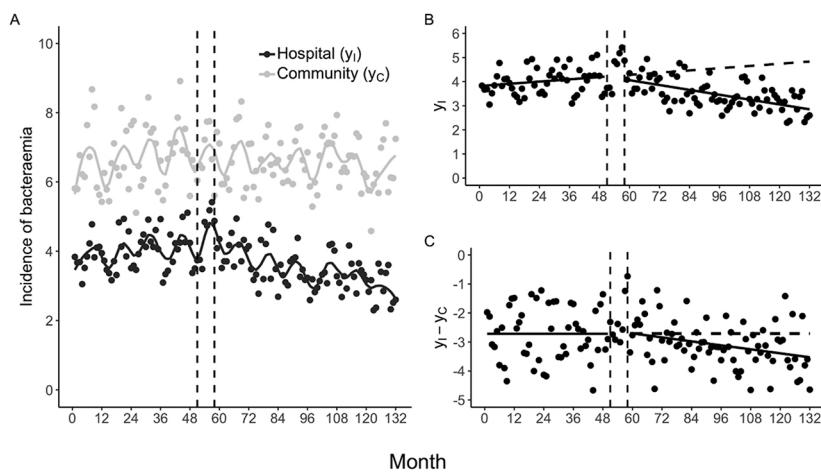
## 8 Including the control as a covariate

In many ITS analyses confounder adjustment is done by including the control series as a covariate—i. e.  $y_{It}$  is regressed on  $y_{Ct}$  and  $x_t$  (Bruhn et al. 2017; Clancy et al. 2002). Superficially this seems reasonable since the control series can be considered a proxy for the combined effect of the confounders. However, it can be seen from eq. (6) that  $y_{Ct}$  and  $g(t)$  are not equivalent;  $y_{Ct}$  includes error whereas  $g(t)$  does not. Thus, as is well known from the theory of measurement error (Carroll et al. 1995), including  $y_{Ct}$  in the model instead of  $g(t)$  causes attenuation bias: the estimated coefficient associated with  $y_{Ct}$  is, on average, closer to zero than the coefficient associated with  $g(t)$  (N.B. under the common trend model the coefficient associated with  $g(t)$  is assumed to be equal to 1). Hence, because of this attenuation, the degree of confounding is underestimated, and the intervention effect estimate is biased. We therefore do not recommend this approach.

## 9 Examples

### 9.1 Example 1: Impact of an infection control programme on gram-negative rod bacteraemia

Goto et al. evaluated a hospital-based infection control programme in 130 Veterans Health Administration facilities in the US using data collected between Jan 2003 and Dec 2013 (Goto et al. 2016). The programme was originally designed to reduce the incidence of Methicillin-Resistant *Staphylococcus aureus* infections, but some components of the programme—e. g. efforts to improve hand hygiene—were thought to be more widely effective. To test this hypothesis, Goto et al. evaluated the intervention impact against hospital-acquired gram-negative rod bacteraemia—specifically cases due to *Escherichia coli*, *Klebsiella* species and *Pseudomonas aeruginosa*—using community-acquired bacteraemia as a control. The data, which consist of monthly incidence rates of hospital-acquired bacteraemia per 10,000 person days (mean = 3.7, range 2.3–5.4) and monthly rates of community-acquired bacteraemia per 10,000 person months (mean = 6.6, range 4.6–8.9), are shown in Figure 2A together with LOESS trend lines.



**Figure 2:** (A) Monthly incidence of hospital-acquired bacteraemia (per 10,000 person days) and community-acquired bacteraemia (per 10,000 person months) before and after the implementation of an infection control programme. Trend curves estimated by LOESS (locally weighted scatter point smoothing). Vertical dashed lines represent the period over which the intervention was rolled out (Mar–Oct 2007). (B) Segmented regression model: the intervention effect is estimated by extrapolating the pre-intervention trend and comparing with the observed incidence post-intervention. (C) Common trend model: the intervention effect is estimated by computing the difference between the two series and comparing the differences pre and post intervention. In both models (B and C) it is assumed the intervention effect increases linearly with time.

Three estimates of the intervention effect at the end of the study were calculated.

The first was an unadjusted estimate (i. e. it does not account for trend). The second was obtained by fitting a segmented regression model (Figure 2B), and the third was obtained by differencing (Figure 2C). All models included an indicator for the transitional period (Mar–Oct 2007) during which the intervention was not fully implemented, and the intervention effect was assumed to increase linearly over time. The segmented regression model additionally included calendar month as a categorical variable to account for seasonality. Standard errors

were adjusted for autocorrelation up to lag three using the Newey-West method. The data are provided as supplementary information together with R code and Stata code for conducting the analysis.

According to the segmented regression model, the intervention reduced the incidence of gram-negative bacteraemia by 1.98 (95 % CI 1.30, 2.67) cases per 10,000 person days by the end of the study (Table 1) and, according to the common trend model, it reduced the incidence by 0.81 cases (95 % CI 0.40, 1.23).

**Table 1:** Estimated impact of a hospital infection control programme on the incidence of gram-negative rod bacteraemia.

	<i>Difference*</i>	<i>95 % CI</i>
No adjustment	-1.09	-1.38, -0.80
Segmented regression model	-1.98	-2.67, -1.30
Common trend model, differencing	-0.81	-1.23, -0.40

\*Estimated difference in incidence per 10,000 person days at study end (Dec 2013).

**Table 2:** Estimated impact of rotavirus vaccine introduction.

	<i>RR*</i>	<i>95 % CI</i>
No adjustment	0.939	0.919, 0.959
Segmented regression model	0.962	0.926, 0.998
Common trend model, control used as offset	0.964	0.951, 0.977

\*Rate ratio for change in rotavirus-positive diarrhoea incidence per month post intervention.

There is a substantial difference between these estimates because the two models make opposite predictions about the counterfactual no intervention scenario: the segmented regression model predicts increasing incidence in the post-intervention period, whereas the common trend model predicts decreasing incidence.

Which estimate is better? In general, if the control satisfies the requirements of the common trend model then the estimate from the common trend model is better because it incorporates the additional information provided by the control.

In this example, biological justification for the common trend model is strong. Firstly, the control is unlikely to be affected by the hospital-based intervention. Secondly the two series are likely to share confounders because they share the same population, same outcome and same mode of surveillance. As well as having a strong biological justification, the model is supported by the fact that the pre-intervention trend lines are approximately parallel (Figure 2A), and by the fact that there is no evidence of linear trend in the pre-intervention differences ( $p = 0.87$ , see also Figure 2B).

Overall, the common trend model would therefore seem to be plausible. Nonetheless the intervention effect estimate from this model is not immune from bias. One possible bias, which was identified by Goto et al. is that the proportion of community-acquired bacteraemia treated in an outpatient setting, as opposed to a hospital setting, may have changed over the course of the study. A change of this kind that is limited to just one of the two series violates the shared confounders assumption.

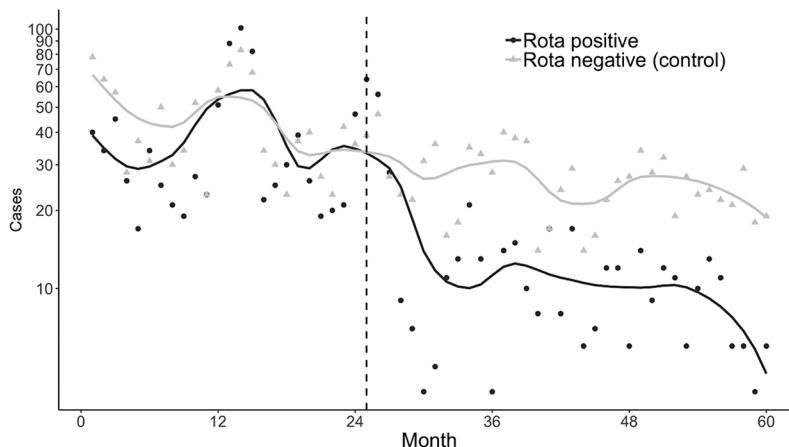
## 9.2 Example 2: Rotavirus vaccine introduction

Rotavirus vaccine was introduced into the Ghanaian childhood vaccination programme in April 2012. To evaluate the impact of its introduction, Armah et al. used surveillance collected between January 2010 and December 2014 from two large tertiary care hospitals (Armah et al. 2016). During surveillance, all children  $< 5$  years admitted with diarrhoea were tested for rotavirus. In their analysis, Armah et al. used rotavirus-negative diarrhoea cases as the control.

As in Example 1, we estimated vaccine impact using both the segmented regression model and common trend model. However, because the data consist of counts, the models were fitted using Poisson regression instead of linear regression.

Both models included an intervention  $\times$  time interaction but no intervention main effect—i. e. we assumed no vaccine effect immediately after introduction followed by an increasing effect over time. The segmented regression model additionally included month as a categorical variable to account for seasonality. In both models, Newey-West standard errors were used to account for autocorrelation up to lag two. The data as well as R code and Stata code for the analysis are provided as supplementary information.

The data, which consist of monthly rotavirus-positive counts (mean = 23, range: 4–101) and rotavirus-negative counts (mean = 34, range: 14–83), are plotted on the log-scale in Figure 3 and on the original scale in Supplementary Figure 1.



**Figure 3:** Monthly rotavirus-positive and rotavirus-negative diarrhoea cases in children < 5 years before and after the introduction of rotavirus vaccine (vertical dashed line). The counts and LOESS trend lines are presented on a log-scale.

Estimates of vaccine impact are similar for the segmented regression and common trend model (Table 2). Based on the segmented regression model, incidence of rotavirus-positive diarrhoea was reduced by 4.1% (95% CI 0.7, 7.5) per month and based on the common trend model it was reduced by 3.6% (95% CI 2.3, 4.8) per month. Interestingly, the confidence interval was narrower for the estimate from the common trend model, which suggests that this model explains more of the variation in incidence.

Both confidence intervals were calculated using Newey-West standard errors and therefore account for overdispersion and autocorrelation. We can see the importance of using the correct standard errors by comparing the Newey-West standard errors with the naïve standard errors from the Poisson regression. For the segmented regression model, the Newey-West standard error is almost three times as large as the naïve standard error (0.0180 versus 0.0067), and for the common trend model it is almost twice as large (0.0067 versus 0.0035). Thus, in both models the naïve standard error substantially underestimates the uncertainty.

Broadly speaking, the assumptions of the common trend model seem reasonable: the control is unaffected by the intervention (the vaccine does not protect against rotavirus negative diarrhoea), and it is plausible that both series share confounders. The argument for shared confounders is reasonably strong because the method of surveillance is the same and, although the pathogens differ between the series, they share the faecal-oral mode of transmission.

Despite the reasonably strong biological justification, when the model was tested through the inclusion of time as a covariate, there was evidence of model misspecification ( $p < 0.001$ ). Consistent with this result, the pre-intervention log trend lines are not parallel on the log scale (Figure 3), though the lines suggest only modest misspecification.

In this example it is unclear whether we should prefer the common trend model or the segmented regression model. Fortunately the choice is not critical here because the trend in the control is approximately log-linear and the models are therefore roughly equivalent.

## 10 Conclusions

ITS are unusual in that the degree of confounding is immediately apparent from a time series plot of the data. When there is no trend, and hence no confounding, ITS can be analysed by comparing pre and post intervention means, accounting for autocorrelation if necessary. On the other hand, when a trend is apparent the analysis must allow for confounding.

The segmented regression model and the common trend model can both be used to adjust for confounding. The latter is potentially the more convincing of the two models, but only if the control has a strong biological justification and is statistically consistent with the model. In studies where no control is available, or where the quality of the control is in doubt, the segmented regression model may provide a useful alternative. This model, however, is based on the strong assumption that the pre-intervention trend would have persisted in the absence of intervention. It should therefore be used cautiously, particularly if the impact of trend over the study period is large relative to the intervention effect.

## Acknowledgements

We would like to thank Michihiko Goto for sharing the gram-negative rod bacteraemia data and Ben Lopman for sharing the rotavirus data. This work was supported by an award that is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, which is also part of the EDCTP2 programme supported by the European Union (grant reference MR/K012126/1).

## Appendix

### A Using Poisson regression to estimate the common trend model

Here we show that the intervention effect parameter in a log-linear common trend model can be estimated by including the control series as an offset in a Poisson regression. This regression model is not equivalent to the log-linear common trend model (i.e., it is misspecified). However, it turns out, as shown below, that this model can nonetheless be used to obtain a consistent estimate of the intervention effect.

Under the log-linear common trend model, and assuming a constant intervention effect, the means of the intervention series,  $\mu_{It}$ , and the control series,  $\mu_{Ct}$ , are denoted by

$$\begin{aligned}\log(\mu_{It}) &= \alpha_I + \beta x_t + g(t) \\ \log(\mu_{Ct}) &= \alpha_C + g(t),\end{aligned}$$

where  $g(t)$  represents the common trend.

However, if we fit a Poisson regression that includes the control series as an offset, then we are assuming the intervention series follows a Poisson distribution with mean

$$\log \tilde{\mu}_{It} = \tilde{\alpha}_I + \tilde{\beta} x_t + \log(y_{Ct} + \delta),$$

where the positive constant  $\delta$  is included to ensure  $\log \tilde{\mu}_{It}$  is defined when  $y_{Ct} = 0$ .

The log likelihood for this assumed model is

$$l(\tilde{\beta}, \tilde{\alpha}_I) = \sum_t \log[e^{-\tilde{\mu}_{It}} \tilde{\mu}_{It}^{y_{It}} / y_{It}!] \propto \sum_t -(y_{Ct} + \delta) e^{\tilde{\alpha}_I + \tilde{\beta} x_t} + y_{It} [\tilde{\alpha}_I + \tilde{\beta} x_t + \log(y_{Ct} + \delta)],$$

and, assuming  $y_{Ct}$  and  $y_{It}$  are independent, the parameters being estimated are the values of  $\tilde{\beta}$  and  $\tilde{\alpha}_I$  that maximise the average log likelihood

$$E[l(\tilde{\beta}, \tilde{\alpha}_I)] = \sum_t - (e^{\alpha_C + g(t)} + \delta) e^{\tilde{\alpha}_I + \tilde{\beta} x_t} + e^{\alpha_I + \beta x_t + g(t)} (\tilde{\alpha}_I + \tilde{\beta} x_t + E[\log(y_{Ct} + \delta)]),$$

where the expectation is taken with respect to the true model (Pawitan 2013).

By solving  $\frac{\partial E[l(\tilde{\beta}, \tilde{\alpha}_I)]}{\partial \tilde{\beta}} = 0$  and  $\frac{\partial E[l(\tilde{\beta}, \tilde{\alpha}_I)]}{\partial \tilde{\alpha}_I} = 0$  it is apparent that as  $\delta \rightarrow 0$  the maximum is achieved when  $\tilde{\alpha}_I = \alpha_I - \alpha_C$  and  $\tilde{\beta} = \beta$ . Thus the Poisson regression model can be used to obtain a consistent estimate of the intervention effect.

Although the estimate is consistent, standard Poisson regression confidence intervals and p-values are not valid because the model is misspecified. Even if we assume both the intervention and control series follow a Poisson distribution, the variance of the residuals obtained from the Poisson regression will still be greater than that predicted by the Poisson model. This is because by using the control series to account for trend we are including additional error (control series = trend + error) that is not accounted for in the standard Poisson regression analysis. Fortunately, the problem can be overcome by using robust standard errors or, if it is necessary to account for autocorrelation, Newey-West standard errors.

## References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control programme. *Journal of the American Statistical Association*, 105(490):493–505.

Angrist, J. D., and Pischke J.-S. (2009). Chapter 5. In: *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Armah, G., Pringle, K., Enweronu-Laryea, C. C., Ansong, D., Mwenda, J. M., Diamenu, S. K. et al. (2016). Impact and effectiveness of monovalent rotavirus vaccine against severe rotavirus diarrhea in Ghana. *Clinical Infectious Diseases*, 62(Suppl 2):S200–S207.

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

Bottomley, C., Kirby, M. J., Lindsay, S. W., and Alexander, N. (2016). Can the buck always be passed to the highest level of clustering? *BMC Medical Research Methodology*, 16:29.

Box, G. E. P., and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349):70–79.

Bruhn, C. A. W., Hetterich, S., Schuck-Paim, C., Kürüm, E., Taylor, R. J., Lustig, R. et al. (2017). Estimating the population-level impact of vaccines using synthetic controls. *Proceedings of the National Academy of Sciences of the United States of America*, 114(7):1524–1529.

Cameron, A. C., and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *The Journal of Human Resources*, 50(2):317–372.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). Chapter 2. In: *Measurement Error in Nonlinear Models*, D.R. Cox, D.V. Hinkley, N. Keiding, N. Reid, D.B. Rubin and B.W. Silverman (Eds.), 21–39. London: Chapman & Hall.

Clancy, L., Goodman, P., Sinclair, H., and Dockery, D. W. (2002). Effect of air-pollution control on death rates in Dublin, Ireland: An intervention study. *Lancet*, 360(9341):1210–1214.

Fowler, S., Webber, A., Cooper, B. S., Phimister, A., Price, K., Carter, Y. et al. (2007). Successful use of feedback to improve antibiotic prescribing and reduce Clostridium difficile infection: A controlled interrupted time series. *The Journal of Antimicrobial Chemotherapy*, 59(5):990–995.

Goto, M., O'Shea, A. M. J., Livorsi, D. J., McDanel, J. S., Jones, M. M., Richardson, K. K. et al. (2016). The effect of a nationwide infection control program expansion on hospital-onset gram-negative rod bacteremia in 130 veterans health administration medical centers: An interrupted time-series analysis. *Clinical Infectious Diseases*, 63(5):642–650.

Huber, P. J. (1967). The Behaviour of Maximum Likelihood Estimates Under Nonstandard Conditions. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1. Berkeley, CA: University of California Press, 221–233.

Hyndman, R. J., and Athanasopoulos, G. (2017). Chapter 9. In: *Forecasting: Principles and Practice*. 2nd Edition. Melbourne, Australia: OTexts. <https://otexts.org/fpp2/>.

Jandoc, R., Burden, A. M., Mamdani, M., Lévesque, L. E., and Cadarette, S. M. (2015). Interrupted time series analysis in drug utilization research is increasing: Systematic review and recommendations. *Journal of Clinical Epidemiology*, 68(8):950–956.

Lopez Bernal, J., Cummins, S., and Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, 46(1):348–355.

Mackenzie, G. A., Hill, P. C., Jeffries, D. J., Hossain, I., Uchendu, U., Ameh, D. et al. (2016). Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: A population-based surveillance study. *The Lancet Infectious Diseases*, 16(6):703–711.

Morgan, O. W., Griffiths, C., and Majeed, A. (2007). Interrupted time-series analysis of regulations to reduce paracetamol (acetaminophen) poisoning. *PLoS Medicine*, 4(4):e105.

Newey, W. K., and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Pawitan, Y. (2013). Chapter 13. In: *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, 365–383. Oxford: Oxford University Press.

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

StataCorp. (2017). Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

Wagner, A. K., Soumerai, S. B., Zhang, F., and Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics*, 27(4):299–309.

Wasserman, L. (2006). Chapter 5. In: *All of Nonparametric Statistics* (Springer Texts in Statistics), G. Casella, S. Fienberg, and I. Olkin (Eds.), 61–123. New York: Springer-Verlag.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

Wooldridge, J. M. (2009). Chapter 12. In: *Introductory Econometrics: A Modern Approach*. 4th Edition, 408–442. Mason, OH: Thomson/South-Western.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

**Supplementary Material:** The online version of this article offers supplementary material (DOI:<https://doi.org/10.1515/em-2018-0010>).