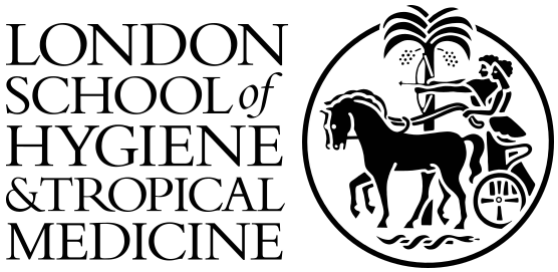
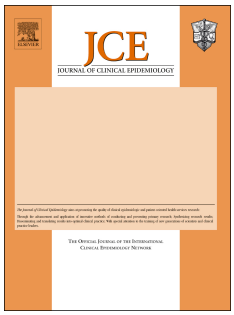


Accepted Manuscript



A Methodological Framework for Model Selection in Interrupted Time Series Studies

J. Lopez Bernal, S. Soumerai, A. Gasparrini

PII: S0895-4356(17)31411-7

DOI: 10.1016/j.jclinepi.2018.05.026

Reference: JCE 9677

To appear in: *Journal of Clinical Epidemiology*

Received Date: 8 January 2018

Revised Date: 17 May 2018

Accepted Date: 30 May 2018

Please cite this article as: Lopez Bernal J, Soumerai S, Gasparrini A, A Methodological Framework for Model Selection in Interrupted Time Series Studies, *Journal of Clinical Epidemiology* (2018), doi: 10.1016/j.jclinepi.2018.05.026.

A METHODOLOGICAL FRAMEWORK FOR MODEL SELECTION IN INTERRUPTED TIME SERIES STUDIES

J. Lopez Bernal,[1,2] S. Soumerai,[2] A. Gasparrini[1]

1. Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, London, United Kingdom
2. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, United States of America

Declarations of interest: none

Funding: This study was funded by a UK Medical Research Council Population Health Scientist Fellowship awarded to JLB – Grant Ref: MR/L011891/1,

ABSTRACT:

Interrupted time series is a powerful and increasingly popular design for evaluating public health and health service interventions. The design involves analysing trends in the outcome of interest and estimating the change in trend following an intervention relative to the counterfactual (the expected ongoing trend if the intervention had not occurred). There are two key components to modelling this effect: first, defining the counterfactual; second, defining the type of effect that the intervention is expected to have on the outcome, known as the impact model. The counterfactual is defined by extrapolating the underlying trends observed before the intervention to the post-intervention period. In doing this, authors must consider the pre-intervention period that will be included, any time varying confounders, whether trends may vary within different subgroups of the population and whether trends are linear or non-linear. Defining the impact model involves specifying the parameters that model the intervention, including for instance whether to allow for an abrupt level change or a gradual slope change, whether to allow for a lag before any effect on the outcome, whether to allow a transition period during which the intervention is being implemented and whether a ceiling or floor effect might be expected. Inappropriate model specification can bias the results of an interrupted time series analysis and using a model that is not closely tailored to the intervention or testing multiple models increases the risk of false positives being detected. It is important that authors use substantive knowledge to customise their interrupted time series model *a priori* to the intervention and outcome under study. Where there is uncertainty in model specification, authors should consider using separate data sources to define the intervention, running limited sensitivity analyses or undertaking initial exploratory studies.

What is new?

- Interrupted time series is one of the strongest quasi-experimental designs for evaluating the effect of health interventions. However, this design requires careful specification of several modelling features, for which little guidance is offered in the literature
- We demonstrate how incorrectly modelling either the trend or the type of impact model can generate misleading results and offer a methodological framework for making modelling choices in interrupted time series analyses.
- Researchers must be transparent in providing a clear and objective justification for the choices they make in defining an interrupted time series model which is tailored to the specific intervention and outcome under study.

INTRODUCTION:

Interrupted time series (ITS) has become a core study design for the evaluation of public health interventions and health policies.(1) The design takes advantage of natural experiments whereby an intervention is introduced at a known point in time and a series of observations on the outcome of interest exist both before and after the intervention. The effect of the intervention is estimated by examining any change following the intervention compared to the 'counterfactual', represented by the expected ongoing trend in the absence of the intervention (Figure 1).(2) ITS involves a pre-post comparison, controlling for the counterfactual baseline trend, within the same population; therefore, it can be used in situations where no control population is available.(3, 4) This also has the advantage that selection bias and confounding due to group differences, which threaten the reliability of non-randomised controlled designs, are rarely a problem in ITS studies.(2, 3) Furthermore, because ITS incorporates the underlying trend it controls for short term fluctuations, secular trends and regression to the mean.(3, 4)) The basic ITS design also has limitations; for example there is the potential for *history bias* whereby other events concurrent to the intervention may be responsible for an observed effect. Also, *instrumentation* effects can occur if there are changes in the way the outcome is measured over time.(3) Previous studies have described these strengths and limitations of ITS in more detail and have provided guidance on its application.(2, 4, 5) Furthermore, methodological publications have discussed effective approaches for limiting the risk of history bias, including controlled ITS designs and multiple baseline designs.(6-8)

One area that has not been covered in detail in the existing literature is how researchers should approach specifying the ITS model used in the analysis. As discussed above, the ITS design involves making a comparison between the outcome observed following the intervention and the counterfactual. This comparison reduces to two key questions that define the estimated effect of the intervention.(2) First, how is the counterfactual defined? This involves modelling the pre-intervention trend. Second, how is the impact model of the intervention defined? That is, what type of effect do we hypothesise that the intervention will have on the outcome (such as whether the effect is gradual or abrupt, immediate or lagged)? This involves parameterizing the effect of the intervention relative to the counterfactual. Multiple alternative approaches exist to defining the counterfactual and the intervention impact model and inappropriate model selection could bias results, yet ITS studies often fail to provide a clear justification for their choice of modelling approach.(9)

In this paper we suggest approaches to ensure model specification is objective and appropriate to the intervention and outcome under investigation. The first section discusses the factors that contribute to defining the counterfactual and the second the factors that contribute to defining the impact model. For each of these sections we use illustrative examples from a recent ITS study of the impact of major reforms to the English National Health Service on hospital activity (described in Box 1)(10) to highlight the pitfalls of incorrect model specification and then provide a framework for a suggested approach to select the model. Finally, we also discuss sensitivity analysis and other approaches to dealing with uncertainty in model specification.

>> insert Figure 1: the interrupted time series design
Solid line = modelled trend; dashed line = counterfactual; vertical line = intervention implementation. This shows a step decrease and decrease in the slope following the intervention.<<

>>insert Box 1<<

DEFINING THE COUNTERFACTUAL

A key step in ITS analysis is to predict how the outcome would have continued over time if no intervention had been implemented, referred to as the 'counterfactual' scenario. It is not possible to observe the intervention both being implemented and not being implemented in the same population at the same time. The true counterfactual is therefore never known and therefore inferring causality is rarely possible. Evaluation design centres on creating the best approximation of

the true counterfactual. This requires both the study population and the counterfactual to share the same characteristics as far as possible In ITS studies this involves modelling the underlying trend in the outcome of interest within a single population. Since the effect of an intervention is a measure of its deviation from the counterfactual it is essential that the counterfactual is defined as accurately as possible. Incorrect definition of the counterfactual can lead to either overestimation or underestimation of the intervention's effect. When estimating the baseline trend, it is necessary to consider both the data that will be included and the way the trend is modelled.

The pre-intervention time period

Routine data sources now often span many years; weekly or monthly time series with hundreds of data points are possible. For example, Swedish data on maternal mortality dates back to the mid eighteenth century.(12) Trends may change over time, therefore, how the counterfactual is predicted can vary depending on the range of data that is included. If the time period is too short, this increases uncertainty as there may be too little data to model the trend.(13) If a very long pre-intervention period is included, there is a risk that trends may have historically differed from current trends which raises doubts about the validity of the comparison. The minimum number of data points is a decision driven by the statistical requirements for the analysis and will depend on the variability of the data and the type of statistical model used. For example, to model a seasonal effect, a minimum of 12 data points will be required if using monthly data and complex autoregressive moving average (ARIMA) models often requires hundreds of data points.(2-4) The maximum amount of data to include is much more of a researcher driven decision and there are therefore risks that the data range can be manipulated to produce different outcomes. Researchers may choose to include the full dataset; nevertheless, the selection should focus on defining a valid counterfactual for the post-intervention measurements. Therefore, periods characterised by external factors affecting the underlying trends, such as changes to data collection procedures or previous interventions targeting the outcome of interest, should be excluded, or the effects of these factors appropriately modelled. Researchers should adopt an objective approach *a priori* to selecting the data which is to be included in the study and any decision to restrict the range of data used in the analysis should be clearly justified and reported transparently.

Expected changes in the trend

Changes may be expected in the trend that are unrelated to the intervention and these should be taken into account when defining the counterfactual. For example, there may be a limit to how long a trend could continue increasing or decreasing if the outcome is constrained by other factors, this can result in a floor or ceiling effect.(14) For example, vaccine uptake is limited to a level below 100% due to a small proportion of patients having allergies or other contraindications. Conversely, hospital length of stay could have a floor effect which will differ depending on the type of patient, disease or treatment being evaluated. The possibility of floor or ceiling effects should be anticipated

a priori and incorporated into the ITS model. Possible approaches to dealing with this include: allowing for a trend change at the anticipated floor or ceiling, restricting the analysis to above or below the floor or ceiling or transforming the data, for example to a log scale.

Time-varying confounders

Under a simple linear ITS model it is assumed that population characteristics associated with the outcome either remain relatively constant throughout the study period, or that they change only slowly and are captured by the underlying linear trend. This may not be the case and irregular fluctuations in the baseline trend may be explained by changes over time in covariates associated with the outcome.(15) Epidemiologists are accustomed to identifying potential confounding variables *a priori* and using multivariate regression models to adjust for these potential confounders.(16) A similar approach can be adopted using segmented regression for ITS studies by including potential time-varying confounders (explanatory variables that could affect the outcome and may change substantially and unpredictably over time).(2) Examples might include meteorological events,(17) population age distribution,(18) ethnicity or levels of deprivation.(19) Adjusting for time-varying confounders may result in irregular trends becoming linear thus conforming to this basic ITS assumption. Seasonality, can also be considered a time varying confounder and accounts for fluctuations in many health outcomes, such as infectious disease rates or hospital admissions.(2, 20) A range of methods exist for controlling for seasonality in time-series regression models which have been described in more detail elsewhere.(2, 21)

Multiple groups

ITS studies commonly use aggregate outcome data for the whole study population and define the underlying trend based on this aggregated data. This assumes that there is a uniform trend within the whole population. Nevertheless, different sub-groups or even different individuals within the study population may follow different trends that can result in irregular, non-linear trends when the data is aggregated together. More sophisticated ITS models can allow for these different trends and should be considered where sub-group or individual level data is available. For example, Steinbach et al evaluated the impact of changes to street lighting on casualties from road traffic collisions at night and used data on trends from individual road segments.(22)

Linearity

The above factors can be defined *a* priori by the researcher and may explain any non-linear trends. Nevertheless, the assumption linearity should be checked both by visual inspection of the data and residuals and through statistical goodness of fit tests such as the Pearson test.(15) If a linear trend exists, prediction of the counterfactual, and thus the isolation of an intervention is relatively

straightforward. However, if the baseline trend is non-linear it can be harder to predict the counterfactual and difficult to disentangle intervention effects.(15) In particular, although flexible methods exist to describe non-linearities in regression models,(23) in the ITS framework it is often the case that the estimates are highly sensitive to the degree of smoothing, so that the intervention effect cannot be distinguished from underlying fluctuations, or is artifactually created by the extreme flexibility of the model.(24) Researchers should therefore be cautious about using ITS if the data follows a non-linear baseline trend which cannot be explained by other factors. Furthermore, it should be recognised that introducing non-linear terms post-hoc is a data driven approach and the underlying reason for these trends is unknown. It therefore must be assumed that the unknown underlying variables that explain this trend in the outcome follows the same pattern in both the pre-intervention and the post-intervention period.

Illustrative example

Figure 2 shows a time series from our case study (described in Box 1). Here we look at the impact of the policy on the number of outpatient specialist visits in England. Our full dataset had data on all outpatient visits between 2007 and 2015. To illustrate how changing the way that the pre-intervention trend is defined can affect the results of an ITS study we demonstrate different approaches to defining the trend. Each of the models allows for a change in the slope of the trend following the intervention. Figures 2a and 2b demonstrate the effects of choosing different pre-intervention time periods. In figure 2a the complete data series is used and there is no significant effect; in Figure 2b the data is restricted prior to April 2010 and there is a clear increase in the rate of specialist visits following the policy. Figures 2c and 2d show two different non-linear models, the second allowing a greater degree of smoothing in the data series. Again the results of the models differ with Figure 2c showing an increase in the outcome following the policy, whilst Figure 2d, which is a more flexible model, shows no change. The differing effects seen in these four models highlight the need for careful model selection in order to accurately estimate the effect of the intervention. In this example, in fact, a data quality issue was identified whereby a misclassification resulted in possible errors in outpatient numbers prior to 2010, therefore it was inappropriate to include data prior to this point.(25) Figure 2b was therefore considered to be the most appropriate model *a priori*.

>>insert Figure 2: Four different approaches to modelling the trend

a: simple linear trend; b: excluding data prior to April 2010; c: non-linear allowing one inflection point; d: non-linear trend allowing two inflection points. Dots= data observations; solid line = fitted model; dashed line = counterfactual; vertical line = policy implementation.<<

Framework for defining the trend

Box 2 outlines a suggested approach to defining the trend in the outcome.

>>insert Box 2<<

DEFINING THE IMPACT OF THE INTERVENTION

As described above the effect estimate in an ITS study is a measure of the level and/or trend change in the outcome following an intervention. We have discussed how the trend is defined, the next step is to define how the intervention and its potential impacts are modelled. Different interventions can have different impacts on an outcome: for example, mandatory helmet legislation might be expected to have an abrupt effect on cycle head injuries, whereas an educational programme on cycle safety might be expected to have a more gradual effect.(26, 27) Likewise, different outcomes can be expected to respond differently to the same intervention, for example policies restricting alcohol availability may be expected to have a relatively rapid effect on alcohol related road traffic casualties but a longer lag before any effect on liver cirrhosis.(28) Different model parameters can be used to allow different effects to be expressed following the intervention. Less specific models can be used which allows a whole range of possible intervention effects to be detected. Nevertheless, this also increases the likelihood of false positive effects being detected due to other confounding events, data errors or chance resulting in type I errors.(2) It is therefore preferable that researchers select a more precise impact model for the intervention *a priori*, taking into account substantive knowledge on the nature of the intervention and how it was implemented, as well as the outcome of interest. There are a number of factors to consider in defining the impact model, including: whether the impact will be abrupt or gradual, whether any lag is expected, whether a ceiling or floor effect can be expected, and whether there was a transition period during which the intervention was implemented. These are discussed below.

Abrupt or gradual effects

The effect of the intervention may either be abrupt or gradual or both. An abrupt effect would result in an immediate or rapid change in the level of the outcome – observed as a step change in the time series (figure 3a(i)). A gradual effect would result in the level of the outcome changing slowly over time – observed as a change in the gradient of the trend (a slope change) (figure 3a(ii)). An intervention that is introduced at a precise point in time with an outcome that could respond rapidly would be expected to follow a step change model, for example the impact of restricting Medicaid funding for prescriptions on the number of prescriptions filled per month.(29) Conversely, interventions that results in a more gradual process of change with an outcome that could respond at a variable rate would be expected to follow a slope change model. This includes complex health policies that require large scale institutional changes such as the example in our case study (Box 1).(10) It is also important to consider the time interval of the time series when deciding whether to include a step change and/or a slope change model, a gradual slope change on a weekly time scale may appear as a step change on an annual time scale. It is important to underline that these two types of effects are not mutually exclusive, interventions may lead to an initial step change followed

by a more gradual slope change in either direction.(4, 29) Nevertheless, modelling both can be problematic and prone to artefacts in the absence of a strong signal in the data. This is particularly an issue where both exhibit a small effect in the same direction as they 'rob' each other of significance.

Immediate or lagged effects

Following the intervention, the effect on the outcome, whether it is a step change or a slope change, or both, may occur immediately or may be delayed (figure 3b). This typically depends on the outcome and how rapidly it could respond to the intervention. Many public health interventions are ultimately targeting disease morbidity or mortality, but they often do so through behaviour changes. An intervention might have an immediate impact on the behaviour but a lagged effect on any health outcome. For example, tobacco control policies might be expected to have an immediate impact on maternal smoking levels but a lag of approximately nine months before any impact on small for gestational age births and a much longer lag before any impact on lung cancer.(30)

Transition period

Interventions may be introduced over a prolonged period of time or may result in a short period of adjustment before the lasting impact on the outcome is manifested.(10) Furthermore, effects can begin prior to the intervention as an anticipatory response to a new policy.(31) This can be accounted for by dividing the time series into three phases: a pre-intervention phase, a transition period (which may or may not be included in the analysis) and a post-intervention phase (figure 3c).(20, 32) For example, Landrigan et al evaluated the impact of introducing a hospitalist system (employing physicians with a primary focus on caring for hospitalised inpatients) on length of stay in a paediatric hospital.(33) They allowed a transition from when the policy was first announced to when hospitalists fully took over patient care in order to allow for the effects of gradual system changes to prepare for the new policy.

Waning effects

The effect of an intervention may change over time. In particular, there may be a more notable effect of the intervention when the intervention is first introduced but with the effect waning over time. This is often due to greater publicity of the intervention when it was first implemented, as was observed when examining the effect of widely publicised warnings about a possible increased risk of suicidality with antidepressant use on antidepressant use.(34) If the initial effect is expected to be abrupt, this could be modelled as a step change to model the effect and a slope change to model the waning of this effect. If a gradual effect is expected, a non-linear term may be included to model both the effect and the waning (Figure 3d).(34)

>>insert Figure 3: Interrupted time series impact models

X-axis represents time, y-axis represents the outcome. The vertical blue line is the time when the intervention was implemented; the red line is the ITS regression model. a(i) abrupt step change effect, a(ii) gradual slope change effect; b(i) immediate effect following the intervention, b(ii) lagged effect; c(i) intervention at a specific time point, c(ii) transition period (blue box) excluded from the model; d(i) waning effect following a step change, d(ii) gradual effect with gradual waning.<<

Illustrative example

Figure 4 is again taken from our case study evaluation of the GP commissioning policy. This time we look at the effect in Wales, a control population.(10) A control series can be added to an ITS study to help control for confounding events occurring around the time of the intervention.(8) Because the control population was not subject to the intervention, we do not expect to see an effect in the control series.(8) We demonstrate three approaches to modelling the impact of the intervention: In Figure 4a we use one of the most commonly used impact models which allows for a step and slope change at the point of the intervention and is therefore less specific about the intervention effect, here there is no significant change following the intervention. However, we have not taken into consideration either our knowledge of the intervention nor how we consider *a priori* that it would impact upon the outcome if effective. In Figure 4b, we instead select what we would consider *a priori* to be the most appropriate model. We know that the policy was enacted in April 2012 but that there was then a period of one year during which the new GP-led Clinical Commissioning Groups worked alongside the existing commissioners, we therefore allow a one year transition period. We also do not expect the policy to have an abrupt effect as existing secondary care contracts would only expire gradually and complex institutional changes would be required to establish new models of care, therefore a slope change model was selected. Again, there is no significant effect of the intervention. Finally, we select a model that provides the best fit to our data (using the Akaike Information Criterion),(35) here we find a highly significant reduction in both the level and the slope associated with the intervention. In this example, however, we know that the intervention did not cause the level and slope change as this was taken from a control population that did not receive the intervention. This highlights the danger that using a data driven approach to select the impact model can lead to spurious results due to factors other than the intervention .

>>insert Figure 4: Four different approaches to modelling the impact of the intervention

a) Step and slope change model; b) slope change only with a one year intervention phase; d) step and slope change with a one year intervention phase.<<

Framework for defining the impact model

Box 3 outlines a suggested approach to modelling the impact of the intervention.

>>insert Box 3<<

DEALING WITH UNCERTAINTY IN MODEL SELECTION

So far in this paper we have emphasised the need to carefully define the pre-intervention trend and the intervention impact model according to the specific intervention, outcome and data being used in the study. Often, however, the single best approach is difficult to define, in particular for novel interventions that have not previously been studied and when analysing the public health effects of unplanned events. Below we discuss some approaches to dealing with uncertainty in model selection:

Modelling unplanned events

While ITS is most commonly used for studies of pre-meditated health interventions or health policies, it can also be used to evaluate the health impacts of unplanned events.(36) If the timing of the event is clearly defined, for example: a natural disaster,(37) a chemical spill,(38) or a terrorist attack,(39) then the same modelling process can be used as for planned interventions. Nevertheless, the timing of many unplanned events is harder to define including: political or economic changes,(40, 41) war,(42) or interruptions in the supply of illicit drugs.(36) Under such circumstances an independent data source (unrelated to the outcome under investigation) should be used to establish the timing of the "intervention" period.(36) For example, Lopez Bernal et al used the widely acknowledged definition for recession of two successive quarters of contracting GDP to establish the timing of the late 2000s financial crisis in Spain in their evaluation of the effect of the financial crisis on suicides.(41)

Multiple models and sensitivity analysis

While we would encourage authors to select the most appropriate model *a priori*, there may be differing underlying theories that may wish to be tested for wish the authors wish to test different models. These should still be defined in advance and it is important that if multiple models are tested, that multiplicity corrections (such as the Bonferroni correction) are applied.(43) Sensitivity analyses can also be used to define a range of possible models or to test different assumptions.(2) For example, different ranges of outcome data or different lag periods may be selected.(41, 44) If the same effect is detected under different assumptions, this can increase confidence in the results.

However, without multiplicity corrections, running multiple different models to test a wide range of assumptions increases the likelihood of false positive effects being detected. As with the primary model, it is therefore important that sensitivity analyses are pre-specified. Where there is a lot of uncertainty about the nature of potential effect of an intervention it may be necessary to run various exploratory analyses in the first instance, rather than regarding the study as an explanatory evaluation. Methods for undertaking exploratory analyses, such as identifying change points have been discussed elsewhere.(45)

Model diagnostics

Regardless of the impact model selected assumptions should always be checked, various model checking techniques exist, such as assessment of residuals.(2) One particular assumption that should be checked in ITS models is that interventions are independent from one another. In time series data, observations close to each other in time tend to be more similar, this is known as autocorrelation. Often autocorrelation is explained by other factors such as time-varying confounders, and, in particular, seasonality. Nevertheless, residual autocorrelation should be checked after adjusting for these factors and where present, several methods exist for adjusting for autocorrelation. Methods for checking for and adjusting for autocorrelation are discussed in more detail elsewhere.(2)

CONCLUSION

Interrupted time series is one of the most rigorous quasi-experimental designs and avoids many of the sources of bias and confounding of other observational studies.(1, 3) Nevertheless, we have demonstrated the risk that incorrect modelling of either the underlying trend or the impact of the intervention has for generating misleading results. The threat to validity is greatest when more flexible or data driven models are chosen as this increases the likelihood of detecting false positive effects due to confounding events or random noise. Therefore, the most appropriate model for a given intervention and outcome should be carefully considered and we have outlined an objective approach for this. Where there is uncertainty over model choices, clearly defined sensitivity analyses can be added. If a flexible model is required, other design adaptations, such as controlled interrupted time series or multiple baseline designs,(46, 47) should be applied to help exclude alternative explanations for any effects.

Given the range of possible model choices in ITS analysis, it is important that researchers are transparent in providing clear and objective justification for any modelling decisions when reporting

ITS studies. The methods section should include a statement on the amount of data available, any data restrictions and the reasons for these. We would also suggest providing a scatter plot of the complete data series as a supplementary appendix so that readers and reviewers can scrutinise any data restrictions or model choices. The primary model for the baseline trend should be clearly justified, including the reasons for including or excluding any time varying confounders and the reasons for any non-linear trends. Similarly, authors should defend their chosen impact model, including a clear description of the nature of the intervention and the nature of its expected effect on the outcome. Finally, authors should acknowledge any uncertainty in model selection in the limitations and any sensitivity analyses should again be fully justified.

REFERENCES

1. Soumerai SB, Starr D, Majumdar SR. How Do You Know Which Health Care Effectiveness Research You Can Trust? A Guide to Study Design for the Perplexed. Preventing Chronic Disease. 2015;12.
2. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2016.
3. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
4. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002;27(4):299-309.
5. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ. 2015;350:h2750.
6. Linden A. Conducting interrupted time series analysis for single and multiple group comparisons. The Stata Journal. 2015;15(2):480-500.
7. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci. 2000;1(1):31-49.
8. Lopez Bernal JA, Cummins S, Gasparrini A. The Use of Controls in Interrupted Time Series Studies of Public Health Interventions. Int J Epidemiol. 2018 [In Press].
9. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. Int J Technol Assess Health Care. 2003;19(4):613-23.
10. Lopez Bernal JA, Lu CY, Gasparrini A, Cummins S, Wharham JF, Soumerai SB. Association between the 2012 Health and Social Care Act and specialist visits and hospitalisations in England: A controlled interrupted time series analysis. PLOS Medicine. 2017;14(11):e1002427.
11. Smith JA, Mays N. GP led commissioning: time for a cool appraisal. BMJ. 2012;344.
12. Högberg U, Wall S. Secular trends in maternal mortality in Sweden from 1750 to 1980. Bull World Health Organ. 1986;64(1):79-84.
13. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. J Clin Epidemiol. 2011;64(11):1252-61.
14. Twisk J, Rijmen F. Longitudinal tobit regression: A new approach to analyze outcome variables with floor or ceiling effects. J Clin Epidemiol. 2009;62(9):953-8.
15. Gasparrini A, Gorini G, Barchielli A. On the relationship between smoking bans and incidence of acute myocardial infarction. Eur J Epidemiol. 2009;24(10):597-602.
16. Coggon D, Barker D, Rose G. Epidemiology for the Uninitiated: John Wiley & Sons; 2009.

17. Hall V, Charlett A, Hughes G, Brook G, Maguire H, Mercer CH, et al. Olympics and Paralympics 2012 mass gathering in London: time-series analysis shows no increase in attendances at sexual health clinics. Sex Transm Infect. 2015;91(8):592-7.
18. Barone-Adesi F, Gasparrini A, Vizzini L, Merletti F, Richiardi L. Effects of Italian smoking regulation on rates of hospital admission for acute coronary events: a country-wide study. PLoS One. 2011;6(3):e17419.
19. Walley AY, Xuan Z, Hackman HH, Quinn E, Doe-Simkins M, Sorensen-Alawad A, et al. Opioid overdose rates and implementation of overdose education and nasal naloxone distribution in Massachusetts: interrupted time series analysis. BMJ. 2013;346.
20. Grijalva CG, Nuorti JP, Arbogast PG, Martin SW, Edwards KM, Griffin MR. Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis. The Lancet. 2007;369(9568):1179-86.
21. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. Int J Epidemiol. 2013;42(4):1187-95.
22. Steinbach R, Perkins C, Tompson L, Johnson S, Armstrong B, Green J, et al. The effect of reduced street lighting on road casualties and crime in England and Wales: controlled interrupted time series analysis. J Epidemiol Community Health. 2015.
23. Wood S. Generalized additive models: an introduction with R: CRC press; 2006.
24. Verbyla AP, Cullis BR, Kenward MG, Welham SJ. The Analysis of Designed Experiments and Longitudinal Data by Using Smoothing Splines. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1999;48(3):269-311.
25. Health and Social Care Information Centre. HES 2010-11 Outpatient Data Quality Note: Health and Social Care Information Centre; 2011 [Available from: http://www.hscic.gov.uk/catalogue/PUB03032/host-outp-acti-2010-11-qual.pdf.
26. Walter SR, Olivier J, Churches T, Grzebieta R. The impact of compulsory cycle helmet legislation on cyclist head injuries in New South Wales, Australia. Accid Anal Prev. 2011;43(6):2064-71.
27. Macarthur C, Parkin PC, Sidky M, Wallace W. Evaluation of a bicycle skills training program for young children: a randomized controlled trial. Inj Prev. 1998;4(2):116-21.
28. Pridemore WA, Chamlin MB, Kaylen MT, Andreev E. The Effects of the 2006 Russian Alcohol Policy on Alcohol‐Related Mortality: An Interrupted Time Series Analysis. Alcoholism: clinical and experimental research. 2014;38(1):257-66.
29. Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment Restrictions for Prescription Drugs under Medicaid. N Engl J Med. 1987;317(9):550-6.
30. Been JV, Mackenbach JP, Millett C, Basu S, Sheikh A. Tobacco control policies and perinatal and child health: a systematic review and meta-analysis protocol. BMJ Open. 2015;5(9).
31. Leopold C, Zhang F, Mantel-Teeuwisse AK, Vogler S, Valkova S, Ross-Degnan D, et al. Impact of pharmaceutical policy interventions on utilization of antipsychotic

medicines in Finland and Portugal in times of economic recession: interrupted time series analyses. International Journal for Equity in Health. 2014;13(1):1-9.
32. Feldstein AC, Smith DH, Perrin N, et al. Reducing warfarin medication interactions: An interrupted time series evaluation. Arch Intern Med. 2006;166(9):1009-15.
33. Landrigan CP, Srivastava R, Muret-Wagstaff S, Soumerai SB, Ross-Degnan D, Graef JW, et al. Impact of a Health Maintenance Organization Hospitalist System in Academic Pediatrics. Pediatrics. 2002;110(4):720-8.
34. Lu CY, Zhang F, Lakoma MD, Madden JM, Rusinak D, Penfold RB, et al. Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: quasi-experimental study. BMJ : British Medical Journal. 2014;348.
35. Akaike H. Information theory and the extension of the maximum likelihood principle. In: Petrov N, Czaki F, editors. Proceedings of the International Symposium on Information Theory. Budapest1973.
36. Gilmour S, Degenhardt L, Hall W, Day C. Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: a general method and example. BMC Medical Research Methodology. 2006;6:16.
37. Yang CH, Xirasagar S, Chung HC, Huang YT, Lin HC. Suicide trends following the Taiwan earthquake of 1999: empirical evidence and policy implications. Acta Psychiatr Scand. 2005;112(6):442-8.
38. Runkle JD, Zhang H, Karmaus W, Martin AB, Svendsen ER. Prediction of Unmet Primary Care Needs for the Medically Vulnerable Post-Disaster: An Interrupted Time-Series Analysis of Health System Responses. International Journal of Environmental Research and Public Health. 2012;9(10):3384.
39. Rodgers JL, John CAS, Coleman R. Did fertility go up after the oklahoma city bombing? An analysis of births in metropolitan counties in Oklahoma, 1990–1999. Demography. 2005;42(4):675-92.
40. Pridemore WA, Chamlin MB, Cochran JK. An Interrupted Time-Series Analysis of Durkheim's Social Deregulation Thesis: The Case of the Russian Federation. Justice quarterly : JQ / Academy of Criminal Justice Sciences. 2007;24(2):271-.
41. Lopez Bernal J, Gasparrini A, Artundo C, McKee M. The effect of the late 2000s financial crisis on suicides in Spain: an interrupted time-series analysis. EJPH. 2013;25(5):732-6.
42. Page A, Morrell S, Taylor R. Suicide and political regime in New South Wales and Australia during the 20th century. J Epidemiol Community Health. 2002;56(10):766-72.
43. Dunn OJ. Multiple comparisons among means. Journal of the American Statistical Association. 1961;56(293):52-64.
44. Grundy C, Steinbach R, Edwards P, Green J, Armstrong B, Wilkinson P. Effect of 20 mph traffic speed zones on road injuries in London, 1986-2006: controlled interrupted time series analysis. BMJ. 2009;339:b4469.
45. Cruz M, Bender M, Ombao H. A robust interrupted time series model for analyzing complex health care intervention data. Stat Med. 2017;36(29):4660-76.

46. St Clair T, Cook TD, Hallberg K. Examining the Internal Validity and Statistical Precision of the Comparative Interrupted Time Series Design by Comparison With a Randomized Experiment. American Journal of Evaluation. 2014:1098214014527337.
47. Hawkins NG, Sanson-Fisher RW, Shakeshaft A, D'Este C, Green LW. The Multiple Baseline Design for Evaluating Population-Based Research. Am J Prev Med. 2007;33(2):162-8.

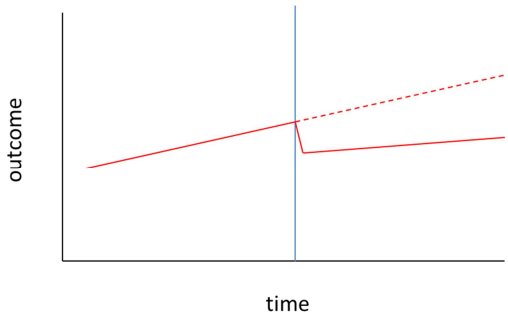


Figure 1: the interrupted time series design
Solid line = modelled trend; dashed line = counterfactual; vertical line = intervention implementation. This shows a step decrease and decrease in the slope following the intervention.

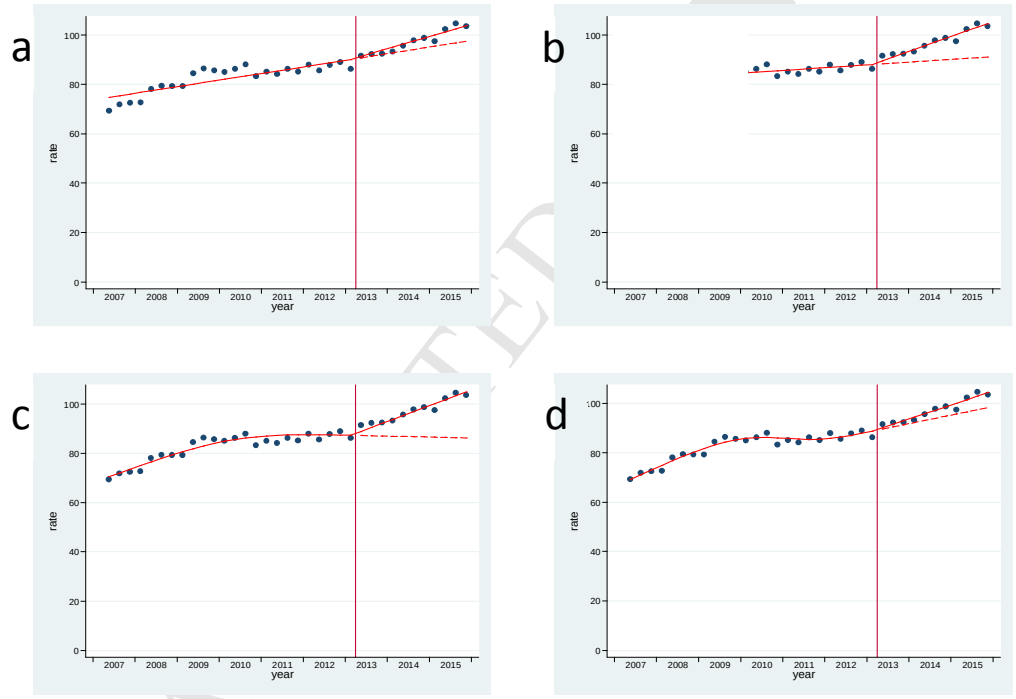


Figure 2: Four different approaches to modelling the trend
a: simple linear trend; b: excluding data prior to April 2010; c: non-linear allowing one inflection point; d: non-linear trend allowing two inflection points. Dots= data observations; solid line = fitted model; dashed line = counterfactual; vertical line = policy implementation.

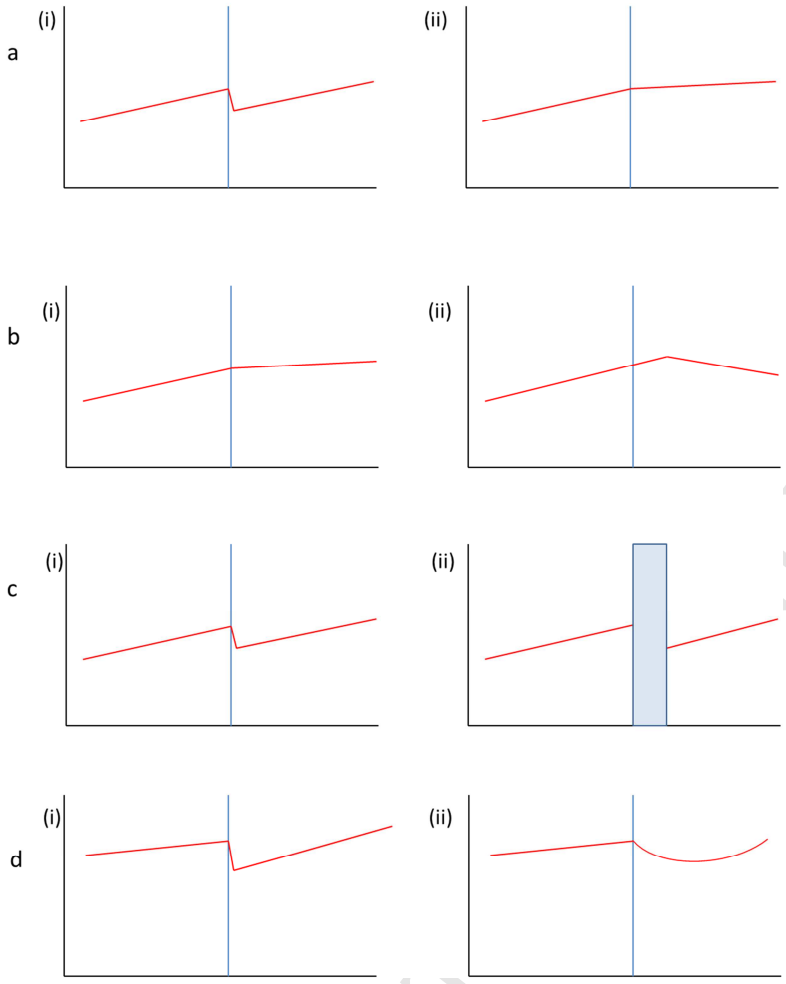


Figure 3: Interrupted time series impact models

X-axis represents time, y-axis represents the outcome. The vertical blue line is the time when the intervention was implemented; the red line is the ITS regression model. a(i) abrupt step change effect, a(ii) gradual slope change effect; b(i) immediate effect following the intervention, b(ii) lagged effect; c(i) intervention at a specific time point, c(ii) transition period (blue box) excluded from the model; d(i) waning effect following a step change, d(ii) gradual effect with gradual waning.

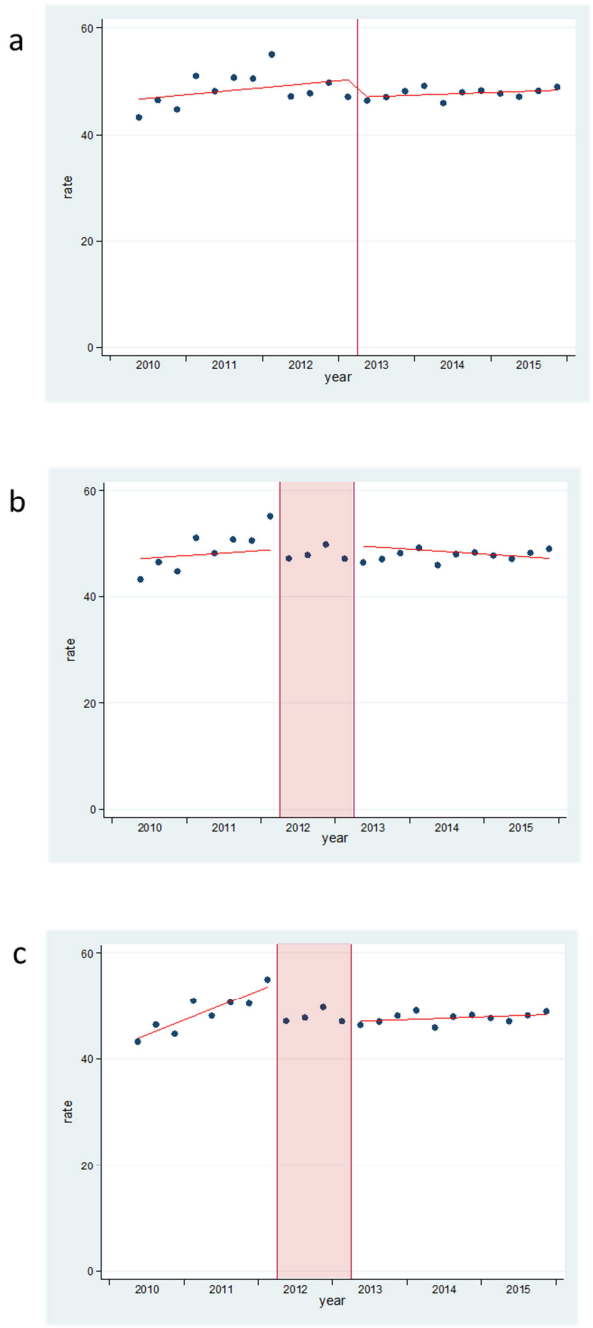


Figure 4: Four different approaches to modelling the impact of the intervention
a) Step and slope change model; b) slope change only with a one year intervention phase; d) step and slope change with a one year intervention phase.

Box 1: Case study

To illustrate the strengths and limitations of different approaches to model specification we use data from a recent study evaluating the impact of the 2012 Health and Social Care Act in England on hospital admissions and outpatient specialist visits.(10) This policy aimed to involve general practitioners (GPs) in commissioning (planning and purchasing) secondary care through the establishment of GP-led Clinical Commissioning Groups. GP-led commissioning is expected to reduce healthcare costs by shifting care away from secondary care to primary and community settings.(11) We therefore hypothesized that the reforms would result in a relative reduction in secondary care activity (inpatient admissions and outpatient visits). The health and social care act was enacted in April 2012, there was then a 12 month period during which the Clinical Commissioning Groups worked alongside the existing healthcare administrative bodies before taking over fully independent commissioning in April 2013. We had quarterly data on all NHS hospital admissions and outpatient visits between the second quarter of 2007 and the final quarter of 2015. More details about the intervention and the data can be found in the original study.(10)

Box 2: Framework for defining the trend in the outcome

1. Find out maximum time range of the dataset.
2. Check with data provider and review data quality notes for changes to data collection and any data errors requiring the data to be truncated.
3. Examine the literature for and seek expert advice on previous interventions or events that may have affected the outcome of interest.
4. Consider whether the outcome is likely to have a seasonal pattern or other known cyclical patterns and adjust for these.
5. Consider whether there are other measurable variables that could influence the outcome and may change substantially over time. If so include the variables within the model.
6. Examine the pre-intervention data graphically for linearity and any obvious cyclical patterns or trend changes.
7. Fit a model with linear trend on the pre-intervention data only and examine the fit.
8. If a linear model is a poor fit, consider a non-linear model. Discuss reasons for any non-linear trend and acknowledge in limitations if these are unknown or unmeasurable.
9. Run relevant sensitivity analyses if there is any uncertainty over model selection

Box 3: Framework for defining the impact model of the intervention

1. Consider whether the intervention was implemented gradually or abruptly.
2. Consider whether the outcome would respond quickly or slowly if the intervention were effective.
3. Consider whether the intervention would be expected to have an immediate or delayed impact on the outcome.
4. Examine existing evidence on the duration of the lag with similar interventions or outcomes.
5. Consider whether the intervention was introduced at a specific point in time or over a prolonged period.
6. Check when the intervention was announced, marketed, implemented or removed – consider whether each of these stages could have affected the outcome.
7. Consider whether there could be a ceiling or floor effect on the outcome.
8. Run relevant sensitivity analyses if there is any uncertainty over model selection