The use of interrupted time series for the evaluation of public health interventions

James Andrés López Bernal

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy
University of London

February 2018

Department of Social and Environmental Health Research

Faculty of Public Health and Policy

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Declaration

'I, James Lopez Bernal, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis'

28/01/2018

Abstract

Robust evaluation of public health interventions is required to ensure that interventions that lead to the greatest health benefit are adopted. However, traditional experimental evaluative designs are rarely possible for public health evaluation. Furthermore, alternative "quasi-experimental" designs are underused, are seldom covered in detail in epidemiology courses and are excluded from many guidelines and reviews. As a result population level health interventions have suffered from an "evaluative bias" whereby interventions not amenable to randomised control trials are often either poorly evaluated or not evaluated at all.

Interrupted time series (ITS) analysis is one of the most powerful quasi-experimental designs for evaluating the effectiveness of population level health interventions. It is increasingly being used to evaluate the effectiveness of interventions ranging from clinical guidelines to national public health legislation. The basic design involves comparing the outcome of interest before and after an intervention, whilst accounting for any underlying trend. Nevertheless, ITS studies, like other quasi-experiments have more inherent threats to their internal validity than experimental designs, many of which have not been adequately addressed in the existing literature. Further guidance is needed on these threats and how they are best addressed in the design, application and appraisal of ITS studies.

The overarching aims of this thesis are to improve the way that interrupted time series studies of public health interventions are designed in order to reduce the risk of bias and to make robust ITS designs more accessible to evaluators of public health interventions. This will be achieved through a range of methodological and applied studies using ITS designs.

Acknowledgements

List of abbreviations

ACEs   Acute Coronary Events
ARIMA   Autoregressive integrated moving average
CBA   Controlled before and after study
CCG   Clinical commissioning group
CI   Confidence interval
CITS   Controlled interrupted time series
CONSORT   Consolidated Standards of Reporting Trials
CRCT   Cluster randomised control trial
DH   UK Department of Health
DiD   Difference in difference
EPOC   Cochrane Effective Practice and Organisation of Care Review Group
FERITS   Framework for Enhanced Reporting of Interrupted Time Series
GDP   Gross Domestic Produce
GP   General practitioner
HES   Hospital Episode Statistics
HSCA   2012 Health and Social Care Act
ICD-10   International Statistical Classification of Diseases and Related Health Problems 10th Revision
INE   Instituto National de Estadisticas (Spain's national statistics institute)
IRR   Incident rate ratio
ISD   Information Services Division NHS Scotland
ITS   Interrupted time series
LANTERNS   Local Authority Collaborators' National Evaluation of Reduced Nightime Streetlight
MRC   UK Medical Research Council
MsSH   Medical Subject Headings
NHS   UK National Health Service
OECD   Organisation for Economic Cooperation and Development
ONS   Office for National Statistics
PCT   Primary Care Trust
PEDW   Patient Episode Database for Wales
PH   Public health
PHE   Public Health England
PSA   Prostate specific antigen
RCT   Randomised control trial
RDD   Regression discontinuity design
RECORD   REporting of studies Conducted using Observational Routinely-collected health Data
RR   Relative risk
SD   Standard deviation
STROBE   Strengthening the Reporting of Observational Studies in Epidemiology
TIDieR   Template for Intervention Description and Replication
TREND   Transparent Reporting of Evaluations with Nonrandomized Designs

Note on formatting

Published papers are presented in their published format and those that have been submitted or prepared for submission are presented in a ready to submit format. Because each paper has its own set of references I present references at the end of each chapter rather than in a single bibliography at the end.

Table of Contents

1. Background

1.1   Introduction

This thesis examines the use of interrupted time series (ITS) for the evaluation of public health interventions. ITS is one of a range of possible evaluative study designs and one that is particularly appropriate to public health interventions.(1) In order to put the design in to context, in this background chapter I will introduce the broader topic of public health evaluation. I will begin by discussing what a public health intervention is and why robust evaluation is important. Next, I will discuss some of the features of public health interventions that make them complex to evaluate. I will then briefly introduce different types of evaluation, focussing on that to which ITS is most commonly applied: quantitative outcome evaluations.

The remainder of the chapter will be focussed on the design of outcome evaluation studies. This will include, a discussion on counterfactuals and how they may be approximated in order to provide a comparator for estimating the effect of the intervention. The way the counterfactual is defined determines the internal validity of an evaluation study and the extent to which causal inferences can be made. I will discuss internal validity and the possible threats to validity in the context of evaluation design. The evaluative study design with the strongest internal validity is the randomised control trial (RCT), or, more commonly in public health evaluation, the cluster randomised control trial (CRCT), therefore this is considered the gold standard. However, I will explain why RCTs and CRCTs are not always possible when evaluating public health interventions and therefore, why other, quasi-experimental, designs such as ITS must be considered. Finally, I will discuss the features of quasi-experimental designs, the broad group into which ITS fits and will introduce a number of other quasi-experimental designs to which I will be drawing comparisons later on in the thesis.

1.2 Public health evaluation

A wide range of programmes are described as public health interventions from infection prevention and control initiatives, to community development programmes, physical activity programmes to government legislation. Furthermore, many non-health interventions (for example those with educational, economic or environmental objectives) and even unplanned events may affect aspects of public health. Robust evaluation of such interventions is important for all stakeholders. Most notably, it is in the interest of the beneficiary population that interventions that they receive are effective in improving their health, and that such interventions do not cause them harm. This is of

course, also of interest to public health practitioners and policymakers who develop and implement these interventions, furthermore evaluations of previous initiatives will also be of interest to public health practitioners to inform the development of any new programmes or policies. Lastly, robust evaluation is important for the funders of health interventions be it taxpayers, private organisations or third sector organisations, each has scarce resources and it is important that these are allocated efficiently to fund the most effective interventions and not to those that lead to little or no improvement in health outcomes.(2, 3)

In clinical practice, evidence based medicine has now become well established. Multiple randomised controlled trials (RCTs) to evaluate the effectiveness of each intervention are common and the establishment of organisations such as the Cochrane Collaboration facilitate the synthesis of this evidence so that it leads to changes in practice. "Evidence for population health", however, lags behind its clinical counterpart.(4) The gold standard RCT is not always possible, or even appropriate for public health interventions, yet alternative designs have not become well established and are excluded from many guidelines and reviews that aim to provide evidence based recommendations. As a result population level health interventions have suffered from an "evaluative bias" whereby recommendations are biased towards those interventions that are easier to evaluate but that may not necessarily be the most effective.(5) This may mean that resources are not allocated efficiently and the interventions leading to the greatest health benefits are not always chosen.(4-8)

1.3 Features of public health interventions

Public health interventions differ from clinical interventions in a number of ways and have several commonly occurring features which can complicate their evaluation:

Population level

Most obviously, by their nature, public health interventions target a population rather than an individual. In clinical medicine an individual patient presents with a health problem, an intervention (be it medication, surgery, lifestyle changes or a wide range of other interventions) is then targeted at that individual with the desired outcome being an improvement in the health of that individual. In public health, the health problem is detected at a population level, this could range from a small community to a regional, national or even multinational population, the intervention is targeted at the population, with the aim of improving health outcomes (for example rates of a disease) within the population as a whole.

Complex

The UK Medical Research Council (MRC) describe complex interventions as "interventions that contain several interacting components".(9) Public health interventions are often complex, for example, the Healthy Towns programme in England involved hundreds of different individual component interventions (such as "family health hubs", healthy urban planning, healthy eating and physical activity initiatives) in nine different towns implemented at different points in time.(6)

Multiple outcomes

Public health interventions may affect a range of health outcomes. Furthermore, interventions may also seek to target non-health outcomes, for example a workplace health programme may seek to improve various health outcomes in employees but also to save an organisation money by reducing absence, street lighting interventions may seek to reduce road traffic accidents but also to reduce crime, the introduction of cycle lanes may seek to increase rates of physical activity but also to reduce carbon dioxide emissions.(5, 9)

Long-term impacts

Many public health interventions target outcomes that would occur months, years or even decades into the future. For example programmes targeting pregnant women and young children such as the UK "Healthy Start" scheme may impact on health outcomes in later childhood or into adulthood.(10)

1.4 Types of evaluation

Evaluation science is a long-standing discipline and a range of theories and frameworks have been developed to explain the purpose and process of evaluation.(11-13) Perhaps the most common distinction is between outcome evaluation and process evaluation. There have been numerous definitions of these two types of evaluation; broadly speaking, outcome evaluation assesses whether an intervention works (its efficacy or effectiveness), whereas process evaluation assesses how and why it works.(14, 15) Outcomes may be intermediate, for example changes in behaviours such as smoking or physical activity, or they may be final, for example changes in disease rates or mortality. Process evaluation may examine how an intervention is implemented or received, why it is or is not effective and whether it could be implemented in other settings.(16)

A further distinction that is often made in the evaluation literature is whether the evidence that is used for the evaluation is quantitative or qualitative. A wide range of data can be used for each of these, for example surveys or routine data in quantitative evaluations and semi-structured interviews or focus groups in qualitative evaluations. Outcome evaluation tends to have more of a quantitative

focus and process evaluation may have more of a qualitative focus, however both methods are used in each type of evaluation.(16, 17)

The focus of this thesis will be on quantitative outcome evaluation.

1.5 Counterfactuals

As described above, the aim of outcome evaluation is to assess whether an intervention works, that is, whether it achieves what it purports to achieve. In order to know whether an intervention has caused an outcome, a comparison needs to be made between what actually happened and the *counterfactual*, that is, what would have happened if the intervention had not taken place. Evaluation of the intervention requires the simulation of an 'experiment' in which these two situations can be observed. Of course, it is not possible to observe the intervention both being implemented and not being implemented in the exact same population at the same time, therefore the true counterfactual is never known. Evaluation design is therefore centred on creating the best approximation of the true counterfactual and then comparing what actually happened to the approximated counterfactual.(18)

1.6 Causal inference and validity

No method for approximating the counterfactual is perfect, therefore it is never possible to infer with certainty that an association found between an intervention and an outcome is causal. Evaluative study designs must reduce the possibility of other factors explaining any observed association. The *internal validity* of a study, in this case the evaluation of an intervention, refers to the extent to which an observed association between the intervention and the outcome (when compared to the approximated counterfactual) reflects a causal effect. Campbell and Stanley identified eight threats to internal validity which have since been expanded upon by Shadish et al (with the addition of ambiguous temporal precedence) (Table 1):(18-20)

Table 1: Threats to internal validity(18-20)

Threat to validity	Description	Epidemiological terminology
Ambiguous temporal precedence	When examining associations between an exposure and an outcome, it may not be clear whether the exposure preceded the outcome. In some cases it may be possible that the outcome caused the exposure, a phenomenon known as reverse causality. This is generally more of a	Reverse causality

	problem in aetiological studies than evaluative studies as the timing of an intervention is normally known. Nevertheless, if the timing of the intervention or even under study is unclear, it may not be obvious whether it preceded the outcome which could affect the study validity.	
Selection	Selection bias occurs when those receiving the intervention differ from a comparison group by some factor or factors that could influence the outcome. Apparent differences in the outcome could therefore simply be due to differences in the groups being compared rather than due to the intervention.	Selection bias and confounding due to population differences
History	History bias refers to any other events that occur at the same time as the intervention and that could affect the outcome. These events could thus provide an alternative explanation for the observed effect.	Confounding – note that history bias is a more specific term than confounding and refers to confounding by other contemporaneous interventions or events that may be associated with the outcome.
Maturation	Maturation changes are those natural changes among participants whether or not an intervention is implemented, such as individuals growing older or more tired and secular changes within a population (such as changes to the economy).	Time-varying confounding – maturation is a general term for confounding variables that may change over time from any pre-intervention observation(s) to any post-intervention observation(s)
Regression	Regression to the mean occurs when an intervention is introduced because of recent extremes in some measure. For example a smoking intervention may be introduced following the recent detection of higher than average smoking rates. In this situation there is a tendency for the measure to be less extreme on subsequent testing even without any intervention because part of the explanation for	Regression to the mean

	the initial extreme measure may have been due to random variation.	
Attrition	Attrition is the loss of participants from the evaluation so that not all participants are measured in each observation. If there are differences in the kind of people lost in different comparison groups then this may explain differences in observed outcomes in the two groups.	Loss to follow-up
Testing	Being tested (or observed) before an intervention may influence the observation on subsequent testing, simply due to being tested rather than due to the intervention. For example, asking somebody details about their smoking status may influence them to reduce or stop smoking.	Research participation effect or Hawthorne effect
Instrumentation	Instrumentation occurs when the way in which an outcome is measured changes during the evaluation. For example in the evaluation of a workplace health intervention looking at the effect on sickness absences, if the way that sickness absences are recorded changes as part of the intervention, this could explain an observed effect on absence rates.	Information bias or measurement bias
Interaction between different threats to validity	Different threats to validity can occur within the same evaluation and sometimes there can be an interaction effect between the two so that their combined effect is greater. For example a selection-maturation interaction may occur if comparison groups come from different communities and then there are different secular economic changes in each of these communities.	

These threats are drawn from the social science literature and the terminology may not always be familiar to epidemiologists. A third column has therefore been added to the table in order to translate Campbell and Stanley's threats into epidemiological terminology. Of particular note, what is typically referred to as 'confounding' in epidemiology, is divided into three different terms in Campbell and Stanley's list: 1) Selection incorporates confounding arising from differences between a control and intervention group according to other covariates (in addition to the typical epidemiological definition of selection bias: differential exclusion of participants from each group); 2) History – specifically describes confounding due to the effects of other interventions or events that are concurrent to the intervention under study; 3) Maturation – describes changes to a population due to time-varying confounding factors which can affect before-after comparisons.(18-21)

While strong internal validity of an evaluation is important in order that more accurate conclusions can be made about the effect of an intervention within the population and setting being studied, an evaluation can be even more useful if those conclusions can be extended to other populations and settings so that they can inform the choice of interventions elsewhere. The e*xternal validity* of an evaluation refers to the extent to which its findings can be extended to other settings and populations. Evaluations conducted on varied populations in a range of settings may facilitate the study of subgroups, if similar effects are found in different subgroups this increases the likelihood that results are generalisable. However, if an evaluation has very stringent inclusion criteria or is conducted in a very specific setting, it is more difficult to draw conclusions about what effect might be expected if the intervention were implemented elsewhere.(18, 19, 22)

1.7 Evaluation of subgroups and effect modification

Interventions may affect different groups in different ways. For example, males may be more affected than females or effects may differ by age. By only looking at the overall effect of an intervention in the whole study population, this can mask effects in certain subgroups. Conversely, observing an effect in the total study population, does not necessarily mean that all subgroups within the population would have been affected or affected equally. Differential effects of the intervention on the outcome according to some third variable (such as sex or age) is known as effect modification.(23, 24) Where differential effects are considered plausible, evaluations should include subgroup analyses to explore effect modification. Formal tests for interaction are also available to test the strength of evidence for such differential effects and whether they may be due to chance.(23) For example, Matthews et al 2016 examined the impact of negative media coverage on statin prescriptions and found that older age groups and those who had been taking statins the longest were most likely to stop taking statins.(25) It is important that such differential effects are examined as this can help to identify groups that should be targeted with future interventions.

1.8 Experimental study designs

Experimental designs involve the random assignment of individuals (randomised control trials [RCTs]) or groups (cluster randomised control trials [CRCTs]) to either an intervention or a control group. Because public health interventions are generally applied at the population level, CRCTs are generally more appropriate in this context.(26, 27) The outcome in the control group provides an approximation

of the counterfactual to which the outcome in the intervention group can be compared in order to estimate the effect of the intervention. RCTs and CRCTs are regarded as the 'gold standard' evaluative design due to their ability to ensure that the intervention and control groups are, on average, the same with regards to all variables (both known and unknown) other than their exposure to an intervention. Both groups would thus be expected to have similar outcomes given the same conditions. Therefore, the control group provides a valid approximation of what would have happened in the intervention group had they not received the intervention.(7, 28) CRCTs have strong internal validity, the randomisation process controls for selection bias, and, because the two groups are on average the same, history, maturation and regression should not affect either group differentially, nor should testing or instrumentation assuming that both groups are treated in the same way throughout the study.(28)

Despite their strong internal validity, RCTs and CRCTs have a number of issues which can limit their applicability in public health evaluation. First, true equipoise may not exist, therefore it would be unethical to exclude some participants from the intervention.(7, 29) This happens, for example when introducing interventions that have already been shown to be effective in other populations or when evaluating secondary outcomes of an intervention in which there is already strong evidence of the effectiveness on primary outcomes.(5, 30, 31) Second, public health practitioners or policy makers introducing the intervention may require certain groups to receive an intervention for other reasons, for example those with the most need. Similarly, it may be necessary that the intervention is delivered to everyone in the population simultaneously, either for legal reasons or because an inherent feature of the intervention is its unanimous adoption, for example new laws, changes to social benefits or reforms of national systems.(7) Third, individuals or groups may have preferences regarding an intervention and disagree with randomisation. Preference trials are a possible solution to this, whereby only those with no preference for or against the intervention are randomised. Nevertheless, this limits generalisability and many public health interventions rely on active participation, therefore those who are indifferent may respond differently to others who take more of an interest in the intervention.(7, 32, 33) Fourth, researchers are often interested in evaluating the effects of interventions that have already been implemented or the health impacts of unplanned events such as natural disasters or political and economic events.(34, 35) In this situation designs that make use of pre-existing observational data will be necessary.(7) Fifth, for evaluations of interventions that target rare outcomes, the numbers needed to detect an effect are often far too large for a trial to be feasible.(32, 36) Finally, RCTs often have low external validity due to stringent inclusion critieria, participants behaving differently because of the knowledge that they are under investigation, and because settings and populations that tend to be involved in trials (for example research conscious

policy makers and participants, and settings with well established relationships universities) may not be representative of the wider population.(32, 37, 38)

Given that experimental designs are often not possible in the context of public health evaluation, other 'quasi-experimental designs' should be considered.

1.9 Quasi-experimental designs

Where a true experiment is not possible, other methods are needed for simulating the counterfactual. Shadish et al define quasi-experiments as "experiments that lack random assignment of units to conditions but that otherwise have similar purposes and structural attributes to randomized experiments".(18) Whereas in experimental designs, the participants or populations are actively assigned to either the intervention group or the control group, quasi-experimental methods often take advantage of exogenous sources assignment to the intervention.(39) Where such exogenous sources of assignment exist, this is also termed a *"natural experiment"*.(5) A whole range of quasi-experimental designs exist, these can broadly be categorised into: designs that use a pre-post comparison, designs that use a control and designs that incorporate trends in the outcome, though designs may fall into more than one category.

1.9.1 Designs that use a control

Where randomisation is not possible, alternative controls may be used to approximate the counterfactual (a cross-sectional non-randomised control design). Here, individuals may be actively assigned to an intervention or control group, or more typically, the intervention has been targeted at one particular population and an unexposed population is selected as the control group in a natural experiment. Selection bias is the main limitation of non-randomised control designs.(18) A number of methods have been developed to minimise selection bias, including adjusting for covariates, matching and propensity score matching can account for known characteristics that differ between the two groups, but cannot control for unmeasured confounders.(7, 29, 32) Furthermore, given that there are no pre-intervention observations, it is not possible to tell whether the two groups already differed with respect to the outcome, even prior to the intervention.(18)

1.9.2 Designs that use a pre-post comparison

An alternative approach to approximating the counterfactual is to use pre-intervention observations within the same population. Here, it is assumed that the outcome of interest would remain the same in the absence of an intervention. The simple before-after design (also known as a one-group pretest-

posttest design) compares observations at a time point prior to the intervention to those at a time point after the intervention. If there is a change in the outcome following the intervention, this may be as a result of the intervention. Nevertheless, this design has significant threats to its validity as there are numerous other potential explanations for a change in the outcome, including: random fluctuations, a pre-existing increasing or decreasing trend in the outcome of interest (maturation), an abnormally high or low observation during the pre-intervention period that simply returns to normal (regression to the mean), or a change due to another simultaneous event (history bias). Interrupted time series designs, also use a pre-post comparison but avoid many of these limitations, as will be discussed in the next chapter.

1.9.3 Designs that use a control and a pre-post comparison

Controlled before and after (CBA) studies (also known as difference in difference designs) combine pre- and post-intervention observations with one or more non-equivalent control groups. The change in outcome in the intervention group is then compared to the change in outcome in the control group (Figure 1). While this design takes into account pre-intervention differences between the intervention and the control groups, it makes the assumption that in the absence of the intervention, observations in the two groups would have followed parallel trends. In order to strengthen the validity of the design it is important to ensure that control groups are as similar as possible to the intervention group. Where differences do exist a range of methods have been developed in order to match groups on individual level and group level variables as well as potential unobserved confounders.(18, 28, 40, 41)
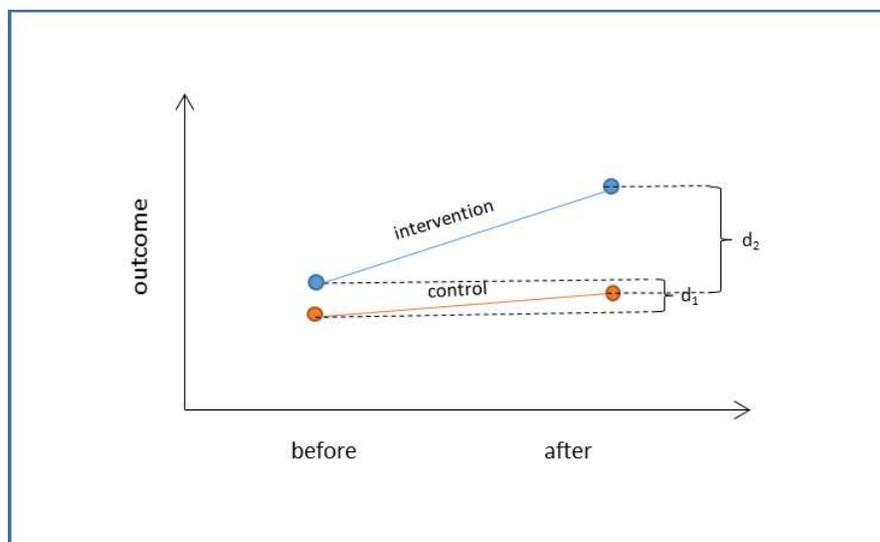
Figure 1: Controlled before and after design
Treatment effect = d_2-d_1

Another design that uses both a control and a pre-post comparison, but does not make the assumption of parallel trends, is the controlled interrupted time series. This is discussed further in Chapters 2 and 6.

1.9.4 Designs that incorporate trends in the outcome

Sometimes interventions are assigned to individuals or populations based on reaching a cut-off threshold in some continuous variable, such as age or income (known as the assignment variable). One approach to analysing the effect of the intervention in this situation would be to compare those below the threshold, for example a group below the threshold age, to a group above the threshold age. This is effectively a non-randomised control design. A regression discontinuity design (RDD) uses a more sophisticated approach incorporating trends in the outcome according to the assignment variable.(39) For example, Shoag et al looked at the impact of prostate biopsy on prostate cancer mortality by looking at prostate specific antigen (PSA) screening.(42) All those above a certain PSA score would be offered a biopsy, whereas those below this threshold would not. RDD involves regression of the outcome among the controls against the assignment variable, by extrapolating this regression line to scores in the assignment variable beyond the threshold a counterfactual can be created for those that received the intervention (Figure 2). The regression line in the intervention group can then be compared to this counterfactual, if the intervention causes an effect there will be a discontinuity in the regression line (either a step change or a slope change). Any threat to the internal validity of RDD would have to cause a discontinuity in the regression line that coincides exactly with the threshold in the assignment variable. There are often few circumstances in which this would be plausible.
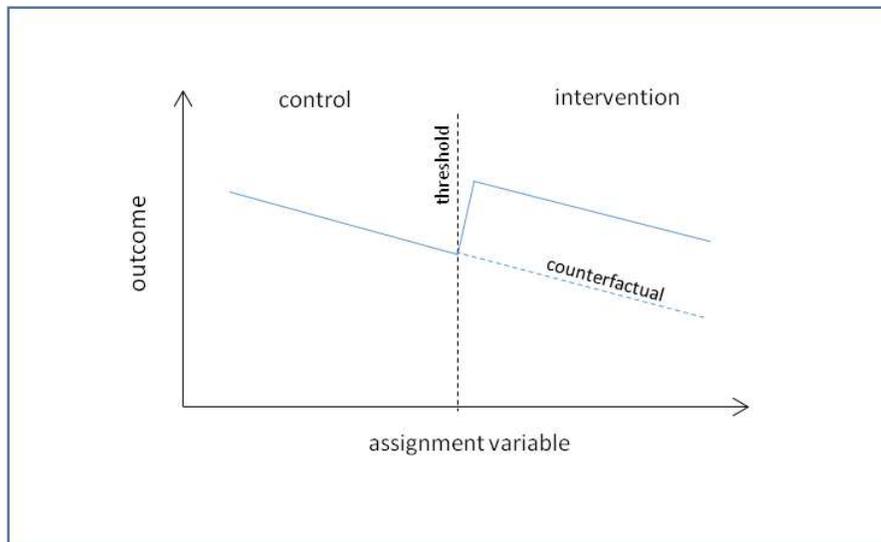
Figure 2: Regression discontinuity design

Interrupted time series can be considered a type of RDD study whereby the assignment variable is time (i.e. before a certain time the participants did not receive the intervention and after that time they did).

The most powerful quasi-experimental designs are able to provide strong evidence on the effectiveness of an intervention, particularly if design adaptations, described in more detail elsewhere, are used to address potential threats to their validity.(18, 19, 43, 44) One or more of the designs can often be used in situations where an RCT is not possible. In particular, quasi-experimental designs can often be applied in natural experiments.(5) Despite their strengths, the potential for bias and confounding is greater in quasi-experimental studies. It is therefore important that researchers have a clear understanding of the threats to validity of such evaluations. Further methodological work on identifying and minimising the threats to validity of quasi-experimental studies is needed, as is further guidance on the transparent reporting of such studies.

1.10 Evaluation guidelines

A range of quality criteria have been developed for reporting and appraising RCTs and the CONSORT statement has now been widely adopted.(45, 46) In recent years the importance of providing guidance on other evaluation designs has become more prominent. The UK Medical Research Council (MRC) has issued guidance on "Developing and evaluating complex interventions" and "Using natural

12

experiments to evaluate population health interventions".(5, 9, 47) The Cochrane Collaboration has also developed criteria on "Including non-Randomized studies" in systematic reviews.(48)

Some reporting recommendations have also been developed for certain quasi-experimental designs. The most widely adopted of these is the "Transparent Reporting of Evaluations with Nonrandomized Designs (TREND)" reporting guidelines for behavioural and public health interventions by the USA Centers for Disease Control.(8) This may go some way towards supporting the inclusion of non-randomised studies in systematic reviews and evidence based guidelines on the choice of interventions, and thus reduce the bias towards only including interventions that are amenable to RCTs.(8) However, such guidance is lacking for other designs such as ITS and RDD.

While further enhancement of existing guidance may be needed, it is encouraging that improving the evaluation of public health interventions is a key focus for applied researchers, systematic reviewers and funding agencies. Such interest should help to minimise the evaluative bias that exists with public health interventions and improve decision making to ensure that the most effective interventions are implemented.

1.11 Summary

Public health interventions are frequently complex and have a number of features which can make them difficult to evaluate. Avoiding the evaluation of interventions where conventional methods such as RCTs are not possible results in evaluative bias and a lack of evidence on the effectiveness of many interventions. This creates difficulties for public health practitioners and policy makers wishing to tackle health problems. In this chapter I have described the features of alternative quasi-experimental designs which may be considered for the evaluation of public health interventions. Nevertheless, further methodological research is needed on identifying the potential threats to the validity of these designs and approaches to dealing with these threats. Further development of reporting criteria for quasi-experimental designs is also required in order to ensure that they are presented transparently and can be easily appraised for inclusion in systematic reviews and guidelines.

1.12 References

1. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci. 2000;1(1):31-49.
2. Habicht J-P, Victora C, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. Int J Epidemiol. 1999;28(1):10-8.
3. Bonnell C, Farah J, Melendez-Torres GJ, Cummins S. "Dark logic" – Theorising the harmful consequences of public health interventions [In Press]. JECH. 2014.
4. Heller R, Page J. A population perspective to evidence based medicine:"evidence for population health". JECH. 2002;56(1):45-7.
5. 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. JECH. 2012;66(12):1182-6.
6. Ogilvie D, Cummins S, Petticrew M, White M, Jones A, Wheeler K. Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers. Milbank Q. 2011;89(2):206-25.
7. Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. JECH. 2009;65(7):582-7.
8. Des Jarlais DC, Lyles C, Crepaz N, Group T. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361.
9. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337.
10. Department of Health. Healthy Start Retailer research summary. In: Health Do, editor. 2012.
11. Donabedian A. Evaluating the quality of medical care. The Milbank memorial fund quarterly. 1966:166-206.
12. Maxwell RJ. Quality assessment in health. Br Med J (Clin Res Ed). 1984;288(6428):1470.
13. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health. 1999;89(9):1322-7.
14. Nutbeam D, Smith C, Catford J. Evaluation in health education. A review of progress, possibilities, and problems. J Epidemiol Community Health. 1990;44(2):83.
15. Potvin L, Haddad S, Frohlich KL. Beyond process and outcome evaluation: a comprehensive approach for evaluating health promotion programmes. WHO Reg Publ Eur Ser. 2001(92):45-62.
16. Macdowall W, Bonnell C, Davies M. Health Promotion Practice: McGraw-Hill Companies,Incorporated; 2006.
17. Nutbeam D. Evaluating health promotion—progress, problems and solutions. Health Promotion International. 1998;13(1):27-44.
18. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
19. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research: Houghton Mifflin Boston; 1963.
20. Campbell DT. Reforms as experiments. Am Psychol. 1969;24(4):409.
21. Haneuse S. Distinguishing selection bias and confounding bias in comparative effectiveness research. Med Care. 2016;54(4):e23-e9.
22. Steckler A, McLeroy KR. The importance of external validity. Am J Public Health. 2008;98(1):9-10.

23. Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandenbroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. Clinical Epidemiology. 2017;9:331-8.

24. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol. 1989;18(1):269-74.

25. Matthews A, Herrett E, Gasparrini A, Van Staa T, Goldacre B, Smeeth L, et al. Impact of statin related media coverage on use of statins: interrupted time series analysis with UK primary care data. The BMJ. 2016;353:i3283.

26. Pinder M, Jawara M, Jarju LBS, Salami K, Jeffries D, Adiamoh M, et al. Efficacy of indoor residual spraying with dichlorodiphenyltrichloroethane against malaria in Gambian communities with high usage of long-lasting insecticidal mosquito nets: a cluster-randomised controlled trial. The Lancet. 2015;385(9976):1436-46.

27. Prins RG, Brug J, van Empelen P, Oenema A. Effectiveness of YouRAction, an Intervention to Promote Adolescent Physical Activity Using Personal and Environmental Feedback: A Cluster RCT. PLOS ONE. 2012;7(3):e32682.

28. Steiner PM, Wroblewski A, Cook TD. Randomized Experiments and Quasi-Experimental Designs in Educational Research. In: Ryan K, Cousins JB, editors. The SAGE International Handbook of Educational Evaluation: SAGE Publications; 2009.

29. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. AJPM. 2007;33(2):155-61.

30. Callinan JE, Clarke A, Doherty K, Kelleher C. Legislative smoking bans for reducing secondhand smoke exposure, smoking prevalence and tobacco consumption. Cochrane Database Syst Rev. 2010;4.

31. 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.

32. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996;312(7040):1215.

33. Torgerson DJ, Sibbald B. Understanding controlled trials. What is a patient preference trial? BMJ: British Medical Journal. 1998;316(7128):360.

34. Milojevic A, Armstrong B, Kovats S, Butler B, Hayes E, Leonardi G, et al. Long-term effects of flooding on mortality in England and Wales, 1994-2005: controlled interrupted time-series analysis. Environmental Health. 2011;10(1):11.

35. 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.

36. Dwyer T, Ponsonby A-L. Sudden infant death syndrome: after the" back to sleep" campaign. BMJ: British Medical Journal. 1996;313(7051):180.

37. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health. 2004;94(3):400-5.

38. Rothwell PM. External validity of randomised controlled trials:"to whom do the results of this trial apply?". The Lancet. 2005;365(9453):82-93.

39. Oldenburg CE, Moscoe E, Bärnighausen T. Regression Discontinuity for Causal Effect Estimation in Epidemiology. Current Epidemiology Reports. 2016;3(3):233-41.

40. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41-55.

41. Qin J. Empirical-likelihood-based difference-in-differences estimators. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2008;70(2):329-49.

42. Shoag J, Halpern J, Eisner B, et al. Efficacy of prostate-specific antigen screening: Use of regression discontinuity in the plco cancer screening trial. JAMA Oncology. 2015;1(7):984-6.

43. 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.

44. West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, et al. Alternatives to the randomized controlled trial. Am J Public Health. 2008;98(8):1359.

45. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials. 1995;16(1):62-73.

46. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.

47. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ: British Medical Journal. 2000;321(7262):694.

48. Reeves B, Deeks J, Higgins J, Wells G. Chapter 13: Including non-randomized studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008]: The Cochrane Collaboration; 2008.

2.Interrupted time series design

2.1 Introduction

The last chapter introduced quasi-experimental designs. I have chosen to focus this thesis on interrupted time series (ITS), one of the more powerful quasi-experimental designs and one that has wide applicability in the evaluation of public health interventions.(1, 2) In this chapter I will provide an overview of the ITS design. I will begin by describing the main features in the context of quasi-experimental design characteristics that were introduced in Chapter 1 and I will explain how the counterfactual is approximated in ITS studies. I will also explain the main analytical approach for ITS and the range of applications. Next, I will consider the main strengths of ITS and summarise the existing empirical evidence on its validity. ITS also has a number of weaknesses and I will discuss how these can threaten the internal validity as well as possible design adaptations that may be used in order to minimise these threats. Finally, I will discuss existing quality criteria that have been used to critically appraise ITS studies.

Throughout the chapter, I will highlight areas where further methodological work is needed. This will be used to inform the aims of the thesis which will be presented in Chapter 3.

2.2 Features of the interrupted time series design

The interrupted time series (ITS) design (also known as the time series experiment) has been described in the social sciences literature since the 1960s, (3, 4) An ITS study involves a set of observations on an object, an individual subject or, as is most commonly the case in public health and social sciences, a population, taken repeatedly over time before and after an intervention. The counterfactual is modelled by extrapolating the pre-intervention trend in the outcome of interest into the post-intervention period (Figure 1). The impact of the intervention is then assessed by examining any change in the trend of the post-intervention observations.(3, 5)
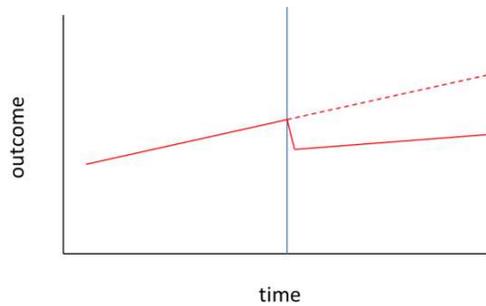
Figure 1: the interrupted time series design
Solid line = modelled trend; dashed line = counterfactual; vertical line = intervention implementation.

ITS is therefore effectively a pre-post comparison. However, the analysis of trends is what differentiates ITS from simple before and after studies where either a single observation before the intervention is compared to a single observation after the intervention or the average of several observations before the intervention is compared to the average of several observations after the intervention. In a before-after design, the trend is assumed to be flat, any change in the outcome after the intervention is therefore considered an effect of the intervention. If this assumption is violated, for example, if there is a pre-existing decreasing trend in the outcome, an effect would be detected whether or not an intervention is introduced (figures 2a and 2b). ITS is able to differentiate an effect of an intervention from that of the underlying trend (figures 2c and 2d).
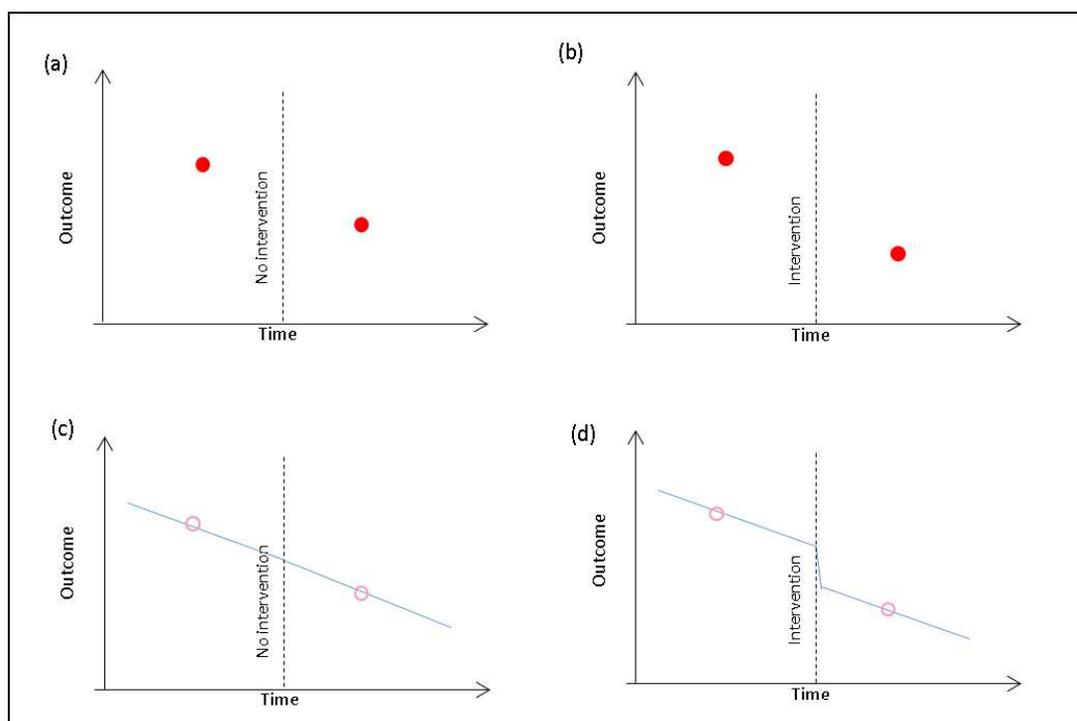
Figure 2 Difference between simple before and after study and interrupted time series

(a) Simple before and after assuming no intervention takes place (shows an effect)

(b) Simple before and after with intervention (shows an effect)

(c) ITS with no intervention (shows no effect) [observations from simple before and after are overlaid]

(d) ITS with an intervention (shows an effect) [observations from simple before and after are overlaid]

2.3 Analytical approach

Data requirements

There are two basic requirements for undertaking an ITS study. First, a time series of the outcome data is required spanning the pre- and post-intervention period. That is: there must be multiple sequential measures of the outcome over time both before and after the intervention. Typically, this involves using routine data sources that are not necessarily collected specifically for the purpose of the evaluation study. Second, the timing of the intervention must be clear in order to separate the pre and post-intervention periods. (1, 5) The Cochrane Effective Practice and Organisation of Care group guidelines suggest a minimum of three data points are included before and three after the

19

intervention.(6) However, others suggest that the number of data points should be sufficient to evaluate seasonal patterns (discussed below), for example a minimum of 12 data points before and after for monthly data.(1)

Describing the effect of the intervention

The impact of an intervention is most commonly described in terms of a change in level (or intercept), a change in slope (or trend), or a change in both of these parameters (Figure 3).(1, 5) Other changes can, however, also be examined, such as changes in variability of the data or changes in cyclical patterns, for example, the UK Winter Fuel Payment might be expected to reduce the seasonal increase in mortality among the elderly that occurs in the winter.(5, 7, 8) Shadish et al describe these parameters (level, slope, variance and cyclicity) as the *form* of the effect, they also discuss two other dimensions by which the effect of an intervention should be described: its *permanence* and its *immediacy*. Permanence refers to whether the effect persists over time (a *continuous* effect) or whether it is temporary and then returns to the preintervention trend (a *discontinuous* effect). The immediacy describes whether an effect occurs straight after the intervention is introduced (an *immediate* effect) or whether there is a lag between the intervention being introduced and any impact (a *delayed* effect).(5)
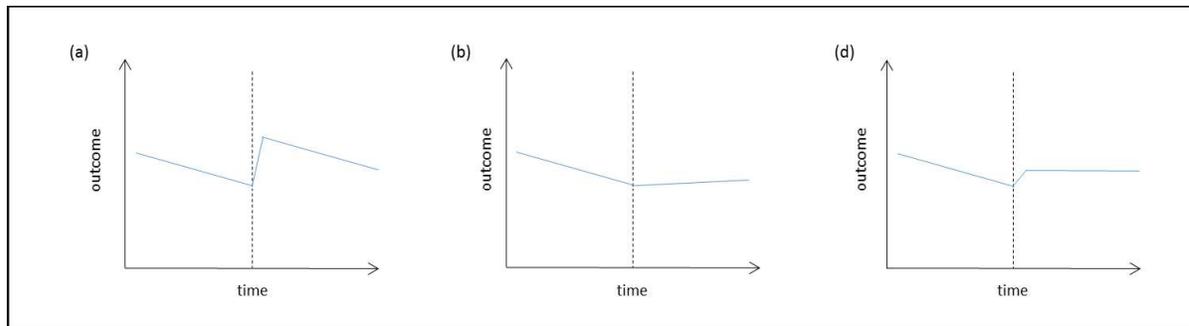
Figure 3 Types of effects in an interrupted time series study

- (a) Change in level
- (b) Change in slope
- (c) Change in level and slope

There is extensive literature on statistical methods for modelling the pre-intervention trend and for quantifying a level or slope change, (9-11). However, there is a lack of guidance on how to decide on the form of effect that should be modelled, whether a lag should be allowed for, and if so, how long the lag should be. Methods often seem to rely on using the data to specify the impact model that fits

best, however, this could increase the likelihood of an effect being detected due to random fluctuations or chance.(12) Further guidance is needed on how researchers can select the most appropriate ITS model *a priori*.

Methods of analysis

The statistical approach to analysis of ITS studies is known as segmented regression. In a basic ITS study with one pre-intervention and one post-intervention period, segmented regression fits a separate least squares linear regression line to each period. The post intervention model can be allowed to change in level, slope or both.(1)

A central assumption to linear regression is that the data are independent, however, this is often not the case in time-series data for two reasons: firstly, consecutive observations tend to be more similar to one another than those that are further apart, a phenomenon known as autocorrelation; secondly, there are frequently seasonal patterns whereby observations in one month are similar to observations in the same month a year previously, in particular when studying health and disease, for example, levels of influenza may be higher in winter, and levels of physical activity higher in summer.(1, 13) A range of techniques exist to test for and adjust for autocorrelation and seasonality.(1, 14-16) In addition to seasonal effects, other time varying confounders can be adjusted for within the segmented regression model. (17, 18)

While some specific aspects of the analytical approach to ITS have been described in detail, for example the use of autoregressive integrated moving average (ARIMA) models to correct for autocorrelation and approaches to deal with seasonality.(11, 19) There is a lack of introductory guidance on ITS analysis, including how to structure the data, how to undertake a segmented regression analysis and the common factors that need to be adjusted for. This may act as a barrier to its wider adoption in public health evaluation. Furthermore, more in depth guidance is needed on selection of the most appropriate impact model and factors that can impact on how the counterfactual is defined.

2.4 Applications of interrupted time series

Some of the earliest examples of ITS come from social policy research, including the impacts of policies on crime, economics and education.(4, 20-22) The design continues to be widely used in these fields but has expanded to evaluations of a broader range of interventions beyond policies such as manufacturing processes (23), war (24), education interventions (25), business practices (22).

ITS is now also being increasingly adopted for the evaluation of health interventions. It has been used for the evaluation of clinical interventions,(26) it has also been used extensively in single subject studies in clinical psychology.(27) However, they are most commonly used for public health interventions that are applied at a population level.(28-30) Part of the reason for this is that ITS designs are well suited to looking at population level outcomes, for which routine data with long time series are commonly available, for example: rates of health service use, disease incidence, mortality rates, smoking prevalence.(31) When examining population level outcomes, ITS studies can be considered an ecological design. Ecological designs are often criticised for the "ecological fallacy" whereby inferences are made at the individual level based on findings at the population level.(32) However, when evaluating the impact of interventions that are applied at a population level, such as smoking legislation, we are primarily interested in outcomes at the population level, such as smoking prevalence, therefore an ecological design is often more appropriate.

Fields of public health for which ITS studies have been used include: a wide variety of health promotion interventions, ranging from national legislation to local community level programmes;(2, 31, 33) communicable disease control, such as new vaccines or antimicrobial stewardship programmes;(28, 34); and health service interventions including the introduction of new services, financial incentives, new guidelines and screening programmes.(30, 35-37) ITS has also been used to evaluate the health impacts of non-health interventions, road changes and mass gathering events.(38, 39) Furthermore, it can also be used to evaluate the impact of unplanned events such as natural disasters and economic events.(40, 41)

2.5 Strengths

Wagner et al describe ITS designs as "the strongest, quasi-experimental designs to estimate intervention effects in non-randomized settings."(1) ITS is rarely subject to many of the threats to internal validity that affect other observational studies. Of Campbell and Stanley's factors that affect the internal validity of a study design (Chapter 1: Table 1), only *history* is identified as a major threat to ITS designs.(3, 5) Having a series of observations before the intervention takes place enables *maturation* and *regression* effects to be detected, if routine data is used then respondents will not be aware that they are part of an experiment so *testing* effects are not plausible, even without routine data testing effects are unlikely in a long time series. *Instrumentation* can be a threat but only if the method of data collection changes over the same period as the intervention. *Selection* and *attrition* are only a threat if the composition of the study population changes after the intervention.(3-5, 42)

The use of longitudinal data in ITS designs also enables researchers to examine the impact of an intervention in much more detail than would be the case with cross sectional data, for example, establishing the permanence and immediacy of any effect.(42) Furthermore, longitudinal designs such as ITS, allow possible reverse causality to be examined, whereby the outcome leads to exposure (or in this case the intervention). For example, a change in a health outcome could lead to an intervention being implemented rather than vice-versa, examination of a time-series will enable the researcher to see which came first.

In common with other observational studies, ITS designs, also have strong external validity given that they are normally undertaken in real world settings, using observational data. Generalisability to other populations or settings will be dependent on the particular intervention, outcome of interest, study population and study setting, however, subgroup analyses may help to establish whether effects are similar in different populations and thus support or reject the likelihood that findings are generalisable.(5)

While designing an ITS study can be complex, another important advantage of ITS studies is that presentation of results in graphical format, and their subsequent interpretation, is simple and can easily be understood by lay readers.(1) This is clearly important in public health evaluations in order to facilitate decision making by public health professionals and policy makers as well as understanding of these decisions by members of the public.

2.6 Limitations

The major threat to the validity of ITS studies is *history bias*, that is, other events occurring at the same time as the intervention which could explain the effect.(1, 3-5) The likelihood of other events impacting on the intervention increases if there is a greater time period between observations (e.g. yearly rather than monthly data) and the greater the lag between intervention and effect.(5) Coinciding events with a potential impact on the outcome should be investigated and controlled for where possible. Other threats to internal validity that occur in some circumstances include *instrumentation*, *selection* and *experimental mortality*. Instrumentation might occur if there is a change to the way the outcome data is collected over the study period, particularly if this coincides with or is as a result of the intervention.(3-5) *Selection* and *experimental mortality* could occur if the intervention leads to a change in the composition of the study population, or the loss of study participants. In this case, where possible, restricting the analysis to participants in all study phases can help.(5, 42)

While a benefit of ITS is that there are a whole range of possible impact models that can be used, selecting the most appropriate model can be challenging. Deciding, for example, the specific timing of the intervention, the form of effect, its permanence and whether there is any lag is complex and requires an in depth understanding of the intervention. An intervention, such as a policy change, could have an impact before it is implemented (due to publicity about the policy), at the point at which it is introduced, or sometime later (if the policy takes time to be implemented or communicated).(43) Similarly, researchers need to decide on the length of the time series to be included in the analysis. Visual inspection of the time-series data under investigation in order to identify the most suitable length of the series and the type of effect and then modelling this in the analysis is unlikely to be appropriate as this could result in any random changes that occur due to chance being interpreted as a significant treatment effect.(5) Evidence based theory or initial exploratory research of other data sources is needed in order to propose the most likely type of effect that the intervention will cause. Nevertheless, there is a lack of methodological literature on how ITS models should be specified. While testing a range of different models can help to deal with uncertainty in model selection it may also increase the risk of false positive effects being detected.

Finally, interpretation of ITS studies can be simple where the impact of the intervention on the outcome occurs very soon after its implementation, however, one of the difficulties in the evaluation of public health interventions, is that often effects would be expected to take many years to materialise. Interventions addressing risk factors for chronic diseases such as diabetes and many cancers may only change disease outcomes decades after they are implemented. As discussed above, the greater the lag between intervention and effect, the greater the risk that history becomes a factor affecting the internal validity of the study, furthermore, the delay between intervention and effect is likely to vary from person to person so it is difficult to pre-specify a lag to model in the analysis.(5)

2.7 Approaches to strengthening the validity of ITS

As discussed above, the principal threat to the validity of ITS is history bias. A number of design adaptations have been developed in order to mitigate history bias. The most common of these is to include a control series from a population that did not receive the intervention. If there is a change in the outcome in the intervention series, but not in a similar control series, this eliminates concurrent events that would have affected both populations as a possible explanation for the change (Figure 4).(5) For example, Dennis et al evaluated the impact of the introduction of helmet legislation in a number of Canadian provinces on cycling related head injuries by comparing outcomes in Canadian

provinces that did not implement helmet legislation.(44) The main limitation of non-randomised control groups in quasi-experimental designs is the possibility of selection bias and confounding due to differences between the groups. This is less of an issue in ITS studies as the pre-intervention trend among the intervention group serves as the primary control and this is derived from the same population as the post-intervention trend. Nevertheless, matching can be used to ensure that the control group is similar to the intervention group and various matching techniques exist to achieve this.(45)
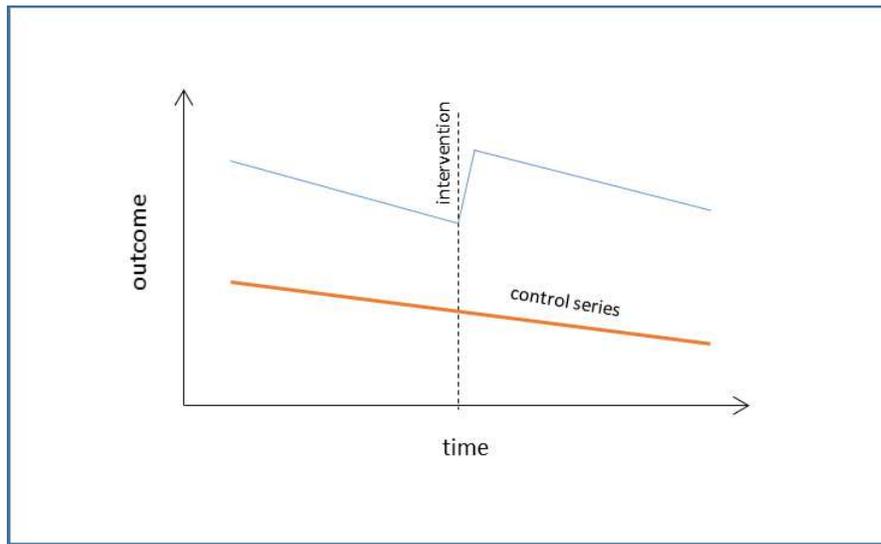
Figure 4: Graphical representation of a controlled ITS

Including a control series can greatly increase the validity of an ITS study. However, in order to do so confounding events must affect both groups. Currently there is a lack of guidance on the different types of control series that can be used in ITS, which types of confounding events they can address and how to approach selecting the most appropriate control series.

An extension of the controlled ITS design is the multiple baseline design. This involves introducing the same intervention to different populations at different points in time (Figure 5). For example, Biglan et al evaluated the impact of an intervention to reduce the sale of tobacco to young people in four communities, after a baseline period the intervention was first introduced in two of the communities and then in the other two communities at a later date.(46) One community acts as a control for the other depending on when it received the intervention. Furthermore any confounding event would

have to occur in two or more populations at two or more different time points. The design also strengthens external validity if findings are concordant in two different settings.(5)
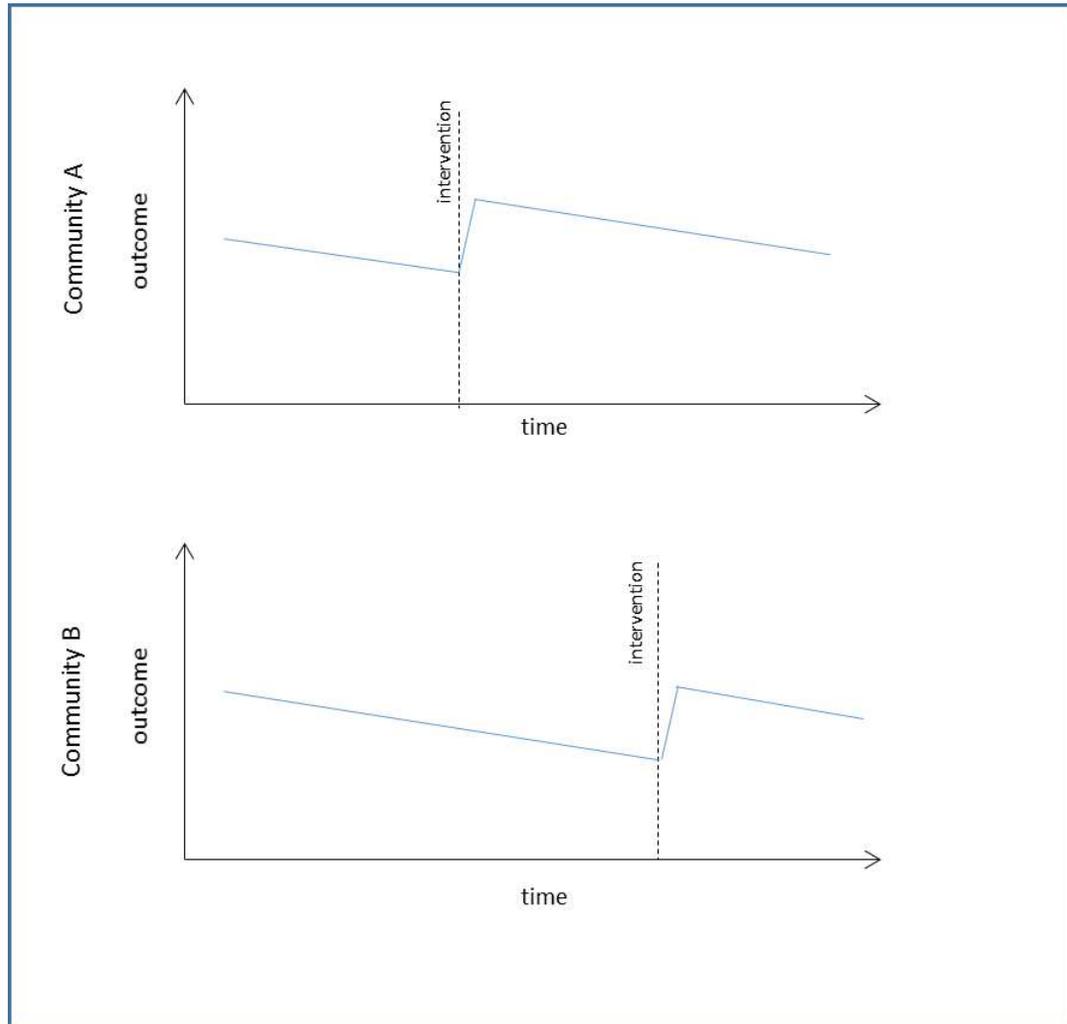
Figure 5: Graphical representation of a multiple baseline design

Another approach to limiting history bias is to introduce the intervention and then withdraw it after a given period (Figure 6). If there is an effect on the outcome of interest when the treatment is in place and this effect disappears when the treatment is removed, this increases the likelihood of a causal link. This design limits the threat of history as a confounding event would have to both be introduced and withdrawn at the same time as the intervention.(2, 5, 47) The design, however, can only be used if, firstly, the intervention can ethically be removed and, secondly, the effects of the intervention do not continue beyond its withdrawal. In practice this may be useful with clinical interventions and single

subject designs, but often not possible with public health interventions that frequently have long term effects.
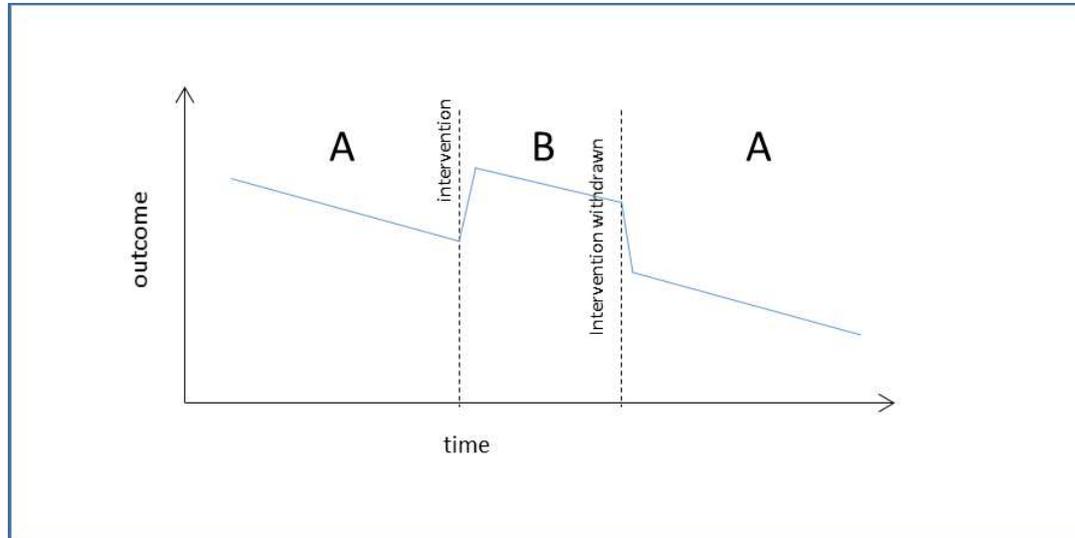
Figure 6: Graphical representation of an ABA ITS design

Finally, another technique to address some of the limitations of ITS studies is to conduct sensitivity analyses. This is primarily used when there is uncertainty about the type of effect that the intervention will have, in order to test different forms of effect (e.g. slope change instead of step change), different permanence (e.g. a discontinuous rather than a continuous effect) and different lags, as well as other model assumptions. No methodological literature was found addressing sensitivity analysis in ITS, however, the technique has been used in applied studies.(5, 16, 48) Whilst a continued effect under different assumptions may increase confidence that the intervention is indeed responsible for the effect, undertaking many hypothesis tests increases the chance of type 1 errors. In this instance there is a risk that researchers could pick the model which shows the greatest effect a posteriori. It seems, therefore, that developing an evidence based theory for the most likely model a priori still remains an important step.

2.8 Evidence of validity of the ITS design

ITS has many theoretical strengths and has been endorsed as one of the strongest designs for evaluating the impact of an intervention where a randomised control trial (RCT) is not possible.(1, 5) Nevertheless, until recently, there was little empirical evidence of the ability of ITS to infer a causal

link between an intervention and an outcome. Three studies were found comparing ITS designs to a RCT:

Fretheim et al compared a cluster RCT of an intervention to improve primary care prescribing of antihypertensives with a simple ITS of the treatment arm analysed using segmented regression.(49) Results were similar with the ITS result (11.5% change in prescribing[95% CI 9.5-13.5%]) within the 95% confidence interval of the cluster RCT result (9.0% [95% CI 4.9-13.1%]). However, there were baseline differences between the intervention and control groups in the cluster RCT, therefore it is difficult to assess whether this was an appropriate comparator.

Schneeweiss et al compared a cluster RCT with a form of controlled ITS in their evaluation of the impact of restriction of state funding of nebulised respiratory therapy on various outcome.(50) The control group in the ITS was the same population but in the two years prior to the intervention, when no restriction was implemented. Results initially differed between the two designs, but it was noted that there was a large amount of crossover from the control arm to the intervention arm in the RCT, after this was adjusted for results for the cluster RCT were concordant with those of the ITS.

St. Clair et al used a comprehensive within study comparison design to compare a cluster RCT of a school based education intervention on academic achievement with a controlled ITS using both non-matched and matched comparison groups.(51) The time-series was relatively short with 6 pre-intervention observations and one post-intervention. Again, they found that the ITS results were concordant with those of the RCT. In addition, the precision of the controlled ITS was found to be higher than that of the RCT. They also compared using grouped pre-test means to an ITS design modelling the slope and found that the ITS design substantially improved the concordance with the RCT result and reduced bias.

A further study by Somers et al examined the validity and precision of a controlled ITS compared to a regression discontinuity design evaluating the impact of a regional school based reading intervention.(52) Results were concordant between the two designs despite the short time series (4 pre-intervention observations). This study may be limited by the fact that it does not use a RCT (the accepted gold standard) as a benchmark, although the authors provide strong arguments for the use of a regression discontinuity design as a valid comparator.

Whilst each of the above studies has certain limitations, together they provide relatively compelling evidence of the validity of the ITS design, in particular when a control group is used. Nevertheless, it is important to be aware of the possibility of publication bias and further within-study comparisons are needed to test the generalisability of these results.

2.9 Existing guidance on use of ITS

Quality criteria and reporting guidelines exist for RCTs and observational studies such as cohort, cross-sectional and case-control studies which both facilitate investigators in conducting high quality research and readers in critically appraising such studies.(53-55) No well-established reporting guidelines exist for ITS studies, nevertheless quality criteria have been used in systematic reviews that include ITS studies.

Perhaps the most widely quoted criteria for ITS designs are the Cochrane Effective Practice and Organisation of Care Review Group (EPOC) risk of bias criteria (Table 1).(6) These guidelines encourage authors to address the threat of history (criterion 1), instrumentation (criterion 3), testing (criterion 4), as well as pre-specify effect models a priori (criterion 2), deal with missing data appropriately (criterion 5) and report all outcomes that were part of the original objectives (criterion 6).

Table 1: Cochrane Effective Practice and Organisation of Care Review Group risk of bias criteria for ITS studies(6)

1. Was the intervention independent of other changes?	Score "Low risk" if there are compelling arguments that the intervention occurred independently of other changes over time and the outcome was not influenced by other confounding variables/historic events during study period. If Events/variables identified, note what they are. Score "High risk" if reported that intervention was not independent of other changes in time.
2. Was the shape of the intervention effect pre-specified?	Score "Low risk" if point of analysis is the point of intervention OR a rational explanation for the shape of intervention effect was given by the author(s). Where appropriate, this should include an explanation if the point of analysis is NOT the point of intervention; Score "High risk" if it is clear that the condition above is not met.
3. Was the intervention unlikely to affect data collection?	Score "Low risk" if reported that intervention itself was unlikely to affect data collection (for example, sources and methods of data collection were the same before and after the intervention); Score "High risk" if the intervention itself was likely to affect data collection (for example, any change in source or method of data collection reported).
4. Was knowledge of the allocated interventions adequately prevented during the study?	Score "Low risk" if the authors state explicitly that the primary outcome variables were assessed blindly, or the outcomes are objective, e.g. length of hospital stay. Primary outcomes are those variables that correspond to the primary hypothesis or question as

	defined by the authors. Score "High risk" if the outcomes were not assessed blindly. Score "Unclear risk" if not specified in the paper.
5. Were incomplete outcome data adequately addressed?	Score "Low risk" if missing outcome measures were unlikely to bias the results (e.g. the proportion of missing data was similar in the pre- and post-intervention periods or the proportion of missing data was less than the effect size i.e. unlikely to overturn the study result). Score "High risk" if missing outcome data was likely to bias the results. Score "Unclear risk" if not specified in the paper (Do not assume 100% follow up unless stated explicitly).
6. Was the study free from selective outcome reporting?	Score "Low risk" if there is no evidence that outcomes were selectively reported (e.g. all relevant outcomes in the methods section are reported in the results section). Score "High risk" if some important outcomes are subsequently omitted from the results. Score "Unclear risk" if not specified in the paper.
7. Was the study free from other risks of bias?	Score "Low risk" if there is no evidence of other risk of biases. e.g. should consider if seasonality is an issue (i.e. if January to June comprises the pre-intervention period and July to December the post, could the "seasons' have caused a spurious effect).

Reproduced from: Cochrane Effective Practice and Organisation of Care Review Group, Suggested risk of bias criteria for EPOC reviews - Risk of bias for interrupted time series (ITS) studies, in EPOC Resources for review authors. 2013, Norwegian Knowledge Centre for the Health Services: Oslo.

Ramsay et al developed a separate set of quality criteria based on an earlier form of EPOC guidance as well as Campbell and Stanley's threats to internal validity (Chapter 1: Table 1), which they used to appraise studies in two systematic reviews of behaviour change strategies (Table 2).(56) Criteria 1-3 and 6 are very similar to EPOC criteria 1-4 (Table 1). The Ramsay et al criteria do not address missing data or selective outcome reporting but are more specific about the reliability of outcome measures (criterion 4), indicate that each data point should cover at least 80% of participants (criterion 5), that the reason for number and spacing of observations in the time series should be stated (criterion 7) and that ARIMA or time series regression should be used for analysis. Interestingly, in their systematic reviews of a total of 58 studies, Ramsay et al found that only 20 had ruled out other concurrent events and none had provided a rationale for the shape of the intervention effect model.

Table 2: Ramsay et al (2003) Quality Criteria for ITS Designs(56)

1. Intervention occurred independently of other changes over time	DONE The intervention occurred independently of other changes over time NOT CLEAR Not specified (will be treated as NOT DONE if information cannot be obtained from the authors) NOT DONE Reported that intervention was not independent of other changes in time
2. Intervention was unlikely to affect data collection	DONE Reported that intervention itself was unlikely to affect data collection (for example, sources and methods of data collection were the same before and after the intervention NOT CLEAR Not specified (treated as NOT DONE if information cannot be obtained from the authors) NOT DONE Intervention itself was likely to affect data collection (for example, any change in source or method of data collection reported)
3. The primary outcome was assessed blindly or was measured objectively	DONE Stated explicitly that primary outcome variables were assessed blindly or outcome variables are objective e.g., length of hospital stay, drug levels assessed by a standardized test NOT CLEAR Not specified (treated as NOT DONE if information cannot be obtained from the authors) NOT DONE Outcomes were not assessed blindly
4. The primary outcome was reliable or was measured objectively	DONE Two or more raters with agreement ≥90% or kappa ≥0.8 or outcome assessment is objective, e.g., length of hospital stay, drug levels assessed by a standardized test NOT CLEAR Reliability not reported for outcome measures obtained by chart extraction or collected by an Individual (will be treated as NOT DONE if information cannot be obtained from the authors) NOT DONE Two or more raters with agreement <90% or kappa <0.8
5. The composition of the data set at each time point covered at least 80% of the total number of participants in the study	DONE Data set covers 80–100% of total number of participants or episodes of care in the study NOT CLEAR Not specified (will be treated as NOT DONE if information cannot be obtained from the authors) NOT DONE Data set covers less than 80% of the total number of participants or episodes of care in the study

6. The shape of the intervention effect was prespecified	DONE A rational explanation for the shape of intervention effect was given by the author(s) NOT CLEAR Not specified NOT DONE Any of the conditions above are not met
7. A rationale for the number and spacing of data points was described	DONE Rationale for the number of points stated (e.g., monthly data for 12 months postintervention was used because the anticipated effect was expected to decay) or sample size calculation performed NOT CLEAR Not specified NOT DONE Any of the conditions above are not met
8. The study was analyzed appropriately using time series techniques	DONE ARIMA models were used or time series regression models were used to analyze the data and serial correlation was adjusted/tested for NOT CLEAR Not specified NOT DONE Any of the conditions above are not met

Reproduced from: Ramsay CR, Matowe L, et al. 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.

These criteria provide useful guidance for those undertaking or appraising ITS studies, and their adoption would improve the validity of many ITS studies. Nevertheless, neither encompasses all sources of bias, and there is limited detail as to how each of the criteria could be addressed. Further advancement of these guidelines and wider adoption of an agreed set of more detailed formal reporting standards may improve the quality of evaluations conducted using an ITS design.

2.10 Summary

Interrupted time series is a powerful quasi experimental design. It involves a pre-post comparison but avoids many of the threats of other pre-post designs by modelling the long term underlying trend in the outcome. It is being increasingly adopted for the evaluation of a broad range of public health interventions. While the design has many strengths, ITS, like other quasi-experiments has more inherent threats to their internal validity than experimental designs and it is important that these are recognised by evaluators and dealt with appropriately. There are a number of gaps in the methodological literature including a lack of guidance on the practicalities of the design and analysis of ITS studies, selecting the most appropriate model for defining the counterfactual and the impact of

an intervention, the use of controls to address history bias, and comprehensive reporting of ITS studies. Without further development of these areas, encouraging the wider adoption of this design and ensuring that it is used appropriately would be challenging.

2.11 References

1. 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.
2. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci. 2000;1(1):31-49.
3. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research: Houghton Mifflin Boston; 1963.
4. Campbell DT. Reforms as experiments. Am Psychol. 1969;24(4):409.
5. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
6. Cochrane Effective Practice and Organisation of Care Review Group. Suggested risk of bias criteria for EPOC reviews - Risk of bias for interrupted time series (ITS) studies. Oslo: Norwegian Knowledge Centre for the Health Services; 2013.
7. GOV.UK. Winter Fuel Payment 2014 [Available from: https://www.gov.uk/winter-fuel-payment/overview.
8. Glass G. Interrupted time series quasi-experiments. In: Jager R, editor. Complementary methods for research in education 2nd edition. 2 ed. Washington DC: American Educational Research Association; 1997.
9. Velicer WF, Fava JL. Time series analysis. Handbook of psychology. 2003.
10. Velicer WF, Harrop J. The reliability and accuracy of time series model identification. Evaluation Review. 1983;7(4):551-60.
11. McDowall D, McCleary R, Meidinger E, Hay RA. Interrupted Time-Series Analysis: Sage Publications; 1980. 96 p.
12. Glass GV, Glass GV, Willson VL, Gottman JM. Design and analysis of time-series experiments: Information Age Pub Incorporated; 2009.
13. Glass GV, Willson VL, Gottman JM. Design and analysis of time-series experiments: Colorado Associated University Press Boulder; 1975.
14. Durbin J, Watson GS. Testing for serial correlation in least squares regression: I. Biometrika. 1950:409-28.
15. Prais SJ, Winsten CB. Trend estimators and serial correlation. Cowles Commission discussion paper; 1954.
16. 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.
17. 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.
18. 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.
19. 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.
20. Campbell DT, Ross HL. The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review. 1968:33-53.
21. Ross HL, Campbell DT, Glass GV. Determining the Social Effects of a Legal Reform The British" Breathalyser" Crackdown of 1967. m Behav Sci. 1970;13(4):493-509.
22. McSweeny AJ. EFFECTS OF RESPONSE COST ON THE BEHAVIOR OF A MILLION PERSONS: CHARGING FOR DIRECTORY ASSISTANCE IN CINCINNATI1. J Appl Behav Anal. 1978;11(1):47-51.

23.	Pour M, Torabizadeh MA. Improved prediction of stability lobes in milling process using time series analysis. J Intell Manuf. 2014:1-13.

24.	Anderton CH, Carter JR. The impact of war on trade: An interrupted times-series study. Journal of Peace Research. 2001;38(4):445-57.

25.	Bloom HS, Ham S, Melton L, O'Brien J. Evaluating the Accelerated Schools Approach: A Look at Early Implementation and Impacts on Student Achievement in Eight Elementary Schools. 2001.

26.	Rotter T, Machotta A, Kinsman L, Champion R. Evaluation of a clinical pathway for 'radical laparoscopic prostatectomy': an interrupted time series study. International Journal of Urological Nursing. 2014;8(2):84-9.

27.	Jones RR, Vaught RS, Weinrott M. TIME-SERIES ANALYSIS IN OPERANT RESEARCH1. J Appl Behav Anal. 1977;10(1):151-66.

28.	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.

29.	Morgan OW, Griffiths C, Majeed A. Interrupted Time-Series Analysis of Regulations to Reduce Paracetamol (Acetaminophen) Poisoning. PLOS Medicine. 2007;4(4):e105.

30.	Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of Pay for Performance on the Quality of Primary Care in England. N Engl J Med. 2009;361(4):368-78.

31.	Johns M, Farley SM, Rajulu DT, Kansagra SM, Juster HR. Smoke-free parks and beaches: an interrupted time-series study of behavioural impact in New York City. Tob Control. 2015;24(5):497-500.

32.	Ben-Shlomo Y. Real epidemiologists don't do ecological studies? Int J Epidemiol. 2005;34(6):1181-2.

33.	Gale M, Muscatello DJ, Dinh M, Byrnes J, Shakeshaft A, Hayen A, et al. Alcopops, taxation and harm: a segmented time series analysis of emergency department presentations. BMC Public Health. 2015;15:468.

34.	DiazGranados CA. Prospective audit for antimicrobial stewardship in intensive care: impact on resistance and clinical outcomes. Am J Infect Control. 2012;40.

35.	Hamilton I, Lloyd C, Bland JM, Savage Grainge A. The impact of assertive outreach teams on hospital admissions for psychosis: a time series analysis. J Psychiatr Ment Health Nurs. 2015;22(7):484-90.

36.	Jiang M, Hughes DR, Duszak R, Jr. Screening Mammography Rates in the Medicare Population before and after the 2009 U.S. Preventive Services Task Force Guideline Change: An Interrupted Time Series Analysis. Womens Health Issues. 2015;25(3):239-45.

37.	Narayan H, Thomas SH, Eddleston M, Dear JW, Sandilands E, Bateman DN. Disproportionate effect on child admissions of the change in Medicines and Healthcare Products Regulatory Agency guidance for management of paracetamol poisoning: an analysis of hospital admissions for paracetamol overdose in England and Scotland. Br J Clin Pharmacol. 2015;80(6):1458-63.

38.	LANTERNS project team. Local Authority Collaborators' National Evaluation of Reduced Night-time Streetlight (LANTERNS) 2014 [Available from: http://lanterns.lshtm.ac.uk/.

39.	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.

40.	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.

41.	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-.

42.   Calsyn RJ, Fergus EO, York JL. Interrupted time series analysis: A research technique for evaluating social programs for the elderly. J Gerontol. 1977;32(1):89-96.

43.   Kleck G, Britt CL, Bordua DJ. Emperor Has No Clothes: Using Interrupted Time Series Designs to Evaluate Social Policy Impact, The. J on Firearms & Pub Pol'y. 2000;12:197.

44.   Dennis J, Ramsay T, Turgeon AF, Zarychanski R. Helmet legislation and admissions to hospital for cycling related head injuries in Canadian provinces and territories: interrupted time series analysis. BMJ: British Medical Journal. 2013;346:f2674.

45.   Betts J, Levin J, Miranda AP, Christenson B, Eaton M, Bos H. An Evaluation of Alternative Matching Techniques for Use in Comparative Interrupted Time Series Analyses: An Application to Elementary Education. American Institutes for Research in collaboration/Berkeley Policy Associates/University of California San Diego; 2010.

46.   Biglan A, Henderson J, Humphrey D, Yasui M, Whisman R, Black C, et al. Mobilising positive reinforcement to reduce youth access to tobacco. Tob Control. 1995;4(1):42.

47.   Reding GR, Raphelson M. Around-the-clock mobile psychiatric crisis intervention: Another effective alternative to psychiatric hospitalization. Community Ment Health J. 1995;31(2):179-87.

48.   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.

49.   Fretheim A, Soumerai SB, Zhang F, Oxman AD, Ross-Degnan D. Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. J Clin Epidemiol. 2013;66(8):883-7.

50.   Schneeweiss S, Maclure M, Carleton B, Glynn RJ, Avorn J. Clinical and economic consequences of a reimbursement restriction of nebulised respiratory therapy in adults: direct comparison of randomised and observational evaluations. BMJ. 2004;328(7439):560.

51.   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.

52.   Somers M-A, Zhu P, Jacob R, Bloom H. The validity and precision of the comparative interrupted time series design and the difference-in-difference design in educational evaluation. 2013.

53.   Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.

54.   Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med. 2007;45(4):247-51.

55.   Des Jarlais DC, Lyles C, Crepaz N, Group T. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361.

56.   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.

3. Aims, objectives and approach

3.1 Introduction

The focus of this thesis is on improving the way that ITS is applied for the evaluation of public health interventions. The last chapter reviewed existing methodological work on ITS in order to identify areas in need of further development. In the remainder of this thesis I intend to build upon this work in order to develop methodological frameworks and guidance that will facilitate researchers in improving the validity of ITS studies. In this chapter I will summarise the gaps in the literature that were identified in the last chapter and use these to inform the aims and objectives of my thesis. I will then explain the methodological framework that I will be using for the thesis and how the thesis will be structured.

3.2 Key gaps in the literature

Four clear gaps in the literature have been identified through the literature review. Firstly, while more traditional observational designs (such as cohort and case-control studies) as well as RCTs receive much attention in epidemiology courses and introductory texts, there is a lack of introductory guidance on the design and analysis of ITS studies. This is required so that the design can easily be adopted by those seeking to evaluate an intervention but with little experience of time series methods. Second, the ITS design is very flexible in terms of how it as applied to different interventions. Researchers need to decide what time series data to include, how to define the counterfactual and how to define the impact model. There is currently a lack of information on how these decisions should be made. There is a risk that if researchers base the decisions on apparent changes in the outcome of interest, detected effects could be due to type 1 errors. It is therefore important to develop guidance on model selection for ITS studies. Third, the potential for control series to strengthen the validity of ITS has been widely recognised, nevertheless traditional population based control groups are not always available or even appropriate. Other potential controls exist, nevertheless there is a lack of methodological literature on the different types of controls, their strengths and limitations and how they should be selected. Fourth, again unlike more widely used designs such as RCTs and cohort studies, there is a lack of formal criteria for reporting ITS studies. This is needed in order to help researchers think about and address the key threats to validity, and to facilitate critical appraisal of ITS studies in evidence syntheses.

3.3 Aims and objectives

The overarching aims of this thesis are to improve the way that interrupted time series studies of public health interventions are designed in order to reduce the risk of bias and to make robust ITS designs more accessible to evaluators of public health interventions. This will be achieved through the following objectives:

1) To provide practical introductory guidance on the use of interrupted time series for the evaluation of public health interventions.
2) To develop a methodological framework for defining the impact model of an intervention in order to strengthen the validity of interrupted time series.
3) To develop a methodological framework for the selection of controls and analysis of controlled interrupted time series to limit the risk of history bias.
4) To demonstrate the use of interrupted time series and controlled interrupted time series designs to evaluate complex interventions through their application in case studies.
5) To develop guidance for transparent reporting of interrupted time series studies.

Making ITS more accessible, will facilitate the choice of ITS over less robust designs, such as simple before and after evaluations or non-randomised controlled post-test only designs. Furthermore, improving understanding of the main threats to validity and methods to address these could lead to improvements in the evaluation of public health interventions using ITS designs, and thus improvements in the evidence base for adopting such interventions. Ultimately, stronger evidence for effective interventions will lead to improvements in the health of targeted populations.

3.4 Approach to thesis

3.4.1 Thesis style

I present this thesis as a paper based PhD with most chapters centred around an individual paper prepared for publication.

3.4.2 Type of studies

Methodological studies

A large proportion of the research contributing to this thesis consists of methodological studies. These consist of new approaches to modelling and adapting the ITS design in order to strengthen the validity of the results, including: an objective approach to defining the specific model to be used in the ITS analysis to ensure this is consistent the intervention and outcome under study and a framework adding an appropriate control series to an ITS study in order to limit history bias. Secondly, I also provide guidance for researchers undertaking ITS studies, including: a tutorial on the design and analysis of ITS, and recommendations for reporting ITS studies. Throughout, I demonstrate the strengths and limitations of different approaches by drawing on examples from the case studies described below and draw on existing methodological work both from ITS literature and wider study design literature.

Case studies

The case studies include my own original evaluations of complex interventions and events on health and health service outcomes using ITS designs. The first investigates the effect of the global financial crisis on suicides in Spain and the second evaluates the effect of National Health Service reforms in England on secondary care activity. The selected interventions and events are intentionally chosen because they cannot feasibly be evaluated using traditional study designs and in order to highlight specific methodological considerations. As well as these case studies, I also draw on other purposively selected existing studies which I reanalyse to illustrate different concepts. These include a study on the effect of the smoking ban in Italy on acute myocardial infarction and a study of the effect of changes to street lighting in the UK on road traffic crash casualties. All of these studies use large routine data sources for the outcome time-series. Finally, the penultimate chapter will include a restricted literature of recent empirical studies to examine how authors are reporting their methods in ITS studies. Specific methods used in the case studies are described within the respective papers.

3.4.3 Structure

I begin by presenting the guidance paper on the practical application of ITS, followed by the new methodological frameworks on model selection for ITS analysis and the use of controls in ITS. I then present the two empirical studies that make up my case studies. Finally, I end by presenting the formal criteria that I have developed for the reporting of ITS studies. This does not always represent the order in which the work was done, therefore, in early chapters, I sometimes refer to work undertaken in later chapters. The papers are interlinked and at the end of each chapter I present a "Contribution of the paper" section in which I discuss how the paper addresses the overall aims and objectives of the

thesis and how it links in with other chapters. Finally, the discussion chapter considers the overall lessons learned from the body of work undertaken, how the ITS methodology field has progressed over the duration of the thesis and areas for further development.

3.4.4 Terminology

Some of the key terms used in this thesis are not used consistently across the existing literature. Below I define how I define some of the terms that are used throughout the thesis. Other terms that relate to more focussed topics are described in more detail within the relevant chapters.

Public health interventions

The UK Faculty of Public Health defines three domains of public health: health improvement (including health promotion and education activities targeting the wider determinants of health), health services (including interventions to improve health service effectiveness, efficiency, planning and equity), and health protection (including communicable disease control and response to chemical, radiological and environmental hazards).(1) In this thesis, I consider public health interventions to include those in all three of these domains that act at a population level, rather than individual level. I also include interventions or unplanned events that do not specifically target health but that nevertheless may affect public health outcomes. I do not include interventions acting at an individual level, such as the effect of a drug or a surgical procedure on patient prognosis.

Interrupted time series

Some authors have described pre-post designs which do not take into account underlying trends as interrupted time series.(2, 3) In common with most study design texts, I use the term interrupted time series exclusively for those studies which incorporate trends over time within the model.(4, 5) Studies with a single pre-intervention baseline observation and a single post-intervention observation or studies in which there are multiple pre- and post-intervention observations but where the time they were measured is not included in the model , in common with most authors, I regard as simple pre-post or simple before-after designs.(4, 5)

Controlled interrupted time series

Similarly I differentiate controlled interrupted time series from difference in difference studies (or controlled before and after studies) in that the former accounts for the trend whereas the latter does not. I use the term controlled interrupted time series but this has also been referred to as comparative interrupted time series and interrupted time series with non-equivalent control elsewhere in the literature.(4, 6) I include all different types of control series within this definition, including control

outcomes (or non-equivalent dependant variables). Different types of controls are discussed in more detail chapter 7.

3.5 References

1. UK Faculty of Public Health. What is public health 2017 [Available from: http://www.fph.org.uk/what_is_public_health.
2. Idemoto LM, Williams BL, Ching JM, Blackmore CC. Implementation of a custom alert to prevent medication-timing errors associated with computerized prescriber order entry. Am J Health-Syst Pharm. 2015;72(17):1481-8.
3. Alfa MJ, Lo E, Olson N, MacRae M, Buelow-Smith L. Use of a daily disinfectant cleaner instead of a daily cleaner reduced hospital-acquired infection rates. Am J Infect Control. 2015;43(2):141-6.
4. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
5. 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.
6. 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.

4. Methodological paper 1: Interrupted time series regression tutorial

4.1 Research paper: Interrupted time series regression for the evaluation of public health interventions: a tutorial

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	James Lopez Bernal
Principal Supervisor	Antonio Gasparrini and Steven Cummins
Thesis Title	The use interrupted time series for the evaluation of public health interventions

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	International Journal of Epidemiology		
When was the work published?	April 2016		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the concept of this paper, all analysis and writing the paper.

Student Signature: ▮▮▮▮▮ **Date:** _____17/01/2018_____

Supervisor Signature: _____   **Date:** __25/1/2018_____

Supervisor Signature: _____   **Date:** __25/1/2018_____

Copyright Clearance Center

Note: Copyright.com supplies permissions but not the copyrighted content itself.

1	2	3
PAYMENT	REVIEW	CONFIRMATION

Step 3: Order Confirmation

Thank you for your order! A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us 24 hrs/day, M-F at +1.855.239.3415 Toll Free, or write to us at info@copyright.com. This is not an invoice.

Confirmation Number: 11693456
Order Date: 01/17/2018

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

James Lopez Bernal
jamie.lopezbernal@gmail.com
+44 7872834737
Payment Method: n/a

Order Details

International journal of epidemiology

Order detail ID:	70937962	
Order License Id:	4271460392718	
ISSN:	0300-5771	
Publication Type:	Journal	
Volume:		
Issue:		
Start page:		
Publisher:	OXFORD UNIVERSITY PRESS	
Author/Editor:	INTERNATIONAL EPIDEMIOLOGICAL ASSOCIATION	

Permission Status: ✓ Granted
Permission type: Republish or display content
Type of use: Thesis/Dissertation

View details

Note: This item will be invoiced or charged separately through CCC's **RightsLink** service. More info $ 0.00

This is not an invoice.

Total order items: 1   **Order Total: 0.00 USD**

Education Corner

Interrupted time series regression for the evaluation of public health interventions: a tutorial

James Lopez Bernal,[1,]* Steven Cummins,[1] Antonio Gasparrini[1]

[1]Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author. Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH, UK. E-mail: james.lopez-bernal@lshtm.ac.uk

Abstract

Interrupted time series (ITS) analysis is a valuable study design for evaluating the effectiveness of population-level health interventions that have been implemented at a clearly defined point in time. It is increasingly being used to evaluate the effectiveness of interventions ranging from clinical therapy to national public health legislation. Whereas the design shares many properties of regression-based approaches in other epidemiological studies, there are a range of unique features of time series data that require additional methodological considerations. In this tutorial we use a worked example to demonstrate a robust approach to ITS analysis using segmented regression. We begin by describing the design and considering when ITS is an appropriate design choice. We then discuss the essential, yet often omitted, step of proposing the impact model a priori. Subsequently, we demonstrate the approach to statistical analysis including the main segmented regression model. Finally we describe the main methodological issues associated with ITS analysis: over-dispersion of time series data, autocorrelation, adjusting for seasonal trends and controlling for time-varying confounders, and we also outline some of the more complex design adaptations that can be used to strengthen the basic ITS design.

Key Messages

- Interrupted time series is a valuable study design for evaluating the effectiveness of population-level health interventions.
- A segmented regression approach can be used to analyse an interrupted time series study by testing the effect of an intervention on the outcome of interest using an appropriately defined impact model.
- Methodological considerations specific to interrupted time series analysis include possible time-varying confounders such as seasonal trends or concurrent events to the intervention, and potential autocorrelation of data.

Introduction

Traditional epidemiological study designs such as cohort and case-control studies can provide important evidence about disease aetiology, but they are less useful as intervention studies, due to limitations such as confounding owing to group differences and, in particular, healthy user bias.[1] Randomized controlled trials (RCTs) have long been considered the gold standard design for evaluating the effectiveness of an intervention, yet RCTs are not always possible, in particular for health policies and programmes targeted at the population level.[2–4] Furthermore, there is often a need to retrospectively evaluate interventions which have already been implemented, often for political reasons, either without randomization or to a whole population and so without any control.[2] The interrupted time series (ITS) study design is increasingly being used for the evaluation of public health interventions; it is particularly suited to interventions introduced at a population level over a clearly defined time period and that target population-level health outcomes.[1,5] ITS has been used for the evaluation of a wide range of public health interventions including new vaccines, cycle helmet legislation, changes to paracetamol packaging, traffic speed zones and precautions against nosocomial infections, as well as in the evaluation of health impacts of unplanned events such as the global financial crisis.[6–11] Other articles have outlined the design and highlighted the strengths and limitations of ITS.[1,5,12,13] Further methodological papers have described some of the more specific in-depth modelling techniques that may be employed by those familiar with the analysis of time series data.[14,15] Nevertheless, there is a lack of introductory guidance for those implementing an ITS evaluation for the first time. Here, we aim to demonstrate a step-by-step ITS analysis including: considering when an ITS might be an appropriate design choice and the data required; hypothesizing the type of impact the intervention will have on the outcome; how to use a regression model to analyse the effect; the main methodological issues that need to be taken into account; and finally, a brief outline of model checking techniques. A worked example is used to illustrate the methods (Box 1) and the supplementary material (available as Supplementary data at *IJE* online) includes the dataset used as well as code for use with the statistical packages Stata and R, so that readers may reproduce the analysis.

The interrupted time series design

A time series is a continuous sequence of observations on a population, taken repeatedly (normally at equal intervals) over time. In an ITS study, a time series of a particular outcome of interest is used to establish an underlying trend, which is 'interrupted' by an intervention at a known point in time. The hypothetical scenario under which the

---

Box 1. The worked example

The example dataset used in this paper is taken from a study by Barone-Adesi *et al.* on the effects of the Italian smoking ban in public places on hospital admissions for acute coronary events (ACEs, ICD10 410-411). In January 2005, Italy introduced regulations to ban smoking in all indoor public places, with the aim of limiting the adverse health effects of second-hand smoke. The subset used here are ACEs in the Sicily region between 2002 and 2006 among those aged 0-69 years. This dataset is available in Supplementary Appendix 1 with an excerpt presented in Table 1 and represented graphically in Figure 1. This example is not meant to contribute to the substantive evidence on the topic, rather to illustrate the methods. Further details on the dataset can be found in the original publication by Barone-Adesi *et al.* 2011.[16]

---

intervention had not taken place and the trend continues unchanged (that is: the 'expected' trend, in the absence of the intervention, given the pre-existing trend) is referred to as the 'counterfactual'. This counterfactual scenario provides a comparison for the evaluation of the impact of the intervention by examining any change occurring in the post-intervention period.[12,17] Figure 1 illustrates the design using the smoking ban example (Box 1): the graph displays the pre-intervention trend of monthly rates of ACE admissions (continuous line), and the counterfactual scenario (dashed line). Given that most of the points lie below the counterfactual line, there is a visual suggestion of a decrease in the ACE admissions in the post-intervention period which is compatible with a possible positive impact of the smoking ban. ITS models, described below, can provide statistical evidence about whether this represents a real decrease.

Step 1: is an interrupted time series design appropriate?

The first decision when considering an ITS is whether it is an appropriate design for the particular evaluation in question. This depends on the nature of both the intervention and the outcome of interest, as well as the type of data available:

The intervention

ITS requires a clear differentiation of the pre-intervention period and the post-intervention period. In some evaluations it may be difficult to define when the intervention began and to differentiate the effects of different
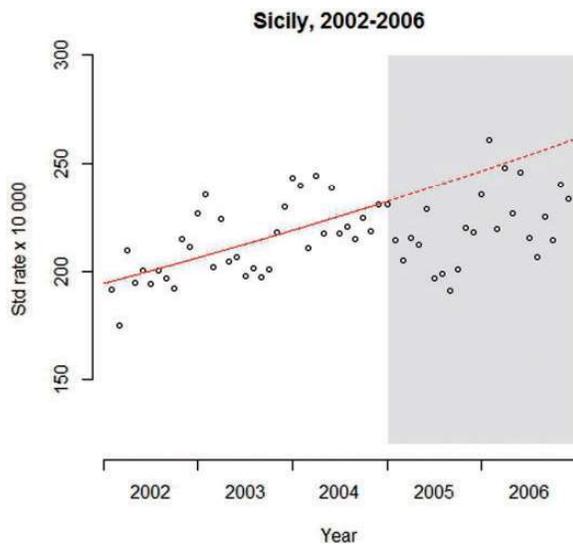
Figure 1 Scatter plot of example dataset. Standardized (Std) rate of ACE over time. White background, pre-intervention period; grey background, post-intervention period; continuous line, pre-intervention trend; dashed line, counterfactual scenario

components. This does not necessarily require the intervention to be introduced overnight but the period of implementation should be well defined so that it can be considered separately.

The implementation of the example intervention was very clear with a ban on smoking in public places throughout Italy from 10 January 2005.[16] Where interventions have a gradual roll-out, the implementation phase can be modelled as a gradual (slope) change (see Step 2).

The outcome

Outcomes may take various forms such as counts, continuous data or binary variables. ITS works best with short-term outcomes that are expected to change either relatively quickly after an intervention is implemented or after a clearly defined lag.

ACEs, the outcome in the worked example, are a short-term outcome with rapid onset, and the authors of the study quote evidence suggesting that the acute effects of both active and passive smoking disappear quickly after the exposure is removed.[16] Other diseases associated with smoking, such as lung cancer, may have been less appropriate as the timing between intervention and outcome is much less clear and can be highly variable. In this situation it may be preferable to use an intermediate outcome such as smoking prevalence.[18]

Data requirements

Sequential measures of the outcome should be available both before and after the intervention. There are no fixed limits regarding the number of data points, as the power depends on various other factors including distribution of data points before and after the intervention, variability within the data, strength of effect, and the presence of confounding effects such as seasonality. Zhang *et al.* conducted simulations with power calculations under different model parameters, and suggest that studies with few time points or with small expected effect sizes should be interpreted with caution as they may be underpowered, and that similar simulations should be conducted a priori under such circumstances.[19] Power increases with the number of time points, but it is not always preferable to have more data points where historical trends have changed substantially, as this would not provide an accurate depiction of the current underlying trends.[20] It is therefore recommended that pre-intervention data are inspected visually. Power is also increased if the numbers of data points are equally distributed before and after the intervention, though this is often not practical.[19] Given the requirement for a relatively long time series, routine data are often most appropriate in ITS studies. As with all study designs, it is important to assess the quality of the data in terms of its validity and reliability. With routine data it is especially important to understand the potential impact of changes to data collection or recording, particularly when these coincide with the implementation of the intervention, as this could bias results.[12]

The example dataset has 59 months of routine hospital admissions data with 600-1100 ACEs at each time point. The large number of time points and minimal variability within the data provides enough power to detect relatively small changes in the hospital admission rate. In practice, as is the case in the worked example, the ITS design is often used in the evaluation of 'natural experiments' occurring in real-world settings and is becoming ever more possible with the increasing availability and quality of routine data spanning before and after interventions.

Step 2: proposing the impact model

Once an ITS design is chosen, the next step is to hypothesize how the intervention would impact on the outcome if it were effective, in particular whether the change will be a gradual change in the gradient of the trend, a change in the level or both, and whether the change will follow the intervention immediately or there will be a lag period before any effect is expected. Examples of some possible impact models are illustrated in Figure 2. It is important that this decision is made a priori based on existing literature and knowledge of
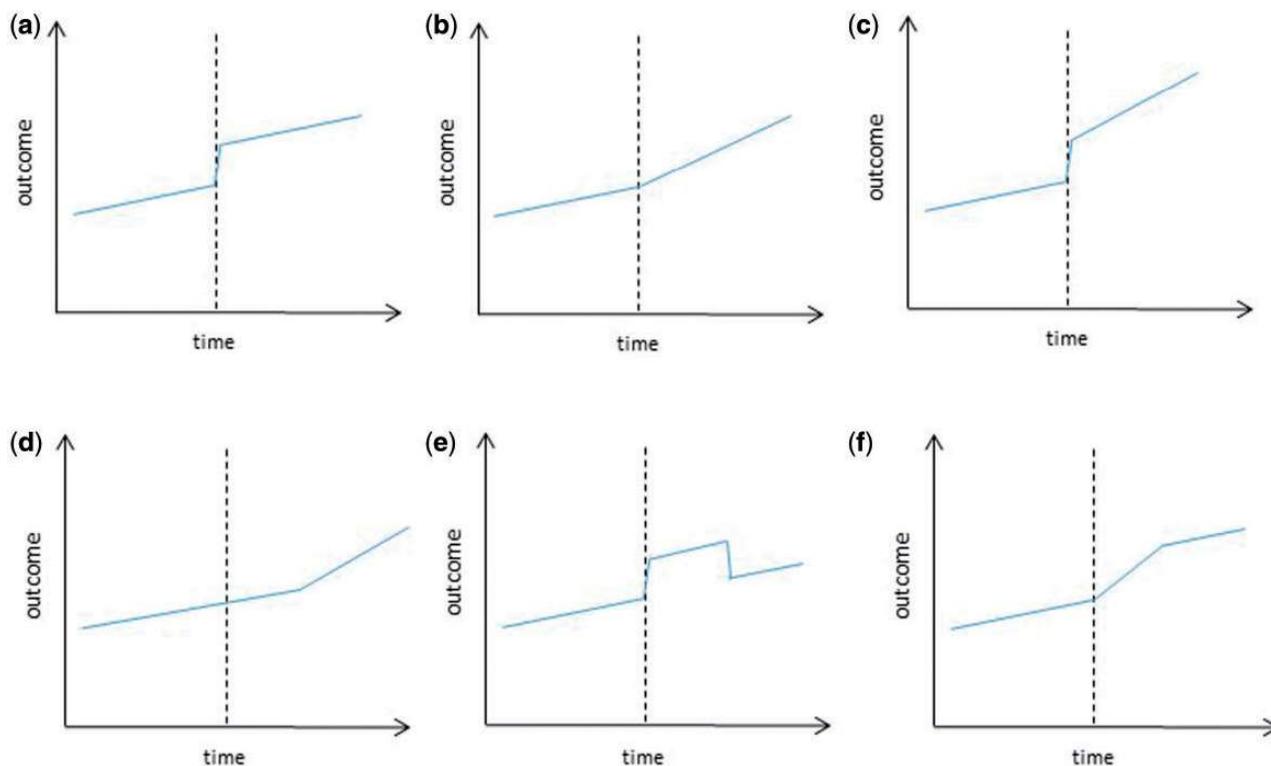
Figure 2 Examples of impact models used in ITS

(a) Level change; (b) Slope change; (c) Level and slope change; (d) Slope change following a lag; (e) Temporary level change; (f) Temporary slope change leading to a level change.

the intervention and the mechanism by which it is expected to act on the outcome. Where existing knowledge of the intervention is limited, selecting the most appropriate impact model can be difficult and may require exploratory analysis of alternative data. Relying on the outcome data to select the best impact model is discouraged as this increases the likelihood of an effect being detected due to random fluctuations or chance, and consequent artefactual conclusions on the effect of the intervention.

Barone-Adesi *et al.* assumed a level change in ACEs occurring with no lag. This assumption was based on existing evidence suggesting that the acute cardiovascular risks from passive smoking disappear within a short time.[16]

Step 3: descriptive analysis

As with all statistical analyses, initial summary statistics and plots should be undertaken to familiarize researchers with the data. This should include a scatter plot of the time series, as displayed in Figure 1, which can help to identify the underlying trend, seasonal patterns and outliers. More traditional descriptive analyses, such as summaries and bivariate comparisons between the outcome and potential time-varying confounders, as well as simple before-and-after comparisons, are recommended.

Step 4: regression analysis

A minimum of three variables are required for an ITS analysis:

i. T: the time elapsed since the start of the study in with the unit representing the frequency with which observations are taken (e.g. month or year);
ii. X_t: a dummy variable indicating the pre-intervention period (coded 0) or the post-intervention period (coded 1);
iii. Y_t: the outcome at time t.

In standard ITS analyses, the following segmented regression model is used:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t$$

where β_0 represents the baseline level at $T = 0$, β_1 is interpreted as the change in outcome associated with a time

Table 1. Excerpt from the example dataset

Year	Month	Time elapsed (*T*)	Smoking ban[a] (*X*)	ACEs (*Y*)	Std popn
2004	1	25	0	914	381656.3
2004	2	26	0	808	383680
2004	3	27	0	937	383504.2
2004	4	28	0	840	386462.9
2004	5	29	0	916	383783.1
2004	6	30	0	828	380836.8
2004	7	31	0	845	383483
2004	8	32	0	818	380906.2
2004	9	33	0	860	382926.8
2004	10	34	0	839	384052.4
2004	11	35	0	887	384449.6
2004	12	36	0	886	383428.4
2005	1	37	1	831	388153.2
2005	2	38	1	796	388373.2
2005	3	39	1	833	386470.1
2005	4	40	1	820	386033.2
2005	5	41	1	877	383686.4
2005	6	42	1	758	385509.3
2005	7	43	1	767	385901.9
2005	8	44	1	738	386516.6
2005	9	45	1	781	388436.5
2005	10	46	1	843	383255.2
2005	11	47	1	850	390148.7
2005	12	48	1	908	385874.9

ACEs, hospital admissions for acute coronary event; Std popn, age-standardized population in person-years.[16]

[a]Smoking ban: 0, smoking ban not in place; 1, smoking ban in place.

Figure 3 Interrupted time series with level change regression model. Line: predicted trend based on the unadjusted regression model

unit increase (representing the underlying pre-intervention trend), β_2 is the level change following the intervention and β_3 indicates the slope change following the intervention (using the interaction between time and intervention: TX_t). The regression model above represents the impact model (c) in Figure 2; models (a) and (b) can easily be specified by excluding the terms $\beta_3 TX_t$ or $\beta_2 X_t$, respectively. Impact models (d)-(f) require slightly more complex variable specifications (Supplementary Appendix 5, available as Supplementary data at *IJE* online).

In our example, T, X and Y are shown in Table 1. As is frequently the case in population health evaluations, here the outcome is a count and, without loss of generality, a Poisson regression model was used. Other regression models can equally be used, such as ordinary least squares (linear) regression for continuous outcomes, for example the duration of cycling trips in an ITS study looking at the impact of public transport strikes on usage of a bicycle share programme in London.[21] Most of the steps described in this tutorial remain the same for the analysis of other types of outcomes, unless specifically stated. Furthermore, the

age-standardized population (in person-years) was included as an offset variable to convert the outcome into a rate and adjust for any potential changes in the population over time (though this is not essential if the population is relatively stable over time, as in this case). Given that a level change model was hypothesized, the interaction term for the slope change is not required in the model. This model, shown using Stata code and R code in Supplementary Appendices 2 and 3 (available as Supplementary data at *IJE* online), suggests that there is very strong evidence of a reduction in ACEs following the smoking ban, with a decrease of 11% [relative risk (RR) 0. 894; 95% confidence interval (CI) 0.864-0.925; $P < 0$. 001) as illustrated in Figure 3.

Step 5: addressing methodological issues

Whereas the basic model implemented so far provides an indication of the potential association between the intervention and the outcome, there are a number of distinctive issues with time series data that may need to be addressed in order to improve the robustness of the analysis.

Seasonality

Many diseases and other outcomes have a seasonal pattern and this is evident in the ACE data in Figure 1. Seasonality can cause two problems: first, if there is an uneven distribution of months before and after the intervention, such as a higher proportion of winter months, this could bias the results, especially in the analysis of short series. Second, outcomes in one month tend to be more similar to those in
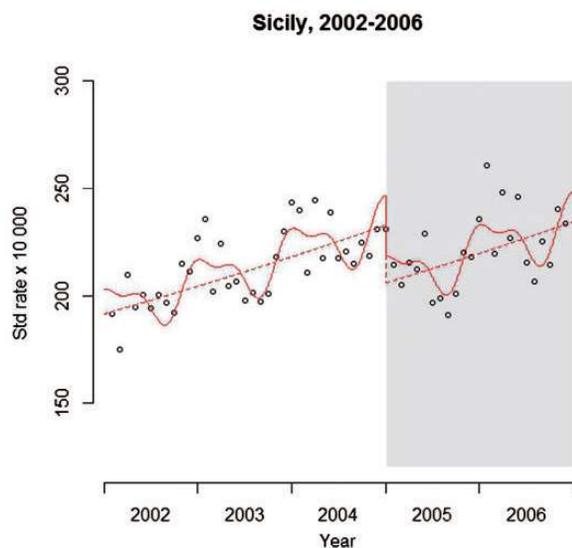
Figure 4 Model adjusted for seasonality. Solid line: predicted trend based on the seasonally adjusted regression model. Dashed line: de-seasonalized trend

neighbouring months within the same time of year, leading to autocorrelation and over-dispersion (discussed below). There are a range of methods for controlling for seasonality and other long-term trends; these include: a model stratified by the calendar month (or other time period); or using more complex functions such as Fourier terms (pairs of sine and cosine functions); or splines. Each of these methods is explained in more detail by Bhaskaran *et al.* 2013.[22] Figure 4 shows the example analysis after adjustment for seasonality through a Fourier term, with results suggesting that the association is largely unaffected: (RR: 0.885; 95% CI 0.839-0.933; $P < 0.001$).

Time-varying confounders

One of the strengths of ITS studies is that they are generally unaffected by typical confounding variables which remain fairly constant, such as population age distribution or socioeconomic status, as these only change relatively slowly over time and are normally taken into account when modelling the underlying long-term trend. Nevertheless, ITS can be affected by time-varying confounders that change more rapidly. Seasonality can be considered a time-varying confounder; others may include levels of a particular infectious disease that is prone to outbreaks, weather events etc. Where such time-varying confounders have been measured, they can be controlled for by including variables representing them in the regression model, as is commonly undertaken in other epidemiological analyses. A special category of time-varying

confounders are other events that occur around the same time as the intervention and that potentially influence the outcome. These might include other simultaneous interventions targeting the same outcome, or risk factors for that outcome, or natural events that could affect the outcome.

Potential time-varying confounders in the smoking ban study might include changes in diagnostic procedures for detecting ACEs, for example a new troponin test had been progressively implemented in Italy since 2000,[23] or interventions targeting other risk factors for cardiovascular disease such as a healthy eating intervention.

Use of controls and other more complex ITS designs

Where time-varying confounders are either unmeasured or unknown, a range of design adaptations can be used to control for possible concurrent events including: adding a control group or control outcome which would not have been affected by the intervention (known as a controlled interrupted time series); using a multiple baseline design whereby the intervention is introduced in different locations at different times; or adding additional phases so that the intervention is first introduced and then withdrawn to establish whether withdrawal of the intervention leads to a reversal of the effect. These methods are described in more detail elsewhere.[12,13,15,24]

Over-dispersion

An assumption of the Poisson distribution is that the variance is equal to the expected count. However, in analyses of real data, the variance frequently tends to be greater (a phenomenon known as over-dispersion) which would lead to incorrect estimation of the standard errors. A scaling adjustment is therefore made to correct to the model to correct this, detailed by Bhaskaran *et al.* and illustrated in Supplementary Appendices 2 and 3 (available as Supplementary data at *IJE* online).[22] This issue does not apply for the analysis of continuous outcomes when a Gaussian distribution, including a residual error to be estimated, is assumed. In the example this widens the 95% confidence interval marginally to 0.839-0.953, yet there is still very strong evidence of an effect ($P = 0.001$).

Autocorrelation

A second assumption of standard regression models is that observations are independent. This assumption is often violated in time series data because consecutive observations tend to be more similar to one another than those that are further apart, a phenomenon known as autocorrelation. Fortunately, in many epidemiological data, autocorrelation is largely explained by other variables, in particular seasonality (discussed above); therefore, after controlling for these factors, residual autocorrelation is

rarely a problem. Nevertheless, autocorrelation should always be assessed by examining the plot of residuals and the partial autocorrelation function and, where data are normally distributed, conducting tests such as the Breusch-Godfrey test.[22,25] Where residual autocorrelation remains, this should be adjusted for using methods such as Prais regression or autoregressive integrated moving average (ARIMA), described in more detail elsewhere.[26,27] There is very little evidence of autocorrelation in the worked example and even less after adjustment for seasonality (Supplementary Appendices 2 and 3, available as Supplementary data at *IJE* online).

Further extensions

Further extensions, as described by Bhaskaran *et al.* (2013) for environmental time series, and in more detail elsewhere, can also be applied to ITS studies, including: stratified analyses according to potential effect-modifying variables; increasing power by allowing different locations to have trends modelled individually rather than relying on the aggregated trend; and modelling non-linear trends.[9,22,23]

Step 6: model-checking and sensitivity analyses

A range of model-checking techniques have been described above including plotting residuals and partial autocorrelation functions. Furthermore, sensitivity analyses can be conducted to test the impact of varying a range of model assumptions, such as different lags, types of impact model or approaches to adjusting for seasonality.[11,22]

Summary

In this article we have introduced the key steps for readers undertaking an ITS study, including highlighting the main methodological considerations and how they may be addressed. ITS analyses are one of the strongest evaluative designs when randomization is not possible; furthermore, they often allow a more detailed assessment of the longitudinal impact of an intervention than may be possible with an RCT and, given that they are frequently undertaken in real-world settings, may have stronger external validity. Another important feature in ensuring that research gets translated into practice is that graphical and numerical presentation of results can be easily understood by those with little expert knowledge of statistical and epidemiological methods. Nevertheless, there are some important threats to the validity of ITS analyses, perhaps the most important of which include the potential for the erroneous conclusion of intervention effectiveness due to data-driven model specification, and lack of control for time-varying confounders. It is therefore essential that any

interpretations regarding the causal effect of any association are undertaken with caution and that some of the key steps to analysis highlighted in this article, such as a priori model specification and methodological extensions to control for confounders, are followed. With carefully planned analyses and handling of potential threats to validity, ITS can provide valuable evidence about the effectiveness of health interventions.

Supplementary Data

Acknowledgements

Funding

References

1. Soumerai SB. How do you know which health care effectiveness research you can trust? A guide to study design for the perplexed. *Prev Chronic Dis* 2015;**12**:E101.
2. Bonell CP, Hargreaves J, Cousens S *et al.* Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. *J Epidemiol Community Health* 2009;**65**:582-87.
3. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;**94**:400–05.
4. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, *et al.* Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;**321**:694.
5. 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**:299-309.
6. Lau WCY, Murray M, El-Turki A *et al.* Impact of pneumococcal conjugate vaccines on childhood otitis media in the United Kingdom. *Vaccine* 2015;**33**(39):5072–9.
7. Dennis J, Ramsay T, Turgeon AF, Zarychanski R. Helmet legislation and admissions to hospital for cycling related head injuries in Canadian provinces and territories: interrupted time series analysis. *BMJ* 2013;**346**;f2674.
8. Hawton K, Bergen H, Simkin S *et al.* Long term effect of reduced pack sizes of paracetamol on poisoning deaths and liver transplant activity in England and Wales: interrupted time series analyses. *BMJ* 2013;**346**;f403.
9. 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.

10. Derde LP, Cooper BS, Goossens H *et al*. Interventions to reduce colonisation and transmission of antimicrobial-resistant bacteria in intensive care units: an interrupted time series study and cluster randomised trial. *Lancet Infect Dis* 2014;**14**:31-39.

11. 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:**23**(5):732–6.

12. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference, Boston, Houghton Mifflin, 2002.

13. Biglan A, Ary D, Wagenaar AC. The value of interrupted time series experiments for community intervention research. *Prev Sci* 2000;**1**:31-49.

14. 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.

15. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. *J Eval Clin Pract* 2011;**17**:1231-38.

16. 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**:e17419.

17. Campbell DT, Stanley JC. *Experimental and Quasi-experimental Designs for Research*. Boston, MA: Houghton Mifflin, 1963.

18. Callinan JE, Clarke A, Doherty K, Kelleher C. Legislative smoking bans for reducing secondhand smoke exposure, smoking prevalence and tobacco consumption. *Cochrane Database Syst Rev* 2010;**4**.

19. 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**:1252-61.

20. Robinson S, Harris H. *Smoking and Drinking Among Adults, 2009: A Report on the 2009 General Lifestyle Survey*. London: Office for National Statistics, 2009.

21. Fuller D, Sahlqvist S, Cummins S, Ogilvie D. The impact of public transportation strikes on use of a bicycle share program in London: interrupted time series design. *Prev Med* 2012;**54**:74-76.

22. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *Int J Epidemio* 2013;**42**:1187-95.

23. Gasparrini A, Gorini G, Barchielli A. On the relationship between smoking bans and incidence of acute myocardial infarction. *Eur J Epidemiol* 2009;**24**:597-602.

24. 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**:162-68.

25. Godfrey L. Misspecification tests and their uses in econometrics. *J Stat Plann Infer* 1996;**49**:241-60.

26. Prais SJ, Winsten CB. Trend estimators and serial correlation. Cowles Commission discussion paper Stat No. 383, Chicago, 1954.

27. Nelson BK. Time series analysis using autoregressive integrated moving average (ARIMA) models. *Acad Emerg Med* 1998;**5**:739-44.

4.2 Summary of appendices for methodological paper 1

All supplementary material for this paper is presented in the appendices (Chapter 11)

11.1.1 Example dataset

11.1.2 Stata code

11.1.3 R code

11.1.4 Further model specifications

11.1.5 Contribution of the candidate to the paper

4.3 Contribution of the paper

Motivation for the paper

The underlying motive for this thesis was to make robust evaluation of public health interventions more accessible to researchers and policy makers. Quasi-experimental designs, such as ITS, rarely feature in epidemiology and public health courses. Furthermore, weak before-after or cross-sectional designs are commonly used for evaluating public health interventions.(1, 2) This is particularly the case in service public health, despite frequent availability of routine data that would allow more powerful ITS designs to be used.(3) Unfortunately, the threats to the validity of these designs are such that effects can rarely be attributed to the intervention with confidence.(4) It was therefore important to me to introduce ITS in a way that would allow researchers to understand the principles of the design and the practicalities of a basic ITS analysis. The main purpose was to encourage wider adoption of ITS and to act as teaching material.

Contribution to the thesis

This paper directly addresses the first objective of the thesis *"To provide practical introductory guidance on the use of interrupted time series for the evaluation of public health interventions."* It also introduces some of the methodological concepts which will be expanded upon in Chapters 5 and 6, namely model selection and controlled interrupted time series analysis. Furthermore, the methods described here form the basis of those that were used in the two case studies in Chapters 7 and 8. By providing an easily reproducible guide to undertaking an ITS study, this paper goes a long way towards achieving one of the primary aims of the thesis – to make ITS accessible. Furthermore, the key threats to the validity of ITS are introduced in order to ensure that these are taken into account by researchers thereby making ITS evaluations more robust.

Outputs and contributions to the literature

The Stata and R code as well as example data layouts for different impact models (appendices 11.1.1-11.1.4) are important outputs of the paper that should facilitate easy adoption of these techniques for researchers new to ITS analysis. The paper is currently the most read article of 2017 in the International Journal of Epidemiology, suggesting that it is encouraging wider adoption of the methods (22/09/2017).(5) Furthermore, this material, along with the stepwise approach to the paper, enable it to be easily adapted for use as teaching material. This is something that I have implemented for teaching of students and public health service professionals.(6-8) Feedback from readers suggests that it is being used in a similar way elsewhere.

Conclusion

This paper is an important contribution to the ITS literature and has been described as "the first detailed guidance on how to conduct ITS studies".(9) It goes some way towards achieving the primary aims of the thesis and introduces those concepts that will be given further attention in later chapters.

References

1. Elhayany A, Regev S, Sherf M, Reuveni H, Shvartzman P. Effects of a fundholding discontinuation. An Israeli health maintenance organization natural experiment. Scandinavian journal of primary health care. 2001;19(4):223-6.
2. Sargent RP, Shepard RM, Glantz SA. Reduced incidence of admissions for myocardial infarction associated with public smoking ban: before and after study. BMJ. 2004;328(7446):977-80.
3. PA Consulting Group. High impact healthy workplace interventions: case studies of good practice. Commissioned by the UK Health Work and Wellbeing Executive. In: Executive UHWaW, editor. London2008.
4. 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.
5. International Epidemiology Association. International Journal of Epidemiology: Most read articles 2017 [Available from: https://academic.oup.com/ije.
6. Hargreaves J, Davey C. Module Specification: Evaluation of Public Health Interventions 2016 [Available from: https://www.lshtm.ac.uk/sites/default/files/2017-02/Evaluation_of_public_health_interventions_2016_17.pdf.
7. Naci H. Executive MSc in Health Economics, Outcomes and Management in Cardiovascular Sciences: HP4D5E Research Design for Evaluating Health Programs and Policies 2017 [Available from: http://www.lse.ac.uk/resources/calendar2017-2018/courseGuides/HP/2017_HP4D5E.htm.
8. Armstrong B, Lopez Bernal J. Marie Stopes International: Interrupted Time Series. Marie Stopes International2016.
9. Polus S, Pieper D, Burns J, Fretheim A, Ramsay C, Higgins JPT, et al. Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews. J Clin Epidemiol. 2017;91(Supplement C):56-69.

5.Methodological paper 2: Model selection in interrupted time series

5.1 Research paper: Model selection in interrupted time series

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	James Lopez Bernal
Principal Supervisor	Antonio Gasparrini and Steven Cummins
Thesis Title	The Use of Interrupted Time Series for the Evaluation of Public Health Interventions

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Journal of Clinical Epidemiology
Please list the paper's authors in the intended authorship order:	James Lopez Bernal, Stephen Soumerai, Antonio Gasparrini
Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the concept of this paper, all analysis and writing the paper.

Student Signature: _____ **Date:** __17/01/2018_____

Supervisor Signature: _____   **Date:** __25/01/2018____

Supervisor Signature: _____   **Date:** __25/01/2018____

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

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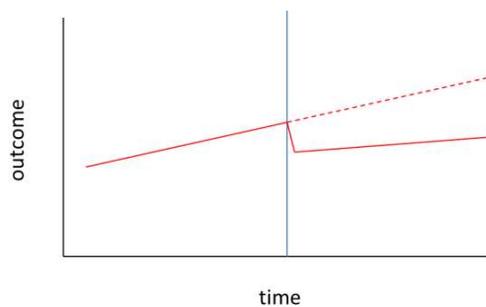
Figure 1: the interrupted time series design
Solid line = modelled trend; dashed line = counterfactual; vertical line = intervention implementation. This shows a step decrease and decrease in the slope following the intervention.

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 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)

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. In practice this involves modelling the underlying trend in the outcome of interest. 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.

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.(14) Epidemiologists are accustomed to identifying potential confounding variables *a priori* and using multivariate regression models to adjust for these potential confounders.(15) 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,(16) population age distribution,(17) ethnicity or levels of deprivation.(18) 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, 19) A range of methods exist for controlling for seasonality in time-series regression models which have been described in more detail elsewhere.(2, 20)

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.(21)

Linearity

The above factors can be defined *a* priori by the researcher and may explain any non-linear trends. Nevertheless, the assumption of 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.(14) 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.(14) In particular, although flexible methods exist to describe non-linearities in regression models,(22) 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.(23) 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.(24) Figure 2b was therefore considered to be the most appropriate model *a priori*.
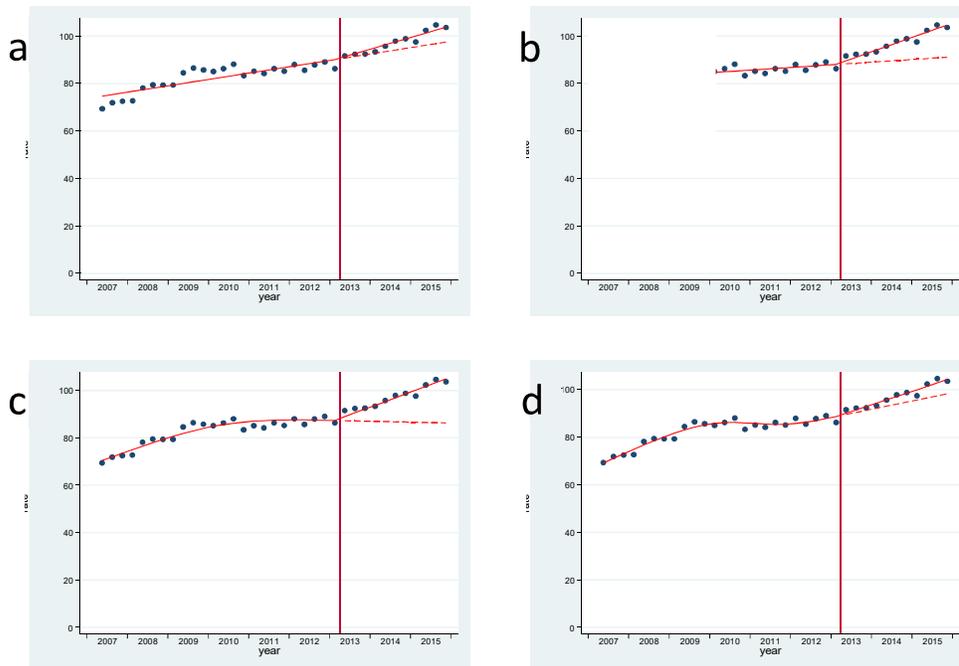
Figure 2: Four different approaches to modelling the trend
Y-axis represents the quarterly total specialist visits in England per 1000. a: simple linear trend; b: excluding data prior to April 2010; c: non-linear allowing one inflection point; d: non-linear trend allowing two inflection points. Dots= 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.

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

---

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.(25, 26) 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.(27) Different model parameters can be used to allow different effects to be expressed following the intervention. Flexible 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.(28) 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, 28) 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.(29)

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.(30) 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).(19, 31) For example, Landrigan et al evaluated the impact of a introducing a hospitalist system (employing physicians with a primary focus on caring for hospitalised inpatients) on length of stay in a paediatric hospital.(32) 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.

Floor and ceiling effects

There is often a limit to how much an intervention could decrease or increase an outcome if the outcome is constrained by other factors, this can result in a floor or ceiling effect (figure 3d).(33) 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, for example by allowing a second slope change at the floor or ceiling.(2) Furthermore, floor and ceiling effects should be considered as a potential reason for trend changes in the discussion.

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)
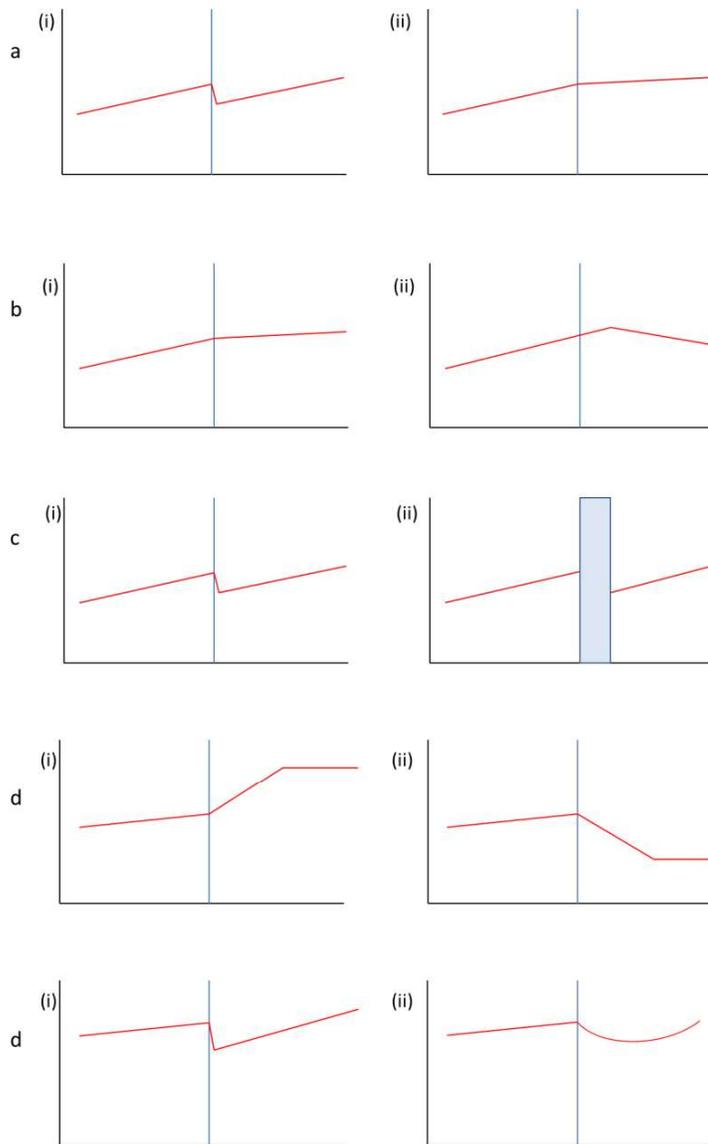
Figure 3: Interrupted time series impact models

70

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) ceiling effect d(ii) floor effect; d(i) waning effect following a step change, d(ii) gradual effect with gradual waning.

Illustrative example

Figure 3 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 3a we use one of the most commonly used flexible impact models which allows for a step and slope change at the point of the intervention, 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 3b, 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 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 .
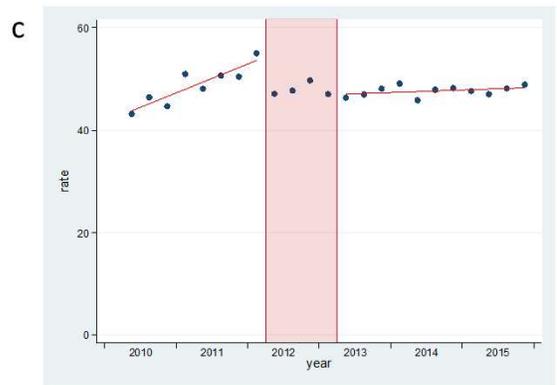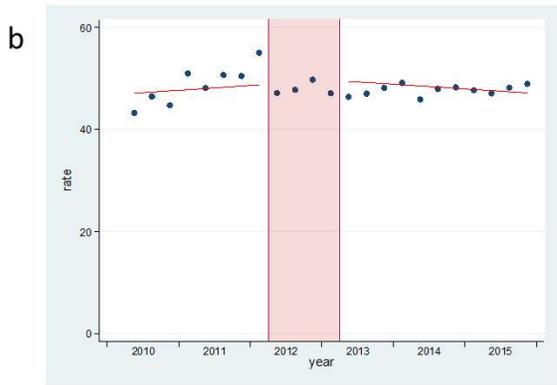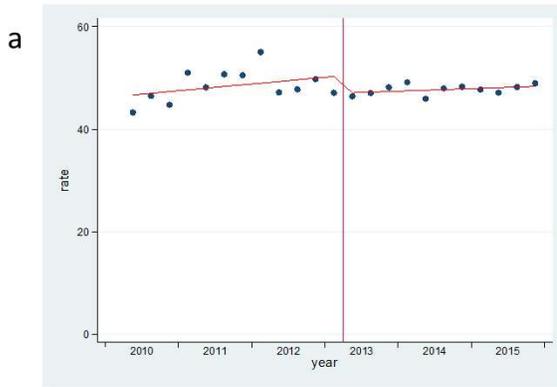
a) Step and slope change model; b) slope change only with a one year intervention phase; d) step and slope change with a one year intervention phase.

Framework for defining the impact model

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

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

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)

Sensitivity analysis

It may be necessary to use sensitivity analyses 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, 43) If the same effect is detected under different assumptions, this can increase confidence in the results. However, 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 to consider any sensitivity analyses *a priori*. Where there is a lot of uncertainty about the nature of potential effect of an intervention it may be necessary to run various exploratory sensitivity analyses in the first instance, rather than regarding the study as an explanatory evaluation.

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,(44, 45) 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 or table 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. 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.
15. Coggon D, Barker D, Rose G. Epidemiology for the Uninitiated: John Wiley & Sons; 2009.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.

21. 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.
22. Wood S. Generalized additive models: an introduction with R: CRC press; 2006.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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).
30. 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.
31. 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.
32. 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.
33. 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.
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.   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.

44.   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.

45.   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.

5.2 Summary of appendices for methodological paper 2

All supplementary material for this paper is presented in the appendices (Chapter 11)

11.2.1 Contribution of the candidate to the paper

5.3 Contribution of the paper

Motivation for the paper

While other limitations of ITS, such as history bias, have been well recognised, the issue of model misspecification has not previously been discussed in the literature. From the literature review, and from undertaking sensitivity analyses in the case studies, it was clear that different assumptions in the way the counterfactual and the impact model were defined could lead to very different results. As shown later in Chapter 9, the majority of ITS studies use a step and slope change model but do not explain why this was chosen and how it fits with the intervention. Neither do they explain the reason for the chosen time series data range. Given the potential impact of these assumptions on the results, it was therefore important for me to highlight the dangers of incorrect model selection and provide a methodological framework for selecting the most appropriate model.

Contribution to the thesis

This paper highlights some of the key underlying modelling decisions that are made, either explicitly or implicitly, when undertaking an ITS study. This directly addresses objective three of the thesis: "To develop a methodological framework for defining the impact model of an intervention in order to strengthen the validity of interrupted time series." As discussed in the paper, modelling the impact of the intervention requires both the counterfactual and the impact model to be accurately and clearly defined and both of these areas are covered in detail. This paper expands on Chapter 4, where impact models were first introduced and draws on examples from the two case studies in Chapters 7 and 8. It will also feed into the reporting guidelines developed in Chapter 9.

Outputs and contributions to the literature

A key message of the paper is that researchers need to adapt the model based on knowledge of the intervention. It is therefore important that researchers have an in depth understanding of the intervention under study, including its timing, how it was rolled out and when it is anticipated to have an impact on the outcome. Modelling decisions should be made a priori and post-hoc changes based on the data should be avoided. Currently authors rarely make modelling choices explicit (see Chapter 9). This can lead to allegations of bias and data manipulation. This paper provides a methodological framework to limit the risk of such bias.

Conclusion

Inappropriate model selection is a previously undescribed threat to the validity of ITS studies. By making model choices clear and explicit and basing these on a good understanding of the intervention and outcome under study, researchers can improve the robustness of ITS studies.

6. Methodological paper 3: The use of controls in ITS studies

6.1 Research paper: The use of controls in interrupted time series studies of public health interventions

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	James Lopez Bernal
Principal Supervisor	Antonio Gasparrini and Steven Cummins
Thesis Title	The use interrupted time series for the evaluation of public health interventions

<u>If the Research Paper has previously been published please complete Section B, if not please move to Section C</u>

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	International Journal of Epidemiology
Please list the paper's authors in the intended authorship order:	James Lopez Bernal, Steven Cummins, Antonio Gasparrini
Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the concept of this paper, all analysis and writing the paper.

Student Signature: _____   **Date:** 17/01/18

Supervisor Signature: _____   **Date:** __25/01/2018_____

Supervisor Signature: _____   **Date:** __25/01/2018_____

82

THE USE OF CONTROLS IN INTERRUPTED TIME SERIES STUDIES OF PUBLIC HEALTH INTERVENTIONS

James Lopez Bernal[1]*

Steven Cummins[1]

Antonio Gasparrini[1]

1. Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH

ABSTRACT

Interrupted time series analysis differs from most other intervention study designs in that it involves a before-after comparison within a single population rather than a comparison with a control group. This has the advantage that selection bias and confounding due to between-group differences are limited. However, the basic interrupted time series design cannot exclude confounding due to co-interventions or other events occurring around the time of the intervention. One approach to minimise potential confounding from such simultaneous events is to add a control series so that there is both a before-after comparison and an intervention-control group comparison. A range of different types of controls can be used with interrupted time series designs, each of which have associated strengths and limitations. Researchers undertaking controlled interrupted time series studies should carefully consider a priori what confounding events may exist and whether different controls can exclude these or if they could introduce new sources of bias to the study. A prudent approach to the design, analysis and interpretation of controlled interrupted time series studies is required to ensure that valid information on the effectiveness of health interventions can be ascertained.

KEY MESSAGES

- History bias due to other interventions or events occurring around the time of the intervention is the primary threat to the validity of interrupted time series studies.

- A wide range of different controls can be used in order to limit history bias and improve the validity of an ITS study.

- Controls should be selected by considering, a priori, the possible sources of history bias and examining for differential changes in covariates between the study series and the control series throughout the study period.

- Researchers should take care in interpreting the results of controlled interrupted time series studies, in particular when the results differ from those of simple (uncontrolled) analysis.

INTRODUCTION

Evaluation of public health interventions normally relies on comparing the outcome of interest in a population exposed to an intervention to that in an external control group not subject to the same intervention.(1) Interrupted time series (ITS) is an increasingly popular design that adopts a different approach whereby comparisons are instead made across time within a single population.(2) This design is generally applied to natural experiments with an intervention introduced at a known point in time. By collecting data at regular intervals over time, a pre-post comparison can be made while accounting for underlying trends in the outcome.(2) Because the evaluation is based on observing a single population over time, the ITS design is free from problems due to between-group differences, such as selection bias or unmeasured confounders. Furthermore, by modelling the underlying trend, ITS also controls for within-group characteristics that tend to change only slowly over time.(3) Nevertheless, ITS studies cannot exclude time-varying confounders which do not form part of the underlying trend, for example other interventions or events occurring around the time of the intervention that may also affect the outcome.(4)

One approach that limits the threat of these other confounding events is to include a control series, a design known as a controlled (or comparative) interrupted time series (CITS) analysis. A lack of effect in a well-chosen control can provide stronger evidence to support a causal relationship between the intervention and outcome. Conversely, the presence of an effect in the control series indicates that the change may be attributable to different factors. Indeed, a number of recent within study comparisons have provided empirical evidence of the validity of the CITS design by demonstrating comparable results to RCT benchmarks.(5-8) Nevertheless, while the basic ITS design has been described in detail elsewhere and reference is made to the inclusion of a control as a method of improving the validity of the design,(2, 9) there is little guidance available on what a control series can and cannot solve and how to select an appropriate control in CITS studies. The purpose of this paper

is to evaluate the use of controls in ITS studies and provide a framework for their selection, analytical approaches and the interpretation of results. We then provide an illustration of the application of this framework using an example from a recent study where alternative types of controls can be selected and compared.

EVALUATIVE STUDY DESIGNS

In order to know whether an intervention has caused an effect, a comparison needs to be made between the observed change in the outcome and the *counterfactual*, that is, what would have happened if the intervention had not taken place. Of course, it is not possible to observe the intervention both being implemented and not being implemented in the same population at the same time, therefore the true counterfactual is never known. Evaluation design is therefore centred on creating the best approximation of the true counterfactual and then comparing what happened in the intervention group to the approximated counterfactual.(3) There are two main approaches to approximating the counterfactual: controlled designs and before-and-after designs.(3)

Controlled designs:
Controlled designs normally compare the same outcome in the intervention group and an external control. Randomised controlled trials, cross sectional studies as well as other designs less commonly used for intervention evaluations (such as cohort and case control studies) all make comparisons between an intervention group and a control. The advantage of this approach is that both intervention and control groups are compared at the same point in time so other time sensitive factors that would affect both populations (such as other interventions or events that might impact on the outcome of interest) can be excluded. Nevertheless, selection bias and differences between the intervention and control population may mean that observed effects could be due to other confounding factors (which may be unknown or difficult to measure) rather than the intervention.(1) Randomisation addresses

this limitation in experimental studies, however this is often not desirable, feasible or practical in studies evaluating public health interventions.(1, 10, 11) Other approaches, such as adjusting for multiple variables in regression models or propensity score matching can account for known characteristics that differ between the two groups, but cannot control for unmeasured confounders.(1, 11, 12)

Before-and-after designs:
Before-after designs involve making a comparison between a period of time after the intervention has occurred and a period of time before the intervention within a single population. Here, the pre-intervention period effectively acts as the control. Simple pre-post designs make before-after comparisons by estimating the change from a single pre-intervention time point to a single post-intervention time point. However, these have poor internal validity as they cannot exclude underlying trends as a cause for any change. Conversely, interrupted time series use multiple pre-intervention and post-intervention observations, thereby allowing the underlying trend to be accounted for. These have the advantage that confounding is rarely a problem as population characteristics tend to only change gradually over time. (3, 13) Nevertheless, such before-after comparisons cannot exclude other events or co-interventions occurring around the same time as the intervention under investigation as the cause of any detected change in the outcome. This phenomenon is known as history bias in Campell and Stanley's classical list of threats to internal validity.(4)

CONTROLLED INTERRUPTED TIME SERIES

Controlled (or comparative) interrupted time series (CITS) involves adding a control series, which was not exposed to the intervention, to the basic ITS design (Figure 1).(8) This results in the definition of a more complex counterfactual based on both a before-and-after comparison and an intervention-control comparison. The primary benefit of this approach is that it can help to control for history bias due to time-varying confounders, in particular co-interventions and other events concurrent to the

intervention.(3) In a CITS, if an effect is detected in the intervention group but not in a well-chosen control (Figure 1a) this suggests that the effect is more likely to be due to the intervention; conversely if an effect is detected in both the intervention and control series (Figure 1b), this suggests that it is due to some confounding event.
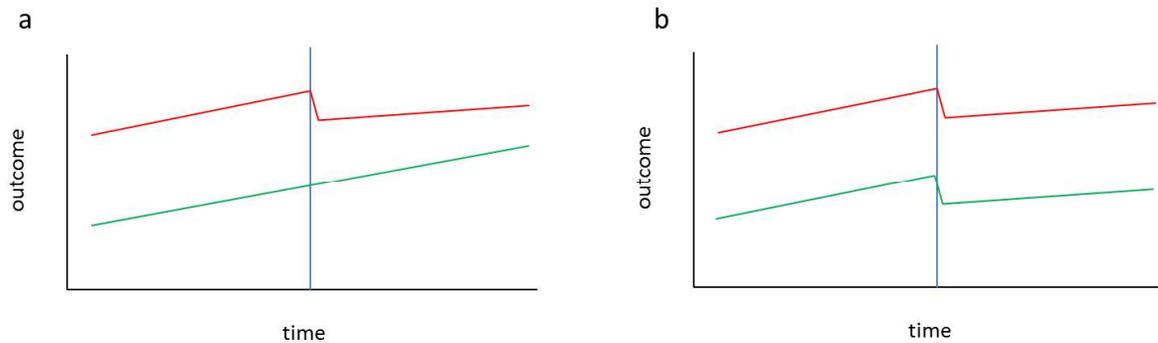
Figure 1: Controlled interrupted time series

Red line = intervention series, green line = control series. (a) Here there is an effect in the intervention series (step and slope decrease) but no effect in the control series which increases confidence that the effect is due to the intervention. (b) Here there is a step and slope decrease in both the intervention and control series suggesting the change is due to some other event or co-intervention that affected both groups.

CITS is related to other study designs applied in evaluation analyses. For instance, the controlled before and after design (CBA) also involves a before-and-after and intervention-control comparison. Nevertheless, the CBA design involves a comparison of a single pre and a single post intervention, or a comparison of pre and post-intervention means. While both CITS and CBA designs involve a difference in difference calculation, CBA designs do not take into account baseline trends and therefore use the control group alone in order to approximate the counterfactual.(3, 14)

An extension of the CITS design is the multiple baseline design. This is similar to a stepped wedge cluster randomised trial but typically does not involve randomisation. Here, following a baseline period, the intervention is first introduced in one group while one or more other groups act as a control.(15, 16) The intervention is subsequently introduced in other groups at different times, with a

different subset acting either as intervention or control groups at each time. In this design, the observation of an effect of similar strength and magnitude following the intervention in multiple different groups at multiple sequential time points, can provide strong evidence that the observed effect is due to the intervention rather than other potential confounding events.(15, 16)

SELECTING A CONTROL

With studies that rely on the control as the sole means of approximating the counterfactual (including RCTs, cross-sectional studies and CBA studies) the central prerequisite when selecting a control is that it is as similar as possible to the intervention group. The ideal control is the same in terms of all variables other than exposure to the intervention.(1, 3) RCTs accomplish this through randomisation. Where randomisation is not possible a range of methods have been developed to achieve covariate balance in cross-sectional and CBA designs including multivariable regression, propensity score matching and synthetic controls.(17-19) Nevertheless, none of these methods can account for systematic differences in unknown variables.(17, 20)

As described above, ITS studies use the pre-intervention trend to predict the counterfactual. The purpose of the control in this case is to exclude time varying confounders, in particular co-interventions or other events occurring around the time of the intervention, as these are generally unpredictable based on modelling pre-intervention trends.(2, 3) It follows that the key attribute of a control series for a CITS study should be its ability to control for known co-interventions or external events that may affect the outcome. Therefore, the control series should be exposed to any such co-interventions or events that might also affect the intervention series, however, it should not be exposed to other interventions or events that could impact on the control series alone (and not the intervention series). The latter could result in artifactual effects being detected in the CITS which are in fact due to independent changes in the control series. Several different types of control series have

been used for CITS analyses; we have broadly classified some of the most commonly used controls as follows: location based control groups, characteristic based control groups, behaviour based control groups, historical cohort controls, control outcomes and control time periods. Table 1 describes these six types of controls, each of which may plausibly control for different sources of confounding events.

Researchers should also consider whether the intervention under study could have an indirect effect on the control series, for example there may be a contamination effect in location based or characteristic based control groups, or a substitution effect with control outcomes.(21, 22) A contamination effect occurs when the effects of the intervention spreads beyond the target population, for example with behaviour change interventions, whereby members of the control population learn about the new behaviour and adopt it themselves.(21) An example of a substitution effect would be an evaluation of the effect of an intervention aimed at reducing the prescription of a certain drug; in this scenario, prescriptions of a similar drug not targeted by the intervention may be considered as a control outcome, however, doctors may substitute the targeted drug with the similar drug so that it is indirectly affected by the intervention.(22) Control series that could be indirectly affected by the intervention should be excluded.

Finally, while covariate balance between the intervention and control series in ITS is not required to predict the counterfactual, and is therefore not the fundamental prerequisite that it is in other controlled designs, it remains important for two reasons: firstly, certain subgroups may be more susceptible to either an intervention or a confounding event than others. If such a subgroup is more concentrated in the intervention group than the control, one would expect a greater effect in the intervention group simply due to the population distribution. Secondly, if certain characteristics are associated with the outcome and these characteristics change differentially over time in the intervention and control groups, the trend in the outcome may change in one group but not the other

simply due to differential changes in the populations under investigation. For example, there is evidence that rates of cycle head injuries are lower in females than in males.(23) In the cycle helmet legislation study by Dennis et al described in Table 1, if the intervention population had a higher proportion of females at baseline than the control population this would not necessarily be a problem.(24) Nevertheless, if the proportion of females increased more rapidly in the intervention group than in the control population following the intervention, this would be a source of confounding as there may be a decrease in head injuries in the intervention group simply due to the population change, rather than any effect of the intervention. Matching techniques, including propensity score matching can be used to ensure balance of known covariates at baseline which can help to limit the effects of differential susceptibility to the intervention by population subgroup.(18, 25) Nevertheless, whether matching at baseline or not, it is still important to check for covariate balance between the control and intervention group throughout the study period. If there are changes over time, variables associated with the outcome can be included in the interrupted time series regression model to adjust for confounding. However, none of these methods can control for unknown confounding and this should be recognised as a limitation of CITS studies in common with other non-randomised controlled designs

Table 1: Types of controls

Type of Control	Description	Examples	Strengths	Limitations
Location based control	The control series is selected from another location similar to the study location but that did not receive the intervention. The type of location depends on the scale of the intervention, for large scale interventions this may be a different geographical area (such as a country, district or city), whereas for smaller scale interventions this could be a different institution or a	Dennis et al (2013) evaluated the impact of the introduction of helmet legislation in a number of Canadian provinces on cycling related head injuries by comparing outcomes in Canadian provinces that did not implement helmet legislation.(24) Lopez Bernal et al (2017) compared the change in hospital activity in England, following major health reforms, to those	Help to control for confounding events that would affect both locations.	Cannot exclude events that are unique to the intervention location. For example, in the study of helmet legislation, reductions in head injuries could be due to a protective effect of helmets (presumably the desired effect) or due to a reduction in the number of cyclists if the need to wear a helmet acts as a deterrent (which may not be a desired

	different ward within a hospital.	in Scotland where the reforms did not apply.(26)		effect), comparing to provinces that did not implement the legislation would not help to distinguish these.(24)
Characteristic based control	Interventions are sometimes targeted according to certain characteristics, for example only males or only females, a certain age group, a specific ethnic minority group or patients with a certain diagnosis. Controls may be chosen from those groups that were not targeted.	Feigl et al (2015) investigated the impact of a national ban on smoking in high schools and selected a control based on age by comparing trends in smoking prevalence among those aged 12-18 years compared to those aged 19-24 years.(27) Kontopantelis et al (2015) examined the impact of a national primary care financial incentive scheme on trends in consultation rates among patients with severe mental illness compared to matched patient controls with no severe mental illness.(28)	In cross sectional or similar designs, this type of control is not ideal as the characteristic that differentiates the two groups is a known confounder that cannot be controlled for, nevertheless in ITS studies, where the pre-intervention trend is the primary control, characteristic control groups can help to exclude concurrent events to the intervention that both groups would have been exposed to.	Interventions may have been targeted at the intervention group because of a detected deviation in the trend, for example in the smoking ban study, high schools may have been targeted because of recent increases in smoking among adolescents therefore trends could differ substantially from the control group.(27)
Behaviour based control	Sometimes the intervention does not affect all of those within the population to whom it is targeted, this tends to occur when the intervention targets a behaviour that some individuals never performed (either prior to the intervention starting or since). Those individuals who never performed the behaviour can therefore be used as a control group.	Ross-Degnan et al (1993) evaluated the impact of the national withdrawal of a non-steroidal anti-inflammatory drug (Zomepirac) on prescribing of other analgesics. They used physicians who never prescribed Zomepirac (and were thus unaffected by its withdrawal) as the control group.(22) Kiseley et al (2011) used a CITS to evaluate the impact of an increase in taxation of "alcopops" on alcohol related harm by comparing the effect in young people aged 15-29 to the effect in those aged 30-49. Alcopops tend to be favoured by young people so it was expected that older groups would be largely unaffected.(29)	Controls can be very similar to the intervention group other than in the specific behaviour targeted by the intervention.	It may be difficult to directly identify those who did not perform the behaviour, therefore, a proxy may have to be used – such as, age, in the alcopops study. This proxy may, however, introduce selection bias for example selecting based on age could bias the alcopops study because age could be independently associated with both the intervention (younger people may be lower earners and thus more affected by a tax increase) and the outcome (if rates of alcohol related harm vary with age).(29)

Historical cohort control	Historical cohorts are commonly used in the evaluation of education interventions but have also been used for healthcare evaluations.(7) This is possible where a cohort periodically progresses to another level (for example moving from one school year to the next) and is replaced by another cohort. The intervention cohort can then be compared to a previous or subsequent cohort.(3)	Schneeweiss et al (2004) evaluated the impact of a restriction of state funding of nebulised respiratory medication.(30) The intervention time series used monthly observations of nebulised drug expenditure, primary care visits and admissions to emergency department for a year (6 months prior to the policy and 6 months after the policy). Control series were taken from the same population one year and two years before.	Historical cohorts help to rule out seasonal effects (such as stockpiling of drugs in the Scheeweiss et al study) and events that occur on an annual basis.(30)	They would not control for events that are unique to the year in which the intervention was implemented.
Control outcome	Where no control group is possible, another option is to compare the effect on the primary outcome to that in a related 'control outcome' (or 'non-equivalent dependant variable') within the same group. Such an outcome should not be affected by the intervention, but would be affected by confounding events.	Walter et al (2011) conducted a study on the impact of helmet legislation on head injuries in Australia (similar to that by Dennis et al described above)(24). Rather than other locations, they used limb injuries as a control outcome to exclude other effects on cycling.(31) Lopez Bernal et al (2016) used accidental deaths as a control outcome in their ITS study of the impact of the financial crisis on suicides in Spain as both suicides and accidental deaths undergo similar judicial review and recording methods.(32) This enabled them to control for other events that could have impacted on these processes.	Uses the same group as an intervention population therefore it is not sensitive to many of the between group differences that can affect other controls. Can often be used to control for potential confounders that would only affect the intervention group. For example by using limb injuries as a control outcome Walter et al were able to control for any changes in the number of cyclists where comparing to different states could not.(24, 31)	Can only control for factors that would affect both the primary outcome and the control outcome.
Control time period	It may be possible to use the primary outcome as its own control for interventions that are only active at certain times (certain times of day or days of the week). In this case the outcome during times in which the	Ross et al (1970) studied the impact of 1967 British Road Safety Act, which increased the use of breathalysers to reduce drink driving, on traffic casualties. They compared the effect on the weekend evenings when pubs are busiest and accidents are more	Uses the same group as the intervention group therefore it is not sensitive to many of the between group differences that can affect other controls.	Can only be used for short-term outcomes with rapid onset. The outcome must be recorded at a sufficiently high time resolution to allow identification of when the intervention is active

	intervention is inactive act as the control.	likely to be due to drink driving to that at commuting hours when pubs are closed and accidents are less likely to be due to drink driving.(33)		and inactive. For example to the nearest hour if the intervention is only active at night time.(34)

ANALYSIS AND INTERPRETATION OF CITS STUDIES

There are a range of analyses that can be employed when undertaking CITS studies. These can broadly be divided into two: separate analysis of the intervention series and the control series; or a single model incorporating both series. Separate analysis is the simpler approach and may be suitable, particularly if there is no change in the control series. A single model can be developed by including indicator variables for the intervention or control series as interaction terms (web appendix 1) or by generating a new series of the ratio or difference between the intervention and control series at each time point.(5, 35) This approach provides a test of the differential effects of the intervention (level or slope change) across the groups. The benefit of this approach is that if there are trend changes in the control series which could be due to some confounding event, any additional effect of the intervention can still be calculated.

Even if a single model combining the intervention and control series is selected, we would recommend starting with a simple (uncontrolled) ITS of the intervention group. Both the uncontrolled ITS and the CITS should always be planned *a priori* and the results reported with equal prominence. If the result of the simple ITS mirrors that of the CITS this provides a greater degree of confidence that any association between intervention and effect is likely to be causal. Results should be interpreted more cautiously if either the simple ITS shows an effect but the CITS shows no effect (or a smaller effect) or if the CITS shows an effect but the simple ITS does not. If the simple ITS shows an effect but the CITS does not, then there may have been a change in both the intervention and the control series – this suggests possible history bias due to some simultaneous event or co-intervention. If the CITS shows

an effect but the simple ITS does not, the change may be due (at least in part) to a change in the

control series, as a result of some other event that affected the control population but not the

intervention group. This framework for analysing and interpreting CITS studies is summarised in Figure

2.
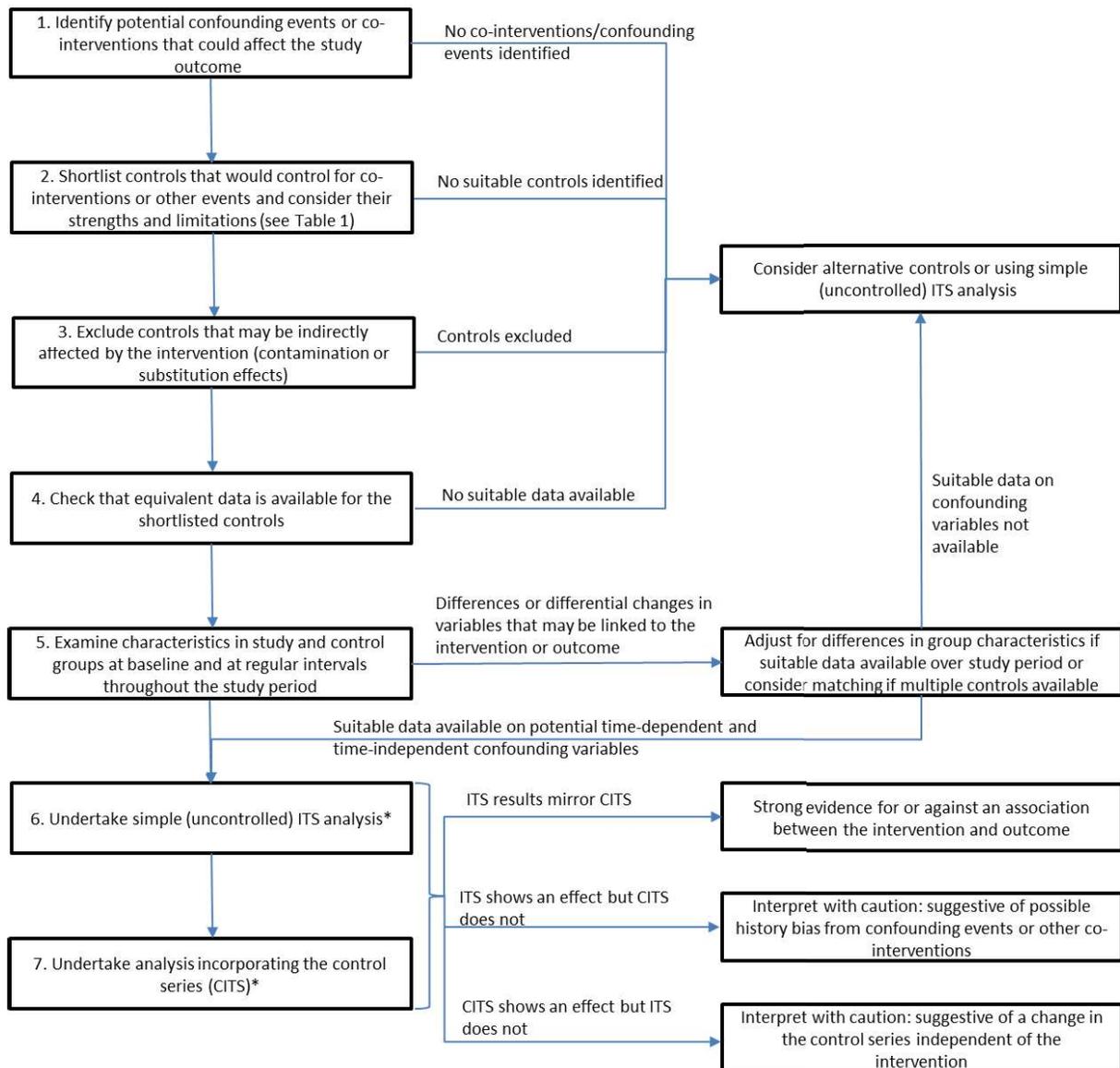
Figure 2: Suggested steps for undertaking a controlled interrupted time series study

*Both analyses should be undertaken and reported

SENSITIVITY ANALYSIS

Different ITS model assumptions can be checked using sensitivity analyses. Specific to CITS designs, different types of controls may control for different sources of bias or confounding events. Therefore, where possible researchers should undertake sensitivity analyses using different types of controls to control for those potential sources of bias that have been identified a priori. Similar to the primary model, sensitivity analyses should be clearly pre-specified to avoid the possibility of 'data dredging'.

ILLUSTRATIVE EXAMPLE

Steinbach et al (2015) recently used a CITS design to evaluate the impact of a range of changes to streetlights in various regions of the UK on road traffic crashes and crime at night.(34, 36) The purpose of the intervention was to save energy and costs. The intervention consisted of reductions in the brightness of streetlights, replacement of bulbs with lower energy consumption bulbs, reducing the hours during which streetlights were turned on at night (i.e. turning on later and turning off earlier) and reducing the ambient light threshold at which sensors would activate streetlights. The authors hypothesised that while the intervention may save costs, reduced street lighting may unintentionally increase road traffic crashes and crime at night. To illustrate the design and interpretation of CITS studies we used an extract of these data on minor roads in the Birmingham and Black Country region to analyse the impact of the intervention (introduced from 2010) on the number of casualties from road traffic crashes. Note that, for simplicity of this illustration, we make the assumption that the intervention was introduced simultaneously in 2010 throughout the region and that it would have a step change effect. A number of different controls can be considered for the analysis and we work through the process of selecting controls and analysing the CITS.

Data on road traffic crash casualties included variables on the region, the road type and the time of the road traffic crash. Therefore, three potential controls could be considered (1) another region as a location based control, (2) comparison of casualties from road traffic crashes on minor roads to those on major roads as a characteristic based control, (3) comparison of road traffic crash casualties at night to road traffic crash casualties during the day when street lights are not in use as a control time period.

Our first step in selecting a control is to identify potential confounding events or co-interventions that would affect the study outcome. In this study other changes to roads, such as changes to road layout or new road safety measures, were identified as a potential confounding event that could impact on road traffic crashes independently of the street lighting interventions. Another potential concern was instrumentation effects due to unidentified changes to data collection. Considering each of the controls in turn: the location based control would not be able to control for the identified confounding factors as road changes may have differed from one region to the next and data collection was separate in each region. The characteristic controls (different road types), would control for changes to data collection processes within a region but would not be able to control for road changes as these are likely to differ between minor and major roads. In this example, the control time period is the most appropriate as this uses the same roads and same data source and should therefore adequately control for all known potential confounders. No other interventions or events that would only affect day time road traffic crashes were identified and it was considered unlikely that the intervention would have any indirect effect on this control. Day time road traffic crashes were therefore selected as the control series.

The next step was to check characteristics of the control and intervention series at baseline and throughout the study period for covariate balance. We know that the data comes from the same roads therefore this will not be different between night and day. However, no data on the characteristics of

the population of night time drivers compared to day time drivers were available. One could assume that there are fewer elderly drivers with visual impairments at night, however this is unlikely to change differentially between the intervention and control group over the study period independently of the intervention.

Figure 3 shows the results of the analysis. First, an uncontrolled ITS analysis (Figure 3a) was undertaken. This shows a significant decrease in road traffic crash casualties following the intervention, contrary to the hypothesised increase. Nevertheless, when a CITS analysis using daytime road traffic crash casualties is run (Figure 3b), the decrease is also present in the controls series and there is no evidence of any additional effect in the intervention series. This suggests that the effect is due to a change occurring at the same time as the intervention and biasing the previously estimated association.

To demonstrate the possible consequences of poor control selection, in figure 3c a location based control is used instead. We select the most closely matched region according to baseline characteristics (including number of roads in the region, population size, age distribution, sex distribution and level of unemployment). There is also no significant difference in baseline trends between the control and intervention group. In this case the results are very similar to the uncontrolled analysis, showing strong evidence of a decrease in road traffic crash casualties following the intervention. Nevertheless, this control group is clearly unable to account for changes to road layout or changes to data collection that are unique to the region, and could result in erroneous conclusions about the effect of the intervention. This highlights the potential pitfalls of selecting controls without first carefully considering potential confounding events or co-interventions specific to the study context, even when there is good covariate balance between the intervention and control group.
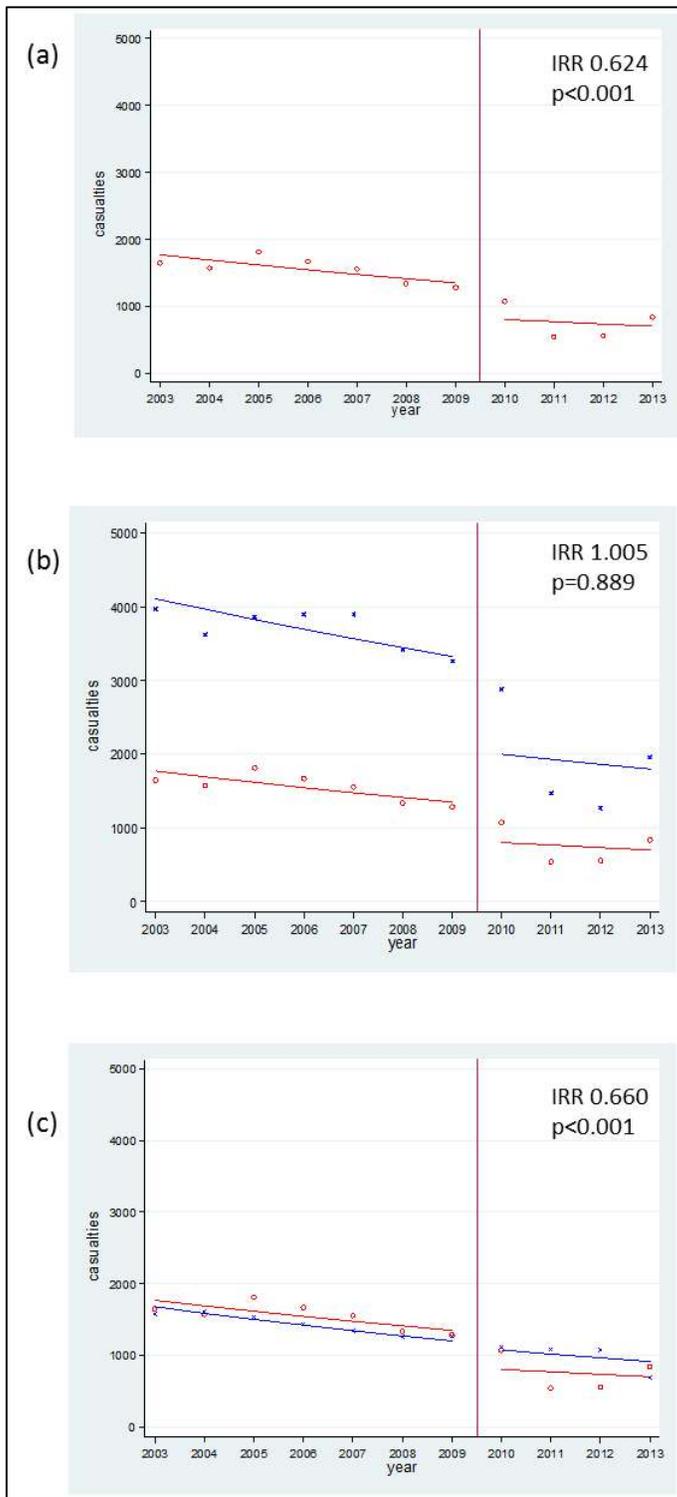
Figure 3: The effect of the Birmingham and Black Country street lighting intervention on road traffic crash casualties

Red regression line is the intervention series (night time road traffic casualties on minor roads in Birmingham and the Black Country); blue regression line is the control series: (a) no control, (b) control time period: day time road traffic crash casualties

on minor roads in Hertfordshire (c) location based control: night time road traffic crash casualties on major roads in West Yorkshire. The vertical red line is the intervention point. The incident rate ratio (IRR) is the step change in road traffic crash casualties following the intervention compared to before the intervention, in figures (b) and (c) the IRR is the step change in the intervention series over and above any step change in the control series.

CONCLUSION

In this paper we have highlighted how ITS studies differ from other evaluation designs by making within group rather than between group comparisons. While this has the advantage of limiting confounding by factors that change only slowly through time history bias can still threaten the validity of ITS studies. A wide range of different controls can be used in order to limit history bias and improve the validity of an ITS study. Nevertheless, it is important to systematically consider a priori the degree of risk of history bias associated with any particular study, what control series are available and whether these will adequately control for history bias. Finally, researchers should take care in interpreting the results of CITS studies, in particular when the results of CITS analysis differ from those of simple (uncontrolled) ITS analysis. If the results of the CITS and the ITS analysis are aligned, CITS studies can provide strong evidence on the effectiveness of public health interventions and when appropriate controls are selected the design ranks second only to randomised controlled designs in terms of their capacity to control for bias.(13)

FUNDING

REFERENCES

1.   Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. JECH. 2009;65(7):582-7.
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.   Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research: Houghton Mifflin Boston; 1963.
5.   Fretheim A, Soumerai SB, Zhang F, Oxman AD, Ross-Degnan D. Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. J Clin Epidemiol. 2013;66(8):883-7.
6.   Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, et al. A re-analysis of cluster randomized trials showed Interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol. 2014.
7.   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.
8.   St.Clair T, Hallberg K, Cook TD. The Validity and Precision of the Comparative Interrupted Time-Series Design: Three Within-Study Comparisons. Journal of Educational and Behavioral Statistics. 2016.
9.   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.
10.   Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, et al. Natural experiments: an underused tool for public health? Public Health. 2005;119(9):751-7.
11.   Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996;312(7040):1215.
12.   Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. AJPM. 2007;33(2):155-61.
13.   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.
14.   Fu AZ, Dow WH, Liu GG. Propensity score and difference-in-difference methods: a study of second-generation antidepressant use in patients with bipolar disorder. Health Services and Outcomes Research Methodology. 2007;7(1):23-38.
15.   Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci. 2000;1(1):31-49.
16.   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.
17.   Cousens S, Hargreaves J, Bonell C, Armstrong B, Thomas J, Kirkwood BR, et al. Alternatives to randomisation in the evaluation of public-health interventions: Statistical analysis and causal inference. J Epidemiol Community Health. 2009;65(7):576-81.
18.   Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research. 2011;46(3):399-424.
19.   Abadie A, Diamond A, Hainmueller J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association. 2010;105(490):493-505.

20. Winkelmayer WC, Kurth T. Propensity scores: help or hype? Nephrology Dialysis Transplantation. 2004;19(7):1671-3.
21. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? BMJ. 2001;322(7282):355-7.
22. Ross-Degnan D, Soumerai SB, Fortess EE, Gurwitz JH. Examining product risk in context: market withdrawal of zomepirac as a case study. JAMA. 1993;270(16):1937-42.
23. Klauber MR, Barrett-Connor E, Marshall LF, Bowers SA. The epidemiology of head injury: a prospective study of an entire community - San Diego County, California, 1978. Am J Epidemiol. 1981;113(5):500-9.
24. Dennis J, Ramsay T, Turgeon AF, Zarychanski R. Helmet legislation and admissions to hospital for cycling related head injuries in Canadian provinces and territories: interrupted time series analysis. BMJ: British Medical Journal. 2013;346:f2674.
25. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. J Eval Clin Pract. 2011;17(6):1231-8.
26. 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.
27. Feigl AB, Salomon JA, Danaei G, Ding EL, Calvo E. Teenage smoking behaviour following a high-school smoking ban in Chile: interrupted time-series analysis. Bull World Health Organ. 2015;93(7):468-75.
28. Kontopantelis E, Olier I, Planner C, Reeves D, Ashcroft DM, Gask L, et al. Primary care consultation rates among people with and without severe mental illness: a UK cohort study using the Clinical Practice Research Datalink. BMJ Open. 2015;5(12).
29. Kisely SR, Pais J, White A, Connor J, Quek L-H, Crilly JL, et al. Effect of the increase in "alcopops" tax on alcohol-related harms in young people: a controlled interrupted time series. Med J Aust. 2011;195(11):690-3.
30. Schneeweiss S, Maclure M, Carleton B, Glynn RJ, Avorn J. Clinical and economic consequences of a reimbursement restriction of nebulised respiratory therapy in adults: direct comparison of randomised and observational evaluations. BMJ. 2004;328(7439):560.
31. 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.
32. 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.
33. Ross HL, Campbell DT, Glass GV. Determining the Social Effects of a Legal Reform The British" Breathalyser" Crackdown of 1967. m Behav Sci. 1970;13(4):493-509.
34. 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.
35. Linden A. Conducting interrupted time series analysis for single and multiple group comparisons. The Stata Journal. 2015;15(2):480-500.
36. Perkins C, Steinbach R, Tompson L, Green J, Johnson S, Grundy C, et al. What is the effect of reduced street lighting on crime and road traffic injuries at night? A mixed-methods study. Southampton (UK): NIHR Journals Library; 2015.

6.2 Summary of appendices for methodological paper 3

Appendices for this paper are presented in Chapter 11:

11.3.1   Web appendix 1: Segmented regression interaction model for a controlled interrupted time series

11.3.2 Contribution of the candidate to the paper

6.3 Contribution of the paper

Motivation for the paper

The topic of CITS was chosen as history bias has been clearly identified as the major threat to the validity of interrupted time series, however, there is a lack of methodological literature on approaches to addressing this limitation. Other approaches exist to address history bias, including multiple phase designs and multiple baseline designs.(1-3) Nevertheless, these approaches require relatively unique circumstances in order to be adopted. Conversely, CITS is more widely applicable and many ITS studies could be strengthened by the inclusion of a control series.

Contribution to the thesis

This paper directly addresses the third objective of the thesis: "To develop a methodological framework for the selection of controls and analysis of controlled interrupted time series to limit the risk of history bias." The paper builds upon the brief introduction of CITS in Chapter 4 and informs the selection of controls in the two case studies in Chapters 7 and 8, in which two different types of controls are chosen. It will also contribute to the reporting guidelines developed in Chapter 9.

Outputs and contribution to the literature

To demonstrate the wide applicability of CITS, the paper, for the first time, identifies and classifies a broad range of controls that can be used and describes the different sources of history bias that these can address. The framework developed provides guidance on how to select controls and analyse CITS studies, but also suggests when controls may be inappropriate and when other designs such as uncontrolled ITS may need to be used. It is always important to consider whether a control will improve the validity of the study or if validity could be impaired by the introduction of new sources of bias. The decision on whether to include a control in an ITS study should be based both on an assessment of the risk of confounding events and the availability of appropriate controls. A well-chosen control strengthens the validity of ITS study when there are concerns that the pre-intervention trend in the intervention population may not be able to accurately predict the counterfactual. This is normally because there is the possibility that confounding events or co-interventions could lead to a deviation in the trend independently of the intervention. It follows that a control series is most helpful when concurrent confounding events are more likely, for example: if there are known co-interventions that need to be excluded, if the intervention is introduced over a prolonged time period (allowing more time during which other confounding events could occur) or if the effect of the intervention on the outcome is lagged (again allowing more time during which other confounding events could have occurred).

There is a lower risk of history bias if there is a long, stable pre-intervention trend, there are no known co-interventions and the intervention is introduced over a short time period and has a rapid effect on the outcome. Under such circumstances there is a limited time during which other events could act and an uncontrolled ITS analysis remains a strong design. If researchers find that no suitable control is available after following the steps that are outlined in the framework. The use of an uncontrolled ITS design may be more appropriate (as indicated in Figure 2 of the paper). A control series can be detrimental to the validity of the study if it introduces selection bias and differs from the intervention series in the way it would respond to interventions or if it has been independently exposed to other interventions or events. Furthermore, using an uncontrolled ITS and stating that a key limitation is that it cannot control for simultaneous events or co-interventions that could impact on the outcome, may be preferable to (and more transparent than) using an inappropriate control that, at best, will not improve the validity of the analysis.

The LANTERNS study provides an excellent example to demonstrate the selection of controls as there are various different types of controls that can be chosen for which data is available.(4) This provides a useful illustration to those considering using a CITS study on how to apply the proposed framework in order to select a control.

Conclusions

Where an appropriate control is available this can facilitate a very powerful quasi-experimental design. This paper encourages the use of CITS where such controls are available and should aid researchers in using this technique to strengthen evaluations of complex public health interventions.

References

1. 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.
2. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci. 2000;1(1):31-49.
3. Reding GR, Raphelson M. Around-the-clock mobile psychiatric crisis intervention: Another effective alternative to psychiatric hospitalization. Community Ment Health J. 1995;31(2):179-87.
4. LANTERNS project team. Local Authority Collaborators' National Evaluation of Reduced Night-time Streetlight (LANTERNS) 2014 [Available from: http://lanterns.lshtm.ac.uk/.

7. Case study 1: The effect of the financial crisis on suicides

7.1 Research paper: The effect of the late 2000's financial crisis on suicides in Spain: an interrupted time series analysis

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	James Lopez Bernal
Principal Supervisor	Antonio Gasparrini and Steven Cummins
Thesis Title	The use interrupted time series for the evaluation of public health interventions

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	European Journal of Public Health
When was the work published?	2013
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	This paper was a follow on from research undertaken during my MSc public health project. It was a core part of the inspiration for my PhD proposal and is included to highlight some of the methodological concepts, including the use of a control series and the use of ITS to evaluate unplanned events

Have you retained the copyright for the work?*	No	Was the work subject to academic peer review?	Yes

If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the concept of this paper, all analysis and writing the paper.

Student Signature: _____ **Date:** _____17/01/2018_____

Supervisor Signature: _____█████████████_____   **Date:** __125/01/2018___

Supervisor Signature: _____████████_____   **Date:** __125/01/2018___

Copyright Clearance Center

Confirmation Number: 11693450
Order Date: 01/17/2018

Customer Information

Customer: James Lopez Bernal
Account Number: 3001238936
Organization: James Lopez Bernal
Email: jamie.lopezbernal@gmail.com
Phone: +44 7872834737
Payment Method: Invoice

<div align="center">

This is not an invoice

</div>

Order Details

European journal of public health

Order detail ID:	70937946
ISSN:	1101-1262
Publication Type:	Journal
Volume:	
Issue:	
Start page:	
Publisher:	OXFORD UNIVERSITY PRESS
Author/Editor:	EUROPEAN PUBLIC HEALTH ASSOCIATION

Permission Status: ✅ Granted

Permission type: Republish or display content
Type of use: Thesis/Dissertation

Order License Id: 4271450861032

Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
Number of pages in chapter/article	5
The requesting person/organization	James Lopez Bernal
Title or numeric reference of the portion(s)	European Journal of Public Health, Volume 23, Issue 5, 1 October 2013, Pages 732–736, https://doi.org/10.1093/eurpub/ckt083
Title of the article or chapter the portion is from	The effect of the late 2000s financial crisis on suicides in Spain: an interrupted time-series analysis
Editor of portion(s)	Peter Allebeck
Author of portion(s)	James Lopez Bernal
Volume of serial or monograph	23
Page range of portion	732–736
Publication date of portion	25 June 2013
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	U.K. and Commonwealth (excluding Canada)
In the following language(s)	Original language of publication
With incidental promotional use	no

Lifetime unit quantity of new product	Up to 499
Title	The Use of Interrupted Time Series for the Evaluation of Public Health Interventions
Instructor name	Antonio Gasparrini and Steven Cummins
Institution name	London School of Hygiene and Tropical Medicine
Expected presentation date	Jan 2018

Note: This item was invoiced separately through our **RightsLink service.** More info $ 0.00

Total order items: 1	**Order Total: $0.00**

About Us | Privacy Policy | Terms & Conditions | Pay an Invoice

....

The effect of the late 2000s financial crisis on suicides in Spain: an interrupted time-series analysis

James A. Lopez Bernal[1], Antonio Gasparrini[2], Carlos M. Artundo[3], Martin McKee[1]

1 Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London WC1H 9SH, UK
2 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London WC1H 9SH, UK
3 Andalusian School of Public Health (ASPH), Granada, Spain

Correspondence: James A. Lopez Bernal, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK, Tel: +44 20 7927 2106, Fax: +44 20 7927 2701, e-mail: james.lopez-bernal@lshtm.ac.uk

Background: The current financial crisis is having a major impact on European economies, especially that of Spain. Past evidence suggests that adverse macro-economic conditions exacerbate mental illness, but evidence from the current crisis is limited. This study analyses the association between the financial crisis and suicide rates in Spain. **Methods:** An interrupted time-series analysis of national suicides data between 2005 and 2010 was used to establish whether there has been any deviation in the underlying trend in suicide rates associated with the financial crisis. Segmented regression with a seasonally adjusted quasi-Poisson model was used for the analysis. Stratified analyses were performed to establish whether the effect of the crisis on suicides varied by region, sex and age group. **Results:** The mean monthly suicide rate in Spain during the study period was 0.61 per 100 000 with an underlying trend of a 0.3% decrease per month. We found an 8.0% increase in the suicide rate above this underlying trend since the financial crisis (95% CI: 1.009–1.156; *P* = 0.03); this was robust to sensitivity analysis. A control analysis showed no change in deaths from accidental falls associated with the crisis. Stratified analyses suggested that the association between the crisis and suicide rates is greatest in the Mediterranean and Northern areas, in males and amongst those of working age. **Conclusions:** The financial crisis in Spain has been associated with a relative increase in suicides. Males and those of working age may be at particular risk of suicide associated with the crisis and may benefit from targeted interventions.

....

Introduction

In his seminal 1897 work, *Le Suicide*, Durkheim proposed that macroeconomic changes may increase suicides.[1] Since then associations have been found between economic instability and both mental ill health and suicide,[2–8] with past economic crises, such as the collapse of the Soviet Union and the Southeast Asian economic crisis in the late 1990s, associated with increases in suicides.[9,10] The current global financial crisis is widely regarded as the worst since the Great Depression of the 1930s and has had a severe deleterious effect on the Spanish economy, culminating in a request for Eurozone support for its banking sector.[11] Following a decade of expansion, economic growth began to slow in 2007 and gross domestic product (GDP) began to contract from the second quarter of 2008, ushering in a recession lasting seven successive quarters (Supplementary Appendix 1).[12] Between 2007 and 2012, unemployment trebled from 8% to 24% (Supplementary Appendix 2), reaching the highest rate in the European Union, with the greatest increases seen in young males (age 20–34): from 11% to 50%.[13] Since the onset of the financial crisis, there have been major cuts to social spending. Savings have included redundancies and salary reductions for health care personnel, changes to drug-prescribing policies, closure of facilities, reductions in opening times and delays in payments to suppliers.[14–16]

Evidence on the influence of the current financial crisis on mental health, either in Spain or elsewhere, is limited, largely because delays in data availability mean that the full impact cannot yet be assessed. The lack of any sign that the crisis is ending makes it important that its effects be understood so that mitigating interventions can be implemented. One previous study has formally examined the effect of the financial crisis on suicides, finding that observed suicide counts in England were above those that would have been expected based on underlying trends, particularly amongst men.[17] There have also been reported increases in suicides in Greece, Italy and the European Union in general and possible increases in mental illness in Spain and Greece, although underlying trends were not accounted for.[18–24] A further study used joinpoint regression as part of their analysis of several mental health indicators in South Australia but found no change in trends of any psychological distress measures associated with the crisis.[25]

In this study, we investigate the relationship between the current financial crisis and suicides in Spain using an interrupted time-series analysis. We also evaluate how this relationship varies by geographical area, sex and age.

Methods

An interrupted time-series analysis was used to compare suicide rates before the financial crisis with those subsequently. A financial crisis may impact on health through a variety of mechanisms, including job loss (or anticipation thereof), reduced working hours and debt. Consequently, it is appropriate to look at the financial crisis as a discrete, if multifaceted, event rather than using one or more intermediary variables. Suicide data were obtained from the 'Instituto Nacional de Estadística' (INE), Spain's national statistics institute.[13,17] Suicides are deaths coded as X60-X84 (ICD-10).[26] In Spain, suicides are determined following judicial review of any deaths that may have a possible accidental or violent cause. Monthly suicide data were used to maximize the data points available since the financial crisis.[27] Full monthly suicide data, disaggregated by region and age-group, were only available from 2005; in addition, the data are published with a 2-year delay, so only data for 72 months were available (January 2005 to December 2010). Population denominators were obtained from the INE official population figures.

111

The 'intervention' of interest in this study was the financial crisis. For unplanned events, the timing of the intervention must be established from data that are independent of the time-series data being analysed.[28] We based the timing on observed changes in GDP, the measure most commonly used in defining a recession.[10,29] In Spain, GDP began to contract from the second quarter of 2008 (Supplementary Appendix 1),[12] so the period up to and including March 2008 was considered pre-financial crisis and April 2008 onwards, post financial crisis.

Statistical analysis

Segmented regression was used to estimate the effect of the financial crisis on suicides.[27] A Poisson distribution of monthly suicide counts was assumed offset by population data to model rates. Adjustments were made for the length of the month and for any seasonal effect, the latter by using a harmonic term based on the month of the year that included two sine/cosine pairs.[30] Initial analyses suggested a moderate degree of overdispersion (dispersion parameter = 1.46), so a more flexible quasi-Poisson model was used for all analyses.[31] Residual autocorrelation was also tested for using the Durbin–Watson test. Further models were tested in a sensitivity analysis, including a model allowing for both a step change and a change in the trend; a two-step model whereby a recession period was modelled as the duration that GDP contracted—second quarter of 2008 until the last quarter of 2009, then returning to growth until the end of the dataset (December 2010) (Supplementary Appendix 1); and a model with the crisis period starting from July 2007, the point where unemployment began to rise (Supplementary Appendix 2), rather than April 2008.

A control analysis was also undertaken using mortality from accidental falls as the outcome to distinguish any association with the financial crisis from other concurrent events.[27] This outcome was chosen as being unlikely to be affected by the financial crisis but has similarities to suicides in that both require judicial review and have short lag times compared with other causes of death such as chronic diseases.

Further stratified analysis was conducted to investigate whether changes in suicides varied by region, sex and age group. The individual autonomous regions of Spain could not be examined separately, as the number of suicides per month was too low (<30 for most regions), which would have led to too much random variability. The regions were therefore grouped into three areas based on geographical location and economic similarities: Northern Spain (comprising Galicia, Asturias, Cantabria, the Basque Country, Navarra, La Rioja and Aragon), central Spain (comprising Castilla and Leon, Castilla La Mancha, Extremadura and Madrid), the Mediterranean area and Canary Islands (comprising Catalonia, Valencia, Murcia, Andalucia, the Balearic Islands and the Canary Islands). Broadly speaking, regions in the Northern area have large manufacturing sectors; the Central area have a relatively large agricultural sector, although Madrid, which relies more on its financial and service sectors, is an anomaly here; and the Mediterranean area has a large service sector, in particular tourism, and a relatively large real estate sector. All regions also have large construction and public service sectors.[13] Age was grouped into younger economically active ages (age 15–39), older economically active ages (age 40–64) and post state retirement age (age 65 plus).

All analyses were conducted using the statistical packages R 2.15.0 and STATA version 11.

Results

Table 1 shows summary data of the average monthly suicide counts and rates over the 72 months of the time series for the different groups studied. The highest monthly suicide rates for the categories of area sex and age group were in Northern Spain, males and those aged ≥65 years, respectively.

Table 1 Mean monthly suicide data for the duration of the time series

Population	Monthly suicide count	Population (millions)	Monthly suicide rate (per 100 000)
All Spain	277 (31)	45.7 (1.1)	0.61 (0.07)
Area			
Northern Spain	66 (10)	8.8 (0.1)	0.75 (0.11)
Central Spain	51 (10)	11.8 (0.3)	0.43 (0.08)
Mediterranean	159 (19)	24.8 (0.7)	0.64 (0.08)
Sex			
Males	213 (26)	22.6 (0.5)	0.95 (0.12)
Females	64 (9)	23.1 (0.5)	0.28 (0.04)
Age group			
15–39	75 (13)	16.8 (0.1)	0.44 (0.08)
40–64	111 (15)	14.6 (0.6)	0.76 (0.09)
65+	91 (16)	7.6 (0.2)	1.20 (0.21)

Mean (SD). Note: Total for all Spain is not equal to the sum of the area totals, as Ceuta and Melilla (two autonomous Spanish cities in North Africa) are not included in the latter.
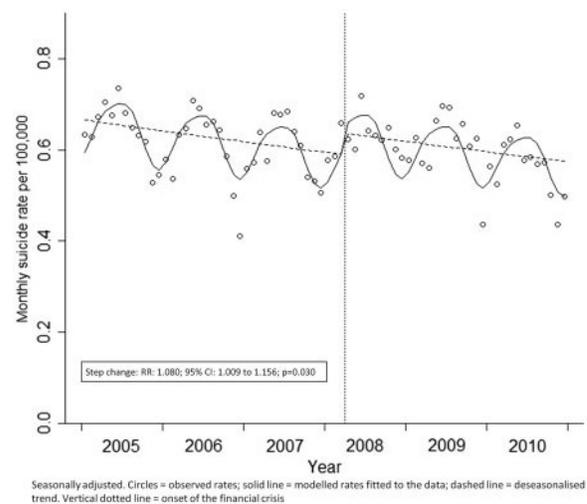
Figure 1 Trend in monthly suicide rates for all of Spain before and since the financial crisis

The principal model for all of Spain (figure 1) suggested that, over the period studied, the underlying trend was of a 0.3% decrease in the suicide rate per month (95% CI: 0.995–0.998; $P < 0.001$). There was an 8.0% step increase in the suicide rate associated with the financial crisis (95% CI: 1.009–1.156; $P = 0.030$). The Durbin–Watson statistic showed no evidence of autocorrelation (DW 2.10, $P = 0.421$). The control analysis showed no evidence of a change in mortality from accidental falls associated with the financial crisis (step change: RR 1.031; 95% CI: 0.939–1.132; $P = 0.525$) (Supplementary Appendix 3).

Results of the stratified analyses are presented in table 2 with the plots in Web Supplementary Appendices 4–6. All results are concordant in suggesting an increase in suicide rates. Although the stratified analyses suggest a greater increase in the Mediterranean and Northern areas, in males and in younger age groups, the low statistical power when testing for interaction and the associated P-values (test for interaction: area $P = 0.868$, sex $P = 0.263$, age $P = 0.923$) prevent any firm conclusions being made on a differential effect by sub-groups.

All models tested in the sensitivity analysis also showed a relative increase in the suicide rate during the financial crisis period (Supplementary Appendix 7). Furthermore, removing three outlying observations (December 2006, December 2009 and November 2010) had little impact on the results.

Table 2 Suicide rates during the financial crisis period compared with before the financial crisis

Population	Step change (RR) 95% CI	P-value
All Spain	1.080 (1.009–1.156)	0.030
Area		
Northern Spain	1.090 (0.967–1.226)	0.160
Central Spain	1.043 (0.911–1.195)	0.538
Mediterranean	1.086 (1.005–1.172)	0.037
Sex		
Males	1.100 (1.026–1.179)	0.007
Females	1.013 (0.894–1.149)	0.834
Age group		
15–39	1.104 (0.980–1.245)	0.110
40–64	1.082 (0.980–1.195)	0.119
65+	1.067 (0.957–1.191)	0.243

RR: rate ratio; 95% CI: 95% confidence interval.
P-values are two-sided. All results based on a step change model, adjusted for seasonality, with the pre-crisis period from January 2005 to March 2008 and the financial crisis period from April 2008 to December 2010.

Discussion

Principal findings

These results suggest that the onset of the financial crisis has been associated with a relative increase in the suicide rate in Spain above the expected underlying trend. If this association was causal, the financial crisis may account for around 21 suicides per month in Spain or around 680 suicides since the crisis so far (up to the end of 2010). Although we were underpowered to test for interaction, the relative increase in the suicide rate appears to have been greatest in the Mediterranean and Northern areas of Spain, amongst males and amongst those of working age.

Comparison with previous studies

The underlying downward trend in suicides seen during the study period is consistent with the long-term trend in most European countries. Numerous explanations have been offered for this downward trend including improved mental health provision, greater use of antidepressants and policies such as changes to paracetamol pack sizes and withdrawal from certain drugs from over the counter sales.[32] Only one previous study has examined the effect of the current financial crisis on suicides while taking into account such underlying trends; this found around 1000 more suicides in England during the recession than would be expected based on historical trends. Similar to our results this increase was found to be greater amongst males.[17] The reported increases in suicides in Greece, Italy and the European Union as a whole also provide some support to our results, although these reports did not account for underlying trends in suicides.[19,21,22] Again, consistent with our results for suicides, the one previous study of the mental health effects of the financial crisis in Spain found increases in primary care attendance for various mental illnesses in 2010–11 compared with 2006–07, including mood, anxiety, somatoform and alcohol-related disorders.[23] Only one study so far, in South Australia, has shown no evidence of an increase in mental health indicators related to suicide since the financial crisis, in this case depression, suicidal ideation and other psychological distress measures.[25] This may perhaps be explained by a low proportion of unemployed people in their sample compared with the population as a whole but also the limited effect that the global financial crisis has so far had on Australia compared with Spain and other countries in Europe.[25] The increase in suicide rates is also supported by evidence of increases during past economic crises in Asia and Russia.[9,10] In addition, other studies find that associations between adverse macro-economic conditions and suicide are strongest amongst males and younger age groups, consistent with our findings.[3–6] No other studies were found that have investigated the differential effect of the current financial crisis on suicide or other mental health indicators between age groups in Spain or abroad; nor were there any studies investigating regional differences in the effect within Spain; therefore, our findings cannot directly be compared with other results.

Strengths and limitations

Our study has a number of strengths. Interrupted time-series is regarded as a powerful quasi-experimental design for assessing the longitudinal impact of an intervention, as it enables both random month on month fluctuations and the underlying trend to be accounted for in the analysis.[27] In addition, it does not suffer from some of the biases and confounders of other observational studies. Other variables associated with suicides that may have changed over the same time period, including seasonal fluctuations, were adjusted for during the analysis. Any residual confounders would have to be events that occurred at the same time as the financial crisis but were not a manifestation of the crisis. While the lack of any association between the financial crisis and deaths from accidental falls cannot exclude all such events, it does help to exclude many, such as unrecognized changes to systems of death registration or classification, or any other events that would have impacted on both causes of death. The results of this study were also robust to sensitivity analysis for the timing of the financial crisis period, with all models that were tested showing a relative increase in suicides above the underlying trend during the financial crisis period. In addition, there was no evidence of autocorrelation in our model. Finally, whilst the use of suicides as an outcome measure may provide a very incomplete picture of the impact of the financial crisis on mental health, it is an objective measure that is less subject to responder bias, observer bias and validity issues compared with other outcomes such as self-reported mental health indicators or suicidal ideation.

There are also potential limitations to this study. Firstly, the number of suicides per month in the stratified analysis was relatively low, leading to relatively wide confidence intervals and requiring the amalgamation of autonomous regions into three larger areas. Secondly, there is no established definition for the timing of the financial crisis in Spain. As recommended for imperfectly identifiable events, an independent indicator was used to determine the onset of the crisis, in this case GDP.[28] Nonetheless, it is clear that some manifestations of the financial crisis in Spain began before this point including the increase in unemployment rates. However, various other definitions of the financial crisis period were modelled in the sensitivity analysis, and all of these also showed evidence of an increase in the suicide rate above the underlying trend during the crisis, so it is unlikely that any incorrect timing of the 'intervention' impacted on the results. Thirdly, no lag period was included in the model, consistent with evidence from elsewhere that there is no significant lag effect.[17,22] However, if a lag did exist, the modelled effect would have been attenuated so the effect of the financial crisis on the suicide would, if anything, have been underestimated. Finally, studies have suggested that deaths from events with undetermined intent (ICD-10: Y10-Y34[26]) should also be included when analysing suicides to avoid underestimation.[33] However, these were not available in a form disaggregated by the variables required (including region and age group). There were only between 1 and 19 deaths per month nationally from events with undetermined intent during the period of study, so even if all had been a suicide, it would have had little impact on the results.

Interpretation and implications

Our study alone cannot establish whether the association found between the financial crisis and suicides is causal; however, this explanation is supported by the relatively large magnitude of effect,
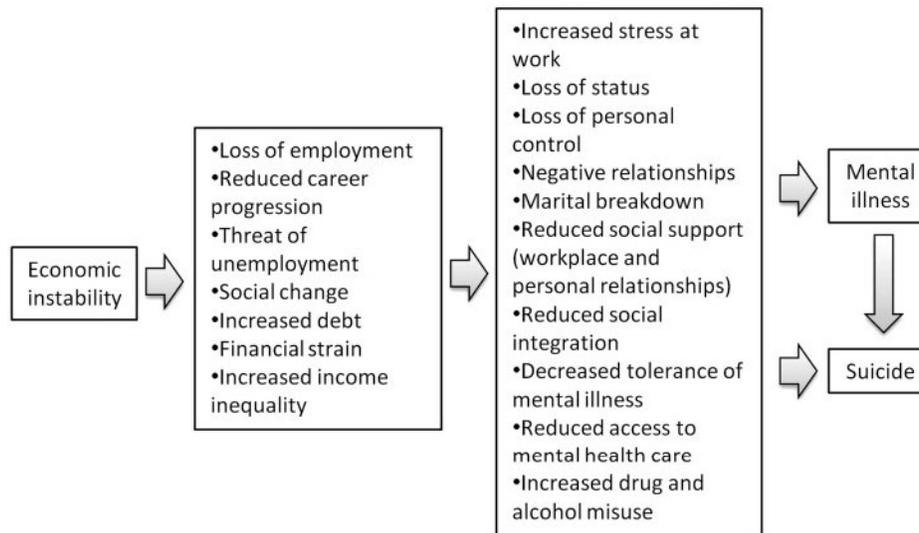
Figure 2 Proposed mechanisms for the relationship between economic instability and mental illness and suicides

consistency with previous studies, coherence with existing theory dating back to Durkheim[1] and the existence of numerous plausible mechanisms for the relationship between financial crises and suicides (figure 2).[34–36] Perhaps the most researched mechanism is through the effects of the financial crisis on unemployment and its association with suicide. Spain has experienced a dramatic rise in unemployment during the financial crisis (Supplementary Appendix 2); numerous studies have found strong associations between increasing unemployment and suicides in the past.[3,5,37] Initial evidence from the current financial crisis in other European countries appears to suggest the same pattern is occurring.[17] Furthermore, although formal statistical analysis would be required, the groups in the stratified analysis that showed evidence of increases in the suicide rate above the underlying trend appear to coincide relatively well with those that have experienced the biggest absolute increases in unemployment rate between the start and the end of the study period, including the Mediterranean region of Spain, males and those of younger working age groups (Supplementary Appendix 8).

This study is the first to show the association between the financial crisis and suicides in Spain; it is also perhaps the strongest evidence to date of the detrimental impact that the current financial crisis may be having on mental health. The association found with suicides is likely to represent only the tip of the iceberg of a possible effect of the financial crisis on wider mental health. In addition, by 2010 the biggest effects of the financial crisis in Spain had yet to manifest themselves, as many of the major social spending cuts had only just been introduced.[14] It may therefore be that, if the association found in this study is causal, a further increase in suicide rates will already have occurred and will continue unless mitigating interventions are introduced promptly. Potential interventions include active labour market programmes, family support programmes and debt relief programmes.[36] The results of the stratified analysis undertaken in this study may help to identify which groups should be targeted with such interventions.

Future research

Although this study has provided important insight into one of the harmful impacts the financial crisis may be having on health, there remains a need for further research. Firstly, a number of potentially at risk groups have been identified during the stratified analysis and possible reasons for the increased risk in these groups have been hypothesized, but we were underpowered to test whether these were true differences or due to chance. Further investigation is needed to try to establish whether these subgroups are truly at greatest risk and if so why this is the case. Secondly, given the global nature of the financial crisis, further international studies are required to establish whether similar effects are being seen with suicide rates elsewhere to help identify whether certain policies implemented by some countries may have a protective effect. Thirdly, further research is needed to identify the effectiveness of mitigating interventions and how they are best implemented.

Supplementary data

Supplementary data are available at *EURPUB* online.

Ethical approval

Ethical approval was obtained from the London School of Hygiene and Tropical Medicine research ethics committee.

Acknowledgements

We would like to thank Isabel Ruiz, Mª José Sánchez and Esther Molina from the Andalusian School of Public Health for their comments and suggestions.

Funding

Conflicts of interest: None declared.

Key points

- In Spain, the financial crisis has been associated with a substantial increase in suicide rates over and above the underlying trend.
- Public health interventions focussing on mitigating the impact of the financial crisis on mental health and suicide should be established.
- The effect of the financial crisis on suicides appears to be greatest amongst men and amongst those of working age; these groups may benefit most from targeted interventions.

References

1 Durkheim E. *Le Suicide*. Felix Alcan: Paris, 1897.

2 Brenner MH. *Mental Illness and the Economy*. Oxford: Harvard U. Press, 1973.

3 Tapia Granados J. Recessions and Mortality in Spain, 1980–1997. *Eur J Popul* 2005;21:393–422.

4 Berk M, Dodd S, Henry M. The effect of macroeconomic variables on suicide. *Psychol Med* 2006;36:181–9.

5 Platt S, Micciolo R, Tansella M. Suicide and unemployment in Italy: description, analysis and interpretation of recent trends. *Soc Sci Med* 1992;34:1191–201.

6 Gerdtham UG, Johannesson M. Business cycles and mortality: results from Swedish microdata. *Soc Sci Med* 2005;60:205–18.

7 Stuckler D, Basu S, Suhrcke M, et al. The public health effect of economic crises and alternative policy responses in Europe: an empirical analysis. *Lancet* 2009;374:315–23.

8 Ruhm CJ. Are recessions good for your health? *Q J Econ* 2000;115:617–50.

9 Pridemore WA, Chamlin MB, Cochran JK. An interrupted time-series analysis of Durkheim's social deregulation thesis: the case of the Russian federation. *Justice Q* 2007;24:271–90.

10 Chang SS, Gunnell D, Sterne JA, et al. Was the economic crisis 1997-1998 responsible for rising suicide rates in East/Southeast Asia? A time-trend analysis for Japan, Hong Kong, South Korea, Taiwan, Singapore and Thailand. *Soc Sci Med* 2009;68:1322–31.

11 Temin P. The great recession and the great depression. *National Bureau of Economic Research, Inc.* Cambridge, Massachusetts, USA: NBER Working Papers: 15645, 2010.

12 Organisation for Economic Cooperation and Development. OECD.StatExtracts, 2012 [cited 2012 04/07/2012]. Available at: http://stats.oecd.org/Index.aspx?QueryName=350.

13 INE. Instituto Nacional de Estadística [National Statistics Institute]. [cited 2012 04/07/2012]. Available at: http://www.ine.es/.

14 Gene-Badia J, Gallo P, Hernandez-Quevedo C, Garcia-Armesto S. Spanish health care cuts: penny wise and pound foolish? *Health Policy* 2012;106:23–8.

15 Houston M, Day M, de Lago M, Zarocostas J. Health services across Europe face cuts as debt crisis begins to bite. *BMJ* 2011;343:d5266.

16 Vogler S, Zimmermann N, Leopold C, de Joncheere K. Pharmaceutical policies in European countries in response to the global financial crisis. *South Med Rev* 2011;4:22–32.

17 Barr B, Taylor-Robinson D, Scott-Samuel A, et al. Suicides associated with the 2008-10 economic recession in England: time trend analysis. *BMJ* 2012;345:e5142.

18 Economou M, Madianos M, Theleritis C, et al. Increased suicidality amid economic crisis in Greece. *Lancet* 2011;378:1459–60.

19 Mattei G, Ferrari S, Rigatelli M. Economic recession in Italy: a review of short-term effects on healthJournal of Psychosomatic Research, Conference: 14th Annual Meeting of the European Association for Consultation Liaison Psychiatry and Psychosomatics, EACLPP 2011 Budapest Hungary. Conference Start: 20110630 Conference End: 20110702. Conference Publication: (var.pagings). *J Psychosom Res* 2011;70:606.

20 Sargent-Cox K, Butterworth P, Anstey KJ. The global financial crisis and psychological health in a sample of Australian older adults: a longitudinal study. *Soc Sci Med* 2011;73:1105–12.

21 Kentikelenis A, Karanikolos M, Papanicolas I, et al. Health effects of financial crisis: omens of a Greek tragedy. *Lancet* 2011;378:1457–8.

22 Stuckler D, Basu S, Suhrcke M, et al. Effects of the 2008 recession on health: a first look at European data. *Lancet* 2011;378:124–5.

23 Gili M, Roca M, Basu S, et al. The mental health risks of economic crisis in Spain: evidence from primary care centres, 2006 and 2010. *Eur J Public Health* 2013;23:103–8.

24 Economou M, Madianos M, Peppou LE, et al. Major depression in the Era of economic crisis: a replication of a cross-sectional study across Greece. *J Affect Disord* 2013;145:308–14.

25 Shi Z, Taylor AW, Goldney R, et al. The use of a surveillance system to measure changes in mental health in Australian adults during the global financial crisis. *Int J Public Health* 2011;56:367–72.

26 WHO. *International Statistical Classification of Diseases and Health Related Problems*. Geneva: World Health Organisation, 2004.

27 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;27:299–309.

28 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 Med Res Methodol* 2006;6:16.

29 Gurfinkel EP, Bozovich GE, Dabbous O, et al. Socio economic crisis and mortality. Epidemiological testimony of the financial collapse of Argentina. *Thromb J* 2005;3:22.

30 Hunsberger S, Albert PS, Follmann DA, Suh E. Parametric and semiparametric approaches to testing for seasonal trend in serial count data. *Biostatistics* 2002;3:289–98.

31 Hoef JMV, Boveng PL. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* 2007;88:2766–72.

32 Mackenbach J, McKee M. *Successes and Failures of Health Policy in Europe Over Four Decades: Diverging Trends, Converging Challenges*. Buckingham: Open University Press, 2012.

33 Ohberg A, Lonnqvist J. Suicides hidden among undetermined deaths. *Acta Psychiatrica Scandinavica* 1998;98:214–8.

34 Zivin K, Paczkowski M, Galea S. Economic downturns and population mental health: research findings, gaps, challenges and priorities. *Psychol Med* 2011;41:1343–8.

35 Lester BY. Learnings from Durkheim and beyond: the economy and suicide. *Suicide Life Threat Behav* 2001;31:15–31.

36 World Health Organization Regional Office for Europe. *Impact of Economic Crises on Mental Health*. Copenhagen, Denmark: WHO Regional Office For Europe, 2011.

37 Yip PSF, Caine ED. Employment status and suicide: the complex relationships between changing unemployment rates and death rates. *J Epidemiol Commun Health* 2011;65:733–6.

7.2 Summary of appendices for case study 1

Appendices for this paper are presented in Chapter 11:

11.4.1 Gross domestic product growth in Spain

11.4.2 Unemployment in Spain

11.4.3 Trend in monthly mortality rate from accidental falls (control) for all of Spain before and since the financial crisis

11.4.4 Trend in monthly suicide rates before and since the financial crisis by area

11.4.5 Trend in monthly suicide rates before and since the financial crisis by sex

11.4.6 Trend in monthly suicide rates before and since the financial crisis by age group

11.4.7 Time series plots of trends in monthly suicide rates for all of Spain based on the alternative models used in sensitivity analysis

11.4.8 Unemployment rates during the first quarter of 2005 and the last quarter of 2010

11.4.9 Contribution of the candidate to the paper

7.4 Contribution of the paper

Motivation for the paper

This paper was chosen as a case study because it highlights how ITS can be used to evaluate a complex intervention. One of the major challenges to evaluating the effects of the financial crisis is that it was an unplanned event, which means both that experimental designs are not possible and that establishing the timing of the intervention is not straightforward. The potential effects of the financial crisis are multifaceted and not only affect individuals but also society as a whole. Furthermore, it had a broad international reach which means that there is no obvious control population that could definitively be assumed to have been unaffected by the crisis. I wanted to highlight how the complexity of the intervention does not preclude evaluation and demonstrate how I dealt with some of these issues.

Contribution to the thesis

This study addresses objective four of the thesis by demonstrating an application of ITS. It demonstrates how ITS can be used in real world settings with complex interventions where outcome evaluation using other robust study designs is unlikely to be possible. It also demonstrates the use of routine data sources in addressing important research questions. Furthermore, this study highlights some of the methodological challenges that link to other sections of this thesis, including how to deal with defining the timing of an unplanned event in evaluative studies which was discussed in more detail in Chapter 5, and the risk of history bias, in particular when the intervention is diffuse, and how to address this when typical controls are not available, this was discussed in more detail in Chapter 6.

Outputs and contribution to the literature

Prior to this study, most previous evaluations of the health effects of the financial crisis relied on simple pre-post designs that did not account for underlying trends or random fluctuations in the outcome and therefore had poor internal validity.(1, 2) I considered a population level ITS as the best design to evaluate the effects of the financial crisis. The lack of any available control group means that other strong evaluative designs are not possible. Furthermore, because the effects of the financial crisis act as much at a population or societal level as an individual level, an ecological study using population level data is more appropriate.

This study provides strong evidence of an association between the financial crisis and suicides and should inform the provision intervention such as mental health and social welfare programmes to mitigate the effects of financial shocks.(3)

Conclusion

This study demonstrates the practical application of ITS to evaluate a very complex intervention and answer an important question. It is difficult to see how other robust designs could be applied in this setting and this highlights the wide applicability of the design in public health evaluation.

References

1.	Economou M, Madianos M, Peppou LE, Theleritis C, Stefanis CN. Suicidality and the economic crisis in Greece. The Lancet. 2012;380(9839):337.
2.	Stuckler D, Basu S, Suhrcke M, Coutts A, McKee M. Effects of the 2008 recession on health: a first look at European data. The Lancet. 2011;378(9786):124-5.
3.	World Health Organization. Impact of economic crises on mental health. Impact of economic crises on mental health. 2011.

8. Case study 2: The effect of the 2012 HSCA on hospitalisations and specialist visits in England

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	James Lopez Bernal
Principal Supervisor	Antonio Gasparrini and Steven Cummins
Thesis Title	The use interrupted time series for the evaluation of public health interventions

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	PLOS Medicine		
When was the work published?	October 2017		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the concept of this paper, all analysis and writing the paper.

Student Signature: _____ **Date:** __17/01/2018___

Supervisor Signature: _____ **Date:** __25/01/2018__

Supervisor Signature: _____ **Date:** __25/01/2018__

From: noreply@salesforce.com <noreply@salesforce.com> on behalf of PLOS Medicine <plosmedicine@plos.org>
Sent: 17 January 2018 17:42
To: James Lopez-Bernal
Subject: Re: permission to include paper in thesis [ref:_00DU0Ifis._5000BgjlI3:ref]

Dear Dr. Lopez Bernal,

Thank you for your email.

All content published with PLOS is done so under the CC-BY Open Access license, which means you as the author agree that anyone can reuse your article in whole or part for any purpose, for free, even for commercial purposes. Anyone may copy, distribute, or reuse the content as long as the author and original source are properly cited. This, of course, includes you!

No permission to use the paper is necessary, you and anyone else are free to use it as you wish.

Please let me know if there's anything else you need!

Kind Regards,

Erin

PLOS | OPEN FOR DISCOVERY
Erin O'Loughlin | Publications Assistant, PLOS Medicine
1160 Battery Street, Suite 100, San Francisco, CA 94111
eoloughlin@plos.org | Main +1 415-624-1200 | Fax +1 415-546-4090

Case Number: 05601408

--------------- Original Message ---------------
From: James Lopez-Bernal [james.lopez-bernal@lshtm.ac.uk]
Sent: 1/17/2018
To: plosmedicine@plos.org
Subject: permission to include paper in thesis

Dear Sir/Madam,

We published the following article (http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002427) in PLOS Medicine last year and I would like to include this in my thesis.

Could you please advise who retains copyright for this? Also are you able to provide permission for me to include this paper in my thesis in its published format?

Many thanks,

Jamie

PLOS | OPEN FOR DISCOVERY
Erin O'Loughlin | Publications Assistant, PLOS Medicine
1160 Battery Street, Suite 100, San Francisco, CA 94111
eoloughlin@plos.org | Main +1 415-624-1200 | Fax +1 415-546-4090
ref:_00DU0Ifis._5000BgjlI3:ref

Association between the 2012 Health and Social Care Act and specialist visits and hospitalisations in England: A controlled interrupted time series analysis

James A. Lopez Bernal[1,2]*, Christine Y. Lu[2], Antonio Gasparrini[1], Steven Cummins[1], J. Frank Wharam[2], Steven B. Soumerai[2]

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

* james.lopez-bernal@lshtm.ac.uk

Abstract

Background

The 2012 Health and Social Care Act (HSCA) in England led to among the largest health-care reforms in the history of the National Health Service (NHS). It gave control of £67 billion of the NHS budget for secondary care to general practitioner (GP) led Clinical Commissioning Groups (CCGs). An expected outcome was that patient care would shift away from expensive hospital and specialist settings, towards less expensive community-based models. However, there is little evidence for the effectiveness of this approach. In this study, we aimed to assess the association between the NHS reforms and hospital admissions and outpatient specialist visits.

Methods and findings

We conducted a controlled interrupted time series analysis to examine rates of outpatient specialist visits and inpatient hospitalisations before and after the implementation of the HSCA. We used national routine hospital administrative data (Hospital Episode Statistics) on all NHS outpatient specialist visits and inpatient hospital admissions in England between 2007 and 2015 (with a mean of 26.8 million new outpatient visits and 14.9 million inpatient admissions per year). As a control series, we used equivalent data on hospital attendances in Scotland. Primary outcomes were: total, elective, and emergency hospitalisations, and total and GP-referred specialist visits. Both countries had stable trends in all outcomes at baseline. In England, after the policy, there was a 1.1% (95% CI 0.7%–1.5%; $p < 0.001$) increase in total specialist visits per quarter and a 1.6% increase in GP-referred specialist visits (95% CI 1.2%–2.0%; $p < 0.001$) per quarter, equivalent to 12.7% (647,000 over the 5,105,000 expected) and 19.1% (507,000 over the 2,658,000 expected) more visits per quarter by the end of 2015, respectively. In Scotland, there was no change in specialist visits. Neither country experienced a change in trends in hospitalisations: change in slope for total, elective, and emergency hospitalisations were −0.2% (95% CI −0.6%–0.2%; $p = 0.257$),

−0.2% (95% CI −0.6%–0.1%; $p = 0.235$), and 0.0% (95% CI −0.5%–0.4%; $p = 0.866$) per quarter in England. We are unable to exclude confounding due to other events occurring around the time of the policy. However, we limited the likelihood of such confounding by including relevant control series, in which no changes were seen.

Conclusions

Our findings suggest that giving control of healthcare budgets to GP-led CCGs was not associated with a reduction in overall hospitalisations and was associated with an increase in specialist visits.

Author summary

Why was this study done?

- In 2012, the government introduced major reforms to the National Health Service (NHS) in England, which handed budgets for specialist care to GP-led organisations known as Clinical Commissioning Groups.

- This gave GPs a major new role in purchasing hospital-based specialist medical care for patients, in addition to their existing role as "gatekeepers" to specialist care.

- An expected effect of this policy was that there would be a shift in care away from expensive hospital-based care and towards the community.

- Our study aimed to evaluate the potential impact of these reforms on levels of hospital activity including outpatient visits to specialists and inpatient admissions.

What did the researchers do and find?

- We examined trends in all NHS specialist visits and hospital admissions between 2007 and 2015 in order to examine changes in trends following the reforms.

- We included equivalent trends in Scotland, where the reforms did not occur, as a control series.

- We found no change in hospital admissions in either country.

- However, in England we found an increase in the trend of outpatient specialist visits following the reforms, equivalent to approximately 3.7 million additional specialist visits between the time the policy was implemented and the end of the study period (compared to expected).

What do these findings mean?

- Our findings suggest that giving control of healthcare budgets to GP-led CCGs was not linked to a decrease in hospital admissions and was associated with an increase in outpatient specialist visits.

124

- Further research is needed to establish the appropriateness of these visits and the reasons for the increase.

- However, these findings suggest that other interventions may be needed in order to shift more patient care into the community.

Introduction

The 2012 Health and Social Care Act (HSCA) in England has been described as "the biggest and most far reaching [reorganisation] in the history of the NHS" [1, 2]. The reforms centred around the introduction of general practitioner (GP) led Clinical Commissioning Groups (CCGs), which received about two-thirds of the National Health Service (NHS) budget (£66.8 billion in 2015–2016) to commission (plan and contract) secondary care, including hospital and specialist services [1]. CCGs represent all GP practices in their local area, and the key difference from the previous commissioning structures was purported to be a major new role for GPs as key decision makers in the commissioning process [1, 3, 4].

Health policy experts and parliamentary and professional bodies have hypothesised that GP-led commissioning could potentially lead to reductions in referrals to specialist care, as either an intended or unintended consequence of the Act [5–10]. They theorise that by giving the gatekeepers, who control access to specialist care, a greater role in budget holding and the purchasing of specialist care, they may be incentivised to reduce referrals [5, 6]. Indeed, Smith and Mays suggest that the primary rationale for GP-led commissioning is to encourage a shift away from expensive secondary care towards more community-based care [6]. Furthermore, 2 out of the 3 main reasons cited by the government for introducing the reforms centred around a need to control costs, although the mechanisms by which GP-led CCGs would achieve cost savings were not made explicit [4]. While the potential for much-needed cost savings in the NHS as a result of reduced secondary care activity has been viewed positively, some—including the National Audit Office and the Royal College of Surgeons—have raised concerns that a reduction in referrals as a consequence of the HSCA and policies introduced by CCGs could result in inequitable rationing of care and missed diagnoses and that their role in commissioning presents GPs with a conflict of interest [7–10].

CCGs and GPs could reduce secondary care activity through various means, including restricting referral criteria, developing community-based care models, investing in preventative healthcare, or promoting services to prevent readmissions [6, 9, 10]. There also exist potential incentives for them to do so: reducing expensive care would allow CCGs to invest savings in other services. Furthermore CCGs are required to maintain a surplus; otherwise, they cannot access additional funding in the form of a "Quality Premium" of up to £5 per person within the population covered by the CCG person [11]. In addition, there are incentives to individual GPs: savings from reduced specialist visits and hospitalisations could allow investment in community-based services provided by GP practices themselves; also, some CCGs have introduced direct payments of £6,000–£11,000 to GP practices for reducing referral rates [7, 8]. Nevertheless, whether these provide a real incentive in practice depends on how engaged GPs feel with the new commissioning organisations, how much responsibility they feel for their budgets, and how much influence they have on the commissioning process. Previous policies that have begun with an intention to place GPs at the centre of commissioning have ultimately resulted in the formation of bureaucratic bodies that have become detached from local

125

practitioners [6]. Furthermore, given existing evidence that increasing GP workload may increase referral rates, it is possible that the increased administrative burden associated with their new commissioning role could instead result in an increase in referrals [12, 13].

We use a controlled interrupted time series (CITS) design to compare changes in the trends of specialist referrals and hospital admissions in England before and after the HSCA with those in Scotland, where the reforms did not occur. We hypothesise that the 2012 HSCA was associated with a reduction in specialist visits and hospitalisations.

Methods

Ethics

Ethical approval was obtained from the London School of Hygiene and Tropical Medicine Observational/Interventions Research Ethics Committee (LSHTM Ethics Ref: 10505).

The intervention

The 2012 HSCA introduced broad ranging and complex reforms to the NHS and public health services in England. These have been described in more detail elsewhere [1, 2, 4, 14, 15]. The principal change was in the way secondary care services were commissioned within the NHS. Prior to 2012, regional healthcare administrative bodies known as primary care trusts (PCTs) and Strategic Health Authorities (SHAs) were responsible for all commissioning. These were abolished as part of the Act and were replaced by CCGs. CCGs are led by a governing body, which includes a representative from each member GP practice, lay members, a secondary care doctor, and a registered nurse [3]. CCGs were first introduced in shadow form (working alongside PCTs) in April 2012 following the enactment of the HSCA; they then took over full budgetary responsibility in March 2013 [1].

Control

While a control is not required in interrupted time series studies, the primary comparison being between preintervention and postintervention trends within the study population, a control population can help to exclude additional confounding events and cointerventions. Healthcare is a largely devolved power in the United Kingdom and the HSCA only applied to England; therefore, we considered the other 3 nations of the UK (Northern Ireland, Scotland, and Wales) as potential controls. These are neighbouring countries with similar population demographics (S1 Table), similar health systems, and shared political structures. Data equivalent to those in England were not available from Northern Ireland; therefore, it was excluded. We chose Scotland as the control, as preintervention data were more stable than for Wales. We also include an analysis as a supplementary appendix with Wales as the control (S2 Table, S1 and S2 Figs).

Data and study population

We obtained quarterly data on all hospital admissions and outpatient specialist visits in NHS hospitals in England between April 2007 and December 2015 from the Health and Social Care Information Centre: Hospital Episode Statistics (HES) [16]. Hospital admissions include all inpatients in NHS hospitals as well as NHS-funded inpatients in the private sector. NHS outpatient activity in England is hospital based; specialist visit data include outpatients in English NHS hospitals and NHS-funded outpatients in the private sector. Our outcomes were total hospital admissions, elective (planned) and emergency (unplanned) admissions, total first specialist visits (excluding follow-up appointments), and GP-referred first specialist visits.

126

Equivalent data for Scotland were obtained from the NHS Scotland Information Services Division: Scottish Medical Records (SMR) [17]. We obtained demographic data for England and Scotland from the Office for National Statistics including midyear population estimates (for denominators), age and sex distribution, crude birth rate, and crude death rate [18]. The Scottish hospital admission data did not include obstetric and psychiatric hospitals and the outpatient visit data did not include visits to nurses, dentists, or other allied health professionals. We therefore excluded these categories from the English data to make the 2 datasets comparable. Data quality reports identified a coding error in the outpatient data prior to April 2010 (3 years before the introduction of the policy); we therefore excluded these data from the analysis [19]. The raw data are provided in the supplementary appendix (S1 and S2 Data). A complete list of the data codes and algorithms used in the data extraction is also provided in the supplementary appendix (S3 and S4 data).

Statistical analysis

We used a CITS design, which allowed us to control both for preintervention trends in the outcome and for potential confounding events that would have affected both the control and the study groups. Although a Poisson distribution is assumed for individual counts, we had very large numbers and the aggregate data were well approximated by a Gaussian distribution (log transformed). Therefore, we used a simple linear segmented regression model to estimate the change in trend in hospital admissions and outpatient visits following the introduction of the HSCA [20]. In order to account for the year during which CCGs were in shadow form, we allowed a one-year phase-in period by excluding the second quarter of 2012 to the first quarter of 2013 from the analysis. We modelled the association as a slope change rather than an immediate level change because choice of providers and referral patterns were likely to change gradually when existing contracts expired and new models of care developed [1]. Adjustments were made for any seasonal effect using a Fourier term [20].

We first estimated the slope changes in England and in Scotland independently. We then used an interaction model for the CITS to estimate the additional trend change in England over and above any change in Scotland, while controlling for any difference in the preintervention trends of the 2 groups (S1 Text). We examined the preintervention data a priori for linearity and autocorrelation at different lags using scatterplots, plots of residuals, and partial autocorrelation functions [20]. A linear trend provided a reasonable fit for all outcomes in the primary model. We included an autoregressive term at the appropriate lag to adjust for any detected autocorrelation. All analyses were conducted using Stata version 14.

Protocol

The original study protocol from the ethics application is available as a supplementary appendix (S1 Protocol). The analysis has only differed from this protocol in that Scotland was selected as the primary control, as it had the most stable data; Wales was included as an additional control following reviewers' recommendations. Northern Ireland was not included as equivalent data to that in England was not available. Furthermore, in this protocol, we also proposed including patient experience measures as secondary outcomes; this was ultimately not included within the current study but we plan to conduct a future study looking at the potential impact on patient experience.

Reporting

This study is reported as per the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement (S1 Checklist).

127

Results

Population characteristics

Age and sex distributions were similar in both England and Scotland (Table 1 and S1 Table). Both populations were slowly aging; the proportion aged 60 or older increased from 21.6% to 23.0% in England and from 22.3% to 24.0% in Scotland. The crude birth rate was consistently about 1.8 per 1,000 higher in England than in Scotland while the crude death rate was consistently about 1.5 per 1,000 lower.

Changes in outpatient specialist visits

Changes in trends of specialist visits are shown in Fig 1 and Table 2. Absolute counts and the rate per 1,000 for each quarter are presented in Table 3. In England, total specialist visits rose slowly by 0.5% per quarter (from 84.7 per 1,000 in quarter 2 [Q2] 2010 to 87.2 per 1,000 in Q1 2012) in the baseline. After the intervention, they rose approximately 3.6 times faster at 1.5% per quarter (from 90.0 per 1,000 in Q2 2013 to 104.6 per 1,000 in Q4 2015). This was equivalent to an increase in slope (additional quarterly increase) of 1.1% (95% CI 0.7%–1.5%), which resulted in a 12.7% higher rate of specialist visits (647,000 additional visits) by the end of the postintervention period in Q4 2015, compared to the underlying (counterfactual) trend. The slope increase was even more marked for GP-referred visits. During the preintervention period, these had a flat trend at 48.3 visits per 1,000 per quarter (trend 1.000, 95% CI 0.998–1.002). After the intervention, this trend increased by 1.6% per quarter (from 49.1 per 1,000 in Q2 2013 to 57.6 per 1,000 in Q4 2015). This was equivalent to an increase in slope of 1.6% (95% CI 1.2%–2.0%) per quarter, which resulted in a 19.1% higher than expected rate of specialist visits (507,000 additional visits) by the end of the study period. For those outcomes that showed strong evidence of a trend change (total and GP-referred specialist visits in England), we have presented observed compared to expected counts in the postintervention period in Table 4. Total specialist visits had weak evidence of seasonal effect with peaks during Q3 (Fourier sin wave $p = 0.055$, cos wave $p = 0.010$).

Specialist visits in Scotland, however, showed no significant change after the policy. Total specialist visits and GP-referred specialist visits almost level at about 72 per 1,000 per quarter (preintervention trend 1.002, 95% CI 0.999–1.006; postintervention trend 1.000, 95% CI 0.996–1.003) and 47 per 1,000 per quarter (preintervention trend 1.002, 95% CI 0.998–1.005, postintervention trend 1.000, 95% CI 0.996–1.003), respectively, throughout the study period.

Table 1. Population characteristics of England and Scotland, 2007–2014.

			England			Scotland		
			2007	2011	2014	2007	2011	2014
		N	51,381,100	53,107,200	54,316,600	5,170,000	5,299,900	5,347,600
Age (%)								
	0–19		24.2	23.9	23.8	23.0	22.3	21.7
	20–39		27.3	26.9	26.6	26.3	25.9	25.9
	40–59		26.8	26.7	26.8	28.5	28.6	28.4
	60–79		17.1	17.8	18.2	18.1	18.9	19.4
	80+		4.5	4.6	4.8	4.2	4.4	4.6
Sex (%)								
	Males		49.2	49.2	49.3	48.3	48.5	48.6
	Females		50.8	50.8	50.7	51.7	51.5	51.4
Crude birth rate (per 1,000)			12.8	13.0	12.2	11.2	11.1	10.6
Crude death rate (per 1,000)			9.2	8.5	8.6	10.8	10.1	10.1

https://doi.org/10.1371/journal.pmed.1002427.t001

128

(a) Total specialist visits:
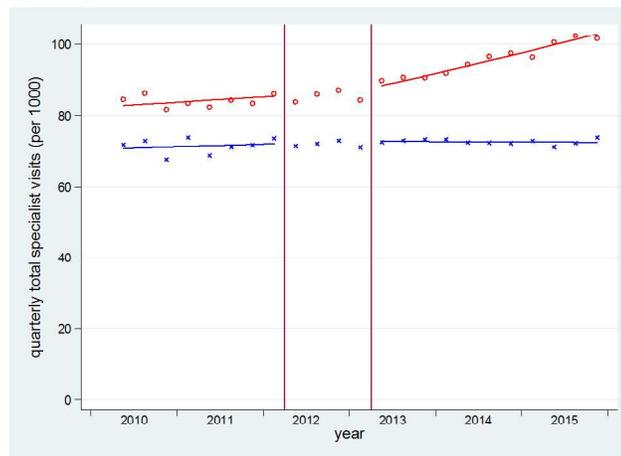
(b) GP referred specialist visits:
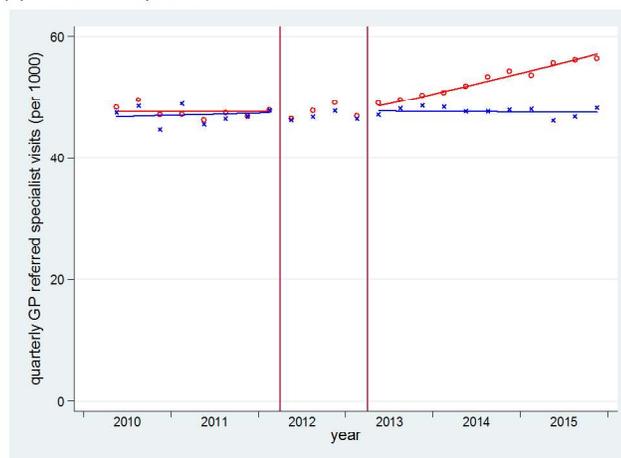
Fig 1. Time series of outpatient specialist visits in England and Scotland. Red o = England, blue x = Scotland. Lines = deseasonalized linear trend. Vertical lines delineate the intervention phase (between quarter 2 [Q2] 2012 and Q2 2013). The data underlying this figure are presented in Table 3. GP, general practitioner.

https://doi.org/10.1371/journal.pmed.1002427.g001

After controlling for trends in Scotland, the CITS analysis produced similar results. The magnitude of the change in slope in total specialist visits in England increased slightly to 1.4% (95% CI 0.6%–2.1%) per quarter (a 15.9% higher rate by the end of the study period). The change in trend in GP-referred specialist visits increased to 1.9% (95% CI 1.1%–2.7%) per quarter (a 22.5% higher rate than expected by the end of the study period).

The magnitude of the differential increase in trend in England was even greater when using Wales as a control series, although this was partly due to an independent reduction in the trend in Wales (S2 Table and S1 Fig).

Changes in inpatient hospitalisations

Changes in trends in hospitalisations following the HSCA are shown in Fig 2 and Table 2. Absolute counts and the rate per 1,000 for each quarter are presented in Table 5. In England, there were slowly increasing trends in all hospitalisations during the baseline period. Total

129

Table 2. Changes in trend in specialist visits and hospitalisations following the intervention.

		Trend change England			Trend change Scotland			Trend change England versus Scotland		
		Effect	95% CI	p-value	Effect	95% CI	p-value	Effect	95% CI	p-value
Outpatient specialist visits										
	Total	**1.011**	**[1.007,1.015]**	**<0.001**	0.997	[0.991,1.003]	0.390	**1.014**	**[1.006,1.021]**	**<0.001**
	GP referred	**1.016**	**[1.012,1.020]**	**<0.001**	0.998	[0.991,1.004]	0.491	**1.018**	**[1.010,1.027]**	**<0.001**
Inpatient hospitalizations										
	Total	0.998	[0.994,1.002]	0.257	0.998	[0.993,1.002]	0.294	1.000	[0.994,1.006]	0.980
	Elective	0.998	[0.994,1.001]	0.235	0.999	[0.993,1.004]	0.625	0.999	[0.992,1.005]	0.670
	Emergency	1.000	[0.995,1.004]	0.866	0.995	[0.989,1.001]	0.114	1.002	[0.995,1.010]	0.497

Coefficients for trend change are relative change in the slope gradient following the intervention. Trend change study versus control is the slope change in England over and above any change in Scotland accounting for differences in baseline trends. All segmented regression models used log transformed Gaussian distribution and p-values were derived from z-tests. Cells in bold indicate strong evidence of an effect (p<0.05).
Abbreviation: GP = general practitioner.

hospitalisations increased by 0.5% per quarter (from 60.1 per 1,000 in Q2 2007 to 65.5 per 1,000 in Q1 2012), elective admissions increased by 0.6% per quarter (from 31.7 to 35.7 per

Table 3. Absolute counts and rates of specialist visits in each quarter.

		England				Scotland			
		Total		GP referred		Total		GP referred	
Year	Quarter	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)
2010	2	44.55	84.7	25.42	48.3	3.73	70.9	2.46	46.8
2010	3	44.84	85.0	25.48	48.3	3.74	71.0	2.47	46.9
2010	4	45.13	85.4	25.54	48.3	3.76	71.2	2.48	47.0
2011	1	45.41	85.8	25.59	48.3	3.77	71.4	2.49	47.1
2011	2	45.71	86.1	25.65	48.3	3.79	71.5	2.50	47.1
2011	3	45.99	86.5	25.70	48.3	3.80	71.7	2.50	47.2
2011	4	46.26	86.9	25.75	48.3	3.81	71.9	2.51	47.3
2012	1	46.54	87.2	25.80	48.3	3.83	72.1	2.52	47.4
2012	2								
2012	3								
2012	4								
2013	1								
2013	2	48.47	90.0	26.44	49.1	3.87	72.7	2.54	47.7
2013	3	49.30	91.4	26.92	49.9	3.88	72.7	2.54	47.7
2013	4	50.15	92.8	27.41	50.7	3.88	72.7	2.54	47.7
2014	1	51.02	94.2	27.91	51.5	3.88	72.7	2.54	47.6
2014	2	51.90	95.6	28.41	52.3	3.88	72.7	2.55	47.6
2014	3	52.80	97.1	28.93	53.2	3.89	72.6	2.55	47.6
2014	4	53.71	98.5	29.46	54.0	3.89	72.6	2.55	47.6
2015	1	54.64	100.0	29.99	54.9	3.89	72.6	2.55	47.5
2015	2	55.58	101.5	30.54	55.8	3.89	72.6	2.55	47.5
2015	3	56.54	103.1	31.09	56.7	3.89	72.6	2.55	47.5
2015	4	57.51	104.6	31.65	57.6	3.90	72.5	2.55	47.5

Abbreviation: GP, general practitioner.
Modelled counts and trends after adjusting for seasonality and autocorrelation.

Table 4. Observed and expected counts for specialist visits in the postintervention period in England.

Year	Quarter	Total visits (x10^5)				GP referred visits (x10^5)			
		Expected	Observed	Difference	Percent difference	Expected	Observed	Difference	Percent difference
2013	2	47.95	48.47	0.52	1.1%	26.03	26.44	0.42	1.6%
2013	3	48.25	49.30	1.06	2.2%	26.08	26.92	0.84	3.2%
2013	4	48.55	50.15	1.60	3.3%	26.13	27.41	1.27	4.9%
2014	1	48.85	51.02	2.17	4.4%	26.19	27.91	1.72	6.6%
2014	2	49.16	51.90	2.74	5.6%	26.24	28.41	2.17	8.3%
2014	3	49.47	52.80	3.32	6.7%	26.30	28.93	2.63	10.0%
2014	4	49.78	53.71	3.93	7.9%	26.36	29.46	3.10	11.8%
2015	1	50.10	54.64	4.54	9.1%	26.41	29.99	3.58	13.5%
2015	2	50.42	55.58	5.17	10.2%	26.47	30.54	4.07	15.4%
2015	3	50.73	56.54	5.81	11.5%	26.53	31.09	4.56	17.2%
2015	4	51.05	57.51	6.47	12.7%	26.58	31.65	5.07	19.1%

Abbreviation: GP, general practitioner.

Expected counts are those expected if there had been no trend change from the pre- to postintervention period (i.e., the counterfactual). Observed counts are those modelled after adjusting for seasonality and autocorrelation. Difference is the additional number of visits in a given quarter over those expected. Percent difference is the additional number of visits expressed as a percentage of the expected number of visits.

https://doi.org/10.1371/journal.pmed.1002427.t004

1,000), and emergency admissions increased by 0.3% per quarter (from 22.9 to 24.5 per 1,000). Total hospitalisations and emergency hospitalisations had a seasonal effect with winter peaks in Q4 (emergency hospitalisations Fourier terms: sin wave $p = 0.002$, cos wave $p = 0.039$). There were no statistically significant changes in any of these trends following the HSCA. Slope changes were −0.2% (95% CI −0.6%–0.2%), −0.2% (95% CI −0.6%–0.1%), and 0.0 (95% CI −0.5%–0.4%) per quarter for total, elective, and emergency hospitalisations, respectively.

Trends in Scotland were flatter during the baseline. Total hospitalisations increased by 0.2% per quarter (from 54.0 to 55.9 per 1,000), elective admissions increased by 0.20% per quarter (from 29.9 to 31.1 per 1,000), and emergency admissions increased by 0.2% per quarter (from 24.1 to 24.8 per 1,000). Again, there was no evidence of any change in these trends after the HSCA: slope changes were −0.3% (95% CI −0.7%–0.2%), −0.1% (95% CI −0.7%–0.4%) and −0.5% (95% CI −1.1%–0.1%), respectively.

The results of the CITS analysis were again similar. The differential slope changes in England (that is, the additional quarterly change following the HSCA after controlling for trends in Scotland) were: 0.0% (95% CI −0.6%–0.6%) per quarter for total hospitalisations, −0.1% (95% CI −0.7%–0.5%) per quarter for elective hospitalisations, and 0.2% (95% CI −0.5%–1.0%) per quarter for emergency hospitalisations.

Results using Wales as a control instead of Scotland were similar (S2 Table and S2 Fig).

Discussion

To our knowledge, this is the first study of the potential impact on secondary care activity of a universal, national policy that gave control of an unprecedented two-thirds of the English NHS budget to GP-led CCGs. Contrary to the underlying hypothesis, we found no evidence of a reduction in hospitalisations or specialist visits in England following the HSCA. Moreover, we found evidence of an increase over and above the underlying trend in specialist visits in England, with no comparable increase in Scotland, where this policy did not occur. This increase was equivalent to approximately 3.7 million additional specialist visits since the policy was implemented (compared to those expected), of which the majority (approximately 2.9 million) were GP referred.

(a) Total hospitalizations:
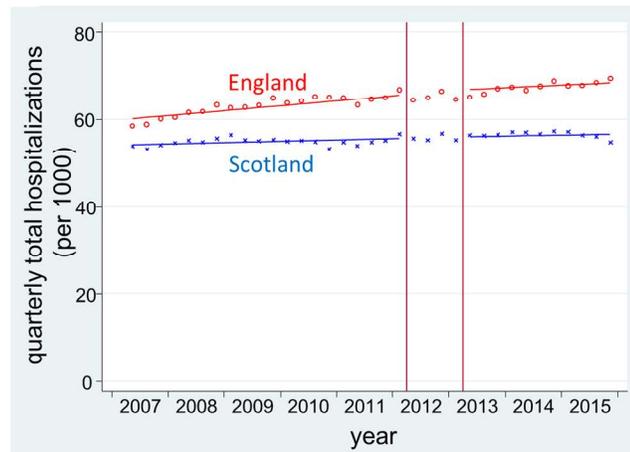
(b) Elective hospitalizations:
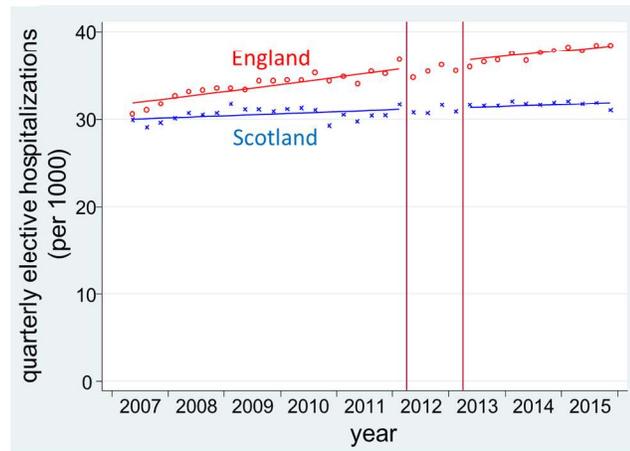
(c) Emergency hospitalizations:

132
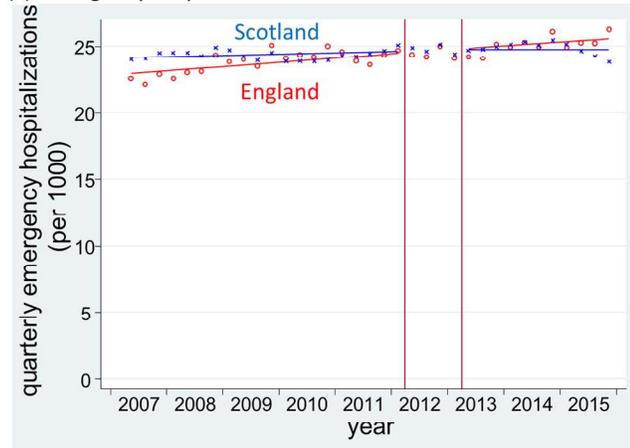
Fig 2. Time series of inpatient hospitalisations in England and Scotland. Red o = England, blue x = Scotland. Lines = deseasonalized linear trend. Vertical lines delineate the intervention phase (between quarter 2 [Q2] 2012 and Q2 2013). The data underlying this figure are presented in Table 5.

https://doi.org/10.1371/journal.pmed.1002427.g002

We used a robust CITS design. By modelling long-term underlying trends, we controlled for secular changes in practice and artefactual changes due to regression to the mean. Selection bias is only an issue in the unlikely event that the population changed suddenly and

Table 5. Absolute counts and rates of hospitalisations in each quarter.

|		England						Scotland					
		Total		Elective		Emergency		Total		Elective		Emergency	
Year	Quarter	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)	Count (x10^5)	Rate (per 1,000)
2007	2	30.87	60.1	16.29	31.7	11.77	22.9	2.79	54.0	1.55	29.9	1.24	24.1
2007	3	31.08	60.4	16.43	31.9	11.84	23.0	2.80	54.1	1.55	30.0	1.25	24.1
2007	4	31.29	60.7	16.57	32.1	11.91	23.1	2.81	54.2	1.56	30.1	1.25	24.2
2008	1	31.50	61.0	16.71	32.3	11.97	23.2	2.82	54.3	1.56	30.1	1.26	24.2
2008	2	31.71	61.2	16.84	32.5	12.04	23.2	2.83	54.4	1.57	30.2	1.26	24.2
2008	3	31.91	61.5	16.98	32.7	12.10	23.3	2.84	54.5	1.58	30.2	1.26	24.3
2008	4	32.12	61.8	17.12	32.9	12.16	23.4	2.85	54.6	1.58	30.3	1.27	24.3
2009	1	32.32	62.1	17.26	33.1	12.23	23.5	2.86	54.7	1.59	30.4	1.27	24.4
2009	2	32.53	62.4	17.40	33.3	12.29	23.6	2.86	54.8	1.59	30.4	1.28	24.4
2009	3	32.74	62.6	17.54	33.6	12.36	23.6	2.87	54.9	1.60	30.5	1.28	24.4
2009	4	32.96	62.9	17.69	33.8	12.43	23.7	2.88	55.0	1.60	30.6	1.28	24.5
2010	1	33.18	63.2	17.83	34.0	12.49	23.8	2.89	55.1	1.61	30.6	1.29	24.5
2010	2	33.40	63.5	17.98	34.2	12.56	23.9	2.90	55.2	1.61	30.7	1.29	24.5
2010	3	33.62	63.8	18.13	34.4	12.63	24.0	2.91	55.3	1.62	30.7	1.29	24.6
2010	4	33.85	64.1	18.29	34.6	12.70	24.0	2.92	55.4	1.63	30.8	1.30	24.6
2011	1	34.08	64.4	18.44	34.8	12.77	24.1	2.93	55.5	1.63	30.9	1.30	24.6
2011	2	34.31	64.6	18.60	35.0	12.85	24.2	2.94	55.6	1.64	30.9	1.31	24.7
2011	3	34.53	64.9	18.75	35.3	12.91	24.3	2.95	55.7	1.64	31.0	1.31	24.7
2011	4	34.75	65.2	18.90	35.5	12.98	24.4	2.96	55.8	1.65	31.1	1.31	24.8
2012	1	34.97	65.5	19.05	35.7	13.05	24.5	2.97	55.9	1.65	31.1	1.32	24.8
2012	2												
2012	3												
2012	4												
2013	1												
2013	2	36.00	66.9	19.78	36.7	13.38	24.9	3.00	56.2	1.67	31.4	1.32	24.9
2013	3	36.16	67.0	19.90	36.9	13.45	24.9	3.00	56.2	1.67	31.4	1.32	24.8
2013	4	36.31	67.2	20.02	37.0	13.52	25.0	3.00	56.2	1.68	31.4	1.32	24.7
2014	1	36.47	67.3	20.14	37.2	13.58	25.1	3.00	56.1	1.68	31.5	1.31	24.6
2014	2	36.63	67.5	20.27	37.3	13.65	25.2	3.00	56.1	1.68	31.5	1.31	24.5
2014	3	36.79	67.6	20.39	37.5	13.72	25.2	3.00	56.0	1.69	31.5	1.31	24.4
2014	4	36.95	67.8	20.52	37.6	13.79	25.3	3.00	56.0	1.69	31.5	1.30	24.3
2015	1	37.11	67.9	20.64	37.8	13.86	25.4	3.00	56.0	1.69	31.5	1.30	24.2
2015	2	37.28	68.1	20.77	37.9	13.93	25.5	3.00	55.9	1.69	31.6	1.30	24.1
2015	3	37.44	68.2	20.90	38.1	14.00	25.5	3.00	55.9	1.70	31.6	1.29	24.1
2015	4	37.60	68.4	21.02	38.2	14.07	25.6	3.00	55.9	1.70	31.6	1.29	24.0

Modelled counts and trends after adjusting for seasonality and autocorrelation.

https://doi.org/10.1371/journal.pmed.1002427.t005

substantially in contrast to the underlying trend and differentially from trends in the control. S1 Table shows that population characteristics maintained stable trends over the study period, suggesting that this is not an alternative explanation for our findings. Furthermore, we controlled for unknown confounding events coincident with the policy by including Scotland as a comparator. Our study is also based on a very large population with stable trends in the outcomes before and after the intervention; therefore, we are well powered to detect effects. Finally, the compulsory nature and large scale of the intervention again limits selection bias and increases both the internal and external validity of our results.

Our study has several limitations. First, it is possible that the observed changes in trends could have been due to other concurrent policies targeting these outcomes but that did not occur in the control population. Following a literature review, we found one such national policy: an "enhanced service" encouraging GPs to provide extra support for patients deemed at risk of unplanned admission to the hospital; however, this was introduced a year after the Act and only targeted 1 of our outcomes (emergency admissions) in which we did not see a change [21]. We also considered the fact that the Act included a broad range of changes alongside GP-led commissioning and that observed changes in trends might be due to other aspects of the reforms. However, most of the other changes were support structures for the changes to commissioning (such as accountability systems and services regulating specialist care providers) that would be considered integral to the intervention itself, or structural changes to public health and preventative services that are likely to have little direct impact on hospitalisations or specialist visits [2, 4]. Second, smaller-scale effects on certain specialties or diagnoses may have been diluted by the scale of our data. However, as the first study of this nationwide policy, and given the government's aim to address rising demands and treatment costs within the NHS as a whole [4], our goal was to examine the association between the policy and trends in specialist visits and hospitalisations. Third, while we have nearly 3 years of postintervention data, it is possible that some effects of GP-led commissioning have not yet become evident. For example, GPs may have chosen to invest more in preventative services, which can take several years to result in population-level reductions in disease. Finally, our study uses routine data that were not specifically created to answer this research question. However, we use the data in high-level aggregate analysis and only use final, rather than provisional, data, which are regarded as complete. Therefore, quarterly changes are unlikely to be due to issues such as data completeness or misclassification [22].

Following the introduction of the HSCA, the Department of Health (DH) called for research to evaluate its impact [23]. Nevertheless, initial proposals were rejected, and, while the DH has published an evaluation focussing on the processes of the reforms, we were unable to find any studies looking at the impact of this policy on hospital activity [23, 24]. There have been studies of previous policies that handed greater budgetary responsibility to GPs in the UK and in Israel [25–32]. However, the results of these studies are mixed and difficult to interpret, as all used simple pre-post designs, which do not take into account underlying trends in hospitalisations or specialist visits, and they examined smaller policies, which were voluntary and subject to volunteer selection bias. The lack of control for underlying trends in these studies is particularly important because study groups often appear to have had unusually high referral rates prior to the intervention (partly because budget allocations based on existing referral rates incentivized practices to inflate referrals before becoming budget holders) [27]. Any reduction could therefore have simply been due to regression to the mean.

Our findings suggest that, on a national scale, the concerns raised around restrictions in access to specialist services and rationing of care have not been realised. However, the lack of decrease in hospitalisations and the unanticipated increase in specialist visits also suggest the theorised shift in care away from hospitals to less expensive community settings does not

134

appear to have occurred and, if anything, the increase in specialist visits may have led to cost increases. There are a number of possible reasons why specialist visits and hospitalisations did not decrease. First, while CCGs intended to increase clinical involvement in commissioning, survey evidence suggests that some GPs do not feel fully engaged with their CCG [33]. For example, the majority of GPs are CCG members but do not have a formal role in the governing body, and this group reported much lower levels of influence and ownership than governing body members. A lack of engagement with members may mean that many GPs feel detached from their CCG and under little pressure to make cost savings or unable to influence the way local health services are managed [33]. Second, the financial incentive for CCGs to reduce costs and GPs to change referral patterns may have been too small or too indirect, and, while practice income may have increased by shifting some care from hospitals to community-based care provided by GPs, concerns about potential conflicts of interest could have discouraged this [7]. Finally, it is also possible that referrals to specialists were already appropriate prior to the intervention, resulting in little scope for further reduction. This is supported by evidence that variations in referral rates in the NHS are primarily explained by characteristics of the patient population and not factors affecting GP services [34].

The increase in specialist visits in our study was surprising and may be an unintended consequence of the policy. We identified annual data on NHS costs for outpatient specialist visits from an independent source (S3 Fig). This also appears to show an increase in costs, corroborating our findings regarding upward trends in specialist visits. One explanation might be that the new responsibility for managing budgets has inadvertently increased administrative workload for GPs, resulting in less time to see patients. Under such circumstances, GPs may reduce their threshold for referral to avoid missing a diagnosis. There is some existing evidence to suggest that increased workload and reduced consultation time is associated with increased referral rates [12, 13], although other studies have shown no effect [35]. We considered decreasing GP numbers or increasing supply of specialists as other potential explanations for this finding. However, although there was a slight decrease in the number of full-time equivalent GPs (from 0.69 to 0.67 per 1,000 population) between 2009 and 2010, this does not coincide with the increase in specialist visits, and the number of GPs remained stable from 2010 and, in fact, increased back to 0.69 per 1,000 population in 2014 [36]. Number of specialists (full-time equivalent consultants) increased gradually over the study period from 0.67 per 1,000 population in 2009 to 0.76 per 1,000 population in 2014 and there was no deviation in this trend around the introduction of the HSCA [37].

In conclusion, we found no evidence that the introduction of GP-led commissioning in England was associated with a reduction in overall hospitalisations or specialist visits. In fact, there was an increase in specialist visits, which appears to have been paralleled by an increase in expenditure. This study begins to decipher the macro effects of these significant reforms to the organisation of NHS commissioning. However, many questions remain unanswered. Examples include the appropriateness of any change in rates of specialist visits and hospitalisations, the effect of this change on health outcomes, whether changes differed according to CCG and why, and the generalizability of our findings to other health systems. This study alone is unable to determine whether the HSCA can be regarded as a good or bad policy, and further research is needed to evaluate other important outcomes such as costs and quality of care. Nevertheless, in the context of similar findings from other large-scale health policy experiments [38], more effort may be needed to target specific costly or poorly evidenced practices (such as tonsillectomy, tympanostomy, or antibiotics prescribed for viral infections) rather than to count on broad, system-wide policy changes that often have unintended consequences.

135

Supporting information

S1 Table. Population characteristics of England and Scotland, 2007–2014.
(DOCX)

S2 Table. Trend changes in specialist visits and hospitalisations following the intervention in England versus Wales. Coefficients for trend change are relative change in the slope gradient following the intervention. Trend change study versus control is the slope change in England over and above any change in Wales accounting for differences in baseline trends. All segmented regression models used log transformed Gaussian distribution.
(DOCX)

S1 Fig. Time series of outpatient specialist visits in England and Wales. Red o = England, blue x = Wales. Lines = deseasonalized linear trend. Vertical lines delineate the intervention phase (between quarter 2 [Q2] 2014 and Q2 2013).
(DOCX)

S2 Fig. Time series of inpatient hospitalisations in England and Wales. Red o = England, blue x = Wales. Lines = deseasonalized linear trend. Vertical lines delineate the intervention phase (between quarter 2 [Q2] 2014 and Q2 2013).
(DOCX)

S3 Fig. National Health Service (NHS) reference costs.
(DOCX)

S1 Text. Controlled interrupted time series model.
(DOCX)

S1 Checklist. REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement.
(DOCX)

S1 Protocol.
(DOCX)

S1 Data.
(XLSX)

S2 Data.
(XLSX)

S3 Data.
(PDF)

S4 Data.
(DOC)

Author Contributions

Conceptualization: James A. Lopez Bernal, Antonio Gasparrini, Steven Cummins, Steven B. Soumerai.

Data curation: James A. Lopez Bernal.

Formal analysis: James A. Lopez Bernal.

Funding acquisition: James A. Lopez Bernal.

136

Investigation: James A. Lopez Bernal, Christine Y. Lu, Antonio Gasparrini, Steven Cummins, J. Frank Wharam, Steven B. Soumerai.

Methodology: James A. Lopez Bernal, Christine Y. Lu, Antonio Gasparrini, Steven Cummins, J. Frank Wharam, Steven B. Soumerai.

Project administration: James A. Lopez Bernal.

Resources: James A. Lopez Bernal.

Software: James A. Lopez Bernal.

Supervision: Antonio Gasparrini, Steven Cummins, Steven B. Soumerai.

Validation: James A. Lopez Bernal, Steven B. Soumerai.

Writing – original draft: James A. Lopez Bernal.

Writing – review & editing: James A. Lopez Bernal, Christine Y. Lu, Antonio Gasparrini, Steven Cummins, J. Frank Wharam, Steven B. Soumerai.

References

1. Ham C, Baird B, Gregory S, Jabbal J, Alderwick H. The NHS under the coalition government. Part one: NHS reform. The King's Fund, 2015.

2. Health and Social Care Act. Health and Social Care Act 2012 [Chapter 7]. Available from: http://www.legislation.gov.uk/ukpga/2012/7/pdfs/ukpga_20120007_en.pdf.

3. NHS Commissioning Board. Clinical commissioning group governing body members: role outlines, attributes and skills. Guidance July. 2012.

4. Department of Health. Health and Social Care Act 2012: fact sheets. https://www.gov.uk/government/publications/health-and-social-care-act-2012-fact-sheets: 2012.

5. Mannion R. General practitioner-led commissioning in the NHS: progress, prospects and pitfalls. British Medical Bulletin. 2011; 97(1):7–15. https://doi.org/10.1093/bmb/ldq042 PMID: 21257619

6. Smith JA, Mays N. GP led commissioning: time for a cool appraisal. BMJ. 2012; 344. https://doi.org/10.1136/bmj.e980 PMID: 22344307

7. National Audit Office. Managing conflicts of interest in NHS clinical commissioning groups. National Audit Office, 2015.

8. Shaw D, Melton P. Should GPs be paid to reduce unnecessary referrals? BMJ. 2015; 351. https://doi.org/10.1136/bmj.h6148 PMID: 26578085

9. Kmietowicz Z. Commissioners defend policies on surgery referrals amid accusations of rationing. BMJ. 2014; 349. https://doi.org/10.1136/bmj.g4644 PMID: 25034351

10. Lacobucci G. GPs put the squeeze on access to hospital care. BMJ. 2013; 347. https://doi.org/10.1136/bmj.f4432 PMID: 23843502

11. NHS England. Quality Premium: 2014/15 guidance for CCGs. 2014.

12. Groenewegen PP, Hutten JBF. Workload and job satisfaction among general practitioners: A review of the literature. Social Science & Medicine. 1991; 32(10):1111–9. http://dx.doi.org/10.1016/0277-9536(91)90087-S.

13. Camasso MJ, Camasso AE. Practitioner productivity and the product content of medical care in publicly supported health centers. Social Science & Medicine. 1994; 38(5):733–48.

14. Naylor C, Curry N, Holder H, Ross S, Marshall L, Tait E. Clinical Commissioning Groups: Supporting Improvements in General Practice? 2013. Kings's Fnd and Nuffield Trust. 2013.

15. Timmins N. Never Again? The story of the Health and Social Care Act 2012: A study in coalition government and policy making. The King's Fund and the Institute for Government, 2012.

16. Health and Social Care Information Centre. Hospital Episode Statistics: Health and Social Care Information Centre; [14th March 2016]. Available from: http://www.hscic.gov.uk/hes.

17. NHS Scotland Information Services Division. SMR Datasets [14th March 2016]. Available from: http://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/.

18. Office for National Statistics. Population estimates [14th March 2016]. Available from: http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates.

137

19. 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.

20. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2016. Epub 2016/06/11. doi: 10.1093/ije/dyw098. PMID: 27283160.

21. NHS England. Avoiding unplanned admissions enhanced service: Guidance and audit requirements 2014 [05/07/2014]. Available from: http://www.nhsemployers.org/~/media/Employers/Publications/Avoiding%20unplanned%20admissions%20guidance%202014-15.pdf.

22. Health and Social Care Information Centre. Hospital Episode Statistics (HES) Analysis Guide 2015. Available from: http://content.digital.nhs.uk/media/1592/HES-analysis-guide/pdf/HES_Analysis_Guide_Jan_2014.pdf.

23. Vittal Katikireddi S, McKee M, Craig P, Stuckler D. The NHS reforms in England: four challenges to evaluating success and failure. Journal of the Royal Society of Medicine. 2014; 107(10):387–92. https://doi.org/10.1177/0141076814550358 PMID: 25271273

24. Department of Health. Post-legislative assessment of the Health and Social Care Act 2012: Memorandum to the House of Commons Health Select Committee. In: Health Do, editor. 2014.

25. Howie JG, Heaney DJ, Maxwell M. Evaluating care of patients reporting pain in fundholding practices. BMJ. 1994; 309(6956):705–10. Epub 1994/09/17. PMID: 7950524; PubMed Central PMCID: PMCPmc2540843.

26. Kammerling RM, Kinnear A. The extent of the two tier service for fundholders. BMJ. 1996; 312 (7043):1399–401. PMID: 8646099

27. Croxson B, Propper C, Perkins A. Do doctors respond to financial incentives? UK family doctors and the GP fundholder scheme. Journal of Public Economics. 2001; 79(2):375–98. http://dx.doi.org/10.1016/S0047-2727(00)00074-8.

28. Dusheiko M, Gravelle H, Jacobs R, Smith P. The effect of financial incentives on gatekeeping doctors: Evidence from a natural experiment. Journal of Health Economics. 2006; 25(3):449–78. http://dx.doi.org/10.1016/j.jhealeco.2005.08.001. https://doi.org/10.1016/j.jhealeco.2005.08.001 PMID: 16188338

29. Coulter A, Bradlow J. Effect of NHS reforms on general practitioners' referral patterns. BMJ. 1993; 306 (6875):433–7. PMID: 8461728

30. Fear CF, Cattell HR. Fund-holding general practices and old age psychiatry. The Psychiatrist. 1994; 18 (5):263–5. https://doi.org/10.1192/pb.18.5.263

31. Gross R, Nirel N, Boussidan S, Zmora I, Elhayany A, Regev S. The influence of budget-holding on cost containment and work procedures in primary care clinics. Social Science & Medicine. 1996; 43(2):173–86. http://dx.doi.org/10.1016/0277-9536(95)00359-2.

32. Elhayany A, Regev S, Sherf M, Reuveni H, Shvartzman P. Effects of a fundholding discontinuation. An Israeli health maintenance organization natural experiment. Scandinavian journal of primary health care. 2001; 19(4):223–6. Epub 2002/02/02. PMID: 11822644.

33. Robertson R, Ross S, Bennett L, Holder H, Gosling J, Curry N. Risk or reward? The changing role of CCGs in general practice. King's Fund and Nuffield Trust, 2015.

34. Reid FDA, Cook DG, Majeed A. Explaining variation in hospital admission rates between general practices: cross sectional study. BMJ. 1999; 319(7202):98–103. https://doi.org/10.1136/bmj.319.7202.98 PMID: 10398636

35. Wilson A, Childs S. The effect of interventions to alter the consultation length of family physicians: a systematic review. British Journal of General Practice. 2006; 56(532):876–82. PMID: 17132356

36. NHS Digital. General and Personal Medical Services, England September 2015—March 2016. 2016.

37. NHS Digital. NHS Workforce Statistics—September 2016, Provisional statistics 2016 [cited 2017 24/07/2017]. Available from: http://www.content.digital.nhs.uk/catalogue/PUB22716.

38. Naci H, Soumerai SB. History, Bias, Study Design, and the Unfulfilled Promise of Pay-for-Performance Policies in Health Care. Preventing Chronic Disease. 2016; 13. https://doi.org/10.5888/pcd13.160133 PMID: 27337559

138

8.2 Summary of appendices for case study 2

Appendices for this paper are presented in Chapter 11:

11.5.1   S1 Table: Population characteristics: England and Scotland 2007-2014 [1]

11.5.2   S2 Table: Trend changes in specialist visits and hospitalisations following the intervention

11.5.3   S1 Figure: Time series of outpatient specialist visits in England and Wales

11.5.4   S2 Figure: Time series of inpatient hospitalisations in England and Wales

11.5.5   S3 Figure: NHS reference costs

11.5.6   S1 Text: Controlled interrupted time series model

11.5.7   S1 Checklist. REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement.

11.5.8   S1 Protocol

11.5.9   S1 Data: English data

11.5.10 S2 Data: Scottish data

11.5.11 S3 Data: Algorithms used for extraction of English data.

11.5.12 S4 Data: Algorithms used for extraction of Scottish data.

11.5.13 Contribution of the candidate to the paper

8.3 Contribution of the paper

Motivation for the paper

The second case study evaluates the impact of the 2012 Health and Social Care Act in England on inpatient hospitalisations and outpatient services. This was a major health policy that resulted in extensive changes to the way healthcare is purchased by giving GP led clinical commissioning groups control of secondary care budgets.(1) It has similarities to the previous case study in that they both evaluate large scale interventions and are undertaken using national population level routine data sources. However, it differs from the previous study in that this was a planned health service policy intervention with clear implementation dates.

This case study was chosen as the evaluation presents several methodological challenges. In their paper entitled *"The NHS reforms in England: four challenges to evaluating success and failure"* Vittal Katikireddi et al summarise some of the main difficulties in evaluating this policy.(2) The four challenges they describe are: 1) *"defining the intervention – what are the reforms to be evaluated?"* Here, they discuss how some of the policies can be considered a continuation of previous initiatives, citing the Private Finance Initiative, the introduction of hospital Foundation Trusts and the Payment by Results system. 2) *"defining the outcomes – what are the outcomes that are likely to change?"* Here, the authors raise a number of issues in defining the outcomes for evaluation, first that changes in outcomes may be confounded by pre-existing trends; second, data quality may vary as a consequence of the policy; third, baseline data may not be available for some outcomes prior to the reforms; fourth, they question the meaning of some outcomes such as avoidable admissions and mortality amenable to healthcare. 3) *"defining the lag time – when will the reforms have an effect?"* Here, they highlight that the timing of the reforms is not completely clear some changes began before the reforms were fully implemented. Furthermore, they highlight that some outcomes may take a long time before becoming apparent. 4) *"defining the counterfactual – what would have happened if the reforms were not introduced?"* Finally, the authors argue that it is difficult to find a comparator in order to model a counterfactual, they suggest a pre-post design may not be suitable as a change may reflect trends in determinants of health rather than the health policy. They also propose comparing England with Scotland or Wales (where the reforms did not occur) as a potential solution and critique the Department of Health for excluding this type of analysis in their call for research on the policy.

Contribution to the thesis

As with the previous case study, this case study directly addresses objective four of the thesis by demonstrating the use of ITS to evaluate a complex public health intervention. In doing so it highlights some of the limitations of routine data in terms of data quality issues. The study also links into

objective three which relates to CITS analysis, the study uses a location based control and was undertaken alongside developing a framework for CITS analysis which is described Chapter 6. Finally, analysis from this study is used as the primary example in developing a framework for model selection in ITS (objective 3), this is illustrated in Chapter 5. The analytical model used here was designed to closely align with how the intervention was implemented and the nature of the expected effect on the outcome. Re-analysis using less appropriate models is demonstrated in Chapter 5 and I highlight the impact that this can have on the results.

Outputs and contribution to the literature

An ideal approach to evaluating this type of health policy would be to pilot it as a cluster randomised trial or to randomise the order in which it is implemented in different regions using a stepped wedge design.(3) This would have good internal validity and improve the ability of researchers to infer causality with any effects. Nevertheless, for practical and political reasons, implementing policies in this way is not always possible, furthermore, this would not address all of the issues outlined above, including defining the outcomes and lagged effects.(4) Given the way that the policy was implemented, possible evaluative designs that could be used include a simple pre-post design, a cross-sectional non-randomised controlled study (with Scotland or Wales as the control), a controlled before and after design (difference in difference), or an interrupted time series (with or without a control series). Of these, a controlled ITS study (CITS) is the most powerful. The pre-post and cross-sectional controlled designs are inherently weak, neither takes into account trends, the former has no control and the latter has no baseline measurements.(5) The controlled before and after study is more powerful, nevertheless, it still does not take into account trends and with the limited controls available for this analysis accurately matching is not possible. Uncontrolled ITS does take into account underlying trends, nevertheless, because the policy was introduced over a relatively long time period (one year), there is a greater risk of history bias. Adding a control series helps to mitigate the risk of history bias and strengthens the design.

The CITS design in itself addresses the fourth issue identified by Vittal et al (defining the counterfactual) by allowing the counterfactual to take into account underlying trends in the outcomes and by including a control which helps to exclude other changes occurring around the time of the policy. In this study I also attempted to address the other issues that were raised by Vittal et al: In defining the intervention I focussed on one of the major changes of the policy which was the introduction of GP-led commissioning groups, this was purported as a major change in that clinicians became the decision makers with regard to secondary care budgets.(4) This was a clear difference from previous policies, nevertheless there is evidence that clinicians felt less empowered than anticipated which may mean the change was not as radical as expected, I addressed this issue in the

discussion of the paper. Second, in defining the outcome I used a routine data series that was in place well before the policy was introduced and that has continued since the policy, I also reviewed data quality reports and where data quality issues could have impacted the data, I excluded this data from the analysis. Finally, I addressed the issue around the timing of the intervention by including a 'transition phase' in the analytical model which excluded the year between when the policy was first enacted and when Clinical Commissioning Groups took full responsibility for commissioning, and by using a slope change model which allowed for a gradual change in the outcome. Nevertheless, effects with a long lag will not be identifiable for some time and this is something that I highlighted in the discussion.

Conclusion

This study again demonstrates the flexible use of ITS for evaluating complex interventions. However, it contrasts from the previous study in the type of impact model used and the type of controls. The study both informs and was informed by the methodological work in the previous chapters and, as such shows a progressive improvement in the robustness of the design and analysis since the previous case study. Nevertheless, the publication was required to adopt established reporting criteria, of which none currently exist for ITS studies of public health interventions. The STROBE and RECORD criteria used do not align well to ITS and this highlights the need for reporting criteria specific to ITS as proposed in Chapter 9.(6, 7)

References

1.   Timmins N. Never Again? The story of the Health and Social Care Act 2012: A study in coalition government and policy making. The King's Fund and the Institute for Government; 2012.
2.   Vittal Katikireddi S, McKee M, Craig P, Stuckler D. The NHS reforms in England: four challenges to evaluating success and failure. Journal of the Royal Society of Medicine. 2014;107(10):387-92.
3.   Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, et al. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. Trials. 2015;16(1):358.
4.   Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. JECH. 2009;65(7):582-7.
5.   Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
6.   Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med. 2007;45(4):247-51.
7.   Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLOS Medicine. 2015;12(10):e1001885.

9. Methodological paper 4: Reporting recommendations for ITS studies

9.1 Research paper: Framework for Enhanced Reporting of Interrupted Time Series (FERITS): Reporting Recommendations for Interrupted Time Series Evaluations of Public Health Interventions and Review of Existing Practice

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	James Lopez Bernal
Principal Supervisor	Antonio Gasparrini and Steven Cummins
Thesis Title	The use of interrupted time series for the evaluation of public health interventions

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Journal of Epidiemiology and Community Health
Please list the paper's authors in the intended authorship order:	James Lopez Bernal, Antonio Gasparrini, Steven Cummins
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the concept of this paper, all analysis and writing the paper.

Student Signature: _____ **Date:** __17/01/2018__

Supervisor Signature: _____　　**Date:** __25/01/2018____

Supervisor Signature: _____　　**Date:** __25/01/2018____

FRAMEWORK FOR ENHANCED REPORTING OF INTERRUPTED TIME SERIES (FERITS): REPORTING RECOMMENDATIONS FOR INTERRUPTED TIME SERIES EVALUATIONS OF PUBLIC HEALTH INTERVENTIONS AND REVIEW OF EXISTING PRACTICE

ABSTRACT

Background

Interrupted time series (ITS) is an increasingly popular design for evaluation of public health interventions. It has a number of methodological features and potential sources of bias that are not addressed in existing quality criteria. We propose a new set of reporting recommendations for ITS studies. To assess the need for such recommendations, we also review recent reporting practice in the field of public health.

Methods

A Framework for Enhanced Reporting of Interrupted Time Series (FERITS) was developed by adapting the Transparent Reporting of Evaluations with Nonrandomised Designs (TREND) statement to include methodological features and potential sources of bias of ITS studies.

The literature review examined the Medline database for ITS studies of public health interventions published in 2015. Data was extracted on interventions, outcomes, data source, and how methods and sources of bias were reported.

Results

104 studies were included in the analysis. Studies evaluated a broad range of interventions and outcomes, primarily using routine data. Authors were generally comprehensive in reporting characteristics of the data series (e.g. 100% clearly defined the timing of the intervention) and the model used (e.g. 98.1% reported whether a level or slope change model). However, they often failed to report sources of bias and how these were addressed including: checking for changes to data collection (20.2%), considering history bias (66.3%) and considering seasonality (47.1%).

Conclusion

There is a need for improved reporting of ITS studies in public health. Researchers are encouraged to use formal reporting criteria such as the proposed FERITS statement when reporting ITS studies.

INTRODUCTION

Interrupted time series (ITS) is a powerful quasi-experimental study design for intervention evaluation.(1, 2) The design involves estimating the effect of an intervention by modelling the underlying trend in the outcome of interest and examining the change in the trend following the introduction of the intervention.(3, 4) It has a number of advantages: accounting for the underlying trend controls for secular trends, regression to the mean and confounding by variables that change relatively slowly in time. Furthermore, because the evaluation is based on a comparison within the same population, selection bias is rarely a problem and a control group is not an essential requirement.(2, 3) The use of ITS in health research has increased exponentially in recent years (Figure 1). It is particularly applicable to evaluations of public health interventions and for examining the public health impacts of non-health sector interventions as these are often not amenable to evaluation through traditional methods such as randomised controlled trials.(5, 6) Furthermore, the increasing availability of large routine public health databases (so called "Big Data") and the implementation of interventions as natural experiments mean that ITS is often a pragmatic study design in this context.(7) Examples of the use of ITS in public health include evaluations of health policies, health promotion programmes, infectious disease interventions and health service reforms.(8-12)
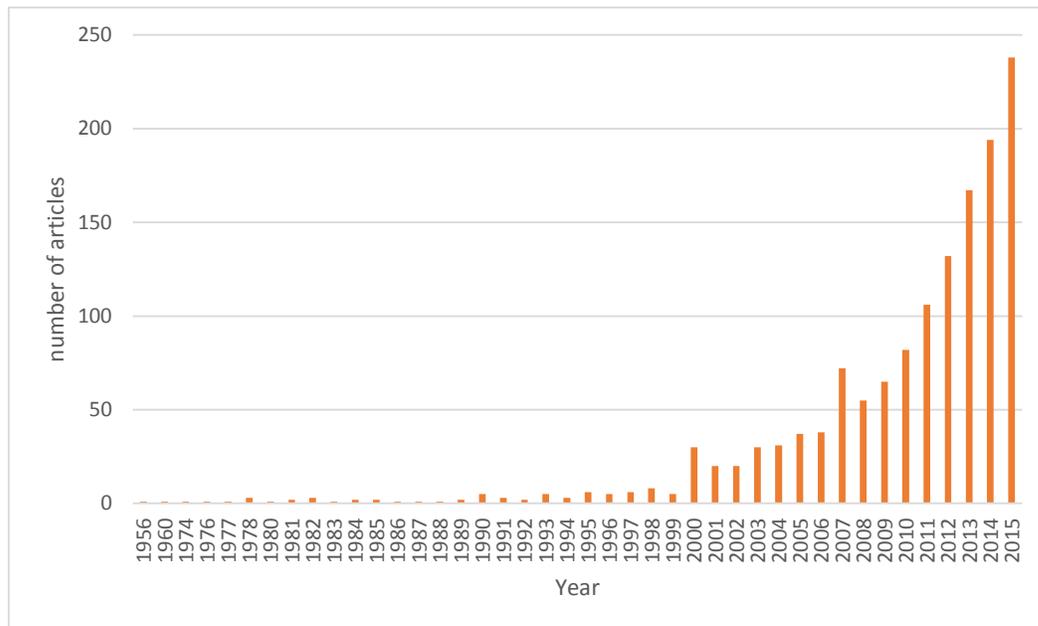
FIGURE 1: NUMBER OF PUBMED ARTICLES WITH THE SEARCH TERM "INTERRUPTED TIME SERIES" BY YEAR

Alongside the increasing popularity of the design, there has been a proliferation of methodological research concerning the design, analysis and validity of ITS studies.(3, 6, 13-20) This research has identified factors that strengthen the validity of ITS and provided recommendations on approaches to undertaking ITS analysis. Because of their longitudinal nature and because the population acts as its own control there are a number of unique methodological considerations in ITS studies that differ from other evaluative designs.(3) For instance, some of the key unique methodological questions that authors need to address include: Have there been any changes to the way outcome data was collected or reported over time (leading to potential instrumentation bias)?(2, 20) How should the underlying trend be defined?(20) What is the nature of the hypothesised effect (i.e. the impact model) of the intervention on the outcome?(20) Was the intervention independent of other changes?(2, 19) Nevertheless, these questions are not always clearly addressed in published ITS studies and potential sources of bias appear to be poorly understood. Without such questions being explicitly answered, comprehensive quality assessment of studies using ITS is not possible.

For other study designs there are now well established quality criteria, such as the CONSORT statement for randomised controlled trials and STROBE for observational studies.(21, 22) These provide guidance for researchers to ensure that the key sources of bias in each design are openly addressed. This facilitates critical appraisal of the quality of studies by reviewers and readers. A number of quality criteria have been developed for systematic reviews of ITS studies in specific fields of health research, including: professional practice and the organisation of health care,(23) health technology assessment,(24) and drug utilisation research.(25) However, these do not incorporate more recent methodological work nor are they focused on the field of public health evaluation. The aims of this paper are twofold: first, to propose a set of reporting recommendations for authors of ITS studies in order to enhance clarity and transparency in the reporting of such studies in the scientific literature. Second, to review current practice in reporting of ITS methods in the field of public health. The first section of the paper will present the proposed new reporting recommendations and explain their scope and how they were developed; the second section will present the methods and results of an illustrative literature review of the applications and methods of recent ITS studies in public health; finally, the discussion will bring these two sections together in order to highlight the major gaps in current reporting practice and how these may be addressed with the proposed reporting framework.

REPORTING RECOMMENDATIONS: FRAMEWORK FOR ENHANCED REPORTING OF INTERRUPTED TIME SERIES (FERITS)

Development of FERITS statement

We reviewed existing established reporting guidelines for observational and quasi-experimental study designs as well as methodological literature on ITS design. The most relevant reporting statement is the Transparent Reporting of Evaluations with Nonrandomised Designs (TREND) statement.(26) TREND has become the most well established set or reporting recommendations for quasi-experimental evaluative studies. As such, the TREND statement forms the basis for developing our reporting recommendations for ITS studies. Nevertheless, the focus of TREND is on non-randomised controlled studies such as difference in difference designs. It does not cover many of the key methodological considerations of time series data. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement also has some items relevant to ITS, however, it's primary focus is on aetiological (rather than evaluative) studies using cohort, case-control or cross-sectional study designs.(22) As evaluative designs, ITS studies must have a clear description of the intervention under study, we therefore also build on recommendations from the Template for Intervention Description and Replication (TIDieR) checklist.(27) Finally, ITS studies of public health interventions typically use routine data sources and we therefore draw on items from the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement.(28) However, these existing reporting statements do not cover some of the important methodological issues outlined earlier in the paper that are specific to ITS. We have identified a number of key areas that comprehensive reporting recommendations should cover, including:

- Sources of bias: history bias and measurement bias have been identified as the primary sources of bias in ITS studies.(2-4) History bias arises when other changes occur around the time of the intervention and could have an effect on the outcome.(2, 19) Because ITS studies examine the change in the outcome before and after the intervention within a single population, it may be difficult to distinguish intervention effects from those of other changes. Measurement bias occurs when there are differences in the way outcomes are measured between groups or over time. As discussed above, ITS studies of public health interventions often rely on routine data sources and the way that routine data is collected or processed can change over time resulting in artefactual changes in the outcome.(2) In particular, instrumentation may occur whereby the intervention itself results in a change in the way outcome data is measured.

- Interrupted time series model specification: ITS analysis involves modelling the underlying trend in the outcome and estimating the intervention effect as the change in this trend following the intervention. Different intervention models can be used to model different effects (for example gradual or abrupt changes in the outcome).(3) The type of model used should be specified according to the data series used and the nature of the intervention and outcome and authors should justify their model selection *a priori*.(20)

- Statistical considerations of time series data: There are a number of statistical considerations specific to time series data which authors should address including, linearity of the trend, whether there is any seasonal variation in the outcome, whether there are time-varying confounders that could influence the outcome and whether data are autocorrelated (a phenomenon whereby data points close together in time are more closely correlated than those that are further apart).(3, 29)

Scope of FERITS statement

The purpose of the proposed FERITS statement is to encourage comprehensive and transparent reporting by authors ITS studies in public health. Many of the items will also be relevant to clinical ITS studies and studies in other fields but there is a focus on studies that use routine data sources. The checklist should also facilitate appraisal of ITS studies for synthesising evidence in systematic reviews or guidelines development.

We present the statement as a basis for further discussion of comprehensive reporting of ITS studies, therefore they should not be regarded as a definitive statement. We hope to receive input from researchers and publishers so that the statement can be improved and developed as a collaborative effort with formal guidance to be generated and adopted in the future.

Items in the FERITS statement

Table 1 presents the items in the FERITS statement.

TABLE 1: FRAMEWORK FOR ENHANCED REPORTING OF INTERRUPTED TIME SERIES STUDIES (FERITS) STATEMENT

Title and Abstract		
Title and Abstract	1	Study design *(interrupted time series)* indicated in the title or the abstract
		Structured abstract recommended
		Information on target population or study sample
Introduction		
Background	2	Scientific background and explanation of rationale
Objectives	3	Specific objectives and hypotheses
Methods		
Study Population	4	Eligibility criteria for participants, including criteria at different levels in recruitment/sampling plan (e.g., cities, clinics, subjects)
		Methods of study population selection (such as codes or algorithms used to identify subjects from routine datasets)
		Settings and locations where the data were collected
Study time-period	5	Start and end dates of the data included in the study, including reasons for selecting this date range and whether this was the full dataset available or if data was restricted. (Presentation of the full time series as a web appendix recommended)
		Time intervals used (e.g. daily, monthly, annual) and the reason for selecting this interval
		Clear definition of the preintervention period, the intervention point (including any transition period) and the post-intervention period
Intervention	6	Details of the intervention(s) and how and when they were actually administered, specifically including:*
		_ Why: Description of the rationale, theory, or goal of the elements essential to the intervention
		_ What: Description of what was done including details of any policy changes, procedures, activities and information provided to participants
		_ Who: Who developed, implemented and/or provided the intervention
		_ How: Description of the modes of delivery of the intervention and whether it was provided individually, in a group or to a whole population
		_ Where: Description of the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features
		_ When: Description of when the intervention was first announced, marketed and delivered, the number of times it was delivered and the duration. Did all groups receive the intervention at the same time
		_ Tailoring and modifications: Description of any adaptations or modifications to the intervention during the course of the study
		_ Adherence: Report on whether the intervention was compulsory, whether adherence was assessed and any activities to increase compliance or adherence
Outcomes	7	Clearly defined outcome measures
		Data source(s)
		Methods used to collect, process, record and extract data; any changes in data collection, processing or recording over time
		Information on validity and reliability of outcome measures
		Information on data quality, coverage and completeness over the duration of the study period; any changes in quality, coverage or completeness over time
History bias	8	Identification of co-interventions or other concurrent events that might affect the outcome; if no such events exist, clear statement that the intervention was independent of other changes
		Description of design adaptations to mitigate the risk of history bias e.g: adding a control series, using multiple phases or a using a multiple baseline design

Unit of Analysis	9	Description of the smallest unit that is being analysed to assess intervention effects (e.g., individual, group, or community)
Statistical Methods	10	Statistical methods used (e.g. segmented Poisson regression, ARIMA etc)
		Appropriateness of a linear model, including description of any tests for linearity and any non-linear terms included
		Detailed description of the *a priori* impact model and why this was chosen, including allowance for: step or slope change effects, lagged effects, transition phase, floor or ceiling effects. Describe why this model is appropriate for the intervention and outcome under study
		Adjustments for time varying confounders (including seasonality)
		Assessment of autocorrelation and how this was handled
		Description of any stratified or subgroup analyses
		Explanation of how missing data were addressed
		Discussion of uncertainty in the primary statistical model and description of any additional sensitivity analyses
		Statistical software or programs used
Results		
Numbers analysed	11	Report on the number of participants (denominator) included in each analysis for each study condition throughout the study period, particularly when the denominators change for different outcomes.
Population characteristics	12	Description of the baseline demographic and clinical characteristics of the intervention group and any control groups at baseline and throughout the study period (a table is recommended)
		Report on study group equivalence at baseline and statistical methods used to control for baseline differences
		Identification of differential changes in population characteristics between study groups throughout the study period and description of statistical methods used to control for differential changes
Outcomes and estimation	13	For each primary and secondary outcome, a summary of results for each estimation study condition, and the estimated effect size and a confidence interval to indicate the precision
		Report on both relative and absolute changes in the study outcomes following the intervention
		Inclusion of null and negative findings
		Inclusion of results from testing pre-specified causal pathways through which the intervention was intended to operate, if any
		Graphical presentation of the time series for each outcome with the regression line, pre-intervention time period, intervention points and post-intervention period clearly indicated
Ancillary analyses	14	Summary of other analyses performed, including subgroup or restricted analyses and sensitivity analyses, indicating which are pre-specified or exploratory
Adverse events	15	Summary of all important adverse events or unintended effects in each study condition (including summary measures, effect size estimates, and confidence intervals)
DISCUSSION		
Interpretation	16	Interpretation of the results, taking into account study hypotheses, sources of potential bias, imprecision of measures, multiplicative analyses, and other limitations or weaknesses of the study
		Discussion of results taking into account the mechanism by which the intervention was intended to work (causal pathways) or alternative mechanisms or explanations
		Discussion of the success of and barriers to implementing the intervention, fidelity of implementation
		Discussion of research, programmatic, or policy implications
Generalizability	17	Generalizability (external validity) of the findings, taking into account the study population, the characteristics of the intervention, length of follow-up, incentives, compliance rates, specific sites/settings involved in the study, and other contextual issues

Overall Evidence	18	General interpretation of the results in the context of current evidence and current theory

LITERATURE REVIEW

In order to provide an illustrative example of current reporting we undertook a literature review of ITS studies of public health interventions published in 2015. Here, we aim to provide a picture of the specific research areas where ITS is applied within public health and the type of data that is used. We then aim to look at how well the studies report the ITS specific methodological issues that feed into the FERITS framework.

Methods

The UK Faculty of Public Health defines three domains of public health: health improvement (including health promotion and education activities targeting the wider determinants of health), health services (including interventions to improve health service effectiveness, efficiency, planning and equity), and health protection (including communicable disease control and response to chemical, radiological and environmental hazards).(30) We include population level studies of interventions in any of these areas as well as studies evaluating the impact of non-health interventions on health outcomes. We exclude studies evaluating clinical interventions that act at an individual level. We searched Medline for all studies published in 2015 using Interrupted Time Series Analysis as a MESH term, or any of the following keywords: "Interrupted" AND "time" AND "series", "segmented regression", or "autoregressive integrated moving average".

Studies were screened using the title and abstract and we excluded studies that were not original research papers, not ITS studies or did not evaluate public health interventions. Further exclusions on this basis were made when reviewing full papers. From eligible studies, we extracted information on the characteristics of the study and the methodological reporting of the study. A detailed description of the data extracted is provided in Appendix 1. For methodological reporting, we focussed on items of the FERITS checklist that are specific to ITS studies, including description and justification of the study time period (item 5), changes in outcome measurement over time (item 7), history bias (item 8), the impact model and statistical considerations of time-series data (item 10).

Results

Figure 2 shows the flow diagram of the literature search. The Medline search identified 304 unique articles. 80 of these were not original empirical research (49 review articles; 17 commentaries, letters or protocols; 14 methodological articles). 81 used other study designs (primarily traditional

time-series correlation or forecasting studies identified through the "autoregressive integrated moving average" search term but that were not ITS), six studies claimed to be ITS studies but did not adjust for trend and would therefore more commonly be regarded as a simple pre-post design. 32 of the remaining studies did not evaluate a public health intervention or outcome. Of the 105 eligible articles (Appendix 2) we were unable to access one and could not contact the author of this study.

Records identified through Medline search (n=322)

↓

Records after duplicates removed (n=304)

→ Records excluded (n=199)
- Review articles (n=49)
- Commentaries/protocols/letters (n=17)
- Methodological articles (n=14)
- Not ITS method (n=81)
- Did not adjust for trend (n=6)
- Not a public health intervention or outcome (n=32)

↓

Eligible records (n=105)

→ Unable to access full text or contact authors (n=1)

↓

Records included in analysis (n=104)

FIGURE 2: FLOW DIAGRAM OF LITERATURE SEARCH

STUDY CHARACTERISTICS

Table 2 shows the characteristics of the studies included in the analysis, including the interventions and outcomes studied and the data sources used. There were a broad range of interventions

evaluated. The most common application of ITS (63.5% of studies) was in health services research, especially health service financing (15.4%, including incentives and penalties for healthcare professionals and changes in user fees) and evaluations of pharmaceutical policies or guidance (15.4%). 16.3% of studies evaluated health protection interventions all of which related to communicable disease control, 10.6% were evaluations of interventions targeting the use of antimicrobials in order to limit antimicrobial resistance and/or infections such as *Clostridium dificile*. 15.4% of studies evaluated health promotion interventions such as those targeting smoking, alcohol and illicit drug use as well as interventions promoting physical activity and road safety. Many of these evaluated national policies or legislation but there were also studies of smaller scale educational programmes.(31-35)

In line with the propensity towards health service interventions, many of the outcomes were health service outcomes including service or treatment uptake (56.7%) and treatment outcomes or service quality measures (13.5%). Measures of health or disease were also common outcomes (24%), in particular disease incidence (17.3%). Other outcomes included mortality measures (4.8%), health behaviours (5.8%) and non-health outcomes of public health interventions (3.8%).

Again, the range of data sources used was broad. Almost all studies used routine data sources, with just two studies using prospective data collection for their time series: one using pedometers,(34) and another collecting new surveillance data.(36) The most common sources of data were health service administrative data (23.1%), for example Hospital Episode Statistics from NHS hospitals in England),(37) electronic health records (18.3%), and health insurance claims data (17.3%). A number of studies used data sources not traditionally associated with health research including police data, transport data, a flight database and a child protection database.

TABLE 2: CHARACTERISTICS OF IDENTIFIED STUDIES (N=104)

Intervention	n	%
Health promotion	**16**	**15.4%**
Alcohol interventions	3	2.9%
Illicit drug use interventions	1	1.0%
Physical activity interventions	1	1.0%
Smoking interventions	8	7.7%
Road safety interventions	3	2.9%
Health protection	**17**	**16.3%**
Antimicrobial intervention	11	10.6%
Infection control	5	4.8%
Vaccine introduction	1	1.0%
Health services	**66**	**63.5%**
Health service financing	16	15.4%
New services/service withdrawal	7	6.7%
Pharmaceutical policy/guidance	16	15.4%
Treatment guidelines	9	8.7%
Screening	4	3.8%
Other health service interventions	14	13.5%
Non-health interventions/events	**5**	**4.8%**
Economic policy/event	2	1.9%
Mass gathering event	1	1.0%
Road changes	1	1.0%
TV programme	1	1.0%

Outcomes	n	%
Health/disease measures	**25**	**24.0%**
disease incidence	18	17.3%
disease prevalence	1	1.0%
health index	1	1.0%
road traffic crash casualties	4	3.8%
self harm	1	1.0%
Mortality	**5**	**4.8%**
suicides	2	1.9%
infant mortality	1	1.0%
mortality	1	1.0%
drug related deaths	1	1.0%
Health behaviours	**6**	**5.8%**
physical activity measure	1	1.0%
smoking prevalence	4	3.8%
drink driving	1	1.0%
Service/treatment uptake	**59**	**56.7%**
prescriptions	29	27.9%
screening uptake	5	4.8%
diagnostic test uptake	1	1.0%
caesarean sections	1	1.0%
treatment uptake	5	4.8%
hospital admissions	2	1.9%
ED presentations	2	1.9%
referrals	1	1.0%
service utilisation	13	12.5%
Treatment outcomes/service quality	**14**	**13.5%**
service performance/quality indicator	6	5.8%
treatment complications	1	1.0%
adverse drug reactions	1	1.0%
length of stay	1	1.0%
medication adherence	1	1.0%
readmissions	4	3.8%
Non health outcomes	**4**	**3.8%**
child protection referrals	1	1.0%
costs	3	2.9%

Data source	n	%
disease registry	2	1.9%
child protection database	1	1.0%
electronic health records	19	18.3%
flight database	1	1.0%
health insurance claims data	18	17.3%
health service admin data	24	23.1%
laboratory data	1	1.0%
maternity records	1	1.0%
microbiology data	7	6.7%
mortality data	4	3.8%
pedometer	1	1.0%
pharmacy data	15	14.4%
police data	3	2.9%
population survey	3	2.9%
surveillance data	2	1.9%
transport data	2	1.9%
worker health examinations	1	1.0%

NB: some studies looked at more than one primary outcome and used more than one data source.

METHODS REPORTING

Table 3 shows how methods were reported by the identified ITS studies. Almost all studies (99.0%) clearly stated the study period (the time range of the data series) and the time interval (97.1%) and all studies clearly defined the timing of the intervention. However, only 20.2% of studies reported

checking for changes to data collection or processing (including variations in data quality or completeness) over time and only 10.6% provided a clear justification for their choice of study period. 66.3% of studies reported considering potential confounding interventions or events as an alternative explanation for their findings (history bias). 35.6% of studies adapted the ITS design to mitigate the risk of history bias, normally by including a control series, a controlled (or comparative) ITS design (29.8% of studies). Other studies used multiple phase designs whereby the intervention was introduced and then withdrawn (1.9%), multiple baselines (whereby the intervention was introduced in different locations at different times) (1.0%), by including potential confounding interventions as dichotomous variables within their model (1.9%) or by triangulating with another study design (1.0%), in this case a regression discontinuity design. The controlled ITS studies used a range of different types of control series, the most common were location based controls (13.5% of all studies) and control outcomes (or non-equivalent dependent variables, 9.6%).

All bar one study reported their method of analysis, 83.7% used segmented regression and 15.4% used autoregressive integrated moving average (ARIMA) models. The majority of studies (76.9%) allowed both a level and slope change impact model, 13.5% allowed a level change only and 7.7% allowed a slope change only. 8.7% of studies included a transition phase during which the intervention was being implemented,(20) and 1.9% allowed for a lagged effect. No studies modelled possible floor or ceiling effects. While the impact model used was generally described in detail, only 2.9% of studies fully justified their choice of impact model *a priori*, with 4.8% justifying part of their impact model (the transition phase only) but not other parts such as why a level and/or slope change model was appropriate. The majority of studies (87.5%) used a linear model but only 14.4% reported checking the fit of a linear model. 47.1% of studies reported considering seasonality as a potential confounder with 39.4% adjusting for seasonality in their model. 39.4% of studies reported considering other possible time varying confounders with 33.7% adjusting for such confounders. 66.3% of studies considered autocorrelation and 49.0% made adjustments to the model to account for autocorrelation. Finally, 27.9% of studies conducted sensitivity analyses and the majority of these (24 out of 29) were clearly justified *a priori*.

Table 3: Methods reporting by identified ITS studies (n=104)

Methodological issue	n	%
Data series		
Study period clearly stated	103	99.0%
Time interval clearly stated	101	97.1%
Timing of intervention clearly defined	104	100.0%
Checked for changes to data collection/processing	21	20.2%
Data range fully justified	11	10.6%
History bias		
History bias considered	69	66.3%
Design adaptations to mitigate history bias	37	35.6%
Control series	31	29.8%
Multiple phase design	2	1.9%
Multiple baseline design	1	1.0%
Included other interventions in model	2	1.9%
Triangulated with other study design	1	1.0%
Type of control series used		
Behaviour based control	1	1.0%
Characteristic based control	6	5.8%
Control outcome	10	9.6%
Control time period	1	1.0%
Historical control cohort	1	1.0%
Location based control	14	13.5%
Different health plan	1	1.0%
Cluster randomised control	1	1.0%
Method of analysis		
Segmented regression	87	83.7%
ARIMA	16	15.4%
Not stated	1	1.0%
Impact model		
Level or slope change model		
level and slope change	80	76.9%
level change only	14	13.5%
slope change only	8	7.7%
Not stated	2	1.9%
Tranisition phase modelled	9	8.7%
Lagged effect modelled	2	1.9%
Floor/ceiling effect modelled	0	0.0%
Impact model justified a priori		
yes	3	2.9%
no	96	92.3%
partly (transition phase only)	5	4.8%
Linearity		
linearity assessed	15	14.4%
linear model used	91	87.5%
non-linear model used	13	12.5%
Seasonality		
not applicable	5	4.8%
considered	49	47.1%
adjusted for	41	39.4%
Time varying confounders		
considered	41	39.4%

adjusted for	35	33.7%
Autocorrelation		
considered	69	66.3%
adjusted for	51	49.0%
Sensitivity analysis		
Sensitivity analysis conducted	29	27.9%
Sensitivity analysis justified a priori	24	23.1%

DISCUSSION

This literature review has illustrated the broad range of public health interventions and outcomes for which ITS is being applied in evaluative studies. Authors tend to clearly report the methods used such as the time periods covered, the method of analysis, the type of model and any adjustments. Nevertheless, they often fail to report how some of the key threats to validity in ITS studies were addressed such as measurement bias from changes to data collection or processing over time, history bias, confounding due to seasonality and time varying confounders, and autocorrelation. Furthermore, studies rarely justify their analytical choices, such as the data range or the impact model selected, *a priori* based on substantive knowledge. This highlights the need for guidance on the reporting of ITS studies.

Whereas openly discussing potential sources of bias has become standard in traditional study designs, for example blinding and allocation concealment in randomised controlled trials or selection bias and recall bias in case-control studies, we found that discussion of the primary sources of bias in ITS studies is often missing. Because of the pre-post comparison in ITS studies, history bias, is widely regarded as the primary threat to the validity of ITS studies.(1, 2, 19) However, a third of papers in our review did not consider whether other changes could have occurred around the time of the intervention and may offer an alternative explanation for the results. While there may be circumstances where history bias is unlikely, for example if there is an immediate large level change effect, authors should always make clear why they believe other events are unlikely to have caused the effect. Nevertheless, it was encouraging to see that those authors that did address history bias used a wide range of different approaches to limit this threat to validity and a number of innovative control series were used. Measurement bias or instrumentation bias is also a particular concern with ITS studies that rely on routine data sources, however few studies reported checking whether there had been changes to the way data was collected over time, for example changes in data quality or completeness. Similarly, authors should consider whether there are time varying confounders that could influence the trend differently before or after the intervention. Confounders common in other study designs, such as socio-economic status or age and gender distribution, tend to change only relatively slowly and regularly over time and are therefore often not a concern in ITS studies, however, variables that may be associated with the outcome and where changes can be more irregular should be adjusted for, examples might include meteorological events or numbers of hospital admissions for specific diseases.(38, 39) One particular time varying confounder that often needs to be adjusted for is seasonality, an uneven distribution of months before and after the intervention can affect the results if the outcome is expressed more at different times of year.(3)

ITS studies require a number of analytical decisions to be made *a priori* including the range of data to be included in the analysis and the type of impact model to be tested.(20) Only 10% of studies fully justified the data range included in the study. The purpose of modelling the underlying trend in an ITS study is to predict the counterfactual, that is the expected outcome if the intervention had not been implemented. It is therefore important that the data range included in the pre-intervention period is still relevant to post intervention trends. If the data dates back a long time it may include trends that historically differed and thus may bias results. Furthermore, if trends have changed over time, there is a concern that the data range can be manipulated to produce different results. Authors should therefore clearly justify their chosen study period based on data availability but also, for example, on whether there have been historical interventions or events that could have changed trends.(20) Even fewer studies reported the reasons for their choice of impact model *a priori*. The type of impact model should be aligned to substantive knowledge of the intervention and outcome under study.(20) Inappropriate or data driven model selection increases the risk of false positives being detected.(20) For example, an immediate level change effect with an intervention that is implemented only gradually or for an outcome which would be expected to follow a lag, suggests that effect is due to some factor other than the intervention.

Two previous studies have reviewed methodological reporting of ITS studies in other fields: Ramsay et al reviewed mass media campaigns and guidelines implementation, Jandoc et al reviewed drug utilisation studies.(24, 25) Though there are some differences in the aspects of the studies that were extracted there is also some overlap with our study. Ramsay et al found in 2003 that no studies gave a rationale for the number of data points included and no studies gave a rationale for the shape of the intervention effect (the impact model) similar to our low numbers of 10.9% and 2.9% respectively in these categories. They also found that only 39.7% of studies assessed whether the intervention was independent of other changes (i.e. the threat of history bias), it is encouraging to see that in our study this has increased to around two thirds of studies. Jandoc et al did not specify the number of studies that considered history bias but found that 35% of studies included a control series to mitigate this threat which is similar to our finding. Jandoc et al also found that 66.4% of studies considered autocorrelation (essentially the same as our finding of 66.3%), 30.9% of studies considered seasonality (compared to 47.1% in our study) and 20.5% included sensitivity analyses (compared to 27.9% in our study). In general our results are very similar to those of the two previous reviews but point to a slight improvement in reporting. This could be due to slightly better practice among public health researchers or to an improvement over time, given that our review focussed only on very recent ITS studies.

Our review has a number of limitations. This was not an in depth systematic review, only one database was searched and we only looked at a single year. The purpose of this review was to gain an illustrative overview of current practice rather than to obtain a detailed answer to an empirical research question, therefore we feel that our literature review is appropriate. Secondly, our findings are based on what authors have reported. It is likely that many more authors considered the methodological issues that we looked at but did not explicitly report doing so in their manuscript. This links into our proposal that formal reporting criteria are needed.

The FERITS statement that we have proposed is intended to improve and standardise the reporting of methodological aspects of ITS evaluations of public health interventions. Previous quality criteria that have been developed for ITS studies have focussed on systematic reviews in relatively narrow research topics.(23-25) The most comprehensive of these was developed by Jandoc et al for ITS studies in drug utilisation research which is adapted from the STROBE statement.(25) The Jandoc recommendations are commendable and share a number of similarities to our recommendations for public health research. They encourage researchers to clearly report the statistical models used and to consider important adjustments such as autocorrelation and seasonality, as well as design adaptations such as including a control series. There are some areas where we believe these recommendations are deficient however, for example they do not encourage authors to justify the data range included in the time series nor the impact model selected, neither do they require authors to identify potential sources of history bias which is the primary limitation of the ITS design. We also believe authors need to provide more details on the way the intervention was implemented as this is required so that readers can judge whether an appropriate impact model has been applied. Furthermore, given that public health ITS studies commonly rely on routine data sources we encourage greater reporting of how outcome data was collected and processed and any changes over time.

The results of our literature review highlight the need for improved reporting of ITS studies in public health and we hope that the FERITS statement will go some way towards achieving this. We reiterate that this should be regarded as a suggested approach and that we welcome input from other research groups and publishers in order to develop a collaborative set of recommendations. Better reporting of ITS studies in public health will improve the ability of researchers and policymakers to evaluate and synthesise evidence and ultimately improve the application of this evidence in selecting the most effective interventions to address public health challenges.

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. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
3. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2016.
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. Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. JECH. 2009;65(7):582-7.
6. 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.
7. 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. JECH. 2012;66(12):1182-6.
8. Briesacher BA, Madden JM, Zhang F, Fouayzi H, Ross-Degnan D, Gurwitz JH, et al. Did Medicare Part D Affect National Trends in Health Outcomes or Hospitalizations?A Time-Series AnalysisMedicare Part D and Health Outcomes. Ann Intern Med. 2015;162(12):825-33.
9. Dennis J, Ramsay T, Turgeon AF, Zarychanski R. Helmet legislation and admissions to hospital for cycling related head injuries in Canadian provinces and territories: interrupted time series analysis. BMJ: British Medical Journal. 2013;346:f2674.
10. 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.
11. Druetz T, Fregonese F, Bado A, Millogo T, Kouanda S, Diabaté S, et al. Abolishing Fees at Health Centers in the Context of Community Case Management of Malaria: What Effects on Treatment-Seeking Practices for Febrile Children in Rural Burkina Faso? PLoS ONE. 2015;10(10):e0141306.
12. 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.
13. Linden A. Conducting interrupted time series analysis for single and multiple group comparisons. The Stata Journal. 2015;15(2):480-500.
14. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. J Eval Clin Pract. 2011;17(6):1231-8.
15. 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.
16. St.Clair T, Hallberg K, Cook TD. The Validity and Precision of the Comparative Interrupted Time-Series Design: Three Within-Study Comparisons. Journal of Educational and Behavioral Statistics. 2016.
17. Fretheim A, Soumerai SB, Zhang F, Oxman AD, Ross-Degnan D. Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. J Clin Epidemiol. 2013;66(8):883-7.

18. Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, et al. A re-analysis of cluster randomized trials showed Interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol. 2014.

19. 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].

20. Lopez Bernal JA, Soumerai S, Gasparrini A. Model selection in interrupted time series. J Clin Epidemiol. 2018 [In Press].

21. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340.

22. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med. 2007;45(4):247-51.

23. Cochrane Effective Practice and Organisation of Care Review Group. Suggested risk of bias criteria for EPOC reviews - Risk of bias for interrupted time series (ITS) studies. Oslo: Norwegian Knowledge Centre for the Health Services; 2013.

24. 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.

25. Jandoc R, Burden AM, Mamdani M, Lévesque LE, Cadarette SM. Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. J Clin Epidemiol. 2015;68(8):950-6.

26. Des Jarlais DC, Lyles C, Crepaz N, Group T. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361.

27. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687.

28. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLOS Medicine. 2015;12(10):e1001885.

29. 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.

30. UK Faculty of Public Health. What is public health 2017 [Available from: http://www.fph.org.uk/what_is_public_health.

31. Wagenaar AC, Livingston MD, Staras SS. Effects of a 2009 Illinois Alcohol Tax Increase on Fatal Motor Vehicle Crashes. Am J Public Health. 2015;105(9):1880-5.

32. Hinde JM, Bray JW, Aldridge A, Zarkin GA. The Impact of a Mandated Trauma Center Alcohol Intervention on Readmission and Cost per Readmission in Arizona. Med Care. 2015;53(7):639-45.

33. Roberts SC, Zahnd E, Sufrin C, Armstrong MA. Does adopting a prenatal substance use protocol reduce racial disparities in CPS reporting related to maternal drug use? A California case study. J Perinatol. 2015;35(2):146-50.

34. Burns RD, Brusseau TA, Hannon JC. Effect of a Comprehensive School Physical Activity Program on School Day Step Counts in Children. J Phys Act Health. 2015;12(12):1536-42.

35. Kalkhoran S, Sebrie EM, Sandoya E, Glantz SA. Effect of Uruguay's National 100% Smokefree Law on Emergency Visits for Bronchospasm. Am J Prev Med. 2015;49(1):85-8.

36. Ng W, Brown A, Alexander D, Ho MF, Kerr B, Amato M, et al. A multifaceted prevention program to reduce infection after cesarean section: Interventions assessed using an intensive postdischarge surveillance system. Am J Infect Control. 2015;43(8):805-9.

37.   Hamilton I, Lloyd C, Bland JM, Savage Grainge A. The impact of assertive outreach teams on hospital admissions for psychosis: a time series analysis. J Psychiatr Ment Health Nurs. 2015;22(7):484-90.

38.   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.

39.   Taggart LR, Leung E, Muller MP, Matukas LM, Daneman N. Differential outcome of an antimicrobial stewardship audit and feedback program in two intensive care units: a controlled interrupted time series study. BMC Infect Dis. 2015;15:480.

9.2 Summary of appendices for methodological paper 4

Appendices for this paper are presented in Chapter 11:

11.6.1 Data extracted in the literature review

11.6.2 Eligible studies

11.6.3 Contribution of the candidate to the paper

9.3 Contribution of the paper

Motivation for the paper

Like many quasi-experimental study designs, there are no well-established reporting criteria for ITS studies. The literature review conducted as part of this paper highlights the problems with current reporting of ITS studies. In many cases papers did not report on how they dealt with the primary sources of bias that typically affect ITS studies. This suggests that there is a clear need for recognised reporting criteria for ITS studies.

Contribution to the thesis

This final paper of the thesis brings together the methodological work from the previous chapters, along with the wider ITS methodological literature in order to propose a set of formal reporting recommendations for ITS studies. It addresses the final objective of the thesis: *"To develop guidance for transparent reporting of interrupted time series studies."* The criteria that I have proposed include recommendations around detailed description of the intervention and the use of routine data necessary for typical ITS studies of complex public health interventions. The recommendations from Chapter 6 of the thesis, on analysis of ITS studies, feed into section 10 of the FERITS reporting criteria on 'Statistical Methods'; the recommendations from Chapter 7, on the use of controls, feed into sections 8 and 12 of the reporting criteria on 'History Bias' and 'Population Characteristics' respectively; finally, the recommendations from Chapter 8, on ITS models, feed into sections 5: 'Study time-period', 6: 'Intervention', 7: 'Outcomes' and 10: 'Statistical Methods'.

Outputs and contribution to the literature

Quality criteria have previously been developed for ITS studies, however, these come from fields outside public health.(1-3) They also do not include a lot of the methodological issues that have been highlighted in the literature, including those covered in the earlier chapters of this thesis. The literature review included in this paper is the first to demonstrate the broad range of public health interventions and outcomes to which ITS is being applied. With the ever increasing use of ITS in public health, reporting criteria such as these are vital to ensuring that researchers produce robust evaluations that can be easily appraised and allows the inclusion of more information within evidence syntheses.

Conclusion

Improving the quality of reporting of ITS studies will facilitate the inclusion and appraisal of ITS studies in systematic reviews, as well as in evidence summaries and guidelines development used in service

public health. Ultimately this will help public health practitioners and policymakers to select the most effective interventions.

References

1. Jandoc R, Burden AM, Mamdani M, Levesque LE, Cadarette SM. Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. J Clin Epidemiol. 2015;68(8):950-6.
2. 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.
3. Cochrane Effective Practice and Organisation of Care Review Group. Suggested risk of bias criteria for EPOC reviews - Risk of bias for interrupted time series (ITS) studies. Oslo: Norwegian Knowledge Centre for the Health Services; 2013.

10. Summary and conclusions

The aim of this thesis was to improve the way that interrupted time series studies of public health interventions are designed in order to reduce bias and to make the design more accessible to researchers and public health professionals. This has included a combination of case-studies and methodological papers. My primary focus has been on study design rather than analysis and statistical approaches, in particular, because this is a topic that has not previously been covered in detail in the literature. This is in contrast to study design of well-established designs such as randomised controlled trials, cohort studies and case-control studies.(1-4)

The overall aim was addressed through five objectives. The first objective was to provide introductory guidance on the use of ITS for the evaluation of public health interventions. There was a clear lack of guidance on ITS prior to starting this thesis and this objective was therefore key to my overall aim of making ITS more accessible to researchers and public health professionals. Chapter 4 provides guidance on the design and analysis of ITS studies in an easily accessible step by step approach with suggested models, data layouts and code. The second objective was to develop a methodological framework for defining the impact model of an intervention. This was addressed in Chapter 5 and drew on examples from the case study analyses in Chapters 7 and 8. Here I highlighted that defining the impact of an intervention requires a clear specification of the counterfactual and of the impact model, both of which need to be specified a priori. This should take into account the specific intervention and outcomes under study as well as the context in which they are being studied and the data available. The third objective was to develop a methodological framework for the selection of controls and analysis of controlled interrupted time series. This was addressed in Chapter 6, in which a range of innovative types of controls were identified and categorised, and a structured approach was proposed for selecting controls based on their ability to limit the risk of history bias. The two primary case studies provide examples of very complex interventions for which it has been recognised, for differing reasons, that evaluation is challenging.(5, 6) This illustrates how ITS can be applied in settings where evaluation is not amenable to other methods, and still provide a robust analysis of the impact of the intervention. Other applications of ITS were utilised to illustrate particular methodological problems or solutions, in particular to illustrate different types of controls in Chapter 6 and issues with reporting in Chapter 9. The final objective was to develop guidance for transparent reporting of interrupted time series studies. This was, again, integral to my overall aims of reducing the risk of bias and making robust ITS more accessible. Chapter 9 proposes formal reporting criteria with a view to making the presentation of ITS studies more transparent and facilitating their appraisal in evidence syntheses.

In this final chapter, I will discuss some of the broad messages that have emerged from the thesis, put these advances into the context of other literature and discuss directions for future research.

10.1 Key messages

The discussions at the end of each paper and the "Contribution of the Paper" sections in each chapter discuss the main messages related to ITS from each paper. Nevertheless, there are some broader recurring themes throughout the thesis, many of which relate to the wider topic of evaluation study design. These are discussed below:

1) Start with evaluation in mind

An important message for public health professionals and policy makers who are responsible for introducing health interventions, is to begin with evaluation in mind. There is a clear gap in the application of evidence based public health when compared to evidence based medicine in clinical practice.(7) If decisions regarding public health interventions are to become more evidence based an expansion of robust evaluation of interventions is required. In order to facilitate this, interventions should be implemented with evaluation in mind from the start. First, a consideration of how the intervention is rolled out with, ideally, a pilot phase where the intervention is rolled out in certain groups but not others so that it can be used to conduct a cluster randomised controlled trial. Where this is not possible other robust designs can be considered at the implementation stage, including stepped wedge trials or multiple baseline ITS studies by introducing the intervention to different groups at different times. Second, ensure that the outcomes being targeted are explicit from the start and that there are measurable outcomes. A lack of clear primary outcome goals by the government was something that made the evaluation of the 2012 Health and Social Care Act in Chapter 5 particularly challenging.(8) This is often the case in policies for which not achieving the desired outcomes can be politically damaging and may require more cross-party collaboration.(5, 9) Finally, public health professionals and policy makers should consider the availability of baseline data on the outcomes of interest. For pre-post designs such as ITS, baseline data is an essential requirement and even for controlled designs, baseline data greatly strengthens the robustness of the analysis.(10) This means either ensuring well-validated routine data exists for the outcomes of interest, or capturing

new data on the outcomes prior to implementation of the intervention, ideally for a sustained period so that pre-existing trends can be incorporated into the analysis.

2) Plan the evaluation design and analysis a priori

Regardless of whether the evaluation was considered by those implementing the intervention, a continually emerging theme across this thesis is the need for researchers to carefully design the evaluation study *a priori*. This is true of all evaluation study designs but for ITS in particular it includes considering 'how is the baseline trend defined?', 'what is the counterfactual?', 'what is the most appropriate impact model?', 'what are the potential sources of history bias?'. The thesis has highlighted the dangers of making these decision based on examination of the data or statistical fit. Instead, the methodological papers have consistently recommended that researchers make these decisions prior to looking at the data and based on an objective consideration of the intervention and outcomes under study. Where there is uncertainty over the most appropriate approach, separate exploratory analyses should be undertaken or sensitivity analyses may be used.

3) Evaluation design is context dependent

Related to the need to plan the design and analysis *a priori*, is an acknowledgement that there is no one size fits all approach to evaluation design.(11) While the analytical approach to ITS is relatively straightforward, customising the design to complex interventions or events with a potential public health impact is complex. In Chapter 7 I highlighted the broad range of possible controls and how selecting these depends on the specific sources of potential history bias and the availability of appropriate data. Similarly, Chapter 8 highlighted how the most appropriate impact model varies depending on the type of intervention, how it was implemented and over what period, and the purported causal pathway between intervention and outcome. Expanding this concept to evaluation more broadly, it is clear that choice of design is very much context dependent, for example where researchers have control over who receives the intervention and when: a randomised controlled trial may be best, if the intervention is allocated according to a predefined threshold (such as above or below a certain age group): a regression discontinuity design may be most appropriate, and when the intervention is implemented at a clear point in time or where there is a lack of traditional control groups: an ITS design may be considered.

4) Make use of routine data

Although the first message highlighted the need for public health professionals and policy makers to clearly spell out their targeted outcomes and ensure baseline data is available, the reality is that this

is often not done. Fortunately, the availability and quality of routine data has improved dramatically over the past decade. Both of the case studies used in this thesis, as well as most of the examples in the methodological papers used routine data sources. This highlights the potential for robust evaluations of complex interventions and unplanned events using routine data. This is the case for ITS, but also for other quasi-experimental designs such as regression discontinuity designs and controlled before and after studies.

5) Use the strongest study design available

It is widely recognised that randomised controlled trials are not always possible, especially when evaluating past interventions or unplanned events.(12, 13) However, too often, evaluation of PH interventions uses weaker pre-post or cross-sectional designs.(14-16) These are subject to multiple inherent biases and perpetuate the idea that robust evaluation of complex PH interventions is not feasible. However, more robust quasi-experimental designs such as ITS can often be applied in these situations. Both case studies are good examples of the types of interventions that are challenging to evaluate and where most other study designs would not be possible, but where it is not necessary to resort to weaker designs that provide poorer evidence on the effectiveness of the intervention.

6) Think outside the box to strengthen design validity

There are many ways in which the validity of quasi-experimental designs such as ITS can be strengthened. Introducing controls is discussed in detail in Chapter 6, other approaches such as multiple baseline designs and multiple phase designs are discussed in Chapters 4, 6 and 9. ITS is often used in situations where traditional location based controls are not possible, and this may be precisely the reason that ITS is selected. Nevertheless, a broad range of other innovative and unconventional controls, such as control outcomes or control time periods were highlighted. Thinking outside the box about such options can enable researchers to control for threats to validity such as history bias or instrumentation in situations where this may otherwise not have been possible.

10.2 Advances in the literature

Over the course of the PhD there have been several advances in the ITS literature that it is important to put my findings into the context of. Perhaps most notable, has been the ongoing exponential increase in the use of ITS for evaluating public health interventions, as illustrated in Figure 1 of Chapter

9. This further justifies the need for methodological work to ensure that the strengths and limitations of ITS are well recognised and to ensure that ITS studies are designed robustly to minimise the risk of bias. Furthermore, the fact that more ITS evaluations are being published means that systematic reviewers will need to become familiar with how to appraise ITS studies in evidence syntheses. This emphasises the importance of reporting guidelines such as those proposed in Chapter 9.

Another important development has been the publication of studies testing the validity of the ITS design. I had originally considered empirically assessing the validity of ITS through a formal within study comparison against a RCT benchmark as one of the objectives of this thesis. When this was first proposed, only one previous comparison of ITS with RCT existed and there were limitations to the benchmark RCT used.(17) Nevertheless, since then, a number of within study comparisons have been published, including one (Fretheim et al 2015) that compared ITS to RCT using data from 8 different cluster-RCTs.(18-22) While it is important to consider the possibility of publication bias, these studies have repeatedly found that ITS is able to produce findings that are consistent with those of the RCT in a broad range of settings. These findings underpin the importance of making ITS accessible to researchers and public health professionals through introductory methodological articles (such as Chapter 4 of this thesis) and guidance (such as Chapter 9 of this thesis).

Significantly, given the emerging evidence on the strength of the validity of ITS, there have also been efforts to make robust evaluative design more accessible to public health professionals, policy makers and the public. In a pair of papers published through the US Centers for Disease Control and Prevention, Soumerai et al 2015 and Naci et al 2016 present the strengths and limitations of various evaluative study designs in an easily accessible format.(23, 24) They present a hierarchy of design in which CITS features alongside RCTs as "Strong designs" and single ITS as an "Intermediate design". They also provide an in depth discussion on the issue of history bias which complements the more in depth paper on how to address this in Chapter 6. Similarly, Kontopantelis et al 2015, present a methods paper introducing ITS in which they argue that ITS is the next best approach for evaluation of interventions when randomisation is not possible. This paper discusses the advantages and disadvantages of different ITS models with examples.(25) Again, this paper complements some of the work of this thesis in which first the practicalities of design and analysis of ITS are presented in a stepwise fashion (Chapter 4) and, second, a methodological framework for selecting the most appropriate model is proposed (Chapter 5).

Finally, as discussed in Chapter 9, Jandoc et al 2015, published a review of ITS studies in drug utilisation research along with proposed reporting recommendations for ITS studies in this field.(26) The fact that reporting guidelines have simultaneously been developed both in pharmacological research and

in public health underscores the increasing use of ITS and need for better reporting of the design in order to facilitate appraisal. My reporting guidelines differed in that they were based on the Transparent Reporting of Evaluations with Nonrandomised Designs (TREND) statement rather than Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. Furthermore, there is a greater emphasis on potential issues with routine data and the methodological work on controls and model selection from Chapters 5 and 6.

10.3 Directions for future research

The field of ITS methodology has advanced substantially in recent years; nevertheless, there remain areas for further development. First, although I have presented an in depth analysis on the use of controls, there is limited discussion on other approaches to address history bias, such as multiple baseline designs, whereby the intervention is introduced to different groups at different times, and multiple phase designs, whereby the intervention is introduced and then removed.(27, 28) Further work on where these may be best applied and how they should be designed is needed. Secondly, another, area for development is the use of ITS studies that allow for differing trends in different subgroups. I discussed in Chapter 8 how this may explain non-linear trends and lead to better definition of the baseline trend and therefore improve the modelling of the counterfactual. Thirdly, in chapter 10 I discussed the need for consensus guidelines on the reporting of ITS studies. Those that I have proposed aim to form a basis for wider agreement on a definitive statement for ITS studies.

Further work is also needed on sample size calculation for ITS studies and dealing with short data series or small expected effect sizes. In chapter 4, I highlighted that ITS studies can be underpowered under such circumstances. Only one paper that I am aware of has tried to address the issue of power calculation for ITS studies.(4) This was based on running simulations in order to estimate the number of time points that provide sufficient power and the distribution of these time points. Nevertheless, they do not address the sample size required at each time point. Because the power of an ITS study depends on such a broad range of factors it may be difficult to develop a conclusive formula for estimating the required sample size. Where researchers predict that effect sizes are likely to be small other approaches may be adopted such as using surrogate or intermediate outcomes. Intermediate or surrogate outcomes are outcomes that lie along the causal pathway. Examples include antibody levels as a surrogate for risk of disease following a vaccination intervention, smoking as an intermediate outcome for the effect of a smoking cessation campaign on lung cancer. Larger effects may be expected with intermediate outcomes, in particular if the final endpoint is rare or follows a

long lag.(5) Nevertheless, there can be problems with using intermediate outcomes in ITS studies, for example, data on intermediate outcomes is generally less likely to be available routinely: routine data on antibody levels is unlikely to be available in most settings, in particular if repeated measures are required as for ITS, routine data on smoking is also less easily and regularly available compared to data on rates of disease (such as hospital admissions or mortality). Furthermore, surrogates are not always good predictors of the final outcome, therefore care needs to be taken in selecting appropriate surrogates and interpreting the results.(6)In addition to the recommendations for ITS, other powerful quasi-experimental designs would benefit from more in depth methodological work and guidance. In particular, regression discontinuity designs (RDD) have wide applicability in public health in situations where thresholds exist, for example age thresholds for vaccination programmes and morbidity thresholds in screening programmes. Introductory methodological papers that present RDD to researchers and policy makers in an accessible manner may result in wider adoption of this method. Furthermore, reporting recommendations on RDD would facilitate their inclusion in systematic reviews and other evidence syntheses.

10.4 Conclusions

Through this thesis I have undertaken an in depth and critical analysis of the ITS design. Within study comparisons have clearly demonstrated that ITS, in particular when used in combination with a control series, are powerful designs.(18-22) I would therefore like to encourage researchers and especially public health practitioners to implement ITS as a robust method of evaluation of interventions and to urge a move away from weak before-after or cross-sectional evaluations. Nevertheless, evaluators need to be aware of the limitations of ITS in order to ensure that evaluations are as robust as possible. The papers presented in this thesis go some way towards increasing understanding of these limitations and how to address them, as well as disseminating this knowledge in an accessible way.

By improving understanding of the ITS design, I am hopeful that there will be an increase in evaluations of complex interventions that might otherwise be considered unevaluable. This is important in order to address the existing evaluative bias which means that there is a lack of evidence on more complex interventions. In addition to increasing the use of ITS, it is equally important that such studies are considered and adequately appraised in systematic reviews, guidelines development or directly in the generation of policies and public health interventions.

The work undertaken through this thesis benefits at least five groups of people: **1. Those who evaluate the effectiveness of public health interventions.** Greater awareness of the ITS design should lead to more robust evaluations, in particular in situations where RCTs are not possible. Beneficiaries in this group include academics, health think tanks such as the King's Fund and the Nuffield trust and, given that evaluation is a core aspect of developing interventions, public health practitioners and policymakers introducing new interventions. **2. Those who review evidence on the effectiveness of interventions and make recommendations on which interventions to implement.** The guidance developed on ITS methodology and reporting will help those who review evidence to appraise whether evaluations have been conducted appropriately and the level of evidence they provide regarding the effectiveness of an intervention. This will benefit systematic reviewers, such as the Cochrane Collaboration, and those who produce guidelines on public health interventions, such as the National Institute of Health and Clinical Excellence. **3. Those who implement public health interventions.** Public health practitioners are frequently frustrated by a lack of evidence on which interventions are effective to address a given health problem. Improved evaluations of public health interventions will enable public health practitioners, policymakers, health service managers and others who implement population level health interventions to make better informed decisions about which health policies and programmes to employ. **4. Those who pay for interventions.** Improved evaluation of interventions will also help to identify those interventions which are ineffective and thus help to limit inappropriate resource allocation. More efficient use of resources could benefit taxpayers, private organisations or third sector organisations that fund public health interventions. **5. Those who stand to benefit from public health interventions.** Ultimately, the most important beneficiaries are the patients and members of the public. Better informed decisions will result in the implementation of more effective health interventions which lead to the greatest improvements in population health.

10.5 References

1. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.
2. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med. 2007;45(4):247-51.
3. Jadad AR. Randomised Controlled Trials: A User's Guide. Health Technol Assess. 1998;2(13):214.
4. Coggon D, Barker D, Rose G. Epidemiology for the Uninitiated: John Wiley & Sons; 2009.
5. Vittal Katikireddi S, McKee M, Craig P, Stuckler D. The NHS reforms in England: four challenges to evaluating success and failure. Journal of the Royal Society of Medicine. 2014;107(10):387-92.
6. Stuckler D, Reeves A, Karanikolos M, McKee M. The health effects of the global financial crisis: can we reconcile the differing views? A network analysis of literature across disciplines. Health Economics, Policy and Law. 2014;10(1):83-99.
7. Heller R, Page J. A population perspective to evidence based medicine:"evidence for population health". JECH. 2002;56(1):45-7.
8. 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.
9. Sheikh A. Evidence-based restructuring of health and social care. PLOS Medicine. 2017;14(11):e1002426.
10. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2016.
11. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. BMJ : British Medical Journal. 2007;334(7591):455-9.
12. Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. JECH. 2009;65(7):582-7.
13. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996;312(7040):1215.
14. Coulter A, Bradlow J. Effect of NHS reforms on general practitioners' referral patterns. BMJ. 1993;306(6875):433-7.
15. Elhayany A, Regev S, Sherf M, Reuveni H, Shvartzman P. Effects of a fundholding discontinuation. An Israeli health maintenance organization natural experiment. Scandinavian journal of primary health care. 2001;19(4):223-6.
16. Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic Health Records and Quality of Diabetes Care. N Engl J Med. 2011;365(9):825-33.
17. Schneeweiss S, Maclure M, Carleton B, Glynn RJ, Avorn J. Clinical and economic consequences of a reimbursement restriction of nebulised respiratory therapy in adults: direct comparison of randomised and observational evaluations. BMJ. 2004;328(7439):560.
18. Hallberg K, editor Assessing the Validity of Comparative Interrupted Time Series Designs and Practice: Lessons from Two within Study Comparisons. 2015 Fall Conference: The Golden Age of Evidence-Based Policy; 2015: Appam.
19. 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.

20. St.Clair T, Hallberg K, Cook TD. The Validity and Precision of the Comparative Interrupted Time-Series Design: Three Within-Study Comparisons. Journal of Educational and Behavioral Statistics. 2016.

21. Fretheim A, Soumerai SB, Zhang F, Oxman AD, Ross-Degnan D. Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. J Clin Epidemiol. 2013;66(8):883-7.

22. Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, et al. A re-analysis of cluster randomized trials showed Interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol. 2014.

23. 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.

24. Naci H, Soumerai SB. History Bias, Study Design, and the Unfulfilled Promise of Pay-for-Performance Policies in Health Care. Preventing Chronic Disease. 2016;13:E82.

25. 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.

26. Jandoc R, Burden AM, Mamdani M, Lévesque LE, Cadarette SM. Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. J Clin Epidemiol. 2015;68(8):950-6.

27. 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.

28. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci. 2000;1(1):31-49.

11. Appendices

11.1 Methodological paper 1 appendices

11.1.1 Example dataset

11.1.2 Stata code

11.1.3 R code

11.1.4 Further model specifications

11.1.5 Contribution of the candidate to the paper

11.1.1 Supplementary appendix 1: Example dataset*

year	month	aces	time	smokban	pop	stdpop
2002	1	728	1	0	364277.4	379875.3
2002	2	659	2	0	364277.4	376495.5
2002	3	791	3	0	364277.4	377040.8
2002	4	734	4	0	364277.4	377116.4
2002	5	757	5	0	364277.4	377383.4
2002	6	726	6	0	364277.4	374113.1
2002	7	760	7	0	364277.4	379513.3
2002	8	740	8	0	364277.4	376295.5
2002	9	720	9	0	364277.4	374653.2
2002	10	814	10	0	364277.4	378485.6
2002	11	795	11	0	364277.4	375955.5
2002	12	858	12	0	364277.4	378349.7
2003	1	887	13	0	363350.8	376762.4
2003	2	766	14	0	363350.8	379032.3
2003	3	851	15	0	363350.8	379360.4
2003	4	769	16	0	363350.8	376162
2003	5	781	17	0	363350.8	377972.4
2003	6	756	18	0	363350.8	381830.7
2003	7	766	19	0	363350.8	379888.6
2003	8	752	20	0	363350.8	380872.2
2003	9	765	21	0	363350.8	380966.9
2003	10	831	22	0	363350.8	381240.4
2003	11	879	23	0	363350.8	382104.9
2003	12	928	24	0	363350.8	381802.7
2004	1	914	25	0	364700.4	381656.3
2004	2	808	26	0	364700.4	383680
2004	3	937	27	0	364700.4	383504.2
2004	4	840	28	0	364700.4	386462.9
2004	5	916	29	0	364700.4	383783.1
2004	6	828	30	0	364700.4	380836.8
2004	7	845	31	0	364700.4	383483
2004	8	818	32	0	364700.4	380906.2
2004	9	860	33	0	364700.4	382926.8
2004	10	839	34	0	364700.4	384052.4
2004	11	887	35	0	364700.4	384449.6
2004	12	886	36	0	364700.4	383428.4
2005	1	831	37	1	364420.8	388153.2
2005	2	796	38	1	364420.8	388373.2
2005	3	833	39	1	364420.8	386470.1
2005	4	820	40	1	364420.8	386033.2
2005	5	877	41	1	364420.8	383686.4
2005	6	758	42	1	364420.8	385509.3
2005	7	767	43	1	364420.8	385901.9
2005	8	738	44	1	364420.8	386516.6
2005	9	781	45	1	364420.8	388436.5
2005	10	843	46	1	364420.8	383255.2
2005	11	850	47	1	364420.8	390148.7
2005	12	908	48	1	364420.8	385874.9
2006	1	1021	49	1	363832.6	391613.6
2006	2	859	50	1	363832.6	391750.4

2006	3	976	51	1	363832.6	394005.6
2006	4	888	52	1	363832.6	391364.9
2006	5	962	53	1	363832.6	391664.6
2006	6	838	54	1	363832.6	389022.3
2006	7	810	55	1	363832.6	391878.5
2006	8	876	56	1	363832.6	388575.3
2006	9	843	57	1	363832.6	392989
2006	10	936	58	1	363832.6	390018.8
2006	11	912	59	1	363832.6	390712.3

*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:e17419.

11.1.2 Supplementary appendix 2: interrupted time series tutorial Stata code

```
***********************************************************************
* This file provides the Stata code used for the analysis of the example dataset
* used in the paper:
* Interrupted time series regression for the evaluation of public health
* interventions: a tutorial
* IJE 2016
* J. Lopez Bernal, S. Cummins, A. Gasparrini
***********************************************************************


clear
set more off
capture log close


***************


insheet using "sicily.csv", comma
*or: import delimited "sicily.csv"

/* This dataset includes the following variables
year
month
time = elapsed time since the start of the study
aces = count of acute coronary episodes in Sicily per month (the outcome)
smokban = smoking ban (the intervention) coded 0 before the intervention and 1 after
pop = the population of Sicily (in 10000s)
stdpop =  age standardised population
*/



*********************************************
*Step 3: Descriptive analyses
*********************************************
/* Examining the data is an important first step. Looking at the pre-intervention trend can give an
indication of how stable the trend is over time, whether a linear model is likely to be approproate
and whether there appears to be a seasonal trend */


*Here we convert the counts into a rate and examine a scatter plot of the pre-intervention data
gen rate = aces/stdpop*10^5
twoway (scatter rate time) if smokban==0, title("Sicily, 2002-2006") ytitle(Std rate x 10000)
yscale(range(0 .)) ylabel(#5, labsize(small) angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006", noticks labsize(small))
xtitle(year)


*It is also useful to produce summary statistics for before and after the intervention
summ, detail
```

bysort smokban: summ aces
bysort smokban: summ rate

```
*************************************************
*Step 4: Poisson regression model
*************************************************
/* In step 2 (main paper) we chose a step change model and we also use a Poisson model as we are
using count data
In order to do this we model the count data directly (rather than the rate which doesn't follow a
Poisson distribution)
We then use the population (log transformed) as an offset variable in order to transform back to rates
*/


*log transform the standardised population:
gen logstdpop = log(stdpop)


*Poisson with the outcome (aces), intervention (smokban) and time as well as the population offset
offset
glm aces smokban time, family(poisson) link(log) offset(logstdpop) eform


*We generate predicted values based on the model in order to create a plot of the model:
predict pred, nooffset

*This can then be plotted along with a scatter graph:
gen rate1 = aces/stdpop /*to put rate in same scale as count in model */
twoway (scatter rate1 time) (line pred time, lcolor(red)) , title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small) angle(horizontal)) ///
xtick(0.5(12)60.5)  xlabel(6"2002"  18"2003"  30"2004"  42"2005"  54"2006",  noticks  labsize(small))
xtitle(year) ///
xline(36.5)


*Generate the counterfactual by removing the effect of the intervention (_b[smokban]) for the post-
intervention period
gen pred1 = pred/exp(_b[smokban]) if smokban==1


*Add the counterfactual to the plot
twoway (scatter rate1 time) (line pred time, lcolor(red)) (line pred1 time, lcolor(red) lpattern(dash)),
title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small) angle(horizontal)) ///
xtick(0.5(12)60.5)  xlabel(6"2002"  18"2003"  30"2004"  42"2005"  54"2006",  noticks  labsize(small))
xtitle(year) ///
xline(36.5)
```

```
************************************************
*Step 5: methodological issues
************************************************
* (a) Allowing for overdispersion
/*In the model above we have not allowed for overdispersion - in order to do this we can add
the scale(x2) parameter to the model which allows the variance to be proportional rather than
equal to the mean */
glm aces smokban time, family(poisson) link(log) offset(logstdpop) scale(x2) eform

* (b) Model checking and autocorrelation
*Check the residuals by plotting against time
predict res, r
twoway (scatter res time)(lowess res time),yline(0)

*Further check for autocorrelation by examining the autocorrelation and partial autocorrelation
functions
tsset time
ac res
pac res, yw

* (c) Adjust for seasonality
/* installation of the "circular" package. o find packages select Help > SJ and User-written Programs,
and click on search */

*we need to create a degrees variable for time divided by the number of time points in a year (i.e. 12
for months)
gen degrees=(time/12)*360

*we then select the number of sine/cosine pairs to include:
fourier degrees, n(2)

*these can then be included in the model
glm aces smokban cos* sin* time, family(poisson) link(log) offset(logstdpop) scale(x2) eform

*we can again check for autocorrelation
predict res2, r
twoway (scatter res2 time)(lowess res2 time),yline(0)
tsset time
ac res2
pac res2, yw


*predict and plot of seasonally adjusted model**
predict pred2, nooffset
twoway (scatter rate1 time) (line pred2 time, lcolor(red)), title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small) angle(horizontal)) ///
xtick(0.5(12)60.5)  xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006", noticks labsize(small))
xtitle(year) ///
```

xline(36.5)

/*it is sometimes difficult to clearly see the change graphically in the seasonally adjusted model therefore it can be useful to plot a straight line as if all months were the average to produce a 'deseasonalised' trend. */

egen avg_cos_1 = mean(cos_1)
egen avg_sin_1 = mean(sin_1)
egen avg_cos_2 = mean(cos_2)
egen avg_sin_2 = mean(sin_2)

drop cos* sin*

rename avg_cos_1 cos_1
rename avg_sin_1 sin_1
rename avg_cos_2 cos_2
rename avg_sin_2 sin_2

*this can then be added to the plot as a dashed line
predict pred3, nooffset

twoway (scatter rate1 time) (line pred2 time, lcolor(red)) (line pred3 time, lcolor(red) lpattern(dash)), title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small) angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006", noticks labsize(small)) xtitle(year) ///
xline(36.5)

****
** additional material
****
***add a change in slope

*generate interaction term between intervention and time centered at the time of intervention
gen inter_smokbantime = smokban*(time-36)

*restore fourier variables that were previously changed
drop cos* sin* degrees
gen degrees=(time/12)*360
fourier degrees, n(2)

*add the interaction term to the model
glm aces smokban inter_smokbantime cos* sin* time, family(poisson) link(log) offset(logstdpop) scale(x2) eform
*(the coefficient and CI for the interaction term suggests that there is very little slope change)

187

```
*plot seasonally adjusted model with deseasonalised trend**
predict pred4, nooffset

egen avg_cos_1 = mean(cos_1)
egen avg_sin_1 = mean(sin_1)
egen avg_cos_2 = mean(cos_2)
egen avg_sin_2 = mean(sin_2)
drop cos* sin*
rename avg_cos_1 cos_1
rename avg_sin_1 sin_1
rename avg_cos_2 cos_2
rename avg_sin_2 sin_2

predict pred5, nooffset

twoway (scatter rate1 time) (line pred4 time, lcolor(red)) (line pred5 time, lcolor(red) lpattern(dash)),
title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small) angle(horizontal)) ///
xtick(0.5(12)60.5)  xlabel(6"2002"  18"2003"  30"2004"  42"2005"  54"2006",  noticks  labsize(small))
xtitle(year) ///
xline(36.5)
```

11.1.3 Supplementary appendix 2: interrupted time series tutorial R code

```
################################################################################
# This file provides the R code used for the analysis of example dataset used
# used in the paper:
#   Interrupted time series regression for the evaluation of public health
#     interventions: a tutorial
#   IJE 2016
#   J. Lopez Bernal, S. Cummins, A. Gasparrini
################################################################################


# Install packages required for the analysis (uncomment if needed)
#install.packages("lmtest") ; install.packages("Epi")
#install.packages("tsModel"); install.packages("vcd")

# load the packages
library(foreign) ; library(tsModel) ; library("lmtest") ; library("Epi")
library("splines") ; library("vcd")

# read data from csv file
data <- read.csv("sicily.csv")
head(data)
View(data)

# This dataset includes the following variables:
# year
# month
# time = elapsed time since the start of the study
# aces = count of acute coronary episodes in Sicily per month (the outcome)
# smokban = smoking ban (the intervention) coded 0 before intervention, 1 after
# pop = the population of Sicily (in 10000s)
# stdpop =  age standardised population



################################################################################
#Step 3: Descriptive analyses
#######################################
# Examining the data is an important first step
# Looking at the pre-intervention trend can give an indication of how stable the
#   trend is over time, whether a linear model is likely to be appropriate, and
#   whether there appears to be a seasonal trend

## Scatter plot

# compute the standardized rates
data$rate <- with(data, aces/stdpop*10^5)
# start the plot, excluding the points and the x-axis
plot(data$rate,type="n",ylim=c(00,300),xlab="Year", ylab="Std rate x 10,000",
  bty="l",xaxt="n")
# shade the post intervention period grey
rect(36,0,60,300,col=grey(0.9),border=F)
```

```
# plot the observed rate for pre-intervention period
points(data$rate[data$smokban==0],cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:5*12,labels=F)
axis(1,at=0:4*12+6,tick=F,labels=2002:2006)
# add a title
title("Sicily, 2002-2006")

# It is also useful to produce summary statistics
summary(data)

#tabulate aces before and after the smoking ban
summary(data$aces[data$smokban==0])
summary(data$aces[data$smokban==1])

summary(data$rate[data$smokban==0])
summary(data$rate[data$smokban==1])


#############################################################################
#Step 4: Poisson regression model
########################################
# In step 2 (main paper) we chose a step change model and we also used a Poisson
#   model as we are using count data
# In order to do this we model the count data directly (rather than the rate
#   which doesn't follow a Poisson distribution), using the population (log
#   transformed) as an offset variable in order to transform back to rates

#Poisson with the standardised population as an offset
model1 <- glm(aces ~ offset(log(stdpop)) + smokban + time, family=poisson, data)
summary(model1)
summary(model1)$dispersion
round(ci.lin(model1,Exp=T),3)

# create a new dataframe with 0.1 time units to improve the graph
datanew <- data.frame(stdpop=mean(data$stdpop),smokban=rep(c(0,1),c(360,240)),
  time= 1:600/10,month=rep(1:120/10,5))

# We generate predicted values based on the model in order to create a plot
pred1 <- predict(model1,type="response",datanew)/mean(data$stdpop)*10^5

#This can then be plotted along with a scatter graph (see above)
plot(data$rate,type="n",ylim=c(0,300),xlab="Year",ylab="Std rate x 10,000",
  bty="l",xaxt="n")
rect(36,0,60,300,col=grey(0.9),border=F)
points(data$rate,cex=0.7)
axis(1,at=0:5*12,labels=F)
axis(1,at=0:4*12+6,tick=F,labels=2002:2006)
lines((1:600/10),pred1,col=2)
title("Sicily, 2002-2006")
```

```
# to plot the counterfactual scenario we create a data frame as if smokban
#   (the intervention) was never being implemented
datanew <- data.frame(stdpop=mean(data$stdpop),smokban=0,time=1:600/10,
  month=rep(1:120/10,5))

# generate predictions under the counterfactual scenario and add it to the plot
pred1b <- predict(model1,datanew,type="response")/mean(data$stdpop)*10^5
lines(datanew$time,pred1b,col=2,lty=2)

# return the data frame to the scenario including the intervention
datanew <- data.frame(stdpop=mean(data$stdpop),smokban=rep(c(0,1),c(360,240)),
  time= 1:600/10,month=rep(1:120/10,5))


###############################################################################
#Step 5: methodological issues
############################################################################

#a) Overdispersion: Quasi-Poisson model
# In the model above we have not allowed for overdispersion - in order to do
#   this we can use a quasipoisson model, which allows the variance to be
#   proportional rather than equal to the mean

model2 <- glm(aces ~ offset(log(stdpop)) + smokban + time, family=quasipoisson,
  data)
summary(model2)
summary(model2)$dispersion
round(ci.lin(model2,Exp=T),3)

#b) Model checking and autocorrelation

# Check the residuals by plotting against time
res2 <- residuals(model2,type="deviance")
plot(data$time,res2,ylim=c(-5,10),pch=19,cex=0.7,col=grey(0.6),
  main="Residuals over time",ylab="Deviance residuals",xlab="Date")
abline(h=0,lty=2,lwd=2)

# Further check for autocorrelation by examining the autocorrelation and
#   partial autocorrelation functions
acf(res2)
pacf(res2)

#c) adjusting for seasonality
# There are various ways of adjusting for seasonality - here we use harmonic
#   terms specifying the number of sin and cosine pairs to include (in this
#   case 2) and the length of the period (12 months)
model3 <- glm(aces ~ offset(log(stdpop)) + smokban + time +
  harmonic(month,2,12), family=quasipoisson, data)
summary(model3)
summary(model3)$dispersion
round(ci.lin(model3,Exp=T),3)
```

```
# EFFECTS
ci.lin(model3,Exp=T)["smokban",5:7]

# TREND
exp(coef(model3)["time"]*12)

# We again check the model and autocorrelation functions
res3 <- residuals(model3,type="deviance")
plot(res3,ylim=c(-5,10),pch=19,cex=0.7,col=grey(0.6),main="Residuals over time",
  ylab="Deviance residuals",xlab="Date")
abline(h=0,lty=2,lwd=2)
acf(res3)
pacf(res3)

# predict and plot of the seasonally adjusted model
pred3 <- predict(model3,type="response",datanew)/mean(data$stdpop)*10^5
plot(data$rate,type="n",ylim=c(120,300),xlab="Year",ylab="Std rate x 10,000",
  bty="l",xaxt="n")
rect(36,120,60,300,col=grey(0.9),border=F)
points(data$rate,cex=0.7)
axis(1,at=0:5*12,labels=F)
axis(1,at=0:4*12+6,tick=F,labels=2002:2006)
lines(1:600/10,pred3,col=2)
title("Sicily, 2002-2006")

# it is sometimes difficult to clearly see the change graphically in the
#   seasonally adjusted model, therefore it can be useful to plot a straight
#   line representing a 'deseasonalised' trend
# this can be done by predicting all the observations for the same month, in
#   this case we use June
pred3b <- predict(model3,type="response",transform(datanew,month=6))/
  mean(data$stdpop)*10^5

#this can then be added to the plot as a dashed line
lines(1:600/10,pred3b,col=2,lty=2)


################################################################################
# additional material
####################################################################

# add a change-in-slope
# we parameterize it as an interaction between time and the ban indicator
model4 <- glm(aces ~ offset(log(stdpop)) + smokban*time + harmonic(month,2,12),
  family=quasipoisson, data)
summary(model4)
round(ci.lin(model4,Exp=T),3)

# predict and plot the 'deseasonalised' trend
# compare it with the step-change only model
```

```
pred4b <- predict(model4,type="response",transform(datanew,month=6))/
  mean(data$stdpop)*10^5
plot(data$rate,type="n",ylim=c(120,300),xlab="Year",ylab="Std rate x 10,000",
  bty="l",xaxt="n")
rect(36,120,60,300,col=grey(0.9),border=F)
points(data$rate,cex=0.7)
axis(1,at=0:5*12,labels=F)
axis(1,at=0:4*12+6,tick=F,labels=2002:2006)
lines(1:600/10,pred3b,col=2)
lines(1:600/10,pred4b,col=4)
title("Sicily, 2002-2006")
legend("topleft",c("Step-change only","Step-change + change-in-slope"),lty=1,
  col=c(2,4),inset=0.05,bty="n",cex=0.7)

# test if the change-in-slope improve the fit
# the selected test here is an F-test, which accounts for the overdispersion,
#   while in other cases a likelihood ratio or wald test can be applied
anova(model3,model4,test="F")
# not surprisingly, the p-value is similar to that of the interaction term

#
```

11.1.4 Supplementary appendix 4: Further model specifications

Slope change following a lag (figure 2 (d)):

Here, the start of the intervention can be coded as 1 following the lag (rather than immediately after the intervention). For example, in Table 5.1, if we assume a 6 month lag and the intervention was introduced in January 2005, we code the intervention as 1 from July 2005. We then use the usual slope change regression model: $Y_t = \beta_0 + \beta_1 T + \beta_3 TX_t$ with β_3 representing the slope change.

Table 5.1: variable specification for slope change following a lag

Year	Month	Time elapsed (T)	Smoking ban lagged (X)	ACEs (Y)	Std popn
2004	1	25	0	914	381656.3
2004	2	26	0	808	383680
2004	3	27	0	937	383504.2
2004	4	28	0	840	386462.9
2004	5	29	0	916	383783.1
2004	6	30	0	828	380836.8
2004	7	31	0	845	383483
2004	8	32	0	818	380906.2
2004	9	33	0	860	382926.8
2004	10	34	0	839	384052.4
2004	11	35	0	887	384449.6
2004	12	36	0	886	383428.4
2005	1	37	0	831	388153.2
2005	2	38	0	796	388373.2
2005	3	39	0	833	386470.1
2005	4	40	0	820	386033.2
2005	5	41	0	877	383686.4
2005	6	42	0	758	385509.3
2005	7	43	1	767	385901.9
2005	8	44	1	738	386516.6
2005	9	45	1	781	388436.5
2005	10	46	1	843	383255.2
2005	11	47	1	850	390148.7
2005	12	48	1	908	385874.9

Temporary level change (figure 2 (e)):

This model may be used where a reversible intervention is introduced temporarily. For example, if the intervention were introduced for six months, then withdrawn, the variables may be specified as per Table 5.2. A level change regression model could then be run as follows: $Y_t = \beta_0 + \beta_1 T + \beta_2 X_t +$

$\beta_3 W_t$ where β_2 represents the level change at introduction and β_3 represents the remaining level change at withdrawal. Under the assumption that the latter is null, and that the level comes back to the underlying trend, $\beta_3 W_t$ can be excluded.

Table 5.2: variable specification for a temporary level change or temporary slope change

Year	Month	Time elapsed (T)	Smoking ban introduced (X)	Smoking ban withdrawn (W)	ACEs (Y)	Std popn
2004	1	25	0	0	914	381656.3
2004	2	26	0	0	808	383680
2004	3	27	0	0	937	383504.2
2004	4	28	0	0	840	386462.9
2004	5	29	0	0	916	383783.1
2004	6	30	0	0	828	380836.8
2004	7	31	0	0	845	383483
2004	8	32	0	0	818	380906.2
2004	9	33	0	0	860	382926.8
2004	10	34	0	0	839	384052.4
2004	11	35	0	0	887	384449.6
2004	12	36	0	0	886	383428.4
2005	1	37	1	0	831	388153.2
2005	2	38	1	0	796	388373.2
2005	3	39	1	0	833	386470.1
2005	4	40	1	0	820	386033.2
2005	5	41	1	0	877	383686.4
2005	6	42	1	0	758	385509.3
2005	7	43	0	1	767	385901.9
2005	8	44	0	1	738	386516.6
2005	9	45	0	1	781	388436.5
2005	10	46	0	1	843	383255.2
2005	11	47	0	1	850	390148.7
2005	12	48	0	1	908	385874.9

Temporary slope change leading to a level change (Figure 2(f)):

This model could be used to represent a phase during which an intervention was gradually introduced in which case we may be interested in the slope change as the intervention was phased in, as well as the absolute level change following its introduction. The variable specification in Table 1 (main text) can be used with the following regression model: $Y_t = \beta_0 + \beta_1 T + \beta_2 X_t$. For this model if we assume the temporary slope change over 5 months we can use the variable specification in table 5.3, where β_2 represents the full change and X_t is modified accordingly to represent partial changes along time.

Table 5.3: variable specification for a temporary level change or temporary slope change

Year	Month	Time elapsed (T)	Intervention (X)	ACEs (Y)	Std popn
2004	1	25	0	914	381656.3
2004	2	26	0	808	383680
2004	3	27	0	937	383504.2
2004	4	28	0	840	386462.9
2004	5	29	0	916	383783.1
2004	6	30	0	828	380836.8
2004	7	31	0	845	383483
2004	8	32	0	818	380906.2
2004	9	33	0	860	382926.8
2004	10	34	0	839	384052.4
2004	11	35	0	887	384449.6
2004	12	36	0	886	383428.4
2005	1	37	0.2	831	388153.2
2005	2	38	0.4	796	388373.2
2005	3	39	0.6	833	386470.1
2005	4	40	0.8	820	386033.2
2005	5	41	1	877	383686.4
2005	6	42	1	758	385509.3
2005	7	43	1	767	385901.9
2005	8	44	1	738	386516.6
2005	9	45	1	781	388436.5
2005	10	46	1	843	383255.2
2005	11	47	1	850	390148.7
2005	12	48	1	908	385874.9

11.1.5 Contribution of the candidate to the paper

- I identified the need for an introduction to segmented regression analysis as a gap in the literature.
- I decided on the structure of the paper as a tutorial with worked example and appropriate control.
- The bulk of the work in this contribution consisted of identifying, cleaning and analysing the dataset, including:
  - Identifying a dataset and intervention and within this selecting a suitable subpopulation that could be used to illustrate a basic ITS analysis.
  - Identifying a dataset with a seasonal pattern so that I could demonstrate approaches to adjusting for seasonality
  - Ensuring that analysis could be undertaken with the dataset that did not deviate too far from a basic ITS model as this was intended as an introductory paper.
- I cleaned the data and reshaped into time series format in order that readers could understand the appropriate data layout.
- I then selected a range of different impact models and presented data layouts and regression models for each of these (appendix 11.1.4).
- Finally, I wrote and appropriately annotated code for the analysis in both R and Stata to clearly describe how to undertake each of the steps involved in ITS analysis in an appropriate statistical package.
- I sought feedback from my co-authors throughout this process and presented various drafts for comment and revised accordingly. I also checked the R code with AG and adapted this as necessary.

11.2 Methodological paper 2 appendices

11.2.1 Contribution of the candidate to the paper

11.2.1 Contribution of the candidate to the paper

- I initiated the idea for this paper after recognising the range of decisions I was making in deciding on the most suitable way to model both the counterfactual and the impact model in each of my case studies.
- This is not an area that had been covered in the existing literature, but it is clearly key to the appropriate design and analysis of interrupted time series studies.
- For this study I went through the various outcomes and sub-populations that I used within my case studies to identify examples that would demonstrate different impacts depending on how the models were defined.
- One of the decisions I had made in the second case study was to restrict the specialists visits data due to data quality issues. I decided to examine how the counterfactual might have been modelled if we had not taken the data quality issues into account and also trialled various linear and non-linear models to demonstrate that these can produce different results and that such modelling decisions should be made a priori.
- I also wanted to demonstrate how inappropriate impact models could show effects irrespective of the intervention. For this I selected one of my control populations which I knew had not received the intervention and analysed the data using various different models. Here it was clear that the best fitting model showed an effect but we knew that this could not be due to the intervention (as it was non-existent). This nicely demonstrated the need to select the impact model according to a clear a priori understanding of the intervention and outcome under study, rather than post-hoc model fit.
- I undertook each of these analyses myself and wrote the paper and produced the frameworks that I propose.
- Again, I sought feedback from my co-authors throughout this process and they made suggestions regarding additional modelling aspects to consider and the paper went through various revisions.

11.3 Methodological paper 3 appendices

11.3.1   Web appendix 1: Segmented regression interaction model for a controlled interrupted time series

11.3.2 Contribution of the candidate to the paper

11.3.1 Web appendix 1: Segmented regression interaction model for a controlled interrupted time series
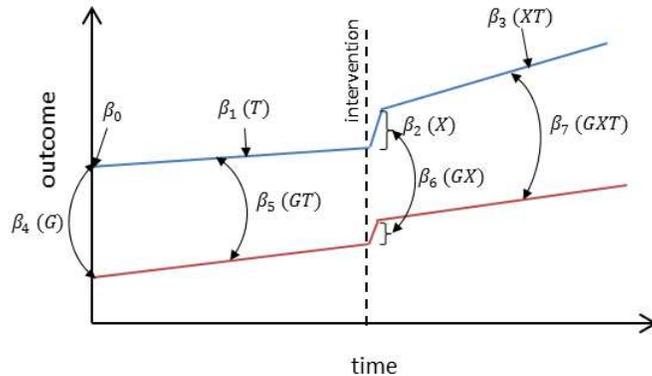

Figure 1: Segmented regression interaction model for a controlled interrupted time series

Intervention group in blue, control group in red. T = time since the start of the study, X = intervention (pre-intervention period = 0, post-intervention period = 1), G = group (control group = 0, intervention group = 1). β0, β2, β4 and β6 relate to intercepts, β1, β3, β5 and β7 relate to slopes. Curved arrows represent differences between the intervention group and control group. (Adapted from Linden and Adams 2011) [1]

Segmented regression equation for slope change with a control series:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 TX_t + \beta_4 G + \beta_5 GT + \beta_6 GX_t + \beta_7 GX_t T$$

Y_t is the outcome variable at time t, T is a variable representing the time since the start of the study and X is a dummy variable indicating the pre- or post-intervention period. G now represents the intervention group ($G = 1$) or control group ($G = 0$). Here β_4 represents the difference in intercept at T=0, β_5 represents the slope difference between the intervention and control group in the pre-intervention period, β_6 represents the difference between the change in level in the control and intervention group associated with the intervention, β_7 represents the difference between the change in slope in the control and intervention group associated with the intervention (Figure 4). Therefore β_6 and β_7 are the parameters of interest for the measures of effect.

Linden, A. and J.L. Adams, *Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation.* J Eval Clin Pract, 2011. **17**(6): p. 1231-8.

11.3.2 Contribution of the candidate to the paper

- The idea for this paper was initiated by me based on a clear gap in the literature.
- I recognised the use of controls as one of the most practical and robust ways for dealing with history bias, the major threat to the validity of ITS. Nevertheless, there was a lack of guidance on how to select controls and design a CITS study.
- I identified and classified 6 different types of controls that can be used in CITS studies (previously only location base controls and control outcomes had been described in the methodological literature).
- I also developed the framework for selecting controls and analysing CITS studies and described where covariate imbalance can bias CITS studies.
- As with the previous methodological studies, much of the work consisted of identifying and exploring a suitable dataset and within this a subpopulation that could clearly demonstrate the concepts that I wanted to introduce.
  - The LANTERNS dataset was a very large dataset and I chose this because of the large number of possible controls and because these included at least three different types of controls from my classification.
  - I found a subpopulation in which the intervention had been introduced around the same time in multiple roads and in which an effect was visible in the uncontrolled ITS.
  - I then used my newly developed framework to select a control and analyse the CITS study.
  - Subsequently I repeated this analysis but this time selecting a control based on covariate balance.
  - This allowed me to demonstrate the importance of first considering potential sources of history bias and whether the control can exclude these.
- The paper was written by me with comments on drafts from my co-authors and subsequent revisions.

11.4 Case study 1 appendices

11.4.1 Gross domestic product growth in Spain

11.4.2 Unemployment in Spain

11.4.3 Trend in monthly mortality rate from accidental falls (control) for all of Spain before and since the financial crisis

11.4.4 Trend in monthly suicide rates before and since the financial crisis by area

11.4.5 Trend in monthly suicide rates before and since the financial crisis by sex

11.4.6 Trend in monthly suicide rates before and since the financial crisis by age group

11.4.7 Time series plots of trends in monthly suicide rates for all of Spain based on the alternative models used in sensitivity analysis

11.4.8 Unemployment rates during the first quarter of 2005 and the last quarter of 2010

11.4.9 Contribution of the candidate to the paper

11.4.1 Web Appendix 1: Gross domestic product growth rate in Spain compared to previous quarter, seasonally adjusted, 2002 to 2012
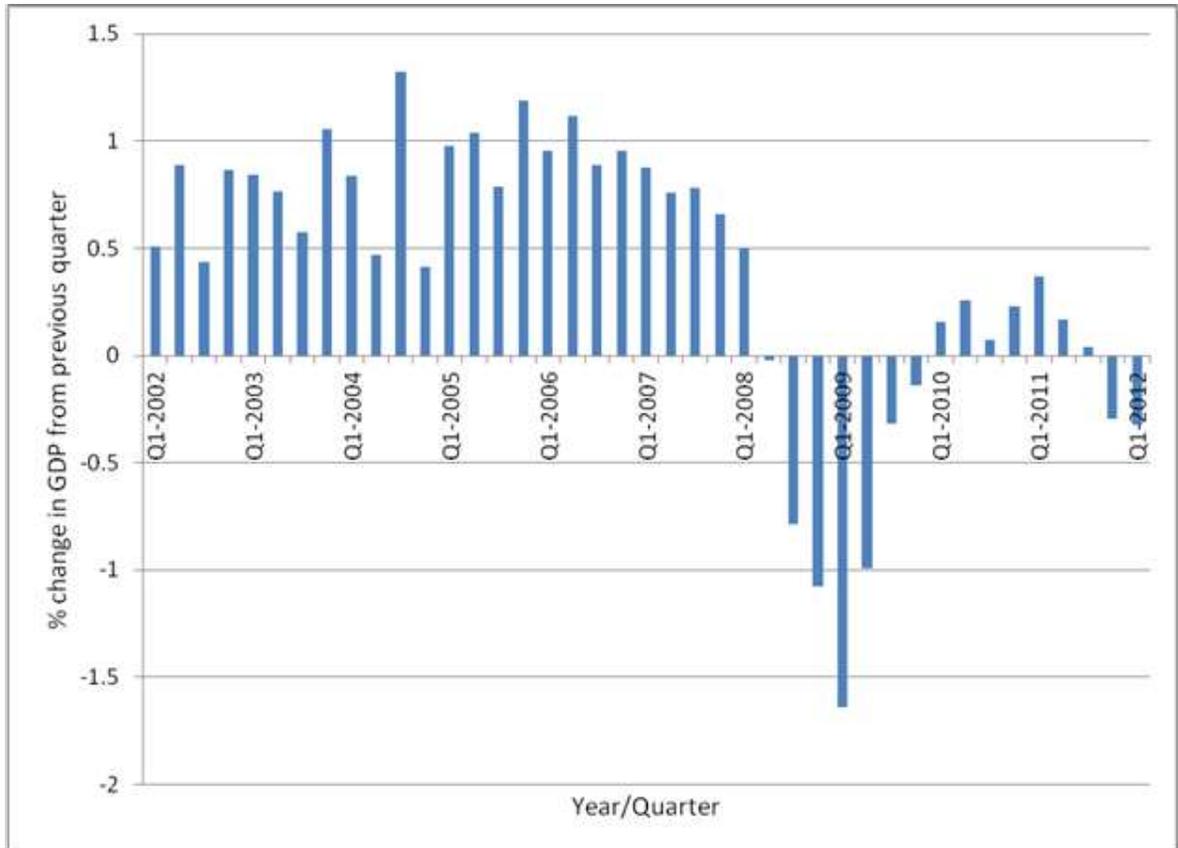
Source: Organisation for Economic Cooperation and Development (OECD) †

†Organisation for Economic Cooperation and Development. *OECD.StatExtracts*. 2012 [cited 2012 04/07/2012]; Available from: http://stats.oecd.org/Index.aspx?QueryName=350.

11.4.2 Web Appendix 2: Unemployment rate in Spain 2005 to 2012
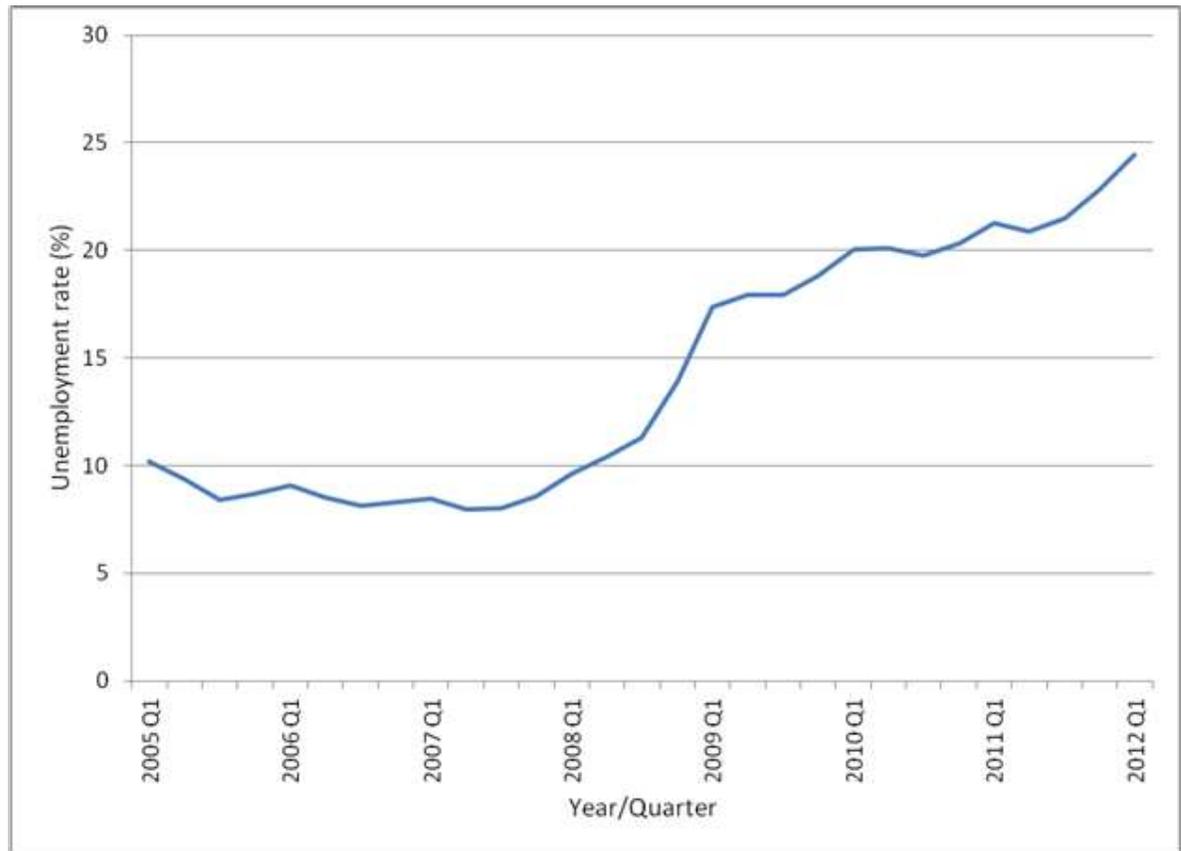
Source: Instituto Nacional de Estadística [National Statistics Institute] †

†INE. *Instituto Nacional de Estadística [National Statistics Institute].*   [cited 2012 04/07/2012]; Available from: http://www.ine.es/.

11.4.3 Web Appendix 3: Trend in monthly mortality rate from accidental falls (control) for all of Spain before and since the financial crisis
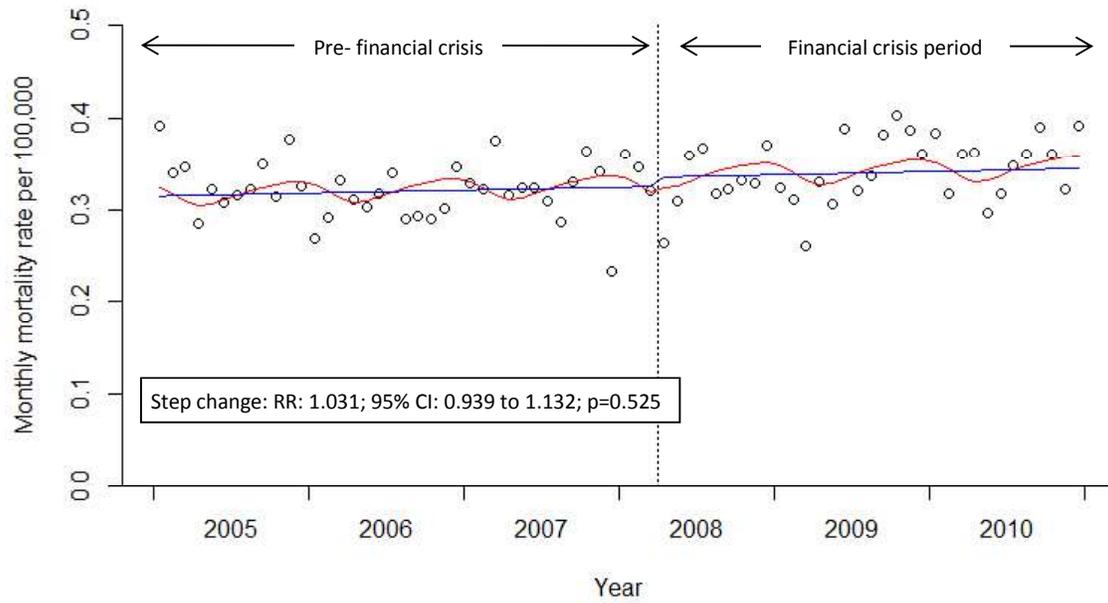
Seasonally adjusted. Circles = observed rates; red line = modelled rates fitted to the data; blue line = deseasonalised trend. Vertical dotted line = onset of the financial crisis

11.4.4 Web Appendix 4: Trend in monthly suicide rates before and since the financial crisis by area

(a) Northern Spain
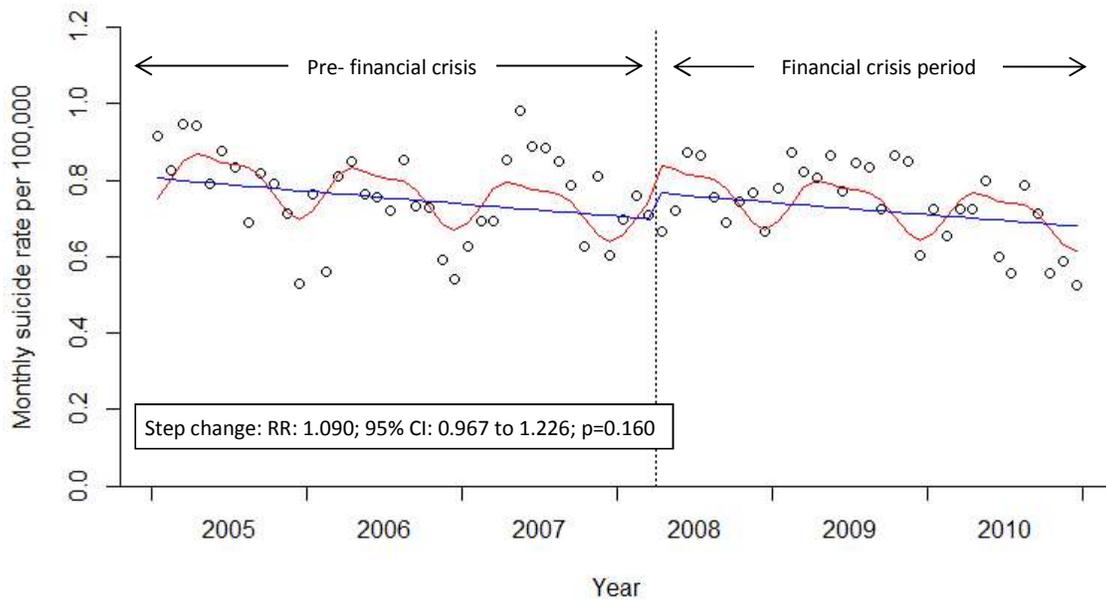
(b) Central Spain
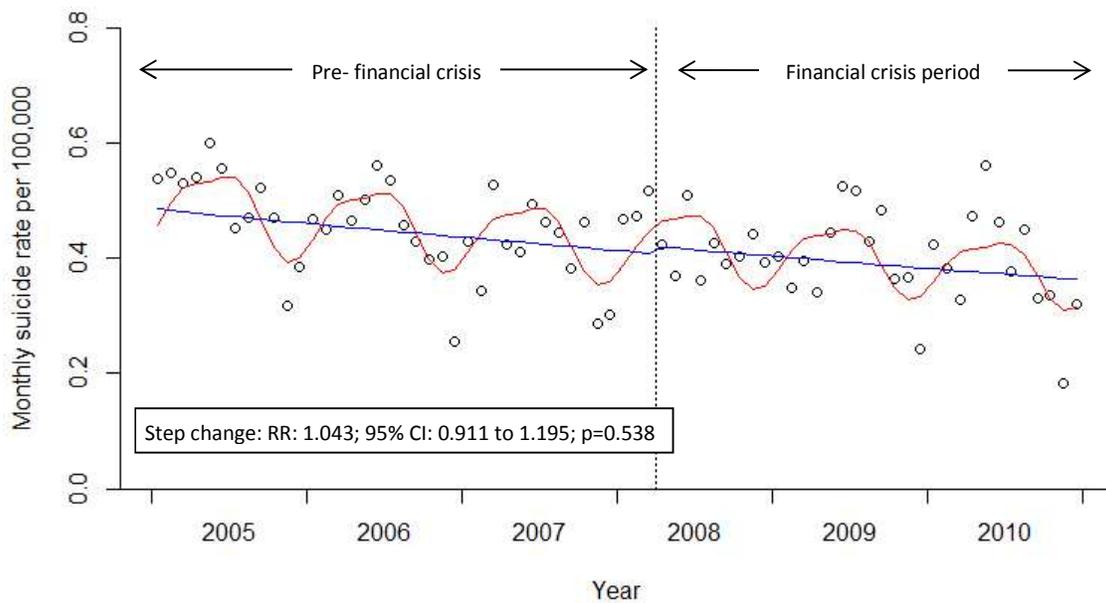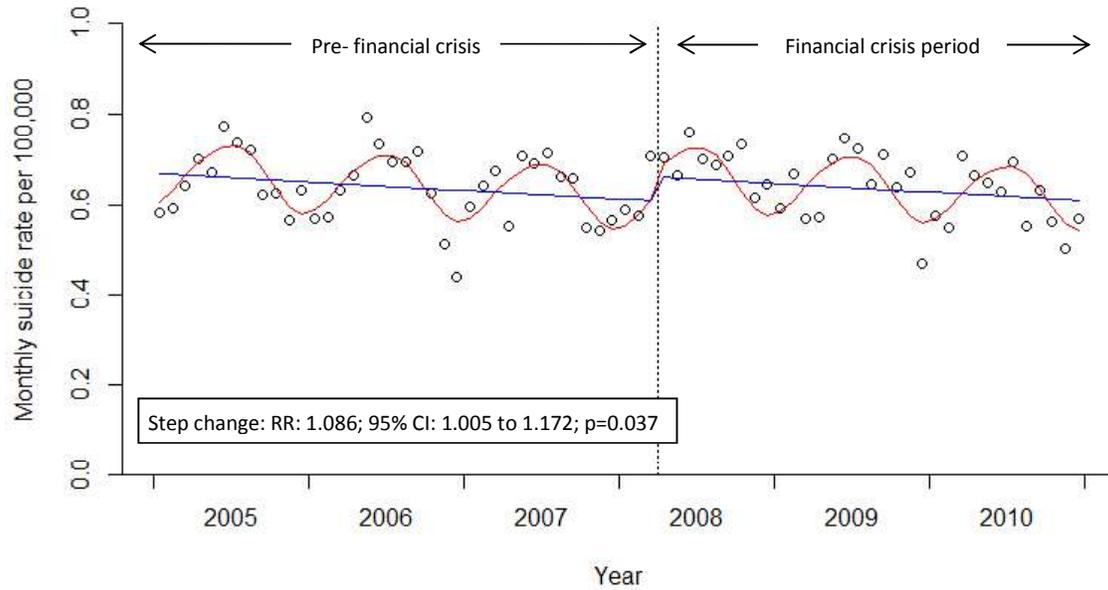
Seasonally adjusted. Circles = observed rates; red line = modelled rates fitted to the data; blue line = deseasonalised trend. Vertical dotted line = onset of the financial crisis. (a) includes Galicia, Asturias, Cantabria, the Basque Country, Navarra, La Rioja and Aragon; (b) includes Castilla and Leon, Castilla La Mancha, Extremadura and Madrid; (c) includes Catalonia, Valencia, Murcia, Andalucia, the Balearic Islands and the Canary Islands.

11.4.5 Web Appendix 5: Trend in monthly suicide rates before and since the financial crisis by sex
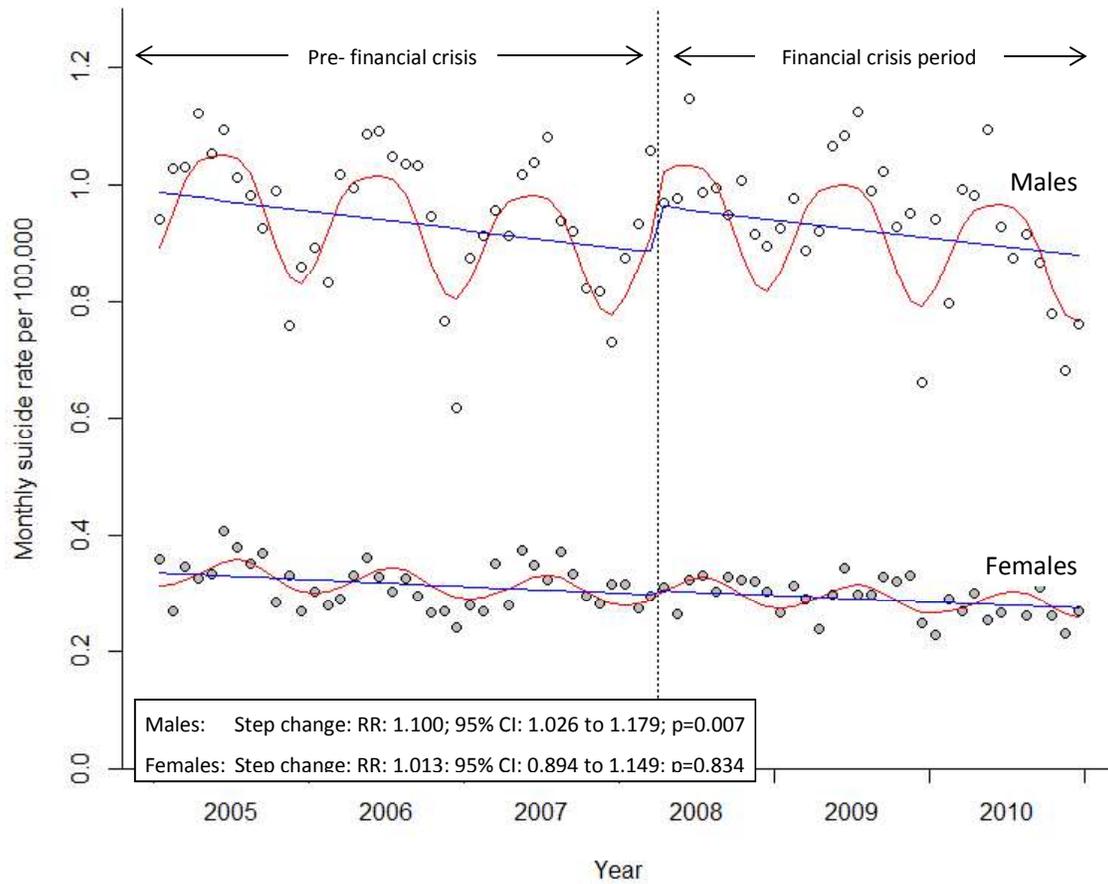
Seasonally adjusted. Circles = observed rates; red line = modelled rates fitted to the data; blue line = deseasonalised trend. Vertical dotted line = onset of the financial crisis.

11.4.6 Web Appendix 6: Trend in monthly suicide rates before and since the financial crisis by age group

(a) **Age 15-39 years**
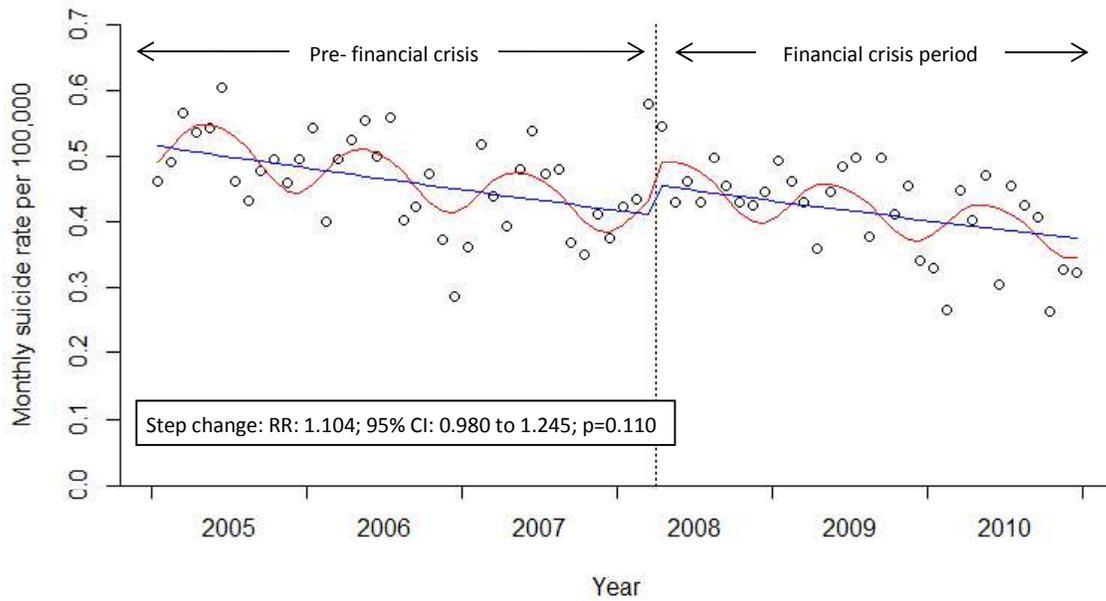

(b) **Age 40-64 years**
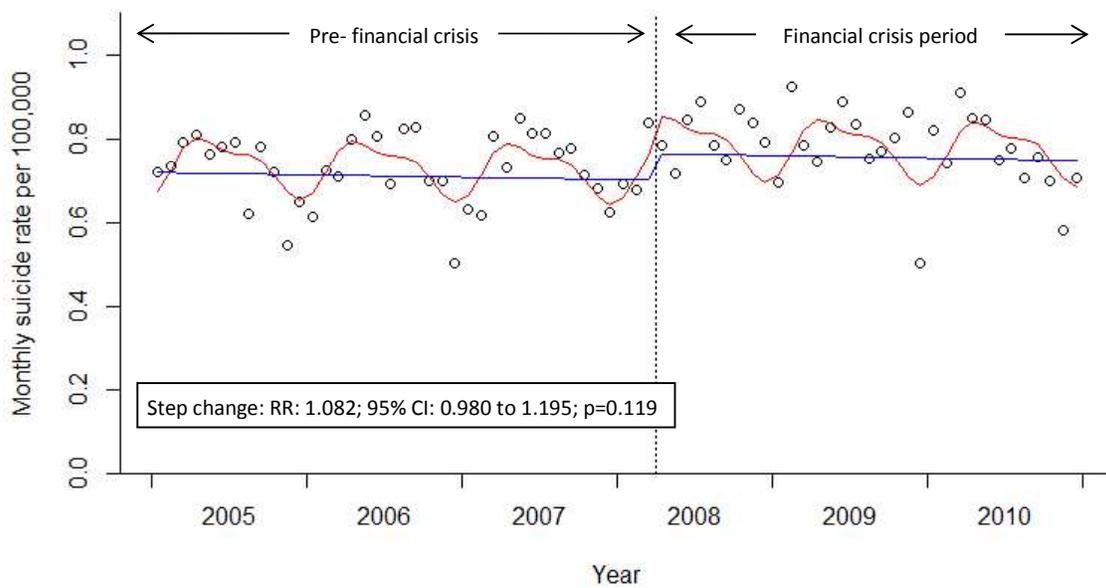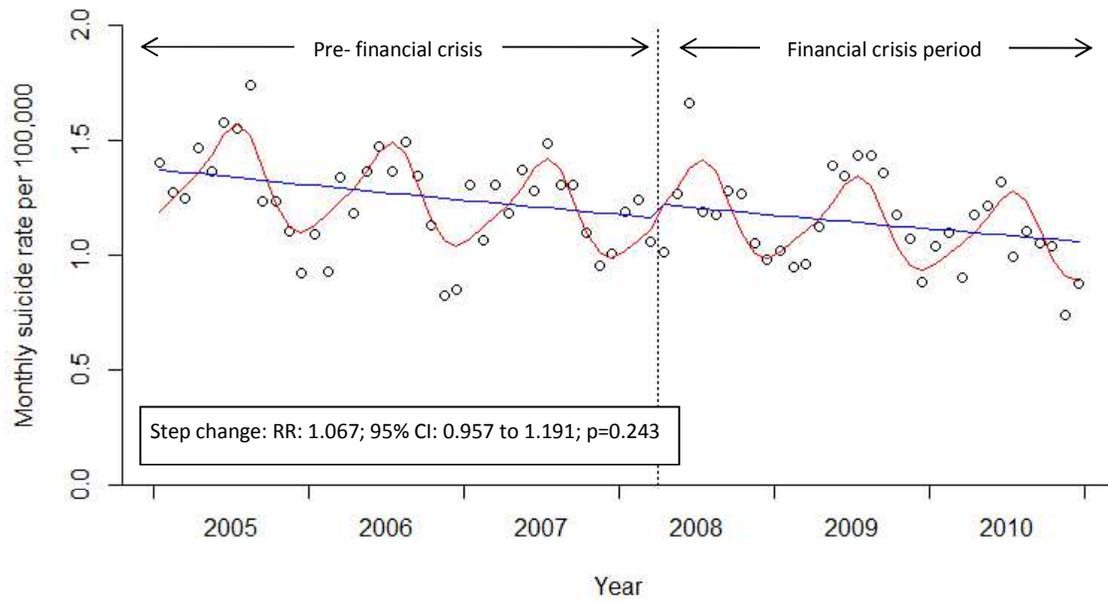
Seasonally adjusted. Circles = observed rates; red line = modelled rates fitted to the data; blue line = deseasonalised trend.
Vertical dotted line = onset of the financial crisis.

11.4.7 Web Appendix 7: Time series plots of trends in monthly suicide rates for all of Spain based on the alternative models used in sensitivity analysis

(a) Model allowing for both a step and a slope change in the financial crisis period

Pre- financial crisis   Financial crisis period

Step change: RR: 1.181; 95% CI: 1.013 to 1.377; p=0.037
Slope change: RR: 0.998; 95% CI: 0.995 to 1.001; p=0.208
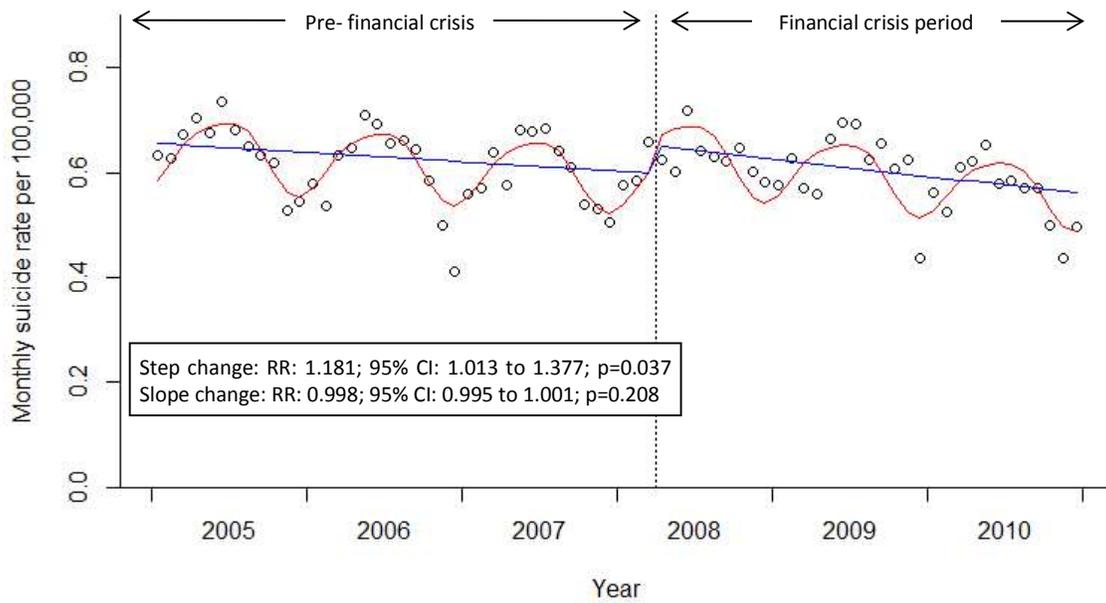
(b) Model with the crisis period lasting only for the duration of GDP contraction (April 2008 to December 2009)

Pre- financial crisis   Financial crisis period   Post-crisis

Step change: RR: 1.063; 95% CI: 1.022 to 1.106; p=0.003
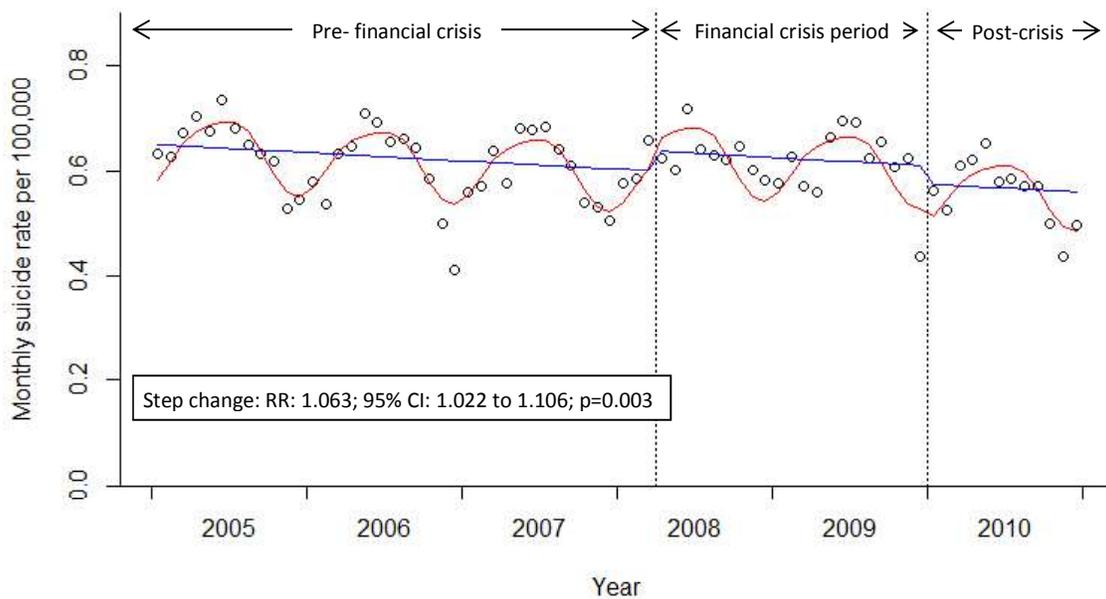
(c) **Model with the onset of the financial crisis at the time when unemployment began to rise (July 2007)**
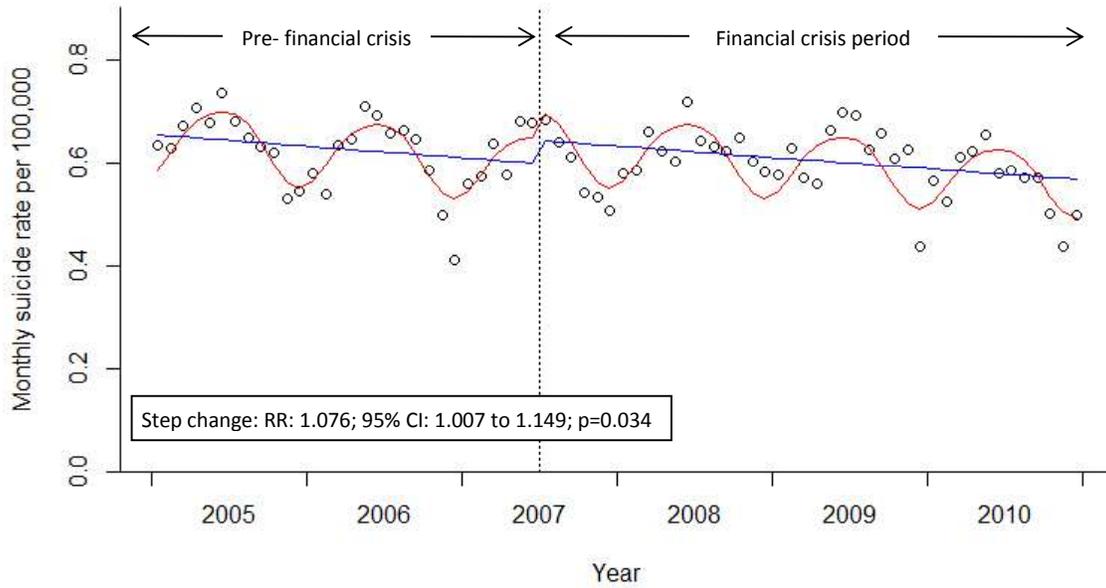
All models seasonally adjusted. Circles = observed rates; red line = modelled rates fitted to the data; blue line = deseasonalised trend. Vertical dotted line = onset of the financial crisis, second dotted line in (b) = end of recession period.

11.4.8 Web Appendix 8: Unemployment rates during the first quarter of 2005 and the last quarter of 2010

	2005 Q1	2010 Q4	Difference
All Spain	10.2%	20.3%	10.1%
Area:			
Northern Spain	8.9%	14.3%	5.4%
Central Spain	9.6%	17.3%	7.7%
Mediterranean	10.9%	23.7%	12.9%
Sex:			
Males	7.8%	20.0%	12.2%
Females	13.6%	20.8%	7.1%
Age group*:			
16-19	39.1%	65.5%	26.4%
20-24	23.2%	36.2%	12.9%
25-54	12.3%	19.6%	7.4%
55+	8.6%	12.8%	4.2%

Unemployment rates as a percentage of the economically active population. Q1 = first quarter, Q4 = last quarter. *Age-group divisions are different to those used in the study as these were the categories published by the Instituto Natcional de Estadística. Source: Instituto Nacional de Estadística [National Statistics Institute] †

†INE. *Instituto Nacional de Estadística [National Statistics Institute].* [cited 2012 04/07/2012]; Available from: http://www.ine.es/.

11.4.9 Contribution of the candidate to the paper

- The concept of this paper was my own. I found that the majority of previous studies of the effects of the financial crisis had used very weak before and after designs and thought that the evidence needed to be improved. I was also interested from a methodological perspective in applying the ITS design to an unplanned event.
- I developed the protocol and sought appropriate ethical approval.
- I identified suitable data sources and extracted the data
- I was responsible for deciding on the methodological approach – this included deciding on the ITS design, defining the timing of the financial crisis, defining the impact model and deciding on the stratification variables.
- I also cleaned and formatted the data and undertook all of the analysis including several sensitivity analyses.
- Finally, I was responsible for writing up the study.
- I received support and advice on the analysis from AG and all of the co-authors contributed to reviewing and providing feedback on the manuscript through several versions.

11.5 Case study 2 appendices

11.5.1 S1 Table: Population characteristics: England and Scotland 2007-2014 [1]

	England								Scotland							
	2007	2008	2009	2010	2011	2012	2013	2014	2007	2008	2009	2010	2011	2012	2013	2014
N	51381100	51815900	52196400	52642500	53107200	53493700	53865800	54316600	5170000	5202900	5231900	5262200	5299900	5313600	5327700	5347600
Age (%)																
0-19	24.2	24.0	23.9	23.8	23.9	23.9	23.8	23.8	23.0	22.8	22.7	22.5	22.3	22.1	21.9	21.7
20-39	27.3	27.2	27.0	26.9	26.9	26.8	26.6	26.6	26.3	26.1	26.0	26.0	25.9	25.8	25.8	25.9
40-59	26.8	26.8	26.8	26.8	26.7	26.8	26.8	26.8	28.5	28.5	28.4	28.4	28.6	28.6	28.5	28.4
60-79	17.1	17.4	17.7	17.8	17.8	17.9	18.0	18.2	18.1	18.4	18.6	18.7	18.9	19.0	19.2	19.4
80+	4.5	4.6	4.6	4.7	4.6	4.7	4.7	4.8	4.2	4.3	4.3	4.4	4.4	4.5	4.5	4.6
Sex (%)																
Males	49.2	49.2	49.2	49.3	49.2	49.2	49.3	49.3	48.3	48.4	48.4	48.5	48.5	48.5	48.6	48.6
Females	50.8	50.8	50.8	50.7	50.8	50.8	50.7	50.7	51.7	51.6	51.6	51.5	51.5	51.5	51.4	51.4
Crude birth rate (per 1000)	12.8	13	12.9	13.1	13	13	12.3	12.2	11.2	11.5	11.3	11.2	11.1	10.9	10.5	10.6
Crude death rate (per 1000)	9.2	9.2	8.8	8.8	8.5	8.7	8.8	8.6	10.8	10.7	10.3	10.3	10.1	10.3	10.3	10.1

References

1.   Office for National Statistics. Population estimates [Available from:
http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates.

11.5.2 S2 Table: Trend changes in specialist visits and hospitalisations following the intervention England vs Wales

|	Trend change England			Trend change Wales			Trend change England v Wales		
	Effect	95% CI	p-value	Effect	95% CI	p-value	Effect	95% CI	p-value
Outpatient specialist visits									
Total	1.011	[1.007,1.015]	<0.001	0.9879	[0.980,0.996]	0.003	1.0233	[1.014,1.033]	<0.001
GP referred	1.016	[1.012,1.020]	<0.001	0.9908	[0.977,1.004]	0.186	1.0254	[1.011,1.040]	0.001
Inpatient hospitalizations									
Total	0.9982	[0.995,1.002]	0.332	0.9988	[0.995,1.003]	0.565	1.000	[0.994,1.005]	0.846
Elective	0.9977	[0.994,1.001]	0.212	0.9983	[0.991,1.006]	0.672	0.999	[0.991,1.007]	0.808
Emergency	1.000	[0.995,1.004]	0.878	0.9987	[0.997,1.001]	0.182	1.000	[0.996,1.005]	0.884

Coefficients for trend change are relative change in the slope gradient following the intervention. Trend change study v control is the slope change in England over and above any change in Wales accounting for differences in baseline trends. All segmented regression models used log transformed Gaussian distribution.

### 11.5.3 S1 Figure: Time series of outpatient specialist visits in England and Wales
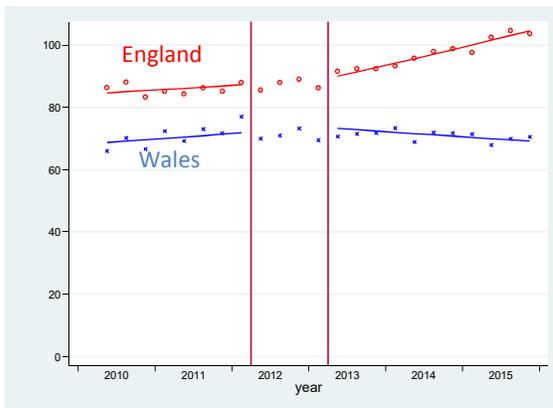
Total

GP referred
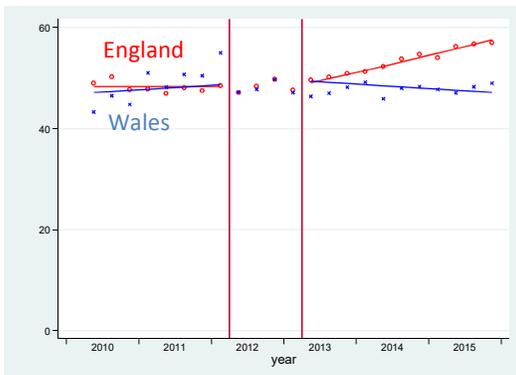
Red o = England, blue x = Wales. Lines = deseasonalized linear trend. Vertical lines delineate the intervention phase (between Q2 2014 and Q2 2013)

11.5.4 S2 Figure: Time series of inpatient hospitalisations in England and Wales

Total
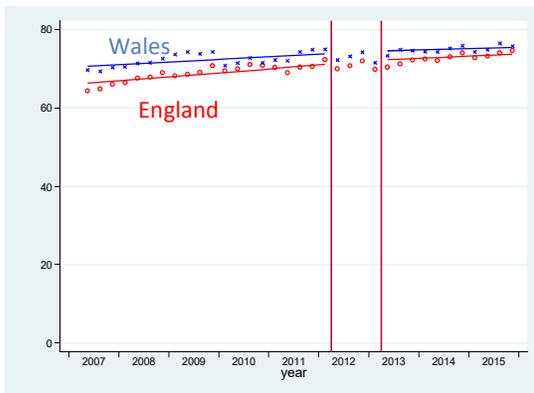
Elective
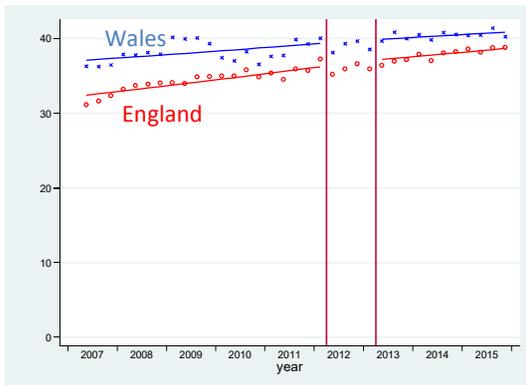
Emergency
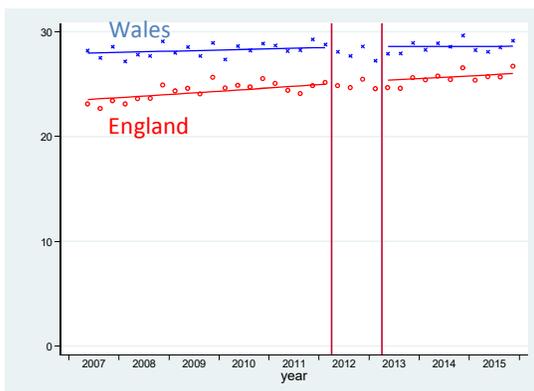
Red o = England, blue x = Wales. Lines = deseasonalized linear trend. Vertical lines delineate the intervention phase (between Q2 2014 and Q2 2013)

11.5.5 S3 Figure: NHS reference costs

A) Reference costs for outpatient specialist visits(1)
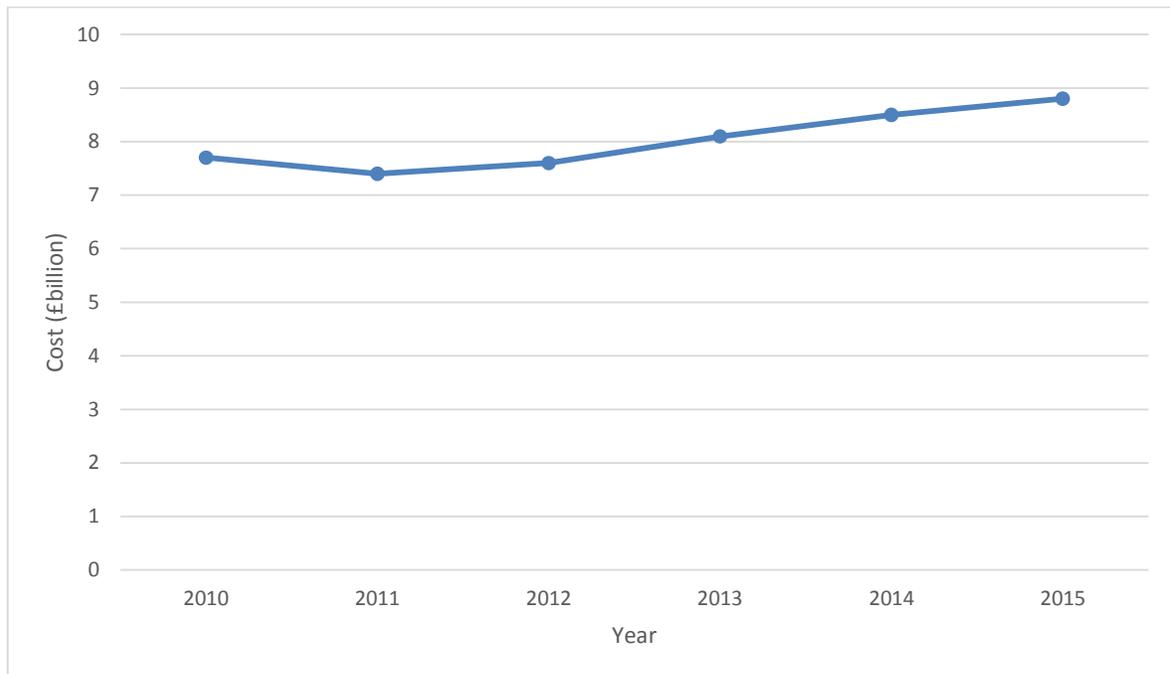
B) Reference costs for inpatient admissions (1)
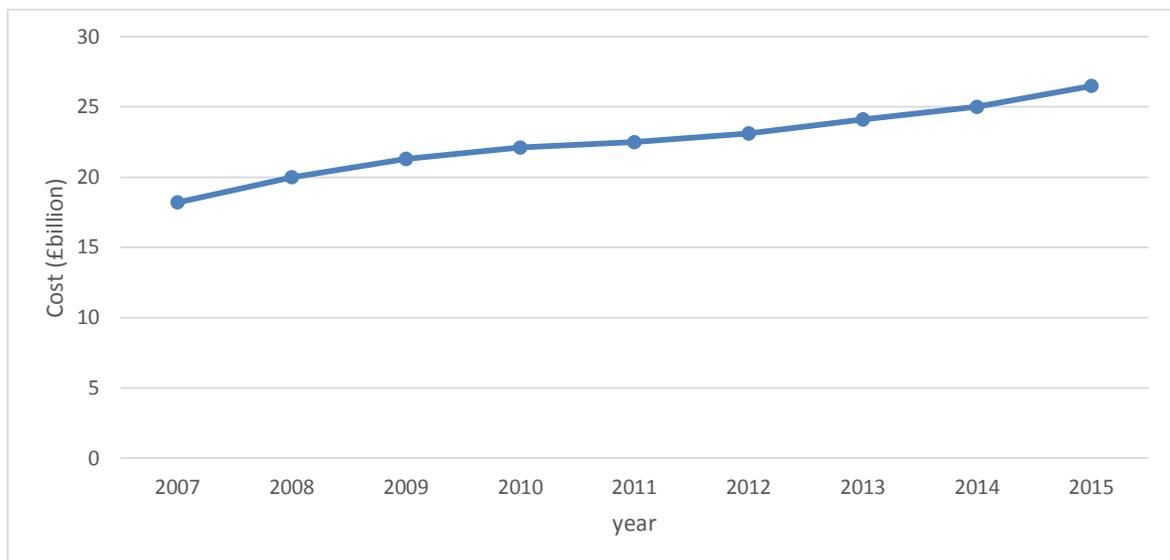
1. Department of Health. Reference costs 2014-15 2015 [Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/477919/2014-15_Reference_costs_publication.pdf.

11.5.6 S1 Text: Controlled interrupted time series model
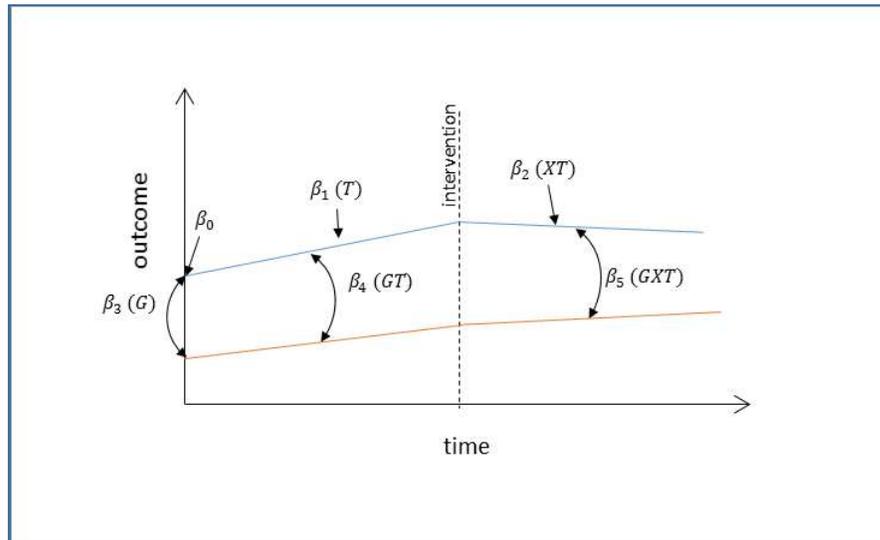
Intervention group in blue, control group in red. T = time since the start of the study, X = intervention (pre-intervention period = 0, post-intervention period = 1), G = group (control group = 0, intervention group = 1). β_0 and β_3 relate to intercepts, β_{1-2} and β_{4-5} relate to slopes. Curved arrows represent differences between the intervention group and control group. (Adapted from Linden and Adams 2011)[1]

Segmented regression equation for slope change with a control series:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 TX_t + \beta_3 G + \beta_4 GT + \beta_5 GX_t T$$

Y_t is the outcome variable at time t, T is a variable representing the time since the start of the study and X is a dummy variable indicating the pre- ($X = 0$) or post-intervention period ($X = 1$). G represents the intervention group ($G = 1$) or control group ($G = 0$). β_0 represents the intercept at T=0, β_1 is the underlying pre-intervention trend (slope), β_2 is the slope change following the intervention, β_3 represents the difference in intercept between the two groups at T=0, β_4 represents the slope difference between the intervention and control group in the pre-intervention period, β_5 represents the difference between the change in slope in the control and intervention group associated with the intervention. Therefore β_5 is the parameter of interest for the measure of effect.

References

1.   Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. J Eval Clin Pract. 2011;17(6):1231-8.

11.5.7 S1 Checklist. REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement.

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Ite m No .	STROBE items	Location in manuscript where items are reported (page)	RECORD items	Location in manuscript where items are reported (page)
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Title and abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.	Abstract paragraph 2
				RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.	Abstract paragraph 2
				RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	n/a
Introduction					
Background rationale	2	Explain the scientific background and	Introduction		

		rationale for the investigation being reported			
Objectives	3	State specific objectives, including any prespecified hypotheses	Introduction final paragraph		
Methods					
Study Design	4	Present key elements of study design early in the paper	Methods: statistical analysis		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods: data and study population		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources	n/a	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.	Methods: data and study population (NB: data for all NHS pts was included other than the exclusions stated on p6) n/a

		and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Methods: data and study population; statistical analysis	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	S1-s2 data extraction
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods: data and study population		
Bias	9	Describe any efforts to address potential sources of bias	Methods: statistical analysis; Discussion paragraphs 2 and 3		

Study size	10	Explain how the study size was arrived at	Methods: data and study population		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Methods: statistical analysis		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	Methods: statistical analysis		
		(b) Describe any methods used to examine subgroups and interactions	Methods: statistical analysis		
		(c) Explain how missing data were addressed	n/a		
		(d) *Cohort study* - If applicable, explain how loss to follow-up was addressed	n/a		
		Case-control study - If applicable, explain how matching of cases and controls was addressed			
		Cross-sectional study - If applicable, describe analytical methods taking account of sampling strategy			
		(e) Describe any sensitivity analyses			

			n/a		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	n/a – all of the database population within the defined study period was included
				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Methods: data and study population
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	n/a
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed)	Table 1 and S1 Table

n/a | RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram. | Methods: data and study population |

		(b) Give reasons for non-participation at each stage.	n/a		
		(c) Consider use of a flow diagram			
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders	Table 1 and S1 Table		
		(b) Indicate the number of participants with missing data for each variable of interest	n/a		
		(c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	n/a		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time	n/a		
		Case-control study - Report numbers in each exposure category, or summary measures of exposure	n/a		
		Cross-sectional study - Report numbers of outcome events or summary measures	n/a		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-	Results: Paragraphs 2-		

		adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8, Table 2, S2 Table		
		(b) Report category boundaries when continuous variables were categorized	n/a		
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Results: Paragraphs 2-8, Table 4		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	S2 Table, S1-2 Fig		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Discussion: paragraph 1		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion: paragraph 3	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they	Discussion: paragraph 3

				pertain to the study being reported.	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion: paragraphs 4-6		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion: paragraph 2		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Financial disclosure		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Methods: data and study population, S1-s2 data extraction

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

11.5.8 S1 Protocol

Evaluation of the impact of GP led commissioning on secondary care activity

Background:

Clinical Commissioning Groups (CCGs) were created following the Health and Social Care Act 2012 (HSCA),(1) and replaced Primary Care Trusts on 1 April 2013 as the main budget holders and commissioners of NHS services in England. CCGs are led by a board, primarily made up of General Practitioners (GPs), and represent all GP practices in the local area. Each is responsible for a population of 100,000-900,000. The theory behind GP led commissioning is, firstly, that GPs understand their patients' needs best so are best placed to commission specialist services on their behalf, and secondly, that by holding the budgets and being given the freedom to reinvest any savings they are incentivized to ration spending. It has been hypothesized that, given the incentive to minimize costs, GP budget holding may lead to a shift away from expensive secondary care activity towards a more community based approach.(2-4) GP commissioning may also influence patient experience of secondary care services as it could incentivize hospitals to reduce waiting times and work to improve patient satisfaction given the threat that commissioners may switch contract to providers that provide a better patient experience.(4) Nevertheless, unintended consequences are also possible, for example, there may be inappropriate reductions in care that should have been in hospital and patient satisfaction could suffer if GPs, as budget holders, prioritize cost over patient experience.

Aim:

To evaluate the impact of GP led commissioning on secondary care activity

Questions

1. Has the introduction of GP led commissioning been associated with a change in rates of hospital admissions?
2. Has the introduction of GP led commissioning been associated with a change in patients' experience of secondary care in the NHS?

Outcomes

Primary outcomes:

Secondary care activity, including:

- GP referrals made
- GP referrals seen
- Total admissions
- elective admissions
- outpatient appointments
- emergency admissions

Secondary outcomes:

Patient experience, including

- waiting times
- patient satisfaction

Analysis

An interrupted time series (ITS) design with segmented regression analysis will be used to evaluate the effect of GP led commissioning on secondary care activity. The intervention was fully implemented in April 2013, yet there was a phase in period during which many CCGs were present in shadow form prior to April 2013 and after full implementation changes to commissioning are likely to only have manifested gradually (existing contracts may have taken some time to expire and CCGs may have chosen to simply renew existing contracts in the early stages before they became well established). This "intervention phase" and gradual change should therefore be reflected in the chosen impact model. Possible impact models are outlined in Figure 1, ultimately one of these will be chosen a priori based on further consideration of existing knowledge of the intervention and/or secondary data, other impact models will be used in a sensitivity analysis.

The greatest limitation of the ITS design is the potential for confounding by events concurrent to the intervention. In order to strengthen the validity of the design we will incorporate control series to exclude effects from possible confounding events. Geographical control groups will include secondary care activity in Scotland, Wales and Northern Ireland (where the HSCA was not implemented). Control outcomes such as critical care activity (a service that would be unlikely to be substituted by changes to community care) will also be included in the analyses.

Stratified analyses by different types of secondary care activity (elective, emergency, outpatient) and by specialty will be undertaken in order to establish the nature of any effect.
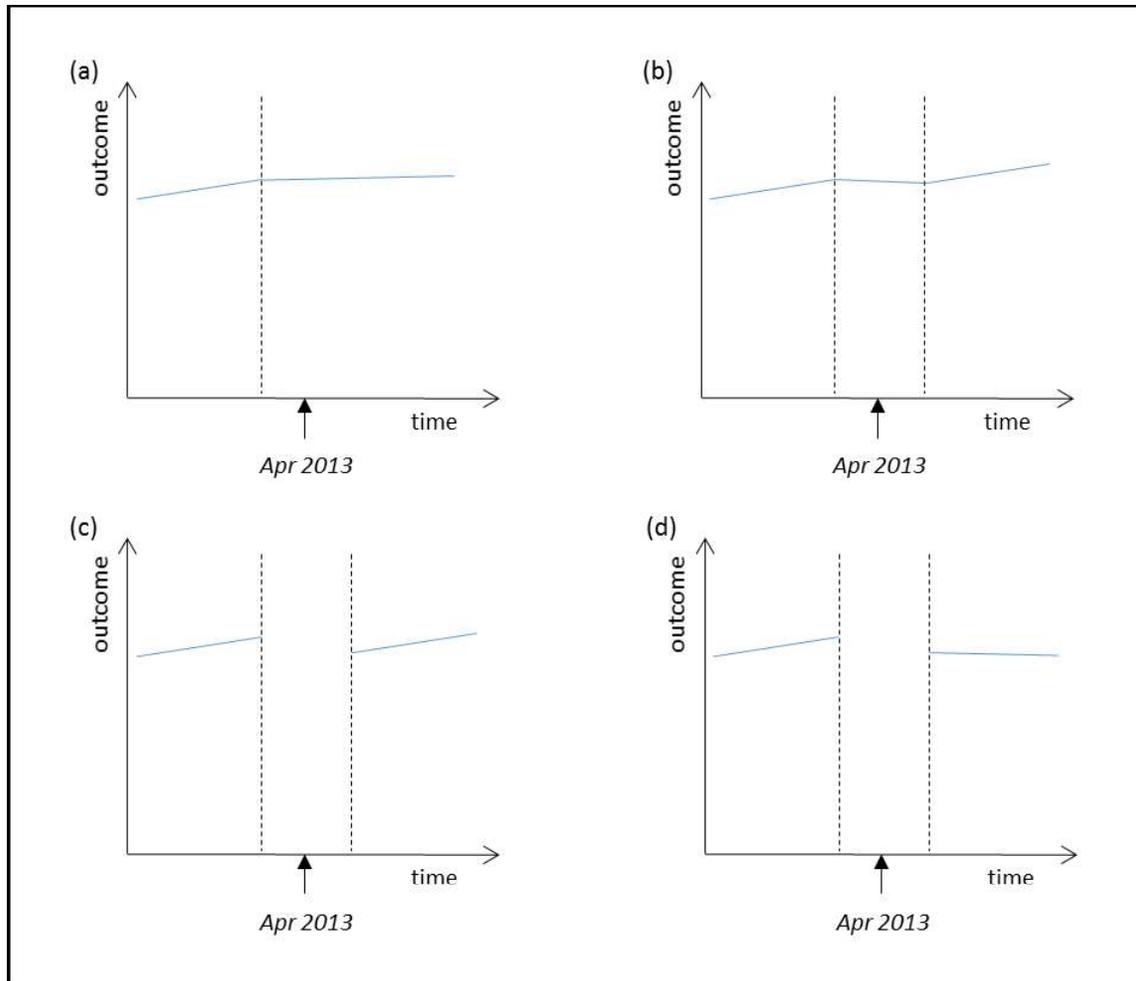

Figure 2: possible impact models

(a) gradual slope change beginning during the phase in of the intervention
(b) gradual slope change during the "intervention period" followed by a step change
(c) step change with exclusion of the "intervention period"
(d) step and slope change with exclusion of the "intervention period"

Data

Monthly data will be needed for each of the following variables – ideally from 2007-2015

Variable	Data source	Range
Outcomes		
Elective admissions (England)	NHS England (HES)	Apr08-Aug15
	HSCIC	Apr07-Jul15
Outpatient appointments (England)	NHS England (HES)	Apr08-Aug15
	HSCIC	Apr07-Jul15
Emergency admissions (England)	NHS England (HES)	Apr08-Aug15
	HSCIC	Apr07-Jul15
GP referrals made (England)	NHS England (HES)	Apr08-Aug15
GP referrals seen (England)	NHS England (HES)	Apr08-Aug15
Denominator data		
Population (England)	ONS	07-14
Stratification variables		
Gender		
Age-group		
Diagnosis		
Specialty		
Controls		
Elective admissions (Scotland)	ISD (NHS Scotland)	04-Jun15
Outpatient appointments (Scotland)	ISD (NHS Scotland)	04-Jun15
Emergency admissions (Scotland)	ISD (NHS Scotland)	04-14
GP referrals made (Scotland)		
GP referrals seen (Scotland)		
Population (Scotland)	ONS	07-14
Elective admissions (Wales)		99-14

Outpatient appointments (Wales)	Statswales	89-12
Emergency admissions (Wales)		99-14
GP referrals made (Wales)	Statswales	Jan05-Sep15
GP referrals seen (Wales)		
Population (Wales)	ONS	07-14
Elective admissions (NI)	dhsspsni	05-15
Outpatient appointments (NI)	dhsspsni	09-15
Emergency admissions (NI)	dhsspsni	05-15
GP referrals made (NI)		
GP referrals seen (NI)		
Population (NI)	ONS	07-14
Critical care admissions (England)	HSCIC	Apr08-Mar14

Secondary outcomes	
Waiting times (England + controls)	
Patient satisfaction score (England + controls)	

Data sources

Secondary care activity:

Data on each inpatient admission, outpatient appointment and accident and emergency attendance in NHS hospitals in England are collected as Hospital Episode Statistics (HES) published by the Health and Social Care Information Centre (HSCIC, http://www.hscic.gov.uk/hes). Overall activity in each of these categories is made publically available on a monthly basis with a four month delay via the health and social care information centre website and dates back to April 2007 (Figure 1). More detailed annual data is also available stratified by various covariates including diagnosis, specialty and hospital provider. Adult Critical care data is also published on an annual basis. In addition to the data which is made publically available HSCIC operate a Data Access Request Service on a cost recovery basis whereby data can be requested for additional covariates (such as age, gender and index of multiple

deprivation) and in greater detail – such as monthly time series of stratified data and even individual patient data. Such requests take between 14 to 60 days to gain access to the data.
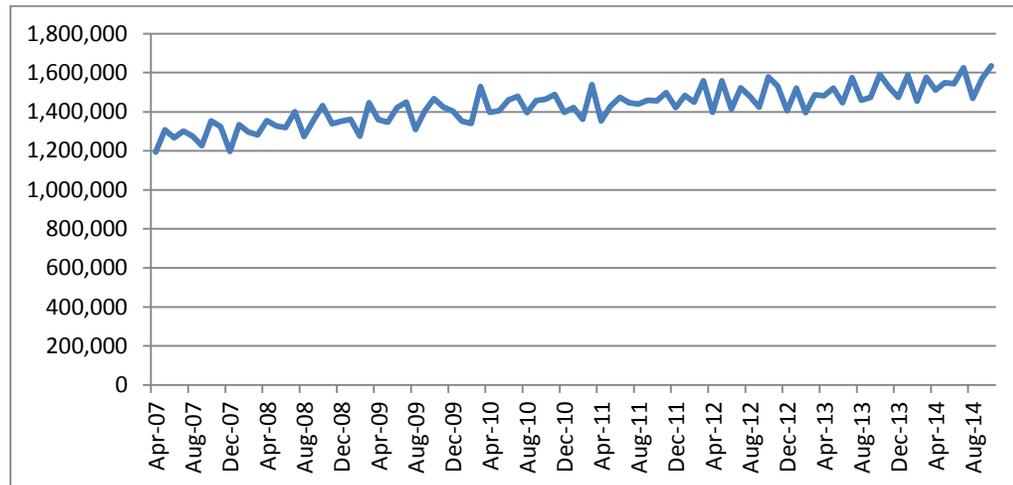

Figure 1: Admitted patients finished consultant episodes in England by month 2007-2014

[More data now available via NHS England, including monthly data stratified by other covariates: http://www.england.nhs.uk/statistics/statistical-work-areas/hospital-activity/monthly-hospital-activity/mar-data/]

Patient experience:

Monthly waiting time data is published by NHS England including:

- Referral to Treatment Waiting Times (http://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/)
- Cancer waiting times (http://www.england.nhs.uk/statistics/statistical-work-areas/cancer-waiting-times/)
- Diagnostic waiting times (http://www.england.nhs.uk/statistics/statistical-work-areas/diagnostics-waiting-times-and-activity/monthly-diagnostics-waiting-times-and-activity/)

The NHS Patient Survey Programme coduct annual surveys of patient experience including the Inpatient, Outpatient, Community Mental Health and Accident & Emergency surveys. (http://www.england.nhs.uk/statistics/statistical-work-areas/pat-exp/ and https://www.gov.uk/government/collections/patient-experience-statistics)

Patient Reported Outcome Measures (PROMs) assess the quality of care delivered to NHS patients from the patient perspective. Currently covering four clinical procedures (hip replacements, knee replacements, groin hernia and varicose veins), PROMs calculate the health gains after surgical treatment using pre- and post-operative surveys.

Control data:

In Scotland quarterly hospital activity data is published online by the NHS Information Services Devision (ISD, http://www.isdscotland.org/Health-Topics/Hospital-Care/) with one quarter delay. Welsh hospital activity data in Wales is collected monthly and collated in the Patient Episode Database for Wales (PEDW) and annual data is published at the end of each year dating back to 1991 by the NHS Wales Informatics Service (http://www.infoandstats.wales.nhs.uk/page.cfm?orgid=869& pid=40977). In Northern Ireland annual hospital activity data is published by the Department of Health, Social Services and Public Safety Information Office dating back to 2005 (http://www.dhsspsni.gov.uk/index/statistics/hospital/hospital-activity.htm).

Threats and potential solutions

Threat	Solution
Some CCGs were present in shadow form for several months prior to becoming the official commissioning body so may have influenced commissioning in the months leading up to the intervention.	Consider modelling the intervention earlier
Changes to commissioning could have taken some time to manifest: existing contracts may have taken some time to expire and CCGs may have chosen to simply renew existing contracts in the early stages before they became well established	Consider modelling a slope change during the changeover period to allow for a gradual change whilst contracts expired then a step change once contracts should have all expired. Alternatively could exclude the intervention period from the model
The HSCA resulted in other changes to the structure of the NHS and public health services (see appendix). Therefore it may be difficult to disentangle whether some effects are due to the introduction of CCGs or due to other changes	It is unlikely that the other changes would have had an impact on secondary care referrals so this should not be an issue for the primary outcome. Some of the other changes such as a purported "greater voice for patients" could result in changes to secondary outcomes relating to patient experience, nevertheless this change was more of a gradual progression rather than a sudden new change with the implementation of the act.

	Using a control outcome such as critical care activity, could help to exclude some of the other changes that could have affected all types of care. Another option is to treat the intervention as the HSCA as a whole rather than focus on a part of the act.

References

1.	Health and Social Care Act. Health and Social Care Act. 2012.
2.	Dusheiko M, Gravelle H, Jacobs R, Smith P. The effect of financial incentives on gatekeeping doctors: Evidence from a natural experiment. Journal of Health Economics. 2006;25(3):449-78.
3.	Smith JA, Mays N. GP led commissioning: time for a cool appraisal. BMJ. 2012;344.
4.	Mannion R. General practitioner-led commissioning in the NHS: progress, prospects and pitfalls. British Medical Bulletin. 2011;97(1):7-15.

Appendix

2012 Health and Social Care Act key policy areas:

1. Clinically led commissioning:
   - Abolition of Primary Care Trusts (PCTs) and Strategic Health Authorities (SHAs) + moving commissioning to GP led Clinical Commissioning Groups (CCGs)
2. Provider regulation to support innovative services:
   - Monitor (a national regulatory body) given greater role in promoting competition and role in licensing providers of NHS services (including private sector and charity providers).
   - Emphasis on choice and competition as a driver of improved patient care
3. Greater voice for patients:
   - Building on previous efforts to increase patient involvement in the NHS
   - Included establishment of Healthwatch (as a committee of the Care Quality Commission) to represent patient views in advising NHS organisations
4. New focus for public health:
   - Establishment of Public Health England with a remit of improving the nation's health at a national level (and abolishment of the Health Protection Agency)
   - Moving public health departments at a local level from PCTs to Local Authorities (establishment of PHE as ne body to drive improvements in PH)
5. Greater accountability locally and nationally:
   - Strengthens and clarifies accountability for and within the NHS
   - CCGs accountable to NHS England and assessed against a commissioning outcomes framework
   - Statutory health and wellbeing boards within local authorities aimed to 'strengthen the local democratic legitimacy of the NHS' with responsibility for joining up the commissioning of local NHS services and social care.
6. Streamlined arms-length bodies:
   - Several bodies abolished with key functions transferred to other bodies (including: General Social Care Council, Office of the Healthcare Professions Adjudicator, Alcohol Education and Research Council, National Patient Safety Agency, NHS Institute for Innovation and Improvement, National Information Governance Board and the Appointments Commission)

11.5.9 S1 Data

English data

Available at:

https://doi.org/10.1371/journal.pmed.1002427.s009

11.5.10   S2 Data

Scottish data

Available at:

https://doi.org/10.1371/journal.pmed.1002427.s010

11.5.11   S3 Data

Algorithms used for extraction of English data.

Available at:

https://doi.org/10.1371/journal.pmed.1002427.s011

11.5.12 S4 Data

Algorithms used for extraction of Scottish data.

Available at:

https://doi.org/10.1371/journal.pmed.1002427.s012

11.5.13   Contribution of the candidate to the paper

- As with the previous case study, the concept of this paper was my own. It was motivated by the lack of quantitative outcome evaluations of this large scale policy and by the fact that it had been identified as difficult to evaluate Vittal Katikireddi[1] et al. I was interested in the methodological challenges posed by this evaluation.
- I developed the proposal for this study, applied for ethical approval and identified and applied for the data from the Health and Social Care Information Centre (England).
- I identified suitable controls and also applied for the data from the NHS Scotland Information Services Division, and the NHS Wales Information Service.
- I spent a significant amount of time investigating the details of the reforms and deciding on the most appropriate impact model in consultation with my co-authors.
- I undertook the data cleaning, formatting and all uncontrolled and controlled analyses for each outcome and the sensitivity analyses.
- I also identified and examined cost data from a separate data source to see if this matched our findings with respect to the increase in outpatient specialist visits.
- I received advice on analytical approaches from SS, JFW, and AG.
- I wrote the paper and received advice and feedback on the paper from all co-authors

1. Vittal Katikireddi S, McKee M, Craig P, Stuckler D. The NHS reforms in England: four challenges to evaluating success and failure. Journal of the Royal Society of Medicine. 2014;107(10):387-92.

11.6 Methodological paper 4 appendices

11.6.1   Data extracted in the literature review

11.6.2   Eligible studies

11.6.3   Contribution of the candidate to the paper

11.6.1 Appendix 1: Data extracted in the literature review

Study characteristics:

- Type of intervention (classified as health promotion interventions, health protection interventions, health service intervention, and non-health interventions or events)
- Primary outcome (classified as health or disease measures, mortality measures, health behaviours, service or treatment uptake, treatment outcomes or service quality, and non-health outcomes) and
- Type of data source

We extracted the following information on the methodological reporting of the study:

- Whether the study time period was clearly stated,
- Whether the time interval was clearly stated,
- Whether the timing of the intervention was clearly defined,
- Whether the authors reported checking for changes to data collection or processing,
- Whether the authors justified their choice of study period,
- Whether the authors considered possible history bias (due to confounding interventions or events around the time of the intervention),
- Whether any design adaptations were implemented to mitigate the risk of history bias,
- Type of control series if a control series was used (classified as location based control, characteristic based control, behaviour based control, historical control cohort, control outcome, control time period and randomised control group).
- Method of analysis (segmented regression or autoregressive integrated moving average model),
- Impact model selected (including whether they allowed for a level change and/or a slope change, whether a transition phase was included, whether they allowed for a lagged effect and whether they allowed for a floor or ceiling effect),
- Whether the selected impact model was justified *a priori*,
- Whether linearity was assessed in the outcome data and Whether a linear or non-linear model was used,
- Whether seasonality was considered and adjusted for,
- Whether time-varying confounders were considered and adjusted for,
- Whether autocorrelation was considered and adjusted for,
- Whether the authors undertook any sensitivity analyses and Whether these were justified *a priori*,

11.6.2 Appendix 2: Eligible studies

1.   Adams AS, Soumerai SB, Zhang F, Gilden D, Burns M, Huskamp HA, et al. Effects of eliminating drug caps on racial differences in antidepressant use among dual enrollees with diabetes and depression. Clin Ther. 2015;37(3):597-609.

2.   Baillie CA, Guevara JP, Boston RC, Hecht TE. A unit-based intervention aimed at improving patient adherence to pharmacological thromboprophylaxis. BMJ Qual Saf. 2015;24(10):654-60.

3.   Balasubramanian D, Prinja S, Aggarwal AK. Effect of User Charges on Secondary Level Surgical Care Utilization and Out-of-Pocket Expenditures in Haryana State, India. PLoS ONE. 2015;10(5):e0125202.

4.   Balicer RD, Hoshen M, Cohen-Stavi C, Shohat-Spitzer S, Kay C, Bitterman H, et al. Sustained Reduction in Health Disparities Achieved through Targeted Quality Improvement: One-Year Follow-up on a Three-Year Intervention. Health Serv Res. 2015;50(6):1891-909.

5.   Barocas DA, Mallin K, Graves AJ, Penson DF, Palis B, Winchester DP, et al. Effect of the USPSTF Grade D Recommendation against Screening for Prostate Cancer on Incident Prostate Cancer Diagnoses in the United States. J Urol. 2015;194(6):1587-93.

6.   Been JV, Mackay DF, Millett C, Pell JP, van Schayck OC, Sheikh A. Impact of smoke-free legislation on perinatal and infant mortality: a national quasi-experimental study. Sci. 2015;5:13020.

7.   Been JV, Szatkowski L, van Staa TP, Leufkens HG, van Schayck OC, Sheikh A, et al. Smoke-free legislation and the incidence of paediatric respiratory infections and wheezing/asthma: interrupted time series analyses in the four UK nations. Sci. 2015;5:15246.

8.   Berkowitz SA, Percac-Lima S, Ashburner JM, Chang Y, Zai AH, He W, et al. Building Equity Improvement into Quality Improvement: Reducing Socioeconomic Disparities in Colorectal Cancer Screening as Part of Population Health Management. J Gen Intern Med. 2015;30(7):942-9.

9.   Bijlsma MJ, Janssen F, Lub R, Bos JH, De Vries FM, Vansteelandt S, et al. Birth cohort appeared to confound effect estimates of guideline changes on statin utilization. J Clin Epidemiol. 2015;68(3):334-40.

10.   Blais E, Carnis L. Improving the safety effect of speed camera programs through innovations: Evidence from the French experience. J Safety Res. 2015;55:135-45.

11.   Blecker S, Goldfeld K, Park H, Radford MJ, Munson S, Francois F, et al. Impact of an Intervention to Improve Weekend Hospital Care at an Academic Medical Center: An Observational Study. J Gen Intern Med. 2015;30(11):1657-64.

12.   Bonander C, Andersson R, Nilson F. The effect of stricter licensing on road traffic injury events involving 15 to 17-year-old moped drivers in Sweden: A time series intervention study. Accid Anal Prev. 2015;83:154-61.

13.   Borde JP, Kern WV, Hug M, Steib-Bauert M, de With K, Busch HJ, et al. Implementation of an intensified antibiotic stewardship programme targeting third-generation cephalosporin and fluoroquinolone use in an emergency medicine department. Emerg Med J. 2015;32(7):509-15.

14.   Borde JP, Litterst S, Ruhnke M, Feik R, Hubner J, deWith K, et al. Implementing an intensified antibiotic stewardship programme targeting cephalosporin and fluoroquinolone use in a 200-bed community hospital in Germany. Infection. 2015;43(1):45-50.

15.   Branas CC, Kastanaki AE, Michalodimitrakis M, Tzougas J, Kranioti EF, Theodorakis PN, et al. The impact of economic austerity and prosperity events on suicide in Greece: a 30-year interrupted time-series analysis. BMJ Open. 2015;5(1):e005619.

16.   Brunette MF, Dzebisashvili N, Xie H, Akerman S, Ferron JC, Bartels S. Expanding Cessation Pharmacotherapy Via Videoconference Educational Outreach to Prescribers. Nicotine Tob Res. 2015;17(8):960-7.

17.   Burden AM, Tadrous M, Calzavara A, Cadarette SM. Uptake and characteristics of zoledronic acid and denosumab patients and physicians in Ontario, Canada: impact of drug formulary access. Osteoporos Int. 2015;26(5):1525-33.

18.   Burns RD, Brusseau TA, Hannon JC. Effect of a Comprehensive School Physical Activity Program on School Day Step Counts in Children. J Phys Act Health. 2015;12(12):1536-42.

19.   Cecil E, Bottle A, Sharland M, Saxena S. Impact of UK Primary Care Policy Reforms on Short-Stay Unplanned Hospital Admissions for Children With Primary Care-Sensitive Conditions. Ann Fam Med. 2015;13(3):214-20.

20.   Chambers EC, Wylie-Rosett J, Blank AE, Ouziel J, Hollingsworth N, Riley RW, et al. Increasing Referrals to a YMCA-Based Diabetes Prevention Program: Effects of Electronic Referral System Modification and Provider Education in Federally Qualified Health Centers. Prev Chronic Dis. 2015;12:E189.

21.   Chung YK, Kim JS, Lee SS, Lee JA, Kim HS, Shin KS, et al. Effect of daily chlorhexidine bathing on acquisition of carbapenem-resistant Acinetobacter baumannii (CRAB) in the medical intensive care unit with CRAB endemicity. Am J Infect Control. 2015;43(11):1171-7.

22.   Corcoran P, Griffin E, Arensman E, Fitzgerald AP, Perry IJ. Impact of the economic recession and subsequent austerity on suicide and self-harm in Ireland: An interrupted time series analysis. Int J Epidemiol. 2015;44(3):969-77.

23.   Dayer MJ, Jones S, Prendergast B, Baddour LM, Lockhart PB, Thornhill MH. Incidence of infective endocarditis in England, 2000-13: a secular trend, interrupted time-series analysis. Lancet. 2015;385(9974):1219-28.

24.   Delcher C, Wagenaar AC, Goldberger BA, Cook RL, Maldonado-Molina MM. Abrupt decline in oxycodone-caused mortality after implementation of Florida's Prescription Drug Monitoring Program. Drug Alcohol Depend. 2015;150:63-8.

25.   Devkaran S, O'Farrell PN. The impact of hospital accreditation on quality measures: an interrupted time series analysis. BMC Health Serv Res. 2015;15:137.

26.   Dinh MM, Green TC, Bein KJ, Lo S, Jones A, Johnson T. Emergency department clinical redesign, team-based care and improvements in hospital performance: A time series analysis. Emerg Med Australas. 2015;27(4):317-22.

27.   Druetz T, Fregonese F, Bado A, Millogo T, Kouanda S, Diabate S, et al. Abolishing Fees at Health Centers in the Context of Community Case Management of Malaria: What Effects on Treatment-Seeking Practices for Febrile Children in Rural Burkina Faso? PLoS ONE. 2015;10(10):e0141306.

28.   Dusetzina SB, Ellis S, Freedman RA, Conti RM, Winn AN, Chambers JD, et al. How Do Payers Respond to Regulatory Actions? The Case of Bevacizumab. J Oncol Pract. 2015;11(4):313-8.

29.   Eichler K, Hess S, Riguzzi M, Can U, Brugger U. Impact evaluation of Swiss Medical Board reports on routine care in Switzerland: a case study of PSA screening and treatment for rupture of anterior cruciate ligament. Swiss Med Wkly. 2015;145:w14140.

30.   Feigl AB, Salomon JA, Danaei G, Ding EL, Calvo E. Teenage smoking behaviour following a high-school smoking ban in Chile: interrupted time-series analysis. Bull World Health Organ. 2015;93(7):468-75.

31.   Flett KB, Ozonoff A, Graham DA, Sandora TJ, Priebe GP. Impact of Mandatory Public Reporting of Central Line-Associated Bloodstream Infections on Blood Culture and Antibiotic Utilization in Pediatric and Neonatal Intensive Care Units. Infect Control Hosp Epidemiol. 2015;36(8):878-85.

32.   Gale M, Muscatello DJ, Dinh M, Byrnes J, Shakeshaft A, Hayen A, et al. Alcopops, taxation and harm: a segmented time series analysis of emergency department presentations. BMC Public Health. 2015;15:468.

33.   Gallagher N, Cardwell C, Hughes C, O'Reilly D. Increase in the pharmacological management of Type 2 diabetes with pay-for-performance in primary care in the UK. Diabet Med. 2015;32(1):62-8.

34.   Gharbi M, Moore LS, Gilchrist M, Thomas CP, Bamford K, Brannigan ET, et al. Forecasting carbapenem resistance from antimicrobial consumption surveillance: Lessons learnt from an OXA-48-producing Klebsiella pneumoniae outbreak in a West London renal unit. Int J Antimicrob Agents. 2015;46(2):150-6.

35.   Giladi AM, Aliu O, Chung KC. The Effect of Medicaid Expansion on Delivery of Finger and Thumb Replantation Care to Medicaid Beneficiaries and the Uninsured. Plast Reconstr Surg. 2015;136(5):640e-7e.

36.   Giladi AM, Chung KC, Aliu O. Changes in use of autologous and prosthetic postmastectomy reconstruction after medicaid expansion in New York state. Plast Reconstr Surg. 2015;135(1):53-62.

37.   Gonzalo JD, Kuperman EF, Chuang CH, Lehman E, Glasser F, Abendroth T. Impact of an Overnight Internal Medicine Academic Hospitalist Program on Patient Outcomes. J Gen Intern Med. 2015;30(12):1795-802.

38.   Hacker KA, Penfold RB, Arsenault LN, Zhang F, Soumerai SB, Wissow LS. Effect of Pediatric Behavioral Health Screening and Colocated Services on Ambulatory and Inpatient Utilization. Psychiatr Serv. 2015;66(11):1141-8.

39.   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.

40.   Hamilton I, Lloyd C, Bland JM, Savage Grainge A. The impact of assertive outreach teams on hospital admissions for psychosis: a time series analysis. J Psychiatr Ment Health Nurs. 2015;22(7):484-90.

41.   Han E, Chae SM, Kim NS, Park S. Effects of pharmaceutical cost containment policies on doctors' prescribing behavior: Focus on antibiotics. Health Policy. 2015;119(9):1245-54.

42.   Hanatani T, Sai K, Tohkin M, Segawa K, Saito Y. Impact of Japanese regulatory action on metformin-associated lactic acidosis in type II diabetes patients. Int J Clin Pharm. 2015;37(3):537-45.

43.   Haren MT, Setchell J, John DL, Daniel M. The impacts of withdrawal and replacement of general practitioner services on aeromedical service trends: a 13-year interrupted time-series study in Tennant Creek, Northern Territory. BMC Health Serv Res. 2015;15:456.

44.   Hartung DM, Middleton L, Markwardt S, Williamson K, Ketchum K. Changes in long-acting beta-agonist utilization after the FDA's 2010 drug safety communication. Clin Ther. 2015;37(1):114-23.e1.

45.   Hernandez-Santiago V, Marwick CA, Patton A, Davey PG, Donnan PT, Guthrie B. Time series analysis of the impact of an intervention in Tayside, Scotland to reduce primary care broad-spectrum antimicrobial use. J Antimicrob Chemother. 2015;70(8):2397-404.

46.   Hersh AL, De Lurgio SA, Thurm C, Lee BR, Weissman SJ, Courter JD, et al. Antimicrobial stewardship programs in freestanding children's hospitals. Pediatrics. 2015;135(1):33-9.

47.   Hinde JM, Bray JW, Aldridge A, Zarkin GA. The Impact of a Mandated Trauma Center Alcohol Intervention on Readmission and Cost per Readmission in Arizona. Med Care. 2015;53(7):639-45.

48.   Hsu JC, Cheng CL, Ross-Degnan D, Wagner AK, Zhang F, Kao Yang YH, et al. Effects of safety warnings and risk management plan for Thiazolidinediones in Taiwan. Pharmacoepidemiol Drug Saf. 2015;24(10):1026-35.

49.   Hsu JC, Ross-Degnan D, Wagner AK, Zhang F, Lu CY. How Did Multiple FDA Actions Affect the Utilization and Reimbursed Costs of Thiazolidinediones in US Medicaid? Clin Ther. 2015;37(7):1420-32.e1.

50.   Hutcheon JA, Strumpf EC, Harper S, Giesbrecht E. Maternal and neonatal outcomes after implementation of a hospital policy to limit low-risk planned caesarean deliveries before 39 weeks of gestation: an interrupted time-series analysis. Bjog. 2015;122(9):1200-6.

51.   Hutchinson AF, Parikh S, Tacey M, Harvey PA, Lim WK. A longitudinal cohort study evaluating the impact of a geriatrician-led residential care outreach service on acute healthcare utilisation. Age Ageing. 2015;44(3):365-70.

52.   Ir P, Korachais C, Chheng K, Horemans D, Van Damme W, Meessen B. Boosting facility deliveries with results-based financing: a mixed-methods evaluation of the government midwifery incentive scheme in Cambodia. BMC Pregnancy Childbirth. 2015;15:170.

53.   Jenkins TC, Knepper BC, Shihadeh K, Haas MK, Sabel AL, Steele AW, et al. Long-term outcomes of an antimicrobial stewardship program implemented in a hospital with low baseline antibiotic use. Infect Control Hosp Epidemiol. 2015;36(6):664-72.

54.   Jiang M, Hughes DR, Duszak R, Jr. Screening Mammography Rates in the Medicare Population before and after the 2009 U.S. Preventive Services Task Force Guideline Change: An Interrupted Time Series Analysis. Womens Health Issues. 2015;25(3):239-45.

55.   Johns M, Farley SM, Rajulu DT, Kansagra SM, Juster HR. Smoke-free parks and beaches: an interrupted time-series study of behavioural impact in New York City. Tob Control. 2015;24(5):497-500.

56.   Judge A, Wallace G, Prieto-Alhambra D, Arden NK, Edwards CJ. Can the publication of guidelines change the management of early rheumatoid arthritis? An interrupted time series analysis from the United Kingdom. Rheumatology (Oxford). 2015;54(12):2244-8.

57.   Kalkhoran S, Sebrie EM, Sandoya E, Glantz SA. Effect of Uruguay's National 100% Smokefree Law on Emergency Visits for Bronchospasm. Am J Prev Med. 2015;49(1):85-8.

58.   Kawai AT, Calderwood MS, Jin R, Soumerai SB, Vaz LE, Goldmann D, et al. Impact of the Centers for Medicare and Medicaid Services Hospital-Acquired Conditions Policy on Billing Rates for 2 Targeted Healthcare-Associated Infections. Infect Control Hosp Epidemiol. 2015;36(8):871-7.

59.   Kim SH, Cho BL, Shin DW, Hwang SS, Lee H, Ahn EM, et al. The Effect of Asthma Clinical Guideline for Adults on Inhaled Corticosteroids PrescriptionTrend: A Quasi-Experimental Study. J Korean Med Sci. 2015;30(8):1048-54.

60.   Kontopantelis E, Olier I, Planner C, Reeves D, Ashcroft DM, Gask L, et al. Primary care consultation rates among people with and without severe mental illness: a UK cohort study using the Clinical Practice Research Datalink. BMJ Open. 2015;5(12):e008650.

61.   Koskinen H, Mikkola H, Saastamoinen LK, Ahola E, Martikainen JE. Time Series Analysis on the Impact of Generic Substitution and Reference Pricing on Antipsychotic Costs in Finland. Value Health. 2015;18(8):1105-12.

62.   Lal S, Ndyomugenyi R, Alexander ND, Lagarde M, Paintain L, Magnussen P, et al. Health Facility Utilisation Changes during the Introduction of Community Case Management of Malaria in South Western Uganda: An Interrupted Time Series Approach. PLoS ONE. 2015;10(9):e0137448.

63.   Laverty AA, Elkin SL, Watt HC, Millett C, Restrick LJ, Williams S, et al. Impact of a COPD discharge care bundle on readmissions following admission with acute exacerbation: interrupted time series analysis. PLoS ONE. 2015;10(2):e0116187.

64.   Liu C, Zhang X, Wan J. Public reporting influences antibiotic and injection prescription in primary care: a segmented regression analysis. J Eval Clin Pract. 2015;21(4):597-603.

65.   Lobos AT, Fernandes R, Willams K, Ramsay C, McNally JD. Routine Medical Emergency Team Assessments of Patients Discharged From the PICU: Description of a Medical Emergency Team Follow-Up Program. Pediatr Crit Care Med. 2015;16(4):359-65.

66.   Lu CY, Zhang F, Lakoma MD, Butler MG, Fung V, Larkin EK, et al. Asthma Treatments and Mental Health Visits After a Food and Drug Administration Label Change for Leukotriene Inhibitors. Clin Ther. 2015;37(6):1280-91.

67.   Ma T, Byrne PA, Haya M, Elzohairy Y. Working in tandem: The contribution of remedial programs and roadside licence suspensions to drinking and driving deterrence in Ontario. Accid Anal Prev. 2015;85:248-56.

68.   Mahant S, Hall M, Ishman SL, Morse R, Mittal V, Mussman GM, et al. Association of National Guidelines With Tonsillectomy Perioperative Care and Outcomes. Pediatrics. 2015;136(1):53-60.

69.   Martin BI, Lurie JD, Tosteson AN, Deyo RA, Farrokhi FR, Mirza SK. Use of bone morphogenetic protein among patients undergoing fusion for degenerative diagnoses in the United States, 2002 to 2012. Spine J. 2015;15(4):692-9.

70.   Marufu O, Desai N, Aldred D, Brown T, Eltringham I. Analysis of interventions to reduce the incidence of Clostridium difficile infection at a London teaching hospital trust, 2003-2011. J Hosp Infect. 2015;89(1):38-45.

71.   Mayer F, Kirchmayer U, Di Martino M, Agabiti N, Fusco D, Davoli M. [Impact of a Decree on ACE inhibitors/ARBs in cardiovascular secondary prevention in the Lazio Region (Central Italy): a pre-post analysis]. Epidemiol Prev. 2015;39(1):19-27.

72.   McGinty EE, Busch SH, Stuart EA, Huskamp HA, Gibson TB, Goldman HH, et al. Federal parity law associated with increased probability of using out-of-network substance use disorder treatment services. Health Aff (Millwood). 2015;34(8):1331-9.

73.   McIlroy G, Thomas SK, Coleman JJ. Second-generation antipsychotic drug use in hospital inpatients with dementia: the impact of a safety warning on rates of prescribing. J Public Health (Oxf). 2015;37(2):346-52.

74.   McKenna B, Furness T, Brown S, Tacey M, Hiam A, Wise M. Police and clinician diversion of people in mental health crisis from the Emergency Department: a trend analysis and cross comparison study. BMC emerg. 2015;15:14.

75.   Mehta SJ, Polsky D, Zhu J, Lewis JD, Kolstad JT, Loewenstein G, et al. ACA-mandated elimination of cost sharing for preventive screening has had limited early impact. Am J Manag Care. 2015;21(7):511-7.

76.   Milder EA, Rizzi MD, Morales KH, Ross RK, Lautenbach E, Gerber JS. Impact of a new practice guideline on antibiotic use with pediatric tonsillectomy. JAMA Otolaryngol Head Neck Surg. 2015;141(5):410-6.

77.   Narayan H, Thomas SH, Eddleston M, Dear JW, Sandilands E, Bateman DN. Disproportionate effect on child admissions of the change in Medicines and Healthcare Products Regulatory Agency guidance for management of paracetamol poisoning: an analysis of hospital admissions for paracetamol overdose in England and Scotland. Br J Clin Pharmacol. 2015;80(6):1458-63.

78.   Newitt S, Myles PR, Birkin JA, Maskell V, Slack RC, Nguyen-Van-Tam JS, et al. Impact of infection control interventions on rates of Staphylococcus aureus bacteraemia in National Health Service acute hospitals, East Midlands, UK, using interrupted time-series analysis. J Hosp Infect. 2015;90(1):28-37.

79.   Ng W, Brown A, Alexander D, Ho MF, Kerr B, Amato M, et al. A multifaceted prevention program to reduce infection after cesarean section: Interventions assessed using an intensive postdischarge surveillance system. Am J Infect Control. 2015;43(8):805-9.

80.   Nguyen HV, Grootendorst P. Intended and unintended effects of restrictions on the sale of cigarillos to youth: evidence from Canada. Tob Control. 2015;24(4):382-8.

81.   Owusu-Edusei K, Jr., Carroll DS, Gift TL. Examining Fluoroquinolone Claims Among Gonorrhea-Associated Prescription Drug Claims, 2000-2010. Am J Prev Med. 2015;49(5):761-4.

82.   Pape UJ, Huckvale K, Car J, Majeed A, Millett C. Impact of 'stretch' targets for cardiovascular disease management within a local pay-for-performance programme. PLoS ONE. 2015;10(3):e0119185.

83.   Perkins RB, Legler A, Hanchate A. Trends in Male and Female Genital Warts Among Adolescents in a Safety-Net Health Care System 2004-2013: Correlation With Introduction of Female and Male Human Papillomavirus Vaccination. Sex Transm Dis. 2015;42(12):665-8.

84.   Petrou P. The Ariadne's thread in co-payment, primary health care usage and financial crisis: findings from Cyprus public health care sector. Public Health. 2015;129(11):1503-9.

85.   Petrou P. An Interrupted Time-Series Analysis to Assess Impact of Introduction of Co-Payment on Emergency Room Visits in Cyprus. Appl Health Econ Health Policy. 2015;13(5):515-23.

86.   Pieroni L, Muzi G, Quercia A, Lanari D, Rundo C, Minelli L, et al. Estimating the Smoking Ban Effects on Smoking Prevalence, Quitting and Cigarette Consumption in a Population Study of Apprentices in Italy. Int J Environ Res Public Health. 2015;12(8):9523-35.

87.   Roberts SC, Zahnd E, Sufrin C, Armstrong MA. Does adopting a prenatal substance use protocol reduce racial disparities in CPS reporting related to maternal drug use? A California case study. J Perinatol. 2015;35(2):146-50.

88.   Robinson TE, Zhou L, Kerse N, Scott JD, Christiansen JP, Holland K, et al. Evaluation of a New Zealand program to improve transition of care for older high risk adults. Australas J Ageing. 2015;34(4):269-74.

89.   Rosenstein MG, Nijagal M, Nakagawa S, Gregorich SE, Kuppermann M. The Association of Expanded Access to a Collaborative Midwifery and Laborist Model With Cesarean Delivery Rates. Obstet Gynecol. 2015;126(4):716-23.

90.   Rutkow L, Chang HY, Daubresse M, Webster DW, Stuart EA, Alexander GC. Effect of Florida's Prescription Drug Monitoring Program and Pill Mill Laws on Opioid Prescribing and Use. JAMA Intern Med. 2015;175(10):1642-9.

91.   Santa-Ana-Tellez Y, Mantel-Teeuwisse AK, Leufkens HG, Wirtz VJ. Seasonal variation in penicillin use in Mexico and Brazil: analysis of the impact of over-the-counter restrictions. Antimicrob Agents Chemother. 2015;59(1):105-10.

92.   Sarma JB, Marshall B, Cleeve V, Tate D, Oswald T, Woolfrey S. Effects of fluoroquinolone restriction (from 2007 to 2012) on Clostridium difficile infections: interrupted time-series analysis. J Hosp Infect. 2015;91(1):74-80.

93.   Sarma JB, Marshall B, Cleeve V, Tate D, Oswald T, Woolfrey S. Effects of fluoroquinolone restriction (from 2007 to 2012) on resistance in Enterobacteriaceae: interrupted time-series analysis. J Hosp Infect. 2015;91(1):68-73.

94.   Schaffer AL, Buckley NA, Dobbins TA, Banks E, Pearson SA. The crux of the matter: Did the ABC's Catalyst program change statin use in Australia? Med J Aust. 2015;202(11):591-5.

95.   Shelton JB, Ochotorena L, Bennett C, Shekelle P, Kwan L, Skolarus T, et al. Reducing PSA-Based Prostate Cancer Screening in Men Aged75 Years and Older with the Use of Highly Specific Computerized Clinical Decision Support. J Gen Intern Med. 2015;30(8):1133-9.

96.	Smith JJ, Johnston JM, Hiratsuka VY, Dillard DA, Tierney S, Driscoll DL. Medical home implementation and trends in diabetes quality measures for AN/AI primary care patients. Prim Care Diabetes. 2015;9(2):120-6.

97.	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;69(11):1118-24.

98.	Stocks SJ, McNamee R, Turner S, Carder M, Agius RM. The impact of national-level interventions to improve hygiene on the incidence of irritant contact dermatitis in healthcare workers: changes in incidence from 1996 to 2012 and interrupted times series analysis. Br J Dermatol. 2015;173(1):165-71.

99.	Taggart LR, Leung E, Muller MP, Matukas LM, Daneman N. Differential outcome of an antimicrobial stewardship audit and feedback program in two intensive care units: a controlled interrupted time series study. BMC Infect Dis. 2015;15:480.

100.	Tippett EC, Chen BK. Association of Attorney Advertising and FDA Action with Prescription Claims: A Time Series Segmented Regression Analysis. Drug Saf. 2015;38(12):1169-78.

101.	Tung YC, Chang GM, Cheng SH. Long-term effect of fee-for-service-based reimbursement cuts on processes and outcomes of care for stroke: interrupted time-series study from Taiwan. Circ Cardiovasc Qual Outcomes. 2015;8(1):30-7.

102.	Valerio M, Munoz P, Rodriguez CG, Caliz B, Padilla B, Fernandez-Cruz A, et al. Antifungal stewardship in a tertiary-care institution: a bedside intervention. Clin Microbiol Infect. 2015;21(5):492.e1-9.

103.	Vaz LE, Kleinman KP, Kawai AT, Jin R, Kassler WJ, Grant PS, et al. Impact of Medicare's Hospital-Acquired Condition policy on infections in safety net and non-safety net hospitals. Infect Control Hosp Epidemiol. 2015;36(6):649-55.

104.	Wagenaar AC, Livingston MD, Staras SS. Effects of a 2009 Illinois Alcohol Tax Increase on Fatal Motor Vehicle Crashes. Am J Public Health. 2015;105(9):1880-5.

105.	Yang JH, Kim M, Park YT, Lee EK, Jung CY, Kim S. The effect of the introduction of a nationwide DUR system where local DUR systems are operating--The Korean experience. Int J Med Inf. 2015;84(11):912-9.

11.6.3 Contribution of the candidate to the paper

- As with the other papers, the concept of this article was my own.
- There were two components to this paper: 1. Development of a new reporting framework, 2. A literature review of the reporting of methods in applied ITS studies.
- I first identified existing reporting frameworks that might be relevant to ITS studies, systematically went through these, considering which components might be included within an ITS reporting framework and collating these.
- I built on existing methodological literature as well as the work that I had developed in chapters 4,5 and 6 to identify other methodological issues pertinent to ITS that were not covered in existing reporting criteria.
- I drew this together to develop ITS specific reporting criteria that were loosely modelled on the TREND (Transparent Reporting of Non-randomised Designs) criteria.
- While these criteria were developed after undertaking the case studies (and therefore not applied to the case studies) I retrospectively checked through the methods reporting of my case studies to identify any further components that should be included within the FERITS statement.
- For the literature review, I defined the search criteria to keep this manageable and appropriate to the aims of the study (providing an illustrative example of current reporting) and decided on what data I wanted to extract from the papers.
- I undertook the literature review including identifying papers, deduplicating, excluding those that were non-eligible and reviewing the methods reporting and extracting the relevant data.
- I then undertook the descriptive analysis of the extracted data and presented the results.
- Finally, I wrote the paper with input from my co-authors on both the reporting criteria and the main body of the manuscript following an initial draft.