Aus dem Institut für Medizinische Informatik, Biometrie und Epidemiologie (IBE)

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.) in

Medical Research - Epidemiology und Public Health

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

ASSESSING THE EFFECTIVENESS OF PUBLIC HEALTH INTERVENTIONS:

A METHODOLOGICAL INVESTIGATION OF CONTROLLED BEFORE-AFTER AND INTERRUPTED TIME SERIES STUDY DESIGNS

Vorgelegt von

Stephanie Assmann-Polus

aus

Heidelberg

am

28.04.2021

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

First supervisor: *Prof. Eva Rehfuess*

Second supervisor*:* *Prof. Ulrich Mansmann*

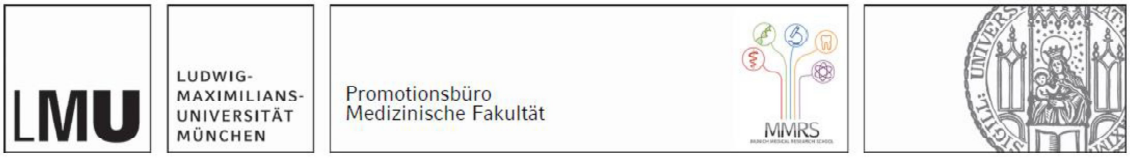
Third supervisor: *Prof. Atle Fretheim*

Dean: Prof. Dr. Thomas Gudermann

Datum der Verteidigung:

28.10.2021

Affidavit



Affidavit

Assmann-Polus, Stephanie
_____
Surname, first name

_____
Street

_____
Zip code, town, country

I hereby declare, that the submitted thesis entitled:

ASSESSING THE EFFECTIVENESS OF PUBLIC HEALTH INTERVENTIONS: A METHODOLOGICAL INVESTIGATION OF CONTROLLED BEFORE-AFTER AND INTERRUPTED TIME SERIES STUDY DESIGNS

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Munich, 12.04.2021 Stephanie Assmann-Polus

_____ _____

place, date Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Assmann-Polus, Stephanie

_____

Surname, first name

_____

Street

_____

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

ASSESSING THE EFFECTIVENESS OF PUBLIC HEALTH INTERVENTIONS: A METHODOLOGICAL INVESTIGATION OF CONTROLLED BEFORE-AFTER AND INTERRUPTED TIME SERIES STUDY DESIGNS

.....

is congruent with the printed version both in content and format.

Munich, 12.04.2021 Stephanie Assmann-Polus

_____ _____

place, date Signature doctoral candidate

Table of contents

List of abbreviations

CBA	Controlled before-after
CBNCP	Community based newborn care package
CI	Confidence Interval
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
DALYs	Disability-adjusted life years
DiD	Difference-in-differences
DRKS	German national trial registry (Deutsches Register Klinischer Studien)
EBPH	Evidence-based public health
EPOC	Effective Practice and Organisation of Care
EU	European Union
FCTC	Framework Convention on Tobacco Control
FERITS	Framework for Enhanced Reporting of Interrupted Time Series
ITS	Interrupted time series
LMICs	Low- and middle income countries
NRCTs	Nonrandomised, controlled trials
OR	Odds ratio
PM	Particulate matter
RCT	Randomised Controlled Trial
ROBINS-I	Cochrane risk of bias tool to assess non-randomized studies of interventions
SARS-CoV-2	Severe acute respiratory syndrome coronavirus type 2
WHO	World Health Organization

List of publications

No.	Title	Authors	Journal	Impact factor	Rank of publication	Reference
I	Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews.	Stephanie Polus Dawid Pieper Jacob Burns Atle Fretheim Craig Ramsay Julian PT Higgins Tim Mathes Lisa M Pfadenhauer Eva A Rehfuess	Journal of Clinical Epidemiology	4.952	20/181	[1]
II	Interrupted time series study found mixed effects of the impact of the Bavarian smoke-free legislation on pregnancy outcomes	Stephanie Polus Jacob Burns Sabine Hoffmann Tim Mathes Ulrich Mansmann Jasper V Been Nicholas Lack Daniela Koller Werner Maier Eva A Rehfuess	Nature Scientific Reports	3.998	17/71	[2]

List of additional publications with reference to the PhD topic

No.	Title	Authors	Journal	Reference
III	The choice of analysis in controlled before-after studies and its impact on effect size and statistical precision	Stephanie Polus Jacob Burns Sara Pedron Deepak Paudel Eva A Rehfuess	Unpublished Manuscript	[3]

Conference presentations and workshops with reference to the PhD topic

Conference	**Title**
Cochrane Colloquium Vienna. *Filtering the information overload for better decisions.* 3-7 October 2015. Vienna, Austria	Shedding light on the maze of study labels and features: analysis of CBA and ITS studies in Cochrane Systematic Reviews
EbM Kongress 2016. *Gemeinsam informiert entscheiden*. 3-5 March 2016, Cologne, Germany	Entsprechen „controlled before-after“ (CBA) und „interrupted time series“ (ITS) Studien in Cochrane Reviews den methodischen Mindestkriterien?
Global Evidence Summit. *Using evidence. Improving lives.* 13-16 September 2017, Cape Town, South Africa	The choice of analysis in controlled before-after studies and its impact on effect size and statistical precision
Method Webinar, Department for Evidence-based Medicine and Evaluation, Danube University Krems, 10 November 2016, Krems, Austria	Interrupted time series and controlled before-after studies
EbM Kongress 2018. Brücken bauen – von der Evidenz zum Patientenwohl. 8-10 March 2018. Graz, Austria	Diese verflixten Studien abseits von RCTs – welche Kriterien können angewendet werden, um den Studientyp richtig zu bestimmen?
19. Deutscher Kongress für Versorgungsforschung 2020. 30 September – 1 October 2020. Berlin, Germany	Interrupted-Time-Series Studien: aktuelle methodische Entwicklungen und Stellenwert in der Versorgungsforschung

1. My contribution to the publications

1.1. Contribution to publication I

Contribution of PhD candidate	Contribution of co-authors
SP conceived the study, coordinated the work at all stages of study conduct; developed the methods; wrote the protocol; conducted the searches, performed selection of studies and data extraction (which was done in duplicate with co-authors); performed data analysis and interpreted the data; wrote the manuscript; and incorporated several rounds of feedback/comments by co-authors; (50%).	DP: development of original idea; co-development of methods, data extraction; comments on protocol and manuscript (5%) JB: data extraction; comments on protocol and manuscript (10%) AF: data extraction; comments on protocol and manuscript (7%) CR: data extraction; comments on protocol and manuscript (5%) JPH: data extraction; comments on protocol and manuscript (5%) TM: data extraction; comments on protocol and manuscript (5%) LMP: data extraction; comments on protocol and manuscript (3%) EAR: co-development of protocol; data extraction; comments on manuscript (10%)

1.2. Contribution to publication II

Contribution of PhD candidate	Contribution of co-authors
SP conceived the study, developed the methods, conducted the data management, conducted and interpreted the statistical analysis, wrote the manuscript, incorporated the	JB: co-conception of study, development of methods, support and interpretation of statistical analysis, comments on manuscript (10%)

co-authors' feedback and coordinated the work at all stages of study conduct (50%).	SH and TM: advise and support of the analyses and their interpretation, comments on manuscript (5% each) UM: advise and support of statistical analyses, comments on manuscript (5%) JVB: support of method development, comments on the manuscript (5%) NL: support of data retrieval analyses, comments on the manuscript (5%) DK and WM: method development for socio-economic subgroup analyses, comments on manuscript (2.5% each) EAR: co-conception of study and development of methods, supervision and interpretation of analyses, comments on manuscript (10%)

1.3. Contribution to additional publication III (Appendix A)

Contribution of PhD candidate	**Contribution of co-authors**
SP conceived the study, developed the methods, conducted and interpreted the statistical analysis, and wrote the manuscript (70%)	JB co-conceived the study, interpreted the analysis, and commented on the manuscript. (10%) SPe conducted the analysis, interpreted the analysis, and commented on the 207 manuscript. (10%) DP aided with the statistical analysis and commented on the manuscript. (3%) EAR co-conceived the study and commented on the manuscript (7%)

2. Introductory summary

2.1. Public health challenges, complexity, and evidence-based public heath

Since the "Declaration of Health in all Policies" in the EU Treaty from 2007 (Treaty of Lisbon), public health – i.e. the health of entire populations – is officially recognized as a key driver for the wellbeing of society [4]. Health indirectly influences most areas of human life and hence should be considered in policies across all sectors, budgets and government priorities [5]. Today, as emphasised during the on-going global SARS-CoV-2/COVID-19 pandemic, public health plays a vital role in society addressing complex political, social, economic and environmental problems through approaches that affect multiple sectors [5]. A problem of major public health significance in industrialised and developing countries alike is, for example, the increase in chronic diseases, such as diabetes, hypertension, or chronic obstructive pulmonary disease [5, 6]. Addressing risk factors for chronic diseases thus plays a major role in prevention efforts as detailed in Box 1.

Box 1.Tobacco smoke as a major risk factor worldwide

Tobacco smoke represents a critical risk factor for public health. Smoking and the exposure to secondhand smoke are responsible for over eight million deaths worldwide every year according to the World Health Organisation (WHO) [7]. In 2015, tobacco smoke caused a disease burden of 150 million disability-adjusted life years (DALYs), mainly for cancers, chronic obstructive pulmonary disease (COPD), and other chronic respiratory disease. It was ranked among the top five risk factors for 109 countries and its disease burden is still growing [8]. Interventions targeting this risk factor thus play a major role in public health. To reduce human consumption and exposure to tobacco smoke and thereby limit its impact on mortality and chronic disease prevalence, governments have implemented different interventions. With the adoption of the WHO Framework Convention on Tobacco Control (FCTC) in 2003, all WHO member states are obliged to implement interventions on a range of different policy areas, such as the monitoring of tobacco use and prevention policies, the protection of the population from tobacco use, taxes on tobacco products, and much more [9].

A typical public health question may ask, for example, how to best address a particular risk factor and thus reduce the associated disease burden at the population-level. This is often addressed through behavioural or environmental interventions, i.e. interventions that aim to modify the behaviour of a population (e.g. through health promotion or prevention programmes, such as a smoking cessation programme) or the environment in which a population lives (e.g. through creating smoke-free environments, such as smoke-free work places). Such public health interventions, including also interventions in the field of health systems, health services, and health policy, are often quite complex [10-13]. This complexity, embodied for example in multiple, long and complex causal pathways, contribute to making an evaluation difficult. The effectiveness of an intervention may vary according to user compliance, the delivery of the intervention, health system characteristics, programming and other policy measures, which are part of the wider geographical, socio-economic, political and cultural context [14, 15]. These aspects, among others, make evaluating the effectiveness of public health interventions challenging. Box 2. illustrates potential challenges by providing an example of an environmental public health intervention, the introduction of a smoking ban.

Box 2. Smoking bans in public places as a means to reduce exposure to tobacco smoke

The introduction of a smoking ban to reduce chronic respiratory diseases represents a public health intervention designed to change the physical and social environment in which people live. There may be, however, many different factors that can influence both the effectiveness of the smoking ban, as well as chronic respiratory disease outcomes.

Examples include:

- Acceptability, compliance and enforcement of the smoking ban
- Other interventions targeting tobacco smoke: e.g. tax increases or the introduction of packaging images
- Other public health measures: e.g. low emission zones to reduce ambient air pollution that may have an influence on chronic respiratory disease outcomes
- Changes to the healthcare system: e.g. improvements in healthcare, such as disease management programmes that may influence disease outcomes
- Demographic changes: e.g. an increase of the age structure of the population.

Policy makers are thus confronted with not only pressing but also complex public health problems and need to make challenging decisions that affect whole populations and require limited resources that may miss on other ends. Therefore, these decisions should be based on the best evidence available. Rigorous primary studies as well as systematic reviews represent an essential basis for policymaking and are the cornerstone of evidence-based public health (EBPH) [15, 16]. Systematic reviews identify and appraise all available evidence of relevance to a specific question. Systematic reviews are characterised by their standardized and reproducible way of searching, screening and selecting the studies, including a statistical, graphical or narrative synthesis of the findings to provide an overall effect of the interventions assessed [17]. They can be quantitative assessments of effectiveness and/or cost-effectiveness but can also be qualitative focusing on other important factors such as acceptability or feasibility of an intervention [18]. Cochrane, an international network of researchers and health professionals has focused on conducting high-quality systematic reviews mostly of effectiveness to inform health decisions. Cochrane reviews provide the highest standard of systematic reviews [17]. The answers that systematic reviews can provide, however, are only as good as the evidence from the studies they include [19]. Rigorous studies are therefore an essential basis to provide reliable answers to public health questions and to help decision makers make clear and safe decisions.

2.2. Nonrandomised studies to assess (complex) public health interventions

In part because of the above-mentioned evaluation difficulties with regards to the complexity of many public health questions, different types of studies may need to be put into practice. In clinical settings, randomised controlled trials (RCTs) are considered the gold standard to assess intervention effectiveness because of their ability to account for observable as well as unobservable differences between study groups [20]. RCTs or cluster RCTs are, however, under certain circumstances not always practically, ethically or financially feasible or applicable particularly to assess environmental public health interventions that are implemented on a large-scale population level [15, 21-24]. It would, for example, be very difficult and potentially unethical to randomly select different cities or regions to assess the effectiveness of a smoking ban. Furthermore, in many cases the intervention has already been implemented as the result of a political decision and a retrospective evaluation is necessary.

Under such circumstances, nonrandomised study designs may thus be the only means to assess the effectiveness of public health interventions in real life [25-27]. In nonrandomised studies, the selection of the study participants is based on approaches other than randomisation, for example by self-selection or researcher selection or in the course of usual healthcare [26, 28].

This implies that the study is more prone, for example, to selection bias and confounding, and that it is more difficult to link the effect causally to the intervention. The risk of drawing misleading conclusions is therefore higher than in RCTs, yet they are easier to implement in real world settings [22, 25, 26, 29, 30]. Nonrandomised studies include a large variety of different study designs, such as cohort studies, case-control studies, controlled before-after (CBA) studies, interrupted time series (ITS) studies and nonrandomised, controlled trials (NRCTs) [26].

Increasing with the need of policymakers to decide on complex public health matters, nonrandomised studies have been more frequently conducted as well as included in systematic reviews of interventions [25, 31]. Alongside, primary researchers as well as systematic reviewers have been presented with methodological challenges [25, 27, 32, 33]. Many studies reveal methodological problems and are often poorly described, which complicates the searching for and screening of the studies for review authors [25, 32, 34, 35]. Additional concerns have been raised regarding how to appraise and synthesize the results derived from nonrandomised studies [36], notably with respect to the higher risk of bias associated with these studies [27].

2.3. Two special types of nonrandomised studies: controlled before-after and interrupted time series studies

Before introducing specific study designs, it is worthwhile to clarify the wide range of terminology used to describe nonrandomised studies in general. Nonrandomised studies, for example, are often referred to as observational studies. However, some, including Cochrane, discourage the use of this term for studies assessing intervention effects as it is most commonly used in researching risk factors and exposures [26]. Specifically in the context of intervention research, researchers have distinguished between experimental (i.e. RCTs) and quasi-experimental studies. Quasi-experimental studies are similar in purpose to “real” experimental studies, in that they aim to make assumptions about causal inferences. Quasi-experimental studies do so, however using non-random methods for assignment into intervention and control group [20]. Furthermore, the term “natural experiment” has been increasingly used, concentrating on studies that investigate interventions that occurred naturally and are not under the control of the researcher. These include mostly (large-scale) policy and programme evaluations [29]. Many quasi-experimental studies are therefore also natural experiments.

Different from cohort studies or case-control studies, which can provide useful evidence about disease causes, quasi-experimental studies, such as ITS and CBA studies can be used for evalu-

ating intervention effects [20, 37]. Indeed, ITS and CBA studies are the most often included non-randomised study designs in Cochrane reviews [25, 38], yet within public health research they are not yet fully implemented and understood [25, 26, 38, 39].

2.3.1. The interrupted time series study

The ITS study is considered among the strongest quasi-experimental designs due to its powerful and flexible method for dealing with trend data [40, 41]. It uses data collected at multiple time points over a longer period of time before and after an intervention (the "interruption") to measure an intervention effect against a pre-intervention trend [42]. It thereby predicts the difference between what would have happened in absence of the intervention (the counterfactual) and the actual data observed (see Figure 1). The study design lends itself to retrospective evaluations using routine data [29, 42] and of interventions and outcomes measured at population level [37, 43]. The ITS study is particularly suitable for the evaluation of natural experiments, that is of naturally occurring events or interventions that were implemented without direct researcher involvement [29]. The intervention effect can be measured as a change in level of the outcome, i.e. an immediate drop or increase of the outcome, as well as a change in slope, i.e. a change in the outcome trend over the post-intervention time period [35].

The ITS study design allows, through appropriate statistical analysis, to control for secular trends, thereby accounting for a natural increase or decrease of the outcome over time that could have otherwise been wrongly attributed to the intervention effect. It also allows for the investigation of seasonal or other regular effects, random fluctuations of the outcome, autocorrelation of the data points, and duration of the intervention effect [35]. As the intervention effect is usually measured within the same population, selection bias and confounding due to group differences do not pose a problem [37].

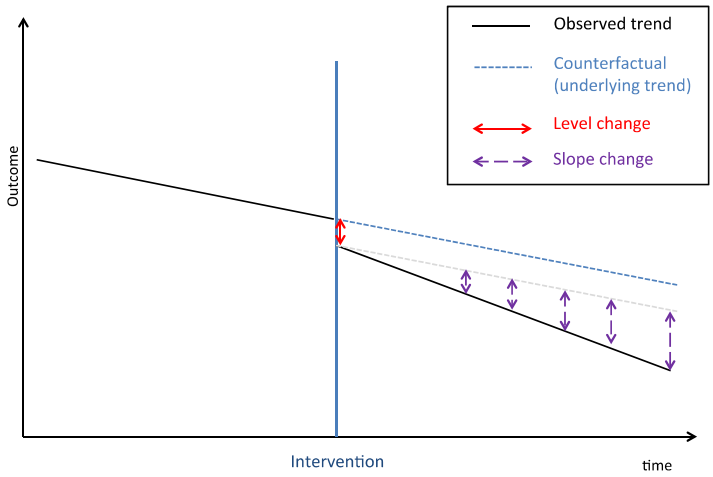


Figure 1. Illustration of the interrupted time series study design (adapted from Grimshaw et al. [44])

Box 3. Example of an ITS study: evaluation of flavoured cigar regulations with wholesale tobacco volumes in Canada

The aim of the study was to investigate the impact of a Canadian policy implemented in 2009 that prohibited flavours (except the flavour menthol) in small cigars on sales of these products. Using an interrupted time series design, the study authors analysed wholesale unit data from early-2001 until the end of 2016, which are reported quarterly to the Canadian government. The analyses were adjusted for seasonal trend and both sales of tobacco products with and without flavour were analysed. The regulations were associated with a decrease of 59 million units (95% CI −86.0 to −32.4) of flavoured cigars. The study shows that the Canadian public health policy restricting flavours in tobacco products had a substantial impact on cigar sales. The impact of a flavour ban could be increased by a comprehensive policy including all flavours and all product types without exemptions.

Chaiton et al. [45]

While public health research has recognized the value of applying ITS studies [23, 25, 46], there is still some discussion going on as to what constitutes an ITS study [47]. Recent studies have focused on clarifying the methodological concepts of the ITS study design [23, 37, 43, 46, 48,

49], looking more closely at the analysis approaches [23, 35] and comparing the effect estimates to those of cluster RCTs [50-52].

Ramsay et al found that most ITS studies were inappropriately analysed, neglecting co-intervention effects and leading in many cases to a statistical overestimation of the effect estimates. Another study found that many ITS studies include an insufficient number of pre-intervention time points, which leads to difficulties in estimating an underlying secular trend and thus to wrong conclusions of significant effects [53]. The Cochrane Effective Practice and Organisation of Care (EPOC), a Cochrane group, which focuses on assessing health care interventions on a population level, has included CBA and ITS studies in their reviews. In the absence of any formal guidance they developed pragmatic methodological criteria for review authors in order to establish a minimum threshold of study quality included in the systematic reviews. For ITS studies, these include a clearly defined intervention time point and a minimum of three time points each before and after the intervention.

2.3.2. The controlled before-after study

CBA studies evaluate an intervention effect by investigating observations, which are made before and after implementation of an intervention, in a group that receives the intervention and in a control group that does not. In CBA studies, clusters such as healthcare centres or hospitals are allocated to an intervention or control group. The analysis compares change scores in the intervention and control group, assuming parallel trends between the groups [54, 55]. The control group should therefore be similar in terms of baseline characteristics and performance to the intervention group, assuming the intervention effect to be caused by the intervention implemented in the intervention group [53, 54].

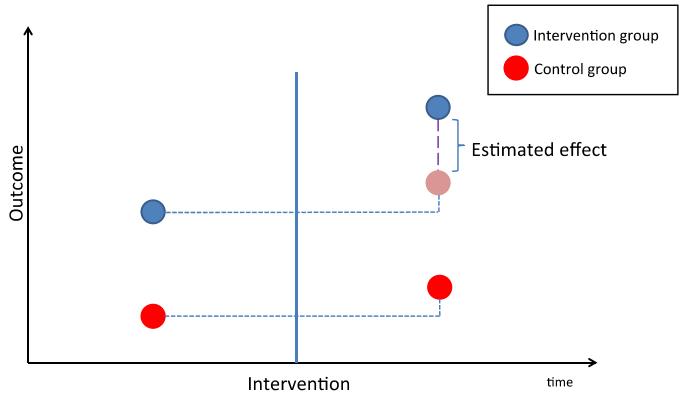


Figure 2. Illustration of the controlled before-after study design (adapted from Grimshaw et al. [44])

Box 4. Example of a CBA study: Evaluation of an educational intervention for smoke-free adolescents

The aim of the study was to assess the association between an educational intervention to prevent the onset of smoking among adolescents and the smoking prevalence in two high schools in La Plata, Argentina, implementing a controlled before-after study design. Adolescents between 12 and 13 years in one of the high schools received the educational intervention. An assessment of smoking prevalence was conducted in both high schools in 2010 immediately prior to the educational intervention and in 2011 and 2012, for follow-up. Multiple regression models were conducted to analyse the association between the intervention and smoking prevalence, adjusting for determinants that were associated with adolescent smoking (smoking family members, sex). The adjusted odds ratio (OR) for smoking at the high school where adolescents received the intervention as compared to the school that did not receive the educational intervention was 0.54 (95% confidence interval (CI) 0.35-0.83) in 2011 and 0.98 (95%CI 0.60-1.61) in 2012. The educational intervention had a positive short-term impact, which was, however, not maintained over time.

Gulayin et al. [56]

The methodological literature on CBA studies is, however, limited. Few studies have looked at CBA studies among other nonrandomised studies and disagree on methodological characteristics, such as investigator control over the intervention allocation and other features that may separate the CBA study design from other designs, such as the nonrandomised controlled trial [22, 24, 39, 54]. Cochrane EPOC defines minimum methodological criteria, such as at least two sites per intervention and control group and contemporaneous data collection in both groups (EPOC). Eccles et al. [53] and Grimshaw et al. [54] emphasize the need for an appropriate control group and appropriate statistical analysis. The challenge for CBA studies is the selection of an appropriate control group and there are several methods, such as propensity score matching, to attempt to have equal baseline measures. However, even in CBA studies with apparently well-matched control groups performance at baseline may differ. In many cases this has led to inappropriate statistical analyses, such as "within group" analyses where intervention and control group differences are not directly compared [53, 54].

2.4. The need for more clarity on interrupted time series and controlled before-after studies

Overall, there is no coherent picture as to how CBA and ITS studies should be designed, conducted and analysed in the most appropriate way. Linked to this problem, systematic reviewers have difficulties to search for and include CBA and ITS studies, while at the same time their higher internal validity compared to other nonrandomised study designs triggers discussions about how to reflect this in systematic review synthesis and quality assessment [12]. This situation is thus challenging both for primary researchers evaluating the effectiveness of public health evaluations, as well as for those synthesizing these studies at the systematic review level.

Researchers would benefit from a clearer picture of CBA and ITS studies considering study design, data collection and analysis in order to properly conduct these studies. Likewise, this clarity would make it easier for systematic reviewers to search for, include and synthesize CBA and ITS study designs in systematic reviews [57]. Ultimately, decision makers would have more certainty in terms of recommending for or against an intervention, if the studies included in a systematic review were of higher quality [58].

2.5. Objectives

The objectives of this doctoral thesis were therefore to

i. investigate how CBA and ITS study designs have been conducted at the primary study level and have been utilized at the systematic review level

ii. contribute to improving the use and conduct of CBA and ITS studies through the conduct of good-practice primary studies and a methodological investigation of study design and analysis characteristics

2.6. Overview of PhD thesis

This PhD thesis comprises two first author publications that were published in highly ranked international journals.

I. 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**: p. 56-69.

II. Polus S, Burns J, Hoffmann S, Mathes T, Mansmann U, Been JV, et al. *Interrupted time series study found mixed effects of the impact of the Bavarian smoke-free legislation on pregnancy outcomes*. Sci Rep, 2021. **11**(1): p. 4209.

It further comprises an additional publication, to be submitted shortly, which is not a formal part of this doctoral thesis but was conducted within the context of the doctoral research.

III. Polus S, Burns J, Pedron S, Paudel D, and Rehfuess E, *The choice of analysis in controlled before-after studies and its impact on effect size and statistical precision*. Unpublished Manuscript.

2.7. Publication I. Application, design, and analysis characteristics for controlled before-after and interrupted time series studies

2.7.1. Background

Given the above outlined challenges related to CBA and ITS studies, the first project of this thesis set out to examine detailed characteristics of CBA and ITS studies. The resulting first publication investigated the use (application characteristics) and the conduct (design and analysis characteristics) of the studies on two levels, the primary research level as well as the systematic review level. Specifically, the study examined for what type of interventions the studies were applied, the study design characteristics themselves and how the analyses were undertaken in the studies. It further looked at the characteristics of the Cochrane reviews that included the CBA and ITS studies and how the review authors as well as the primary researchers defined and labelled the studies. This was done to receive an updated and more detailed picture of the systematic reviews that included nonrandomised studies, following previous study findings of characteristics of systematic reviews including nonrandomised study designs by Ijaz et al. [38].

2.7.2. Methods

For this purpose, a group of experts with methodological expertise in these study designs as well as systematic reviewers were gathered. The study included a two-step process with selection and data extraction at two levels, the systematic review level as well as primary study level. In a first step, Cochrane systematic reviews published between June 2012 and March 2015 that included nonrandomised study design were selected. For these systematic reviews detailed characteristics, such as which types of study designs they included, were examined. In a second step, the systematic reviews were categorized according to ten pre-specified intervention types. Two reviews per intervention type were selected that included at least two studies using the respective study designs. In this way, an assessment of a heterogeneous sample of studies applied in different fields of health research was ensured. The PhD candidate together with one of the co-authors then independently extracted detailed information on the publication (i.e. general aspects, such as publication year), application (e.g. type and level of intervention), and methodological (including information about study design, data collection and data analysis) characteristics, and reported strengths and weaknesses of the study designs. The experienced group of co-authors ensured a detailed and thorough examination of the included studies, which required in many cases extensive scientific discussion.

2.7.3. Results

Out of 136 reviews that considered nonrandomised studies and included 1956 primary studies, 69 explicitly considered CBA and ITS studies. After categorization of these systematic reviews

into intervention type assessed, 21 CBA and 18 ITS studies were selected for primary study analysis. Due to the abundance of information gathered on the characteristics of the systematic reviews as well as primary studies, the final publication includes only the most relevant study findings while the data extraction was much more elaborate. The Cochrane reviews revealed great differences with regards to definitions and criteria of CBA and ITS studies. While most reviews reported to have used the EPOC criteria as a selection criterion, it seemed difficult for many review authors to then correctly apply the criteria to the primary study under consideration. Some included primary studies did not comply with the EPOC criteria or even more basic characteristics that define the study designs, such as a control group for the CBA study and an "interruption", i.e. an intervention for the ITS study. On a primary study level the findings showed a discrepancy between study design and the actual statistical analysis, showing that researchers often did not use the collected data in the statistical analysis. However, poor reporting of the primary studies often prevented identification of the analysis methods.

2.7.4. Discussion

A detailed discussion of key characteristics of both study designs and considered related challenges aims to distil key aspects towards accepted definitions and features of CBA and ITS studies. For example, for CBA studies the advantage of including at least two intervention and two appropriate control sites is pointed out. Further, the potential problem of defining what constitutes the site in case of higher-level clustering is discussed. For ITS studies, for example, the discussion explicitly mentions the importance of the statistical analysis, which should include multiple measurements over time, adjusting for important secular trends. The study findings resulted in and contributed to several further projects that were conducted in the context of this PhD thesis and are therefore discussed below.

2.7.5. Dissemination and further use of study findings

The insights of the first project were disseminated in subsequent years at workshops and presentations in Germany and abroad for primary study researchers and systematic reviewers to increase their understanding and thus conduct of the study designs and their correct use in systematic reviews. In these workshops the study designs were explained in detail and information given on what to consider when (a) conducting them and (b) including them in systematic reviews. A list of presentations and workshops is listed under the publication list of this doctoral thesis.

The insights also informed the new version of the Cochrane Handbook, which advises reviewers on how to include studies not by study label but distinct features [26]. The concentration on study features instead of study labels is also reflected in the recent Cochrane risk of bias tool to assess non-randomized studies of interventions, ROBINS-I [30], which is under development for

specific versions for CBA and ITS study designs, distinguishing more broadly between uncontrolled studies (simple before-after studies, ITS studies) and controlled studies (e.g. controlled ITS, CBA studies) assessing risk of bias according to different study design features [59].

Insights of publication I informed the parallel work on two Cochrane reviews, assessing the effects of complex public health interventions, in which the PhD candidate was a co-author. Burns et al. [60] assessed the effects of air pollution interventions on air quality and health outcomes and von Philipsborn et al. [61] investigated environmental interventions to reduce the consumption of sugar-sweetened beverages. Both included CBA and ITS studies and insights from our methodological work fed into the methodological approach implemented in these systematic reviews. This will be briefly described using the systematic review by Burns et al. [60].

Ambient air pollution, caused by particulate matter (PM) of various sizes that mix in the atmosphere [62], is an important public health risk factor, contributing to a high mortality and morbidity, especially in low- and middle-income countries (LMICs) [63, 64]. The systematic review objective was to assess the effectiveness of interventions to reduce ambient PM air pollution in reducing pollutant concentrations and improving associated health outcomes. The findings of our methodological study led to a much more careful identification, selection and in-depth characterisation of ITS and CBA studies, which was not based on study labels but detailed study features. Particularly problems associated with the conduct and analysis of ITS and CBA studies were confirmed in the review. We were not able to include many ITS studies as such in the review as they were inadequately analysed; more specifically, these studies assessed detailed serial data, yet analysed them in a way that did not make use of this data. Similarly, the analyses of several included CBA studies were conducted in such a way that the quantitative data on effectiveness could not be considered in the evidence synthesis of the review.

2.8. Publication II. Interrupted time series study on Bavarian smoke-free legislation

2.8.1. Background

In a second project, the lessons learned were applied in a rigorously conducted ITS study evaluating a complex public health intervention, the introduction of the smoke-free legislation in Bavaria. Recent research has focused on the impact of smoking and the exposure to secondhand smoke on pregnancy outcomes. Specifically, the exposure of tobacco smoke during pregnancy seems to affect the unborn baby and can result in decreased birth weight, preterm birth, and other complications, which may have life-long consequences [65]. Previous evidence suggests that smoke-free legislation, prohibiting smoking e.g. in public places, such as bars and restaurants, public buildings and institutions does not only improve adult health, but may also improve the health of newborn babies [66, 67]. Our study objective was therefore to assess the impact of the smoke-free legislation on several pregnancy outcomes.

2.8.2. Methods

As a well-designed and carefully prepared ITS study can increase the confidence with which the effect estimate can be attributed to the intervention [53], additional efforts and time went into the preparation of a detailed study protocol. In order to elaborate the complexity of the intervention, it included a system-based logic model outlining the long causal pathway between the smoke-free legislation and the pregnancy outcomes. The logic model further detailed potentially important co-interventions, i.e. simultaneous interventions or natural events that can influence the outcome and thus present the biggest threat to internal validity of an ITS study [35, 37]. After several iterations of the protocol, it was registered with the German national trial registry DRKS (study ID: DRKS00014805). In Bavaria, the implementation process of the smoke-free legislation took over two years, as it was first implemented in 2008 but loosened again in 2009 due to political reasons. The final implementation of the smoke-free legislation was in 2010. For our main impact model, we hypothesized a level as well as a slope change in preterm birth, small for gestational age (primary outcomes), low birth weight, stillbirth, and very preterm birth (secondary outcomes) in 2010. The structure of the data demanded the need for an advanced statistical analysis that accounted for seasonal patterns and remaining autocorrelation structures in a Poisson or more flexible negative binomial distribution. We investigated the effects of co-interventions through sensitivity analyses or ruled them out through choosing an impact model with a shorter time period. Due to uncertainties in the timing of the implementation process of the smoke-free legislation as well as in the occurrence of measurable effects on pregnancy outcomes, we explored different time points of the “interruption” and a two-year transition period from 2008 until 2010 in further sensitivity analyses. As population health interventions are essential to reducing health inequalities [29], the study also investigated the impact of the smoke-

free legislation according to socio-economic status and maternal smoking status in subgroup analyses.

2.8.3. Results

The study found heterogeneous results regarding the impact of the smoke-free legislation. Most outcomes show statistically non-significant, small effects in both directions. For example, for preterm birth, we observed a rate ratio of 1.0163 (95%CI 0.9762, 1.0580) for the level change and a rate ratio of 0.9995 (95%CI 0.9976, 1.0013) for the slope change. We found slightly greater and statistically significant effects for both level and slope changes of the secondary outcome very preterm birth: a rate ratio of 0.8960 (95%CI 0.8413, 0.9542) for the level change and a rate ratio of 0.9954 (95% 0.9928, 0.9982) for the slope change. The majority of sensitivity analyses confirm these results, i.e. small and statistically non-significant effects for level and slope change for most outcomes except very preterm birth where the effects are of a similar magnitude although not statistically significant. Also the subgroup analyses on maternal smoking status and socio-economic status do not show clear effects.

2.8.4. Discussion

The study uses rigorous methods and was carefully prepared to assess the association between the smoke-free legislation and pregnancy outcomes in Bavaria. We suggest several aspects with regards to the complexity of the intervention and the context, in which it was implemented, as well as design-inherent methodological features that may explain why our study did not show a clear impact of the legislation. Particularly the methodological discussion may be important for further research in this field. For example, we believe that “determining the ‘best-fit’ model among a range of alternatives remains at least partially arbitrary”. There are several alternatives regarding how to analyse ITS studies. Among the most popular methods are segmented regression analysis and Auto Regressive Integrated Moving Average (ARIMA) [68]. However, when going into detail, small decisions on which parameters to include in a statistical model may have important implications for the effect estimates [69]. Profound statistical knowledge is a requirement for conducting ITS studies, especially in cases where the data are characterised by non-linear trends. Furthermore, some statistical challenges of ITS studies have not yet been investigated in-depth and detailed statistical guidance for analysing ITS studies is still missing. This may include, for example, aspects, such as the inherent heterogeneity of population characteristics when data are aggregated, e.g. from different healthcare facilities. Ewusie [70] has worked on a weighted segmented regression method that can be more precise and less biased, however, further research needs to investigate remaining challenges regarding how to best analyse such data in ITS studies.

2.8.5. Further applications of quasi-experimental study designs within the context of this doctoral research

Polus et al. [3] (see additional publication III) applies the set of common analysis methods for CBA studies encountered in publication I. and explores their implications for study findings. Many CBA studies do apply a correct design, i.e. they collect the data in control and intervention sites or groups before and after the intervention. In the statistical analysis, however, they do not fully exploit the potential of their collected data in the statistical analysis. This study therefore compared different commonly applied analysis methods to illustrate the consequences of the choice of analysis in terms of effect size and precision of the study findings using a previously conducted CBA study on a community-based newborn health programme in Nepal [71]. In Nepal, neonatal mortality still poses a substantial public health problem with 21 deaths per 1000 live births in 2016 [72]. Simple interventions including, for example, antenatal care visits and home-based treatment can substantially reduce the risk of neonatal mortality. The Nepalese government therefore introduced a community-based newborn care package (CBNCP), comprising several community-based and home-based interventions aiming to reduce neonatal mortality. Paudel et al. [71] investigated the effects of the CBNCP in Nepal comparing ten intervention districts, where the programme had been implemented, with ten control districts where the intervention had not been introduced.

The study results of additional contribution III reveal differences in effect size and associated statistical precision related to the effectiveness of the CBNCP in Nepal according to statistical analysis choice. This paper can be regarded as a "CBA good practice call" for primary researchers as well as systematic reviewers. Primary researchers should carefully choose their analysis method and fully exploit the strengths of the CBA study design. Systematic reviewers should, where needed, re-analyse the data before including findings of a CBA study in the evidence synthesis.

The work by Burns et al. [73] represents a further application of quasi-experimental study designs in primary research. The study, in which the PhD candidate was a co-author, assessed the impact of the COVID-19 mitigation measures in March 2020 on air quality in Munich, Germany, through a controlled ITS study as well as a synthetic control design. The study hypothesized that the mitigation measures implemented due to the emergent SARS-CoV-2/COVID-19 pandemic caused an immediate reduction of traffic leading to a reduction of NO_2 concentrations. The synthetic control design was used as a complementary approach to the controlled ITS study approach. Both approaches compared changes in NO_2 in 2020 to changes occurring in 2014-2019. The c-ITS as well as the SC approach support the hypothesis of an immediate improvement of air quality at traffic sites. This natural experiment suggests that traffic reducing policies for highly trafficked areas could lead to improved air quality.

2.9. Contribution of PhD thesis

The first publication of this doctoral thesis, the methodological review of CBA and ITS study characteristics published in 2017, was one of the first studies that undertook a detailed examination of CBA and ITS study design characteristics and of the characteristics of the systematic reviews that made use of these study designs. Since then, other methodological studies have confirmed our findings and have contributed to providing guidance, notably regarding the ITS study design [23, 37, 48, 49, 68, 74]. Hudson et al. [68] follow up on our work as well as on that of Ramsay et al. [35] and Jandoc et al. [75], including a wide range of systematic reviews published in 2015 and assessing detailed design and analysis characteristics of 116 ITS studies that were included in these reviews. Turner et al. [74] also investigated 200 ITS studies published between 2013 and 2017 that evaluated public health interventions or exposures. They particularly highlight the need for formal reporting guidelines to increase quality of reporting of design and analysis features and results of ITS studies [68, 74]. Indeed, reporting guidelines for ITS studies are currently under development (see Framework for Enhanced Reporting of Interrupted Time Series (FERITS), registered in 2018) [76]. Turner et al. [74] highlight the need for an improved description of the statistical methods and approaches to adjust for and estimate autocorrelation. Hudson et al. [68] investigated that only 6% of their included ITS studies referred to a study protocol while Lopez Bernal et al. [37] emphasize the importance of pre-specifying the methods and especially the impact model to reduce the risk of detecting an effect due to chance when purely relying on the outcome data. Lopez Bernal et al. [37], Lopez Bernal et al. [49] contributed much in providing detailed step-by-step guidance on how best to plan and conduct an ITS study including elaborate considerations on the choice of modelling the effect and the impact model. Publication II. of the doctoral thesis, the ITS study on the Bavarian smoke-free legislation, may serve as a role model regarding the reporting and execution of methods. The study was registered based on a detailed study protocol, which pre-specified the impact model, the main statistical analysis including sensitivity and subgroup analyses, and a detailed logic model displaying the complexity of the intervention and how this informed the study methods. The publication also provides the statistical code for researchers to be able to follow and, as appropriate, replicate our methodological approach.

The methodological literature on CBA studies, on the other hand, is still limited. Publication I, which analyses the methodological characteristics of CBA and ITS studies and additional contribution III thus contribute much by providing some guidance to researchers on how to conduct CBA studies and appraise them at the systematic review level. However, there has been much progress in utilizing and describing other quasi-experimental study designs that can equally provide useful evidence for complex public health interventions. These include both the controlled ITS study, as well as synthetic control, instrumental variable (IV), and regression discontinuity

designs [55]. The scientific discourse has now advanced from that of the beginning of this doctoral research, currently discussing for example the commonalities and differences between the DiD design, the synthetic control and the controlled ITS study design and their terminology [77, 78]. Lopez Bernal et al. [48] explain in detail the controlled ITS study, including potential control types as well as statistical analysis approaches. Particularly the proposed methodological extensions such as the inclusion of a synthetic control have caused confusion as to how this is different from a DiD design. Lopez Bernal et al. [77] argue that the DiD design typically refers to a CBA study where "the outcome is measured at a single baseline (pre-intervention) time point and a single post-intervention time point, or where pre- and post-intervention means are compared but where 'time' is not incorporated into the model" [21, 79, 80]. Thus, a key difference between the DiD and controlled ITS study design is the adjustment for secular trend. The synthetic control design can be used as a complementary approach to the controlled ITS study see, for example, Burns et al. [73], not necessarily only as an alternative [77]. Clearly, the methodological development of quasi-experimental studies to evaluate population level public health interventions continues to be active [73, 81, 82].

This doctoral thesis has thus contributed to the methodological discussion and refinement of CBA and ITS studies. It provides a baseline for the ongoing research of developing, defining, and clarifying quasi-experimental methods for the evaluation of complex public health interventions. With further advancement and clarification of the methods used in and critically needed for public health research, policy-makers will increasingly have the opportunity to make clear, evidence-based decisions on important public health matters.

Publication I

ELSEVIER

Journal of Clinical Epidemiology ■ (2017) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews

Stephanie Polus[a,*], Dawid Pieper[b], Jacob Burns[a], Atle Fretheim[c], Craig Ramsay[d], Julian P.T. Higgins[e], Tim Mathes[b], Lisa M. Pfadenhauer[a], Eva A. Rehfuess[a]

[a]*Institute for Medical Information Processing, Biometry and Epidemiology, Pettenkofer School of Public Health, LMU Munich, Marchioninistr. 15, 81377, Munich, Germany*

[b]*Institute for Research in Operative Medicine, University Witten/Herdecke, Ostmerheimer Str. 200, 51109 Cologne, Germany*

[c]*Norwegian Institute of Public Health, P.O. Box 4404 Nydalen, 0403 Oslo, Norway*

[d]*Health Services Research Unit, University of Aberdeen, Aberdeen AB25 2ZD, UK*

[e]*School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK*

Accepted 21 July 2017; Published online xxxx

Abstract

Objectives: The aim of the study was to examine the application, design, and analysis characteristics of controlled before-after (CBA) and interrupted time series (ITS) studies and their use in Cochrane reviews.

Study Design and Setting: We searched the Cochrane library for reviews including these study designs from May 2012 to March 2015 and purposively selected, where available, two reviews each across 10 prespecified intervention types. We randomly selected two CBA and two ITS studies from each review. Two researchers independently extracted information from the studies and the respective reviews.

Results: Sixty-nine reviews considered CBA and ITS studies for inclusion. We analyzed 21 CBA and 16 ITS studies from 11 to 8 reviews, respectively. Cochrane reviews inconsistently defined and labeled CBA and ITS studies. Many studies did not meet the Cochrane definition or the minimum criteria provided by Cochrane Effective Practice and Organisation of Care. The studies present a heterogeneous set of study features and applied a large variety of analyses.

Conclusion: While CBA and ITS studies represent important study designs to evaluate the effects of interventions, especially on a population or organizational level, unclear study design features challenge unequivocal classification and appropriate use. We discuss options for more specific definitions and explicit criteria for CBA and ITS studies.

Keywords: Controlled before-after studies; Interrupted time series analysis; Review; Methods; Public health; Nonrandomised study designs

1. Introduction

One key element of evidence-informed health care and public health is that treatment and policy decisions are informed by the best available scientific evidence [1]. Decisions are ideally guided by well-conducted systematic reviews that gather evidence from well-conducted primary studies to assess whether an intervention is more effective and preferably also less costly than another intervention.

Interventions in the field of public health, health services, health systems, and health policy tend to be more difficult to evaluate than clinical interventions [2–4]. In these fields especially, it may not be possible to conduct randomized controlled trials (RCTs) for reasons of feasibility (e.g., interventions to reduce ambient air pollution [5]), ethical considerations (e.g., home-based palliative care [6]), or lack of political will [7]. Consequently, assessments of effectiveness in such cases often have to rely on nonrandomized studies [8,9]. Among these, interrupted time series (ITS) and controlled before-after (CBA) studies are the study designs most commonly included in Cochrane reviews [10].

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. S.P. work is funded through a doctoral scholarship by the Heinrich Böll Foundation (Germany). J.P.T.H. is supported in part by Medical Research Council (MRC) grant MR/M025209/1 and is a member of the MRC Integrative Epidemiology Unit at the University of Bristol, which is supported by the MRC and the University of Bristol (grant MC_UU_12013/9).

* Corresponding author. Tel.: +49(0)89/2180 7212 7.
E-mail address: polus@ibe.med.uni-muenchen.de (S. Polus).

http://dx.doi.org/10.1016/j.jclinepi.2017.07.008
0895-4356/

What is new?

Key findings

- CBA and ITS studies are increasingly used but inconsistently labeled and defined in Cochrane reviews.

What this adds to what was known?

- Variable definitions and unclear key characteristics challenge their identification and classification as well as distinction from other study designs.

What is the implication and what should change now?

- We detail and explain CBA and ITS study characteristics and propose steps toward a consensus process to define key characteristics of these two study designs.

A CBA study is defined in the Cochrane Handbook as a study in which observations are made before and after the implementation of an intervention, both in a group that receives the intervention and in a control group that does not [11]. The Cochrane Effective Practice and Organisation of Care (EPOC) Group, based on a long experience in considering nonrandomized studies, has developed criteria for inclusion of CBA studies in systematic reviews, to ensure a minimum level of methodological rigor [12]. They recommend at least two intervention sites and two control sites [12], as well as contemporaneous data collection [13].

While the methodological literature on CBA studies is limited, there is disagreement as to whether a key characteristic of a CBA study is that the investigator has no control over the intervention allocation [11,12,14,15]. Incoherent use of terminology leads to a lack of differentiation between features of CBA studies and other study designs, such as nonrandomized controlled trials (NRCTs) [8,10,15,16].

An ITS study is defined in the Cochrane Handbook as a study that uses observations at multiple time points before and after an intervention (the ''interruption''). The design attempts to detect whether the intervention has had an effect significantly greater than any underlying trend over time [11,12]. The study is frequently conducted retrospectively using routine data [17,18], and usually, there is no investigator control over the allocation of the intervention [11,12]. Cochrane EPOC specifies minimum criteria that ITS studies must use at least three data points before and three after the intervention and clearly define the point in time when the intervention occurred [12].

Several papers have examined the use of ITS studies in health research with respect to methodological aspects [16,19–24]. It was noteworthy that ITS studies applied inappropriate methods for statistical analysis, which led to the frequent judgment of statistically nonsignificant effects as significant [19].

Clarity about CBA and ITS studies in terms of design, data collection, and data analysis would be helpful for researchers wishing to conduct a CBA or ITS study and facilitate a common terminology. Likewise, improved knowledge and transparency about these study designs would make it easier to search for and include these study designs in systematic reviews [25]. Ultimately, decision makers will have more certainty to recommend for or against an intervention based on studies that generate valid findings [26].

The objective of this study was therefore to examine the application, design, and analysis characteristics of CBA and ITS studies included in Cochrane reviews. We based our analysis on the Cochrane database because it is a generally accepted point of reference for evidence-informed decision-making in health and because it applies relatively homogenous standards in terms of study design terminology.

1.1. Primary question

What are the characteristics of CBA and ITS studies included in Cochrane reviews in terms of design, conduct, and analysis?

1.2. Secondary questions

Which types of interventions are assessed by Cochrane reviews that consider and identify CBA and ITS studies? How are CBA and ITS studies defined by review authors? How are CBA and ITS studies (as defined by Cochrane authors) defined and labeled by primary study authors?

2. Methods

2.1. Identification and analysis of systematic reviews including CBA and ITS studies

Ijaz et al. [10] documented the use of CBA and ITS studies as well as other nonrandomized studies in Cochrane reviews up to May 2012. As we expected the conduct of CBA and ITS studies in primary research as well as their inclusion in systematic reviews to have increased in recent years, we updated the search by Ijaz et al. replicating their methods. An a priori protocol of our study is available online. Our search sought to identify reviews published between May 2012 and March 2015, whose authors explicitly used the terms ''controlled before-after'' or CBA and ''interrupted time series'' or ITS studies (i.e., merely ''before-after studies'' or ''time series'' were excluded).

For all reviews including nonrandomized studies, one author (S.P.) extracted information with cross-checks performed by two further authors (E.A.R. and D.P.). Information was extracted on (1) type of study designs included and number of studies identified for each type; (2) responsible Cochrane group; (3) definition of CBA and/or ITS study

by review authors; (4) risk of bias/quality appraisal tool and assessment used by review authors; (5) level of intervention (i.e., population, organizational, and individual level); and (6) type of health intervention. While descriptions and definitions are often used interchangeably, reviews may state criteria without a clear notion of the features of the study design they refer to. This is why we also specifically examined whether reviews provided definitions.

We prespecified and defined 10 intervention types. We based this on a previous publication, where we had made a first pragmatic attempt toward a classification of public health interventions [4]. These included behavioral/educational, clinical, environmental, health policy, health system, nutrition, occupational, pharmaceutical, screening, and vaccination interventions (see Appendix A at www.jclinepi.com for definitions). We examined the labeling and descriptions of CBA and ITS studies as well as their applications to different intervention types across the included reviews.

2.2. Selection and analysis of CBA and ITS studies

As we were interested in obtaining insights regarding the use of CBA and ITS studies across different areas of health, we purposively selected two reviews per intervention type from those reviews that had included at least two CBA or two ITS studies. A minimum of two studies was considered important to ensure a reasonable applicability of the study designs to a given intervention type and a minimum degree of representativeness in study conduct. For those intervention types, where we had to choose among several options (e.g., health systems), we chose reviews from different Cochrane groups and assessing different interventions. For each selected systematic review, we randomly selected two studies, using an online random choice generator [27]. We undertook the selection process separately for CBA and ITS studies.

For the selected CBA and ITS studies, two authors (S.P., J.B., A.F., D.P., T.M., C.R., J.P.T.H., E.A.R., L.M.P.) independently extracted information onto a data extraction form that was specifically developed for the purposes of this study and pretested in five studies. The data extraction form considered (1) publication characteristics (i.e., year of publication, journal, country of study, language of study, funding source, terminology/labeling, and definitions); (2) application characteristics (i.e., study objective, population, intervention, comparison and outcome, type of intervention, level of intervention); and (3) methodological characteristics covering study design (e.g., setting, control, allocation, temporal design), data collection (e.g., number of measurements, outcome assessments, source of data, timing), and data analysis (e.g., statistical methods, unit of analysis) and reported strengths and weaknesses of study design. Results were compared to achieve consensus, and uncertainties and discrepancies were extensively discussed, if necessary with the whole author team.

Using the extracted data across studies, we assembled information on how CBA and ITS studies were defined by primary study authors and, comparing design and analysis features, attempted to define key characteristics of both study designs.

3. Results

3.1. Identification and analysis of systematic reviews

For the period May 2012 to March 2015, we found 136 Cochrane reviews (4.8% of a total of 2,861 Cochrane reviews published in this time period) that considered nonrandomized studies for inclusion (Supplemental Table 1 at www.jclinepi.com). The 136 reviews included a total of 1,956 studies; the most prevalent study designs according to the labels employed by review authors are listed in Table 1. Nineteen reviews identified no studies for inclusion.

Sixty-nine of the reviews explicitly considered CBA and ITS studies for inclusion (see Appendix B at www.jclinepi.com for a complete reference list). Among these, 18 reviews identified at least two CBA studies (range: 2−30 CBA studies) and 16 reviews identified at least two ITS studies (range: 2−52 ITS studies). Altogether, 12 reviews identified both CBA and ITS studies. Additionally, five and three reviews identified only one CBA and ITS study, respectively.

3.1.1. Which types of interventions are assessed by Cochrane reviews that consider and identify CBA and ITS studies?

In our sample of 69 reviews, CBA and ITS studies were most widely considered in reviews of health system

Table 1. Study designs among the 1,956 studies included in Cochrane reviews that considered nonrandomized studies according to the labels employed by review authors (May 2012−March 2015)

Label	Number (percentage)
RCTs	597 (31)
Cohort studies	166 (9)
CBA	168 (9)
ITS	143 (7)
Cross-sectional studies	109 (6)
Controlled clinical trials (CCTs)	91 (5)
Uncontrolled before-after studies	76 (4)
Observational studies	75 (4)
Cluster RCT	65 (3)
Case control	60 (3)
Retrospective cohort studies	55 (3)
NRCT	42 (2)
Prospective controlled cohort studies	26 (1)
Prospective cohort studies	25 (1)

Abbreviations: RCT, randomized controlled trial; NRCT, nonrandomized controlled trial; CBA, controlled before-after; ITS, interrupted time series.

interventions (n = 36), followed by reviews of behavioral (n = 8), environmental (n = 6), occupational (n = 5), clinical (n = 5), and health policy (n = 4) interventions. They were rarely or not at all considered in reviews of vaccination, screening, pharmaceutical, or nutrition interventions. Among the reviews considering CBA and ITS studies, 12 were targeting the population, 43 the organizational, and 14 the individual level. Tables 2 and 3 show how CBA and ITS studies have recently been applied in reviews that actually identified these study designs, suggesting that both study designs are most frequent in reviews of health system interventions directed at an organizational level. We included CBA and ITS studies that derived from the same reviews [28,29].

3.1.2. How are CBA and ITS studies defined by review authors?

About a third of the 69 reviews considering both CBA and ITS studies reported the EPOC criteria of two intervention and two control sites (43%) for CBA studies and three data points before and three after intervention and a clearly defined intervention time point (36%), respectively, for ITS studies (Fig. 1). For CBA studies, many also referred to "contemporaneous data collection" and/or "an appropriate choice of control"; the need for "same time periods before and after the intervention" was also specified repeatedly. For ITS studies, 25% of reviews described one of the two EPOC criteria, mostly omitting a clearly defined intervention time point. Less than a tenth of all reviews (4% for CBA and 7% for ITS studies) referred to EPOC criteria for CBA or ITS studies without stating them.

Among the seven reviews that provide specific definitions for CBA studies, two reported the Cochrane Handbook definition [11] together with the EPOC criteria [30,31] and another two referred to "prospective cohort studies" [32,33]. Four of the five reviews that provided definitions for ITS studies reported them together with the EPOC criteria. Two reviews explicitly referred to the Cochrane Handbook definition for ITS studies [30,31].

3.2. Selection and analysis of primary studies

For the analysis of CBA studies, we purposively selected 11 reviews, covering two reviews each for behavioral, environmental, health policy, and health system interventions and one review each for nutrition, occupational, and screening interventions. Among the 22 CBA studies, we randomly selected from these reviews, one study (from the screening review) was excluded postselection due to an initial misclassification (i.e., it was identified as a CBA study at abstract level, but the review authors labeled the study in the risk of bias assessment as a CCT) [34]. For the analysis of ITS studies, we selected eight reviews, two each concerned with behavioral, health policy, health system, and occupational interventions. Random selection of two studies from each review yielded 16 ITS studies.

3.2.1. How are CBA and ITS studies labeled or defined by primary study authors?

Primary study authors did not label any study "CBA." The descriptions of CBA labels shown in Table 4 were often mentioned in combination with "before and after the intervention." Primary study authors labeled their study as an "ITS" in only one study. All other studies used various descriptions with "time series" mentioned most often in combination with "analysis" or "design." None of the included studies gave a definition of the study design.

Table 2. Characteristics of reviews including CBA studies according to intervention type, level, responsible Cochrane group, and number of studies identified (May 2012–March 2015)

Intervention type (number of reviews)	Cochrane group (number of reviews)	Population level	Organizational level	Individual level	Number of CBA studies
Health systems (11)	EPOC (6) Injuries (1) OSH (3) PH (1)		11		42
Behavioral (3)	Injuries (1) DA (1) PH (1)	2	1		32
Health policy (1)	OSH (1)		1		3
Environmental (2)	PH (1) ARI (1)	1	1		13
Occupational (2)	Injuries (1) OSH (1)		2		20
Clinical (1)	PAPAS (1)			1	2
Nutrition (1)	DPLP (1)			1	11
Screening (1)	PAPAS (1)		1		2
Pharmaceutical, vaccination (0)	—				

Abbreviations: ARI, acute respiratory infections; CBA, controlled before-after; DA, drugs and alcohol; DPLP, developmental, psychosocial and learning problems; EPOC, Effective Practice and Organisation of Care; OSH, occupational safety and health; PH, public health; PAPAS, pain, palliative and supportive care.

Table 3. Characteristics of reviews including ITS studies in terms of intervention type, level, responsible Cochrane group, and number of studies identified (May 2012–March 2015)

Intervention type (number of reviews)	Cochrane group (number of reviews)	Population level	Organizational level	Individual level	Number of ITS studies
Health systems (10)	EPOC (8) OSH (1) PH (1)		10		93
Behavioral (3)	DA (1) EPOC (1) TA (1)	3			32
Health policy (4)	DA (1) EPOC (1) OSH (1) PH (1)	2	2		23
Occupational (2)	Injuries (1) OSH (1)		2		14
Clinical, environmental, nutrition, pharmaceutical, screening, vaccination (0)	—				

Abbreviations: DA, drugs and alcohol; EPOC, Effective Practice and Organisation of Care; OSH, occupational safety and health; PH, public health; TA, tobacco addiction; ITS, interrupted time series.

3.2.2. What are the characteristics of CBA and ITS studies included in Cochrane reviews in terms of design, conduct and analysis?

3.2.2.1. CBA studies. Among the 21 selected CBA studies, five were not CBA studies according to the Cochrane definition (Fig. 2) because they lacked control sites [35,36] or measurements before the intervention [37–39]; in one study, hospital units were randomized into control and intervention group and we therefore classified it as a cluster RCT [40]. Of the 16 actual CBA studies, nine fulfilled the EPOC criteria (i.e., two intervention and two control sites, contemporaneous data collection). Compared to the Cochrane Handbook, Cochrane EPOC provides a more specific definition of a CBA study where the investigators do not have control over the intervention allocation. If we adopt this more specific definition of the selected 16

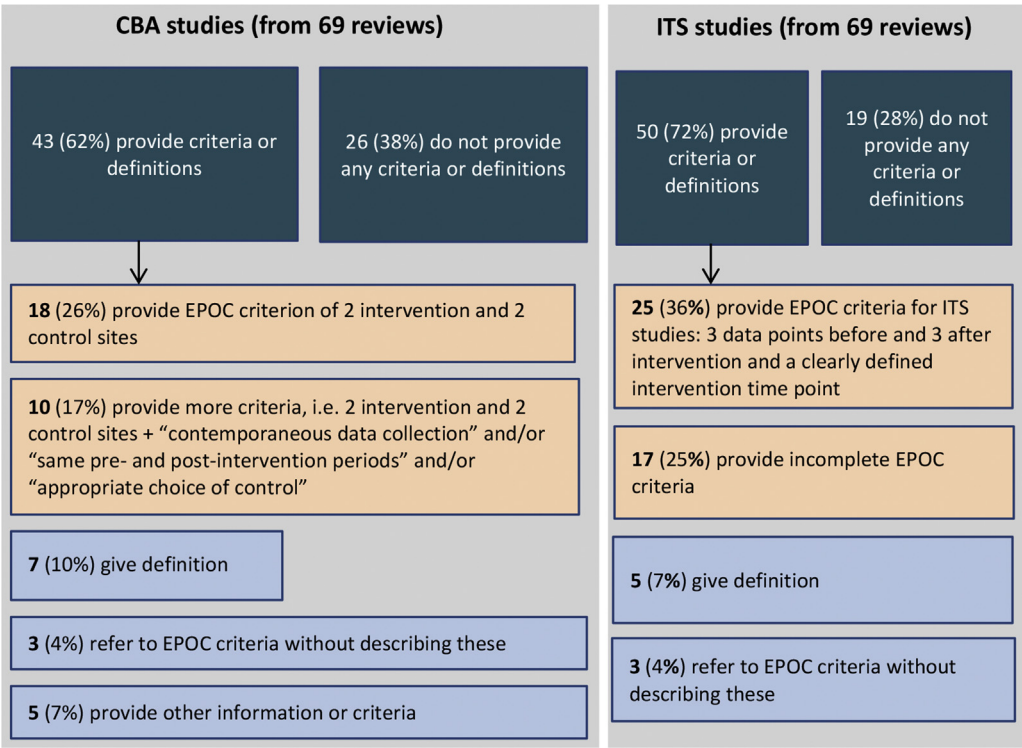


Fig. 1. Criteria for CBA and ITS studies as used in Cochrane reviews. CBA, controlled before-after; ITS, interrupted time series; EPOC, Effective Practice and Organisation of Care.

Table 4. Labels of CBA and ITS studies in primary studies

CBA labels	ITS labels
Quasiexperimental (n = 5)	Time series analysis/design (n = 5)
Survey (n = 4)	Observational (n = 3)
Comparative study (n = 3)	Analysis (n = 3)
Observational (n = 2)	Difference in difference (n = 2)
Cross-sectional (n = 2)	Retrospective (n = 2)
Natural experimental (n = 2)	Surveys (n = 1)
Case control (n = 1)	Interrupted time series (n = 1)
Prospective cohort (n = 1)	Natural experimental research (n = 1)
Difference in difference (n = 1)	

Abbreviations: CBA, controlled before-after; ITS, interrupted time series.

CBA studies, six complied with both the EPOC definition and criteria.

Table 5 presents a selection of further study design and analysis characteristics (see Supplemental Table 2 at www.jclinepi.com for an extended version). There was approximately equal use of CBA studies undertaken in a prospective or retrospective manner. We defined retrospective as a study, in which outcome data collected prior to the study period are used. In contrast, prospective studies collect outcome data during the study period. Defining studies as retrospective or prospective was, however, quite challenging [8] and judgments may vary. For about half of the selected studies, allocation of the intervention was not controlled by the investigators. The median number of sites among studies classified as EPOC CBA studies according to design was 7.5 (range: 3–748) for intervention and 5 (range: 3–8,301) for control sites. In some cases, the definition of "sites" was unclear and appeared to be synonymous with individuals (e.g., [47]). Study authors used a variety of mostly inappropriate or inefficient statistical analysis methods. Many studies applied simple statistical analysis methods, such as simple *t*-tests and did not take clustering into account, leading to unit of analysis errors and imprecision of confidence intervals. Studies performed, for example, a simple before and after comparison in the intervention group only or compared postmeans of individually aggregated data into intervention and control group.

3.2.3. ITS studies

Of the 16 selected ITS studies, two did not meet the Cochrane definition, as they did not include any data before the intervention [56,57] (Fig. 2). Of the 14 actual ITS studies, one did not comply with the EPOC criteria (i.e., at least three data points before and after the interruption and a clearly defined intervention time point), due to an insufficient number of data points before the intervention. Of the 13 ITS studies complying with EPOC design criteria, five did not perform a statistical analysis and merely displayed the results graphically or reported means before and after the intervention. We identified one study, where the intervention was under control of the investigators [58]. This study was, however, different in many ways, as the review authors lumped together several "meth studies" [35,58] and included them as a single ITS study [28].

As shown in Table 6 (see Supplemental Table 3 at www.jclinepi.com for an extended version), one study applied autoregressive integrated moving average (ARIMA) and at least two studies applied segmented regression analysis, although bad reporting impeded a clear identification. A majority of studies conducted some form of regression analysis and some adjusted or tested for autocorrelation (n = 6) and/or reported to adjust for secular trend (n = 6). Eight ITS studies [57,63,65–71] from five reviews were reanalyzed by review authors as recommended by EPOC in case of an inappropriate analysis. For ITS studies adhering to EPOC design criteria, the median number of data points was 12 (range: 3–46) before and 12 (range: 3–86) after the intervention. Three studies had a control group, and we therefore classified them as controlled ITS studies [61,62,71]. All studies were conducted retrospectively.

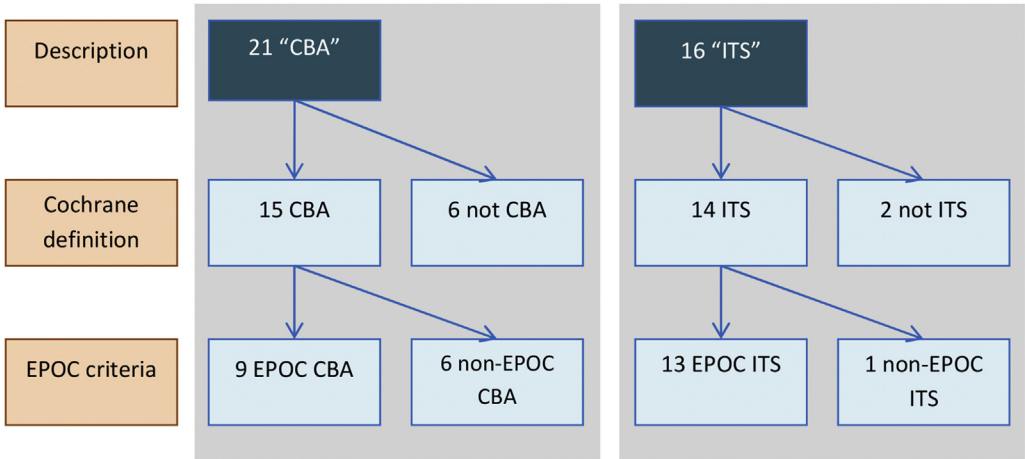


Fig. 2. EPOC criteria assessment for CBA and ITS studies. CBA, controlled before-after; ITS, interrupted time series; EPOC, Effective Practice and Organisation of Care.

Table 5. CBA study characteristics

Study ID	Intervention assessed	Contemporaneous data collection	Appropriate control	Study conduct				
				Temporality	Intervention allocation outside of researcher control	No. of intervention sites	No. of control sites	Unit of allocation
Georgia Meth 2011 [35]	Awareness campaign against drug use	Na	Na	Retr.	Yes	1	0	State
Miller 2000 [41]	Alcohol and Drug Abuse Prevention Program	Probably yes	Probably no	Prosp.	No	1	1	School
Pasco 2012 [36]	Gatekeeper training program for suicide prevention	Unclear	Probably no	Prosp.	No	1	0	Individual
Tompkins 2009 [42]	Gatekeeper training program on suicide prevention	Probably no	Probably yes	Prosp.	No	2 or 3 (unclear)	3	School
Butala 2010 [43]	Slum upgrading intervention	Probably yes	Probably yes	Retr.	Yes	14	NR	Community
Taylor 1987 [44]	Shelter upgrading for the urban poor	Probably yes	Probably no	Retr.	Yes	NR but likely >2	NR but likely >2	Community
Meklin 2005b [45]	Moisture and mold renovations	No	No	Prosp.	No	2	2	School
Shortt 2007 [46]	Housing intervention (energy efficiency measures)	Unclear	No	Prosp.	No	54	46	Household
Levine 2012 [37]	Safety inspections in hospital on injuries and job loss	Unclear	Unclear	Retr.	Yes	409	409	Company
Nelson 1997 [47]	Inspections and citation for violating fall prevention rules	Yes	Unclear	Retr.	Yes	784	8,301	Company
Tucker 2007 [48]	Classroom sexual health education and drop-in clinics	Yes	Probably yes	Prosp.	Yes	10	5	School
Hultberg 2005 [49]	Cofinanced collaboration model of primary care	Probably yes	Probably yes	Prosp.	Yes	3	4	Other: city area
Kaushal 2008 [40]	Unit-based clinical pharmacists to reduce medication errors	Yes	Probably yes	Prosp.	No	3	3	Hospital unit
Morriss 2009 [50]	A barcode scanning system for administrating medication	Probably yes	Probably yes	Prosp.	No	1	1	NICU section
Coyne 1980 [51]	Preschool meals at schools (food program)	Yes	Probably yes	Prosp.	Yes	5	5	Community

(*Continued*)

Table 5. Continued

Study ID	Intervention assessed	Contemporaneous data collection	Appropriate control	Study conduct				
				Temporality	Intervention allocation outside of researcher control	No. of intervention sites	No. of control sites	Unit of allocation
Santos 2005 [52]	Food supplementation program (milk program)	Yes	Probably yes	Prosp.	Yes	10	10	Community
Maizlish 1995 [53]	Targeted and active surveillance model for health care providers	Unclear	Unclear	Prosp.	Yes	10	NR	Hospital
Smits 2008 [54]	In-company workshop on the reporting of occupational diseases	Probably yes	Probably yes	Prosp.	No	NR but likely >2	NR but likely >2	Individual
Meyer 1993 [38]	Detailed follow-up (DFU) audiometric examinations on air force employees	Na	Na	Retr.	Yes	1	1 (same)	Patient
Nilsson 1980 [39]	Employees wearing ear muffs	Na	Na	Retr.	Yes	1	1 (same)	Individual
Jordan 2003 [55]	Application of the nursing nutritional screening tool	Yes	Probably yes	Prosp.	No	1	1	Hospital unit

Abbreviations: CBA, controlled before-after; Retr., retrospective; Prosp., prospective; NR, not reported; Na, not applicable.

4. Discussion

4.1. Key findings

In relation to our primary research question, we found a heterogeneous set of different study designs under the label ''CBA'' and ''ITS'' studies. Not all studies fitted the Cochrane definition of a CBA and ITS study. We found, for example, CBA studies without control sites as well as ITS studies without an intervention (''the interruption'') included in Cochrane reviews. Some CBA studies did not comply with the EPOC criteria, for example, because of an insufficient number of intervention and control sites; one ITS study had insufficient data points. Researchers were involved in the intervention allocation in almost half of all included CBA studies. According to EPOC guidance, such studies should be classified as NRCTs. It is also noteworthy that there is a stark discrepancy between methods employed for data collection vs. data analysis, where researchers often do not fully exploit the strength of the collected data in their analysis. Bad reporting, however, often precluded clear identification of the analysis methods.

In relation to our secondary research questions, most Cochrane reviews that included CBA and ITS studies were concerned with interventions on an organizational level; few addressed interventions on a population level, and as expected, very few took place on an individual level. We did not find many reviews of typical public health interventions, for example, environmental, vaccination, or screening interventions, which would lend themselves to the use of CBA and ITS studies.

There are striking differences among Cochrane reviews with respect to labeling and defining CBA and ITS studies. One-third of the included reviews did not provide any criteria for the study designs. These findings were all the more surprising, given that our sample was drawn from the relatively homogeneous and strongly methodologically influenced Cochrane community. Our analysis thus confirms that the inconsistent use of terminology leads to confusion among systematic reviewers regarding what can be classified as an ITS or CBA study [8,21,72,73].

On the primary study level, the labels ''CBA'' and ''ITS'' appear infrequently; instead, a large variety of terms is used. This suggests that CBA and ITS study labels and the study design characteristics associated with them are hardly used or known among primary study authors.

4.2. Toward clearer CBA and ITS study definitions and criteria

Considering the challenges we faced trying to categorize the CBA and ITS studies included in this analysis and

Table 6. ITS study characteristics

		Study conduct					Study analysis according to primary study authors	
Study ID	**Intervention assessed**	**Clear intervention time point**	**Data points before**	**Data points after**	**Time period data point**	**Control group**	**Analysis characteristics included**	**Reported analysis method**
Grootendorst 2005 [59]	Reference pricing of nonsteroidal anti-inflammatory drugs	Yes	13 (for intervention #2, 31)	86 (for intervention #2, 68)	Monthly	No	Trend, autocorrelation	Linear regression
Puig 2007 [60]	Reference pricing for generics	Yes (but differing between sites and drugs)	16 (varies depending on site and drug)	30 (varies depending on site and drug)	Monthly	No	Trend, autocorrelation	Generalized Least-Squares regression
Smart 1976 [61]	Ban on alcohol advertisements	Yes	>12	>12	Monthly	Yes	Trend	Calculation of geometric mean consumption, *t*-test for comparison
Makowsky 1991 [62]	Lifting of an advertising ban on alcohol	Yes	32	46	Monthly	Yes	Trend, autocorrelation	ARIMA (Box and Jenkins method)
Khan 2003 [63]	Change in antibiotic policy and use of antibiotics	Yes (diff. time points for 2 separate intervention)	6 (1st intervention)	12 (1st intervention)	Quarterly	No	—	No statistical analysis
Mercer 1999 [64]	Antibiotic control policy	Yes	12	12	Monthly	No	—	No statistical analysis
Goldwater 1989 [65]	Introduction of recapping device for needles	Yes	9	36	Monthly	No	—	No statistical analysis
Sossai 2010 [57]	Sharps awareness campaign and needlestick prevention devices	Yes (yr)	0	5	Yearly	No	—	No statistical analysis
Carpenter 2011 [56]	Antidrug media campaign	Yes	0	3	Na	No		Multivariate logistic regression of postintervention time
Idaho Meth 2011 [58]	Messaging campaign on drug use	Yes	1	3	Yearly	No	—	No statistical analysis
Jackevicius 2001 [66]	Publication of scientific evidence on medical practice	Yes	32	28	Monthly	No	Trend, autocorrelation	Segmented regression analysis, linear regression
Lam 2009 [67]	Publishing of large RCT about statins in nephrology	Yes	33	7	Other	No	Autocorrelation	Linear regression to estimate annual increase in statin use and subsequent F-test to assess slope difference
Beal 2007 [68]	Regulation on architectural design for construction sites	Yes	14	10	Yearly	No	—	No statistical analysis

(*Continued*)

Table 6. Continued

		Study conduct					Study analysis according to primary study authors	
Study ID	**Intervention assessed**	**Clear intervention time point**	**Data points before**	**Data points after**	**Time period data point**	**Control group**	**Analysis characteristics included**	**Reported analysis method**
Lipscomb 2003 [69]	Washington State fall standard for the construction industry	Yes	8	31	Quarterly	No	Trend	Poisson regression
Joy 2007 [70]	Permissible exposure level (PEL) for noise exposure in coal mining	Yes	12	5	Yearly	No	—	Linear regression
Rabinowitz 2011 [71]	Mandatory hearing protection program	Yes (yr)	5	4	Yearly	Yes	—	Difference-in-differences analysis based on individual-specific regression coefficients before and after the intervention

Abbreviations: ITS, interrupted time series; RCT, randomized controlled trial.

considering the limited use of the study design labels in primary research, we explain in detail study characteristics and potentially problematic features.

This discussion is intended to help review authors identify these study designs in the screening process; from our experience, the definitions and criteria provided by Cochrane and Cochrane EPOC, while helpful, still leave much space for interpretation, a lack of clarity that is partially responsible for the heterogeneous findings of this study and previous studies [10,15,16,19,20]. This discussion is also intended to offer input toward consensual definitions and features of these study designs, which would eventually be helpful for both review authors and primary researchers.

4.2.1. CBA studies

4.2.1.1. Key characteristics. According to EPOC and with some additional elaboration, a high-quality CBA study (1) uses at least two intervention and two appropriate control sites and (2) employs contemporaneous data collection, whether carried out specifically for this purpose or using existing datasets, at relevant preintervention and postintervention time points at all sites. CBA studies may be prospective or retrospective in nature. The intervention effects can be analyzed at cluster or individual level, but the analysis should compare the difference in prechange and postchange between intervention and control groups.

4.2.1.2. Explanations. Using two intervention and two control sites may be advantageous because study validity increases with more sites being used. With only one site per group, any difference in observed effect between the intervention and control group may simply be due to underlying differences in the characteristics of the two sites, where these characteristics may be measured, known but not measured or unknown. In circumstances, when more than two levels are involved, for example, individuals or classes nested within schools and cities, it may be challenging to decide what constitutes the site [74]. Furthermore, should sites be actual locations (e.g., villages, schools) or can other clusters or groups of people (e.g., family members in a household or employees in a given company) form a site? This may, however, be irrelevant as long as the analysis takes the groups into account. The sites should have similar baseline characteristics, by choice or through matching; in case of baseline differences, an appropriate method of statistical adjustment should be applied.

Whereas Cochrane EPOC [12] and Hartling et al. [15] acknowledge CBA studies as natural experiments, in the Cochrane handbook, investigator control to some extent is not ruled out [11]. Deeks et al. [14] suggest that a CBA "can also be considered an experimental design if the investigator has control over or can deliberately manipulate the introduction of the intervention."

The analysis should take into account the presence of a control group. A simple *t*-test comparing postchanges between the groups may not suffice to show an intervention effect, particularly where baseline differences between the groups exist. The analysis should adjust for potential clustering effects where unit of observation and unit of analysis differ. More advanced methods, such as difference-in-differences analysis, adjusting for differences between the

different sites, may better reflect the design. Such analysis methods have been widely applied in other disciplines, such as economics [75], and it would be beneficial to take on-board lessons learnt.

4.2.1.3. Differences and similarities in relation to other study designs. CBA studies partially overlap with other study designs with implications for how these studies are searched for, described, and appraised as well as synthesized in systematic reviews. The main difficulty lies in differentiating between cluster NRCTs and CBA studies. One possibility is to use active intervention allocation by the investigator as the distinguishing feature between cluster NRCTs (present) and CBA studies (absent; natural experiment); this approach has been adopted by Cochrane EPOC [12]. There are, however, cases where such a differentiation is difficult due to poor reporting and various interpretations of what to consider a natural experiment. Interestingly, Shadish et al. [76] do not distinguish between specific study design labels when describing ''quasiexperimental designs that use both control groups and pretests.'' Acknowledging CBA studies and cluster NRCTs as part of a broader study design group without the necessity to identify the design more specifically may be another way forward. Differences in study design features could thus be articulated as part of the risk of bias assessment rather than as part of the study classification.

4.2.2. ITS studies

4.2.2.1. Key characteristics. ITS studies are usually designed as natural experiments. They may be prospective or retrospective in nature and may include a control group (controlled ITS) [8,15,22]. As mentioned by Cochrane EPOC, ITS studies should (1) use at least three data points before and three after the intervention and (2) clearly define the point in time, when the intervention occurred. An appropriate statistical analysis includes adjustment for secular trend.

4.2.2.2. Explanations. Although ITS studies are usually defined as natural experiments [12,15], ITS studies can be used to assess interventions allocated by the investigators [77]. The EPOC threshold of three data points before and after the intervention is based on the reasoning that drawing a line through any fewer than three data points would estimate trend in a very unreliable way. Indeed, several recent studies suggest that sufficient statistical power is only achieved when at least eight data points are included; even more may be required when using ARIMA or segmented regression analysis [22,24,78]. Generally speaking, the precision of ITS studies increases with the number of data points. An unequivocal distinction between preintervention and postintervention and implementation time periods is critical; this also refers to multiple interventions implemented sequentially or staggered implementation of a given intervention in different groups, institutions, or geographical areas [22]. Adding a control group further enhances the study's validity and minimizes risk of bias [79]: whereas an ITS study compares the postintervention trend with a counterfactual (i.e., the prediction of what would have happened in case the intervention had not taken place estimated from preintervention trends), a controlled ITS study compares preintervention and postintervention time trends between an intervention and control group. Visualization of data can help the reader interpret the study results [22,23,80] but can also be misleading and should therefore not be used routinely as a means of identifying or measuring an effect [81].

The discrepancies between data collection and analysis in the included ITS studies highlight the importance of a statistical analysis that adjusts for secular trend [17,19,22,24]. ARIMA or segmented regression models, which recognize secular trend as well as autocorrelation, are considered highly appropriate for analyzing ITS data [19,24,80,82]; other regression analyses may also be appropriate. Studies whose statistical analysis does not explicitly acknowledge secular trend (e.g., comparison of preintervention and postintervention means) or that merely display results graphically in fact miss the most important strength of the ITS design. EPOC allows ITS studies with inappropriate analysis to be included in systematic reviews, provided the data are reanalyzed. This relies, however, partly on primary study authors providing their original data and is a time-consuming and resource-intensive process.

4.2.2.3. Differences and similarities in relation to other study designs. ITS studies are sometimes interchangeably listed as ''time series.'' However, a time series merely investigates an ordered sequence of values of a variable at equally spaced time intervals [83], whereas an ITS study is characterized by an interruption. ITS studies are also closely related to repeated measures studies, where measurements are made in the same individuals at each time point [12]. A further related study design is the regression discontinuity design, where different temporal occasions can be assigned to different treatment conditions [84].

Especially with respect to controlled ITS studies, it may be hard for systematic reviewers to label these as an ITS study vs. a CBA study. As mentioned above, the essential feature of an ITS study is the statistical analysis, which must reflect multiple measurements over time and adjust for important secular trends. If this is not the case, a controlled ITS, with multiple measurements before and after the intervention, may be considered as a CBA study.

4.3. Strengths and weaknesses of this study

We analyzed a sample of CBA and ITS studies included in recent Cochrane reviews with respect to application and specific methodological characteristics. The sample was intended to be somewhat representative of the prespecified types of interventions. Representativeness of findings is, however, limited, as we only selected two reviews per

intervention type (where available) and two included studies from the selected reviews.

At the primary study level, data extraction was done in duplicate and difficulties were extensively discussed among the data extractors and, where necessary, with the whole author team. As we did not reanalyze the studies, our insights reflect CBA and ITS studies as originally conducted, analyzed, and reported rather than according to their potential. In fact, we did not contact study authors for missing study details. Finally, the generalizability of our findings is probably limited to Cochrane reviews; we would expect to find even more variation in methodological characteristics of CBA and ITS studies outside of Cochrane.

4.4. Recommendations for research and practice

CBA and ITS studies are increasingly being recognized as important study designs that, if conducted and analyzed well, can provide reliable effect estimates of the impacts of interventions, where randomization is not feasible. Importantly, there is a need to further the understanding of the definitions and key characteristics of these study designs among primary researchers and systematic reviewers, including through textbooks of epidemiology and epidemiological curricula and beyond the field of epidemiology. Recently published research provides the first detailed guidance on how to conduct ITS studies [85]. However, CBA studies in particular almost appear to be ''artificial study designs,'' with the label created by systematic reviewers but little used in the primary research world. More specific definitions and key characteristics would be beneficial for systematic review authors inside and outside of Cochrane to facilitate greater clarity with respect to including or excluding study designs. Further discussions should clarify when to include a CBA or ITS study according to EPOC criteria and when to include a study not meeting EPOC criteria but downgrade for risk of bias. This would minimize confusion and improve consistency within Cochrane and beyond.

While we have summarized and explained important features and characteristics of CBA and ITS studies and initiated a discussion about key characteristics, key study design and analysis characteristics should be clarified and their definitions updated through a consensus process, such as a Delphi procedure. There are direct implications for risk of bias assessment for these study designs that will be developed as the ROBINS-I tool [86] is advanced for different study designs. The development of a new reporting guideline, for example, an extension of the Transparent Reporting of Evaluations with Nonrandomized Designs statement [87], could be an important second step. Taken together, this could greatly advance methodological practice at primary study as well as systematic review level and ensure that CBA and ITS studies are put to the best use possible in evaluating the impacts of interventions.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2017.07.008.

References

[1] Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. Annu Rev Public Health 2009;30:175−201.

[2] Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. J Epidemiol Community Health 2002;56:119−27.

[3] Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655.

[4] Rehfuess EA, Akl EA. Current experience with applying the GRADE approach to public health interventions: an empirical study. BMC Public Health 2013;13:9.

[5] Burns J, B H, Turley R, Pfadenhauer LM, van Erp AM, Rohwer AC, et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database Syst Rev 2014.

[6] Brereton L, Wahlster P, Lysdahl KB, Mozygemba K, Burns J, Chilcott JB, et al, On Behalf of the Integrate-HTA Project Team. Integrated assessment of home based palliative care with and without reinforced caregiver support: 'A demonstration of INTEGRATE-HTA methodological guidances' − Executive Summary. [Online] 2016. Available at http://www.integrate-hta.eu/wp-content/uploads/2016/02/Integrated-assessment-of-home-based-palliative-care-with-and-without-reinforced-caregiver-support-Executive-summary.pdf.

[7] Rehfuess EA, Bartram J. Beyond direct impact: evidence synthesis towards a better understanding of effectiveness of environmental health interventions. Int J Hyg Environ Health 2014;217:155−9.

[8] Higgins JPT, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC, et al. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. Res Synth Methods 2013;12−25.

[9] Norris SL, Atkins D. Challenges in using nonrandomized studies in systematic reviews of treatment interventions. Ann Intern Med 2005;142:1112−9.

[10] Ijaz S, Verbeek JH, Mischke C, Ruotsalainen J. Inclusion of non-randomized studies in Cochrane systematic reviews was found to be in need of improvement. J Clin Epidemiol 2014;67:645−53.

[11] Reeves BC, Deeks JJ, Higgins JPT, Wells GA. Chapter 13. Including non-randomized studies. Cochrane handbook for systematic reviews of interventions version 5.1. 0 (updated March 2011). Chichester: The Cochrane Collaboration; 2011.

[12] Effective Practice and Organisation of Care Group (EPOC). What study designs should be included in an EPOC review and what should they be called? 2013. Available at http://epoc.cochrane.org/epoc-specific-resources-review-authors. Accessed April 2016.

[13] Effective Practice and Organisation of Care (EPOC). Data extraction and management. EPOC Resources for review authors 2013: cited 2016; Available at http://epoc.cochrane.org/epoc-specific-resources-review-authors. Accessed June 2016.

[14] Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7. iii-x, 1-173.

[15] Hartling L, Bond K, Santaguida PL, Viswanathan M, Dryden DM. Testing a tool for the classification of study designs in systematic reviews of interventions and exposures showed moderate reliability and low accuracy. J Clin Epidemiol 2011;64:861−71.

[16] Sanson-Fisher RW, D'Este CA, Carey ML, Noble N, Paul CL. Evaluation of systems-oriented public health interventions: alternative research designs. Annu Rev Public Health 2014;35:9−27.

[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;66:883−7.

[18] Bausell RB. The design and conduct of meaningful experiments involving human participant. New York, NY, USA: Oxford University Press; 2015.

[19] 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;19:613−23.

[20] Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Fam Pract 2000;17:S11−6.

[21] Rockers PC, Feigl AB, Rottingen JA, Fretheim A, de Ferranti D, Lavis JN, et al. Study-design selection criteria in systematic reviews of effectiveness of health systems interventions and reforms: a meta-review. Health Policy 2012;104:206−14.

[22] Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. Acad Pediatr 2013;13:S38−44.

[23] 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;68:950−6.

[24] Taljaard M, McKenzie JE, Ramsay CR, Grimshaw JM. The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. Implement Sci 2014;9:77.

[25] Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev 2012;11:MR000030.

[26] Schünemann H, Brożek J, Guyatt G, Oxman A. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach 2013.

[27] TextFixer.com. Random choice generator 2016: cited 2016; Available at TextFixer.com. Accessed April 2015.

[28] Ferri M, Allara E, Bo A, Gasparrini A, Faggiano F. Media campaigns for the prevention of illicit drug use in young people. Cochrane Database Syst Rev 2013;6:CD009287.

[29] Verbeek JH, Kateman E, Morata TC, Dreschler WA, Mischke C. Interventions to prevent occupational noise-induced hearing loss. Cochrane Database Syst Rev 2012;10:CD006396.

[30] Maaskant JM, Vermeulen H, Apampa B, Fernando B, Ghaleb MA, Neubert A, et al. Interventions for reducing medication errors in children in hospital. Cochrane Database Syst Rev 2015;3:CD006208.

[31] Pega F, Carter K, Blakely T, Lucas PJ. In-work tax credits for families and their impact on health status in adults. Cochrane Database Syst Rev 2013;8:CD009963.

[32] Sauni R, Verbeek JH, Uitti J, Jauhiainen M, Kreiss K, Sigsgaard T. Remediating buildings damaged by dampness and mould for preventing or reducing respiratory tract symptoms, infections and asthma. Cochrane Database Syst Rev 2015;2:CD007897.

[33] Mischke C, Verbeek JH, Job J, Morata TC, Alvesalo-Kuusi A, Neuvonen K, et al. Occupational safety and health enforcement tools for preventing occupational diseases and injuries. Cochrane Database Syst Rev 2013;8:CD010183.

[34] Rypkema G, Adang E, Dicke H, Naber T, de Swart B, Disselhorst L, et al. Cost-effectiveness of an interdisciplinary intervention in geriatric inpatients to prevent malnutrition. J Nutr Health Aging 2004;8:122−7.

[35] GfK Roper Public Affairs & Corporate Communications, Georgia Meth Use & Attitudes Survey 2011. Statewide survey measuring attitudes and behaviors towards methamphetamine in Georgia. GfK Roper Public Affairs & Corporate Communications: New York, NY; 2011.

[36] Pasco S, Wallack C, Sartin RM, Dayton R. The impact of experiential exercises on communication and relational skills in a suicide prevention gatekeeper-training program for college resident advisors. J Am Coll Health 2012;60:134−40.

[37] Levine DI, Toffel MW, Johnson MS. Randomized government safety inspections reduce worker injuries with no detectable job loss. Science 2012;336:907−11.

[38] Meyer GD, Wirth DB. An evaluation of the U.S. Air Force's detailed follow-up audiometric examination program. Mil Med 1993;158:603−5.

[39] Nilsson R, Lindgren F. The effect of long term use of hearing protectors in industrial noise. Scand Audiol Suppl 1980;(Suppl 12):204−11.

[40] Kaushal R, Bates DW, Abramson EL, Soukup JR, Goldmann DA. Unit-based clinical pharmacists' prevention of serious medication errors in pediatric inpatients. Am J Health Syst Pharm 2008;65:1254−60.

[41] Miller WR, Toscova RT, Miller JH, Sanchez V. A theory-based motivational approach for reducing alcohol/drug problems in college. Health Educ Behav 2000;27:744−59.

[42] Tompkins TL, Witt J. The short-term effectiveness of a suicide prevention gatekeeper training program in a college setting with residence life advisers. J Prim Prev 2009;30:131−49.

[43] Butala NM, VanRooyen MJ, Patel RB. Improved health outcomes in urban slums through infrastructure upgrading. Soc Sci Med 2010;71:935−40.

[44] Taylor J. Evaluation of the Jakarta kampung improvement programme. In: Skinner RJ, Taylor JL, Wegelin EA, editors. Shelter upgrading for the urban poor: evaluation of Third World experience. Manila: Island Publishing House; 1987.

[45] Meklin T, Potus T, Pekkanen J, Hyvarinen A, Hirvonen MR, Nevalainen A. Effects of moisture-damage repairs on microbial exposure and symptoms in schoolchildren. Indoor Air 2005;15(Suppl 10):40−7.

[46] Shortt N, Rugkasa J. "The walls were so damp and cold" fuel poverty and ill health in Northern Ireland: results from a housing intervention. Health Place 2007;13:99−110.

[47] Nelson NA, Kaufman J, Kalat J, Silverstein B. Falls in construction: injury rates for OSHA-inspected employers before and after citation for violating the Washington State Fall Protection Standard. Am J Ind Med 1997;31:296−302.

[48] Tucker JS, Fitzmaurice AE, Imamura M, Penfold S, Penney GC, Teijlingen E, et al. The effect of the national demonstration project Healthy Respect on teenage sexual health behaviour. Eur J Public Health 2007;17:33−41.

[49] Hultberg EL, Lonnroth K, Allebeck P. Interdisciplinary collaboration between primary care, social insurance and social services in the rehabilitation of people with musculoskeletal disorder: effects on self-rated health and physical performance. J Interprof Care 2005;19:115−24.

[50] Morriss FH Jr, Abramowitz PW, Nelson SP, Milavetz G, Michael SL, Gordon SN, et al. Effectiveness of a barcode medication administration system in reducing preventable adverse drug events in a neonatal intensive care unit: a prospective cohort study. J Pediatr 2009;154:363−8. 368.e1.

[51] Coyne T, Dowling M, Condon-Paoloni D. Evaluation of preschool meals programmes on the nutritional health of Aboriginal children. Med J Aust 1980;2:369−75.

[52] Santos IS, Gigante DP, Coitinho DC, Haisma H, Valle NC, Valente G. Evaluation of the impact of a nutritional program for undernourished children in Brazil. Cad Saude Publica 2005;21:776−85.

[53] Maizlish N, Rudolph L, Dervin K. The surveillance of work-related pesticide illness: an application of the Sentinel event Notification system for occupational risks (SENSOR). Am J Public Health 1995;85:806−11.

[54] Smits PB, de Boer AG, KuijerI PP, Braam I, Spreeuwers D, Lenderink AF, et al. The effectiveness of an educational programme on occupational disease reporting. Occup Med 2008;58:373−5.
[55] Jordan S, Snow D, Hayes C, Williams A. Introducing a nutrition screening tool: an exploratory study in a district general hospital. J Adv Nurs 2003;44:12−23.
[56] Carpenter CS, Pechmann C. Exposure to the above the Influence antidrug advertisements and adolescent marijuana use in the United States, 2006-2008. Am J Public Health 2011;101:948−54.
[57] Sossai D, Puro V, Chiappatoli L, Dagnino G, Odone B, Polimeri A, et al. Using an intravenous catheter system to prevent needlestick injury. Nurs Stand 2010;24:42−6.
[58] GfK Roper Public Affairs & Corporate Communications, Idaho Meth Use & Attitudes Survey 2010. Statewide survey measuring attitudes and behaviors towards methamphetamine in Idaho 2011: [New York, NY].
[59] Grootendorst PV, Marshall JK, Holbrook AM, Dolovich LR, O'Brien BJ, Levy AR. The impact of reference pricing of nonsteroidal anti-inflammatory agents on the use and costs of analgesic drugs. Health Serv Res 2005;40:1297−317.
[60] Puig-Junoy J. The impact of generic reference pricing interventions in the statin market. Health Policy 2007;84:14−29.
[61] Smart RG, Cutler RE. The alcohol advertising ban in British Columbia: problems and effects on beverage consumption. Br J Addict Alcohol Other Drugs 1976;71(1):13−21.
[62] Makowsky CR, Whitehead PC. Advertising and alcohol sales: a legal impact study. J Stud Alcohol 1991;52:555−67.
[63] Khan R, Cheesbrough J. Impact of changes in antibiotic policy on Clostridium difficile-associated diarrhoea (CDAD) over a five-year period in a district general hospital. J Hosp Infect 2003;54:104−8.
[64] Mercer KA, Chintalapudi SR, Visconti EB. Impact of targeted antibiotic restriction on usage and cost in a community hospital. J Pharm Technol 1999;15:79−84.
[65] Goldwater PN, Law R, Nixon AD, Officer JA, Cleland JF. Impact of a recapping device on venepuncture-related needlestick injury. Infect Control Hosp Epidemiol 1989;10(1):21−5.
[66] Jackevicius CA, Anderson GM, Leiter L, Tu JV. Use of the statins in patients after acute myocardial infarction: does evidence change practice? Arch Intern Med 2001;161:183−8.
[67] Lam NN, Jain AK, Hackam DG, Cuerden MS, Suri RS, Huo CY, et al. Results of a randomized controlled trial on statin use in dialysis patients had no influence on statin prescription. Kidney Int 2009;76:1172−9.
[68] Beal A. CDM Regulations: 12 years of pain but little gain. In: Proceedings of ICE. Leeds, UK: Taylor & Francis Group; 2007.
[69] Lipscomb HJ, Li L, Dement J. Work-related falls among union carpenters in Washington state before and after the Vertical Fall Arrest standard. Am J Ind Med 2003;44:157−65.
[70] Joy GJ, Middendorf PJ. Noise exposure and hearing conservation in U.S. coal mines−a surveillance report. J Occup Environ Hyg 2007;4:26−35.
[71] Rabinowitz PM, Galusha D, Kirsche SR, Cullen MR, Slade MD, Dixon-Ernst C. Effect of daily noise exposure monitoring on annual rates of hearing loss in industrial workers. Occup Environ Med 2011;68:414−8.
[72] Peinemann F, Kleijnen J. Development of an algorithm to provide awareness in choosing study designs for inclusion in systematic reviews of healthcare interventions: a method study. BMJ Open 2015;5:e007540.
[73] Reeves BC, Deeks JJ, Higgins JPT, Wells GA, on behalf of the Cochrane Non-Randomised Studies Methods Group. Including non-randomized studies. In: Green S, Higgins JPT, editors. Cochrane handbook for systematic reviews of interventions. Chichester, UK: Cochrane handbook for systematic reviews of interventions; 2008.
[74] Cradock AL, McHugh A, Mont-Ferguson H, Grant L, Barrett JL, Wang YC, et al. Effect of school district policy change on consumption of sugar-sweetened beverages among high school students, Boston, Massachusetts, 2004-2006. Prev Chronic Dis 2011;8:A74.
[75] Angrist JD, Pischke JS. Mostly harmless econometrics: An Empiricist's Companion. Princeton, New Jersey: Princeton University press; 2008.
[76] Shadish W, Cook T, Campbell M. Quasi-experimental designs that use both control groups and pretests. In: Prancan K, editor. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Wadsworth Cengage Learning; 2002.
[77] Svoronos T, Fretheim A. Clarifying the interrupted time series study design. BMJ Qual Saf 2015;24:475.
[78] 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.
[79] Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, et al. A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol 2015;68:324−33.
[80] 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.
[81] Lee DS, Lemieux T. Regression discontinuity designs in economics. In: NBER Working Paper Series. Cambridge, USA: National Bureau of Economic Research; 2009.
[82] Effective Practice and Organisation of Care (EPOC). Analysis in EPOC reviews. EPOC Resources for review authors 2013. Available at http://epoc.cochrane.org/epoc-specific-resources-review-authors. Accessed September 2016.
[83] NIST/SEMATECH. e-Handbook of Statistical Methods 2012. Available at http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc41.htm. Accessed September 2016.
[84] Reichardt CS, Henry GT. Regression-discontinuity designs. In: Cooper H, et al, editors. APA handbook of research methods in psychology. Washington, DC, US: American Psychological Association; 2012:511−26.
[85] Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2016.
[86] Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.
[87] 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;94:361−6.

Supplementary Information

Title: Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews

Stephanie Polus, Dawid Pieper, Jacob Burns, Atle Fretheim, Craig Ramsay, Julian P.T. Higgins, Tim Mathes, Lisa M. Pfadenhauer, Eva A. Rehfuess

Supplemental Table 1 List of Cochrane reviews that included nonrandomised studies

1 Review ID (AU+YR)	Title
de Jong 2012	Mobile phone messaging for facilitating self-management of long-term illnesses
Adams 2015	Lipid-lowering efficacy of atorvastatin
Akl 2013	Educational games for health professionals
Siegfried 2014	Restricting or banning alcohol advertising to reduce alcohol consumption in adults and adolescents
Allen 2013	Diagnostic accuracy of laparoscopy following computed tomography (CT) scanning for assessing the resectability with curative intent in pancreatic and periampullary cancer
Anglemeyer 2013	Antiretroviral therapy for prevention of HIV transmission in HIV-discordant couples
Anglemeyer 2014	Treatment of Kaposi sarcoma in children with HIV-1 infection
Atherton 2012	Email for clinical communication between patients/caregivers and healthcare professionals
Taylor 2014	Computed tomography (CT) angiography for confirmation of the clinical diagnosis of brain death
Aubin 2012	Interventions to improve continuity of care in the follow-up of patients with cancer
Baalbergen 2013	Primary surgery versus primary radiotherapy with or without chemotherapy for early adenocarcinoma of the uterine cervix
Curti 2015	Interventions to increase the reporting of occupational diseases by physicians
Bala 2013	Mass media interventions for smoking cessation in adults
Acosta 2014	Pharmaceutical policies: effects of reference pricing, other pricing, and purchasing policies
Barte 2014	Yellow fever vaccine for patients with HIV infection
Beller 2015	Palliative pharmacological sedation for terminally ill adults
Benathar 2012	Medical and surgical treatment for ocular myasthenia
Brocklehurst 2013	The effect of different methods of remuneration on the behaviour of primary care dentists
Bromley 2014	Treatment for epilepsy in pregnancy: neurodevelopmental outcomes in the child
Henderson 2015	Provision of a surgeon's performance data for people considering elective surgery
Brusamento 2012	Male involvement for increasing the effectiveness of prevention of mother-to-child HIV transmission (PMTCT) programmes
Meyer 2012	Email for communicating results of diagnostic medical investigations to patients
Cirocchi 2012	Non-resection versus resection for an asymptomatic primary tumour in patients with unresectable stage IV colorectal cancer
Cirocchi 2012	Radiofrequency ablation in the treatment of liver metastases from colorectal cancer
Clement 2013	Mass media interventions for reducing mental health-related stigma
Coppo 2014	School policies for preventing smoking among young people
Algie 2015	Interventions for reducing wrong-site surgery and invasive procedures
Dangour 2013	Interventions to improve water quality and supply, sanitation and hygiene practices, and their effects on the nutritional status of children
Davey 2013	Interventions to improve antibiotic prescribing practices for hospital inpatients
Harrod 2014	Interventions for primary prevention of suicide in university and other post-secondary educational settings

Hardt 2013	Lateral pararectal versus transrectal stoma placement for prevention of parastomal herniation
Dyer 2014	Dental auxiliaries for dental care traditionally provided by dentists
Eliakim-Raz 2013	Influenza vaccines in immunosuppressed adults with cancer
Ballini 2015	Interventions to reduce waiting times for elective procedures
Fiander 2015	Interventions to increase the use of electronic health information by healthcare practitioners to improve clinical practice and patient outcomes
Rolfe 2014	Interventions for improving patients' trust in doctors and groups of doctors
Fonner 2012	Voluntary counseling and testing (VCT) for changing HIV-related risk behavior in developing countries
Galvagno 2013	Helicopter emergency medical services for adults with major trauma
Gbabe 2014	Treatment of severe or progressive Kaposi's sarcoma in HIV-infected adults
Gentry 2013	Telephone delivered interventions for reducing morbidity and mortality in people with HIV infection
Kendrick 2012	Home safety education and provision of safety equipment for injury prevention
Flodgren 2013	Interventions to improve professional adherence to guidelines for prevention of device-related infections
Glenny 2013	Antibiotics for the prophylaxis of bacterial endocarditis in dentistry
Kendrick 2013	Parenting interventions for the prevention of unintentional injuries in childhood
Gillaizeau 2012	Computerized advice on drug dosage to improve prescribing practice
Vodopivec-Jamsek 2012	Mobile phone messaging for preventive health care
Goyder 2015	Email for clinical communication between healthcare professionals
Gurumurthy 2014	Effectiveness of different treatment modalities for the management of adult-onset granulosa cell tumours of the ovary (primary and recurrent)
Hanchard 2013	Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement
Muckle 2012	Managed alcohol as a harm reduction intervention for alcohol addiction in populations at high risk for substance abuse
Haroutiiunian 2012	Methadone for chronic non-cancer pain in adults
Gurol-Uranci 2012	Mobile phone messaging for communicating results of medical investigations
Pani 2013	QTc interval screening for cardiac risk in methadone treatment of opioid dependence
McCleery 2015	Dopamine transporter imaging for the diagnosis of dementia with Lewy bodies
Henschke 2013	Red flags to screen for malignancy in patients with low-back pain
Henson 2013	Nutritional interventions for reducing gastrointestinal toxicity in adults undergoing radical pelvic radiotherapy
Gomes 2013	Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers
Hunt 2015	Thromboelastography (TEG) and rotational thromboelastometry (ROTEM) for traumainduced coagulopathy in adult trauma patients with bleeding
Brown 2014	Centre-based day care for children younger than five years of age in low- and middle-income countries
Baker 2015	Community wide interventions for increasing physical activity
Jefferson 2012	Vaccines for preventing influenza in healthy children
Jefferson 2014	Vaccines for preventing influenza in healthy adults
Hughes 2013	Infection control strategies for preventing the transmission of meticillin-resistant Staphylococcus aureus (MRSA) in nursing homes for older people
Jin 2012	Dietary flavonoid for preventing colorectal neoplasms
Tusting 2013	Mosquito larval source management for controlling malaria

Hayes 2012	Collaboration between local health and local government agencies for health improvement
Jia 2014	Strategies for expanding health insurance coverage in vulnerable populations
Kredo 2013	Decentralising HIV treatment in lower- and middle-income countries
Kredo 2014	Task shifting from doctors to non-doctors for initiation and maintenance of antiretroviral therapy
Khangura 2012	Primary care professionals providing non-urgent care in hospital emergency departments
Jayaraman 2014	Advanced training in trauma life support for ambulance crews
Rutebemberwa 2014	Financial interventions and movement restrictions for managing the movement of health workers between public and private organizations in low- and middle-income countries
Kristjansson 2015	Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years
Lazzerini 2013	Specially formulated foods for treating children with moderate acute malnutrition in low- and middle-income countries
Legare 2014	Interventions for improving the adoption of shared decision making by healthcare professionals
Webster 2014	Exit interviews to reduce turnover amongst healthcare professionals
Lindegren 2012	Integration of HIV/AIDS services with maternal, neonatal and child health, nutrition, and family planning services
Lip 2014	Anticoagulation versus placebo for heart failure in sinus rhythm
Liu 2014	Reminder systems to improve patient adherence to tuberculosis clinic appointments for diagnosis and treatment
Lopez 2012	Steroidal contraceptives and bone fractures in women: evidence from observational studies
Lewis 2014	Physician anaesthetists versus non-physician providers of anaesthesia for surgical patients
Maeda 2013	Perianal injectable bulking agents as treatment for faecal incontinence in adults
Maguire 2014	Antidepressants for people with epilepsy and depression
Martlew 2014	Psychological and behavioural treatments for adults with non-epileptic attack disorder
Lawrie 2013	Laparoscopy versus laparotomy for FIGO stage I ovarian cancer
McRobbie 2014	Electronic cigarettes for smoking cessation and reduction
Lavoie 2014	Devices for preventing percutaneous exposure injuries caused by needles in healthcare personnel
Gupta 2012	Electric fans for reducing adverse health impacts in heatwaves
Montero 2014	Steroid avoidance or withdrawal for pancreas and pancreas with kidney transplant recipients
Maaskant 2015	Interventions for reducing medication errors in children in hospital
Moran 2013	Effectiveness of systematic screening for the detection of atrial fibrillation
Rooney 2013	Pharmacological treatment of depression in patients with a primary brain tumour
Atherton 2012	Email for the coordination of healthcare appointments and attendance reminders
Murthy 2012	Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians
Mischke 2013	Occupational safety and health enforcement tools for preventing occupational diseases and injuries
Pande 2013	The effect of pharmacist-provided non-dispensing services on patient outcomes, health service utilisation and costs in low- and middle-income countries
Jayaraman 2014	Advanced trauma life support training for hospital staff
Pavlov 2015	Transient elastography for diagnosis of stages of hepatic fibrosis and cirrhosis in people with alcoholic liver disease
Peckham 2013	Homeopathy for treatment of irritable bowel syndrome
Omidvari 2013	Nutritional screening for improving professional practice for patient outcomes in hospital and primary care settings

Poirot 2013	Mass drug administration for malaria
Reda 2012	Healthcare financing systems for increasing the use of tobacco dependence treatment
Parmelli 2012	Interventions to increase clinical incident reporting in health care
Rizzuto 2013	Risk of ovarian cancer in women treated with ovarian stimulating drugs for infertility
Pega 2013	In-work tax credits for families and their impact on health status in adults
Reeves 2013	Interprofessional education: effects on professional practice and healthcare outcomes (update)
Romano 2012	Exercises for adolescent idiopathic scoliosis
Desapriya 2014	Vision screening of older drivers for preventing road traffic injuries and fatalities
Ferri 2013	Media campaigns for the prevention of illicit drug use in young people
Rockers 2013	Interventions for hiring, retaining and training district health systems managers in low- and middle-income countries
Nava 2014	Biologics, colchicine, corticosteroids, immunosuppressants and interferon-alpha for Neuro-Behcet's Syndrome
Saeterdal 2014	Interventions aimed at communities to inform and/or educate about early childhood vaccination
Ruotsalainen 2014	Preventing occupational stress in healthcare workers
Sawmynaden 2012	Email for the provision of information on disease prevention and health promotion
Siegfried 2013	Optimal time for initiating antiretroviral therapy (ART) in HIV-infected, treatment-naive children aged 2 to 5 years old
Sauni 2015	Remediating buildings damaged by dampness and mould for preventing or reducing respiratory tract symptoms, infections and asthma
Smailagic 2015	(18)F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)
Smart 2014	Early referral to specialist nephrology services for preventing the progression to end-stage kidney disease
Steingart 2013	Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults
Steingart 2014	Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults
Lihua 2013	Spinal cord stimulation for cancer-related pain in adults
Theron 2014	The diagnostic accuracy of the GenoType((R)) MTBDRsl assay for the detection of resistance to second-line anti-tuberculosis drugs
Thomas 2013	Influenza vaccination for healthcare workers who care for people aged 60 or older living in long-term care institutions
Tudor Car 2013	Telephone communication of HIV testing results for improving knowledge of HIV infection status
Turley 2013	Slum upgrading strategies involving physical environment and infrastructure interventions and their effects on health and socio-economic outcomes
Shrestha 2015	Workplace interventions for reducing sitting at work
van der Molen 2012	Interventions to prevent injuries in construction workers
van Ginneken 2013	Non-specialist health worker interventions for the care of mental, neurological and substance-abuse disorders in low- and middle-income countries
van Velthoven 2013	Telephone delivered interventions for preventing HIV infection in HIV-negative persons
Verbeek 2012	Interventions to prevent occupational noise-induced hearing loss
Vinceti 2014	Selenium for preventing cancer
Giguère 2012	Printed educational materials: effects on professional practice and healthcare outcomes (Review)
Walshe 2013	Larvivorous fish for preventing malaria transmission
Chan 2013	End-of-life care pathways for improving outcomes in caring for the dying
Winfield 2014	Non-pharmaceutical management of respiratory morbidity in children with severe global developmental delay
Young 2013	Post-exposure passive immunisation for preventing measles

Supplemental Table 2 CBA study characteristics

The grey shaded area on the left includes the information depicted in Table 5 in the publication. The right part (in white) presents the additional information.

Study ID	Intervention assessed	Contemp. data collection	Appropriate control	Temporality	Intervention allocation outside of researcher control	No. of int. sites	No. of control sites	Unit of allocation	Review ID	Type of intervention	Cochrane def.	Unit of analysis	Reported analysis methods
Georgia Meth 2011	Awareness campaign against drug use	Na	Na	Retr.	Yes	1	0	State	Ferri 2013	Behavioural	No	Individual	No statistical analysis
Miller 2000	Alcohol and Drug Abuse Prevention Program	Probably yes	Probably no	Prosp.	No	1	1	School	Ferri 2013	Behavioural	Yes	School	ANOVA
Pasco 2012	Gatekeeper training program for suicide prevention	Unclear	Probably no	Prosp.	No	1	0	Individual	Harrod 2014	Behavioural	No	Individual	ANOVA
Tompkins 2009	Gatekeeper training program on suicide prevention	Probably no	Probably yes	Prosp.	No	2 or 3 (unclear)	3	School	Harrod 2014	Behavioural	Yes	Individual	ANCOVA/ANOVA (unclear reporting), t-tests
Butala 2010	Slum upgrading intervention	Probably yes	Probably yes	Retr.	Yes	14	NR	Community	Turley 2013	Environmental	Yes	Individual	Difference-in-differences analysis, nonlinear probit regression
Taylor 1987	Shelter upgrading for the urban poor	Probably yes	Probably no	Retr	Yes	NR but likely >2	NR but likely >2	Community	Turley 2013	Environmental	Yes	Household	T-tests
Meklin 2005b	Moisture and mould	No	No	Prosp.	No	2	2	School	Sauni 2015	Environmental	Yes	Individual	Chi-2 test, multiple logistic regression

	renovations												
Shortt 2007	Housing intervention (energy efficiency measures)	Unclear	No	Prosp.	No	54	46	Household	Sauni 2015	Environmen-tal	Yes	Household	Chi-2 tests, McNemars test
Levine 2012	Safety inspections in hospital on injuries and job loss	Unclear	Unclear	Retr	Yes	409	409	Company	Mischke 2013	Occupational	No	Community	Difference-in-differences analysis, Negative binomial regression analysis
Nelson 1997	Inspections and citation for violating fall prevention rules	Yes	Unclear	Retr	Yes	784	8301	Company	Mischke 2013	Occupational	Yes	Individual	Multivariate logistic regression
Tucker 2006	Classroom sexual health education and drop-in clinics	Yes	Probably yes	Prosp.	Yes	10	5	School	Hayes 2012	Behavioural	Yes	Individual	Difference-in-differences multilevel analysis
Hultberg 2005	Co-financed collaboration model of primary care	Probably yes	Probably yes	Prosp.	Yes	3	4	Other: city area	Hayes 2012	Health policy	Yes	Individual	For one outcome Chi-2 test, for the other outcome t-test and multiple linear regression
Kaushal 2008	Unit-based clinical pharmacists to reduce medication errors	Yes	Probably yes	Prosp.	No	3	3	Hospital unit	Maaskant 2015	Health system	No	Hospital unit	NR, probably t-test
Morriss 2009	A barcode scanning system for administrating medication	Probably yes	Probably yes	Prosp.	No	1	1	NICU section	Maaskant 2015	Health system	Yes	Medication dosage	Generalized estimating equation (GEE)
Coyne 1980	Preschool meals at schools (food programme)	Yes	Probably yes	Prosp.	Yes	5	5	Community	Kristjansso n 2015	Nutrition	Yes	Individual	T-tests

Santos 2005	Food supplementation programme (Milk Program)	Yes	Probably yes	Prosp.	Yes	10	10	Community	Kristjansson 2015	Nutrition	Yes	Community	Crude and adjusted linear regression analyses, multilevel modelling
Maizlish 1995	Targeted and active surveillance model for health care providers	Unclear	Unclear	Prosp.	Yes	10	NR	Hospital	Curti 2015	Occupational	Yes	Healthcare practice	No analysis
Smits 2008	In-company workshop on the reporting of occupational diseases	Probably yes	Probably yes	Prosp.	No	NR but likely > 2	NR but likely > 2	Individual	Curti 2015	Occupational	Yes	Individual	Zeroinflated Poisson model for primary outcomes
Meyer 1993	Detailed follow-up (DFU) audiometric examinations on air force employees	Na	Na	Retr	Yes	1	1 (same)	Patient	Verbeek 2012	Occupational	No	Individual	Relative risk calculation
Nilsson 1980	Employees wearing ear muffs	Na	Na	Retr	Yes	1	1 (same)	Individual	Verbeek 2012	Occupational	No	Individual	Relative risk calculation with confidence intervals and one-sided chi-square tests
Jordan 2003	Application of the nursing nutritional screening tool	Yes	Probably yes	Prosp.	No	1	1	Hospital unit	Omidvari 2013	Other (monitoring)	Yes	Individual	Frequency counts and cross-tabulations (chi²-test)

Supplemental Table 2 References

1. GfK Roper Public Affairs & Corporate Communications, *Georgia Meth Use & Attitudes Survey 2011. Statewide survey measuring attitudes and behaviors towards methamphetamine in Georgia* 2011: New York, NY.
2. Ferri, M., E. Allara, A. Bo, A. Gasparrini, and F. Faggiano *Media campaigns for the prevention of illicit drug use in young people*. Cochrane Database Syst Rev, 2013. **6**, CD009287 DOI: 10.1002/14651858.CD009287.pub2.
3. Miller, W.R., R.T. Toscova, J.H. Miller, and V. Sanchez *A theory-based motivational approach for reducing alcohol/drug problems in college*. Health Educ Behav, 2000. **27**, 744-59.
4. Pasco, S., C. Wallack, R.M. Sartin, and R. Dayton *The impact of experiential exercises on communication and relational skills in a suicide prevention gatekeeper-training program for college resident advisors*. J Am Coll Health, 2012. **60**, 134-40 DOI: 10.1080/07448481.2011.623489.
5. Harrod, C.S., C.W. Goss, L. Stallones, and C. DiGuiseppi *Interventions for primary prevention of suicide in university and other post-secondary educational settings*. Cochrane Database Syst Rev, 2014. **10**, CD009439 DOI: 10.1002/14651858.CD009439.pub2.
6. Tompkins, T.L. and J. Witt *The short-term effectiveness of a suicide prevention gatekeeper training program in a college setting with residence life advisers*. J Prim Prev, 2009. **30**, 131-49 DOI: 10.1007/s10935-009-0171-2.
7. Butala, N.M., M.J. VanRooyen, and R.B. Patel *Improved health outcomes in urban slums through infrastructure upgrading*. Soc Sci Med, 2010. **71**, 935-40 DOI: 10.1016/j.socscimed.2010.05.037.
8. Turley, R., R. Saith, N. Bhan, E. Rehfuess, and B. Carter *Slum upgrading strategies involving physical environment and infrastructure interventions and their effects on health and socio-economic outcomes*. Cochrane Database Syst Rev, 2013. **1**, CD010067 DOI: 10.1002/14651858.CD010067.pub2.
9. Taylor, J., *Evaluation of the Jakarta kampung improvement programme*, in *Shelter upgrading for the urban poor: evaluation of Third World experience*, Skinner RJ, Taylor JL, and Wegelin EA, Editors. 1987, Island Publishing House: Manila.
10. Meklin, T., T. Potus, J. Pekkanen, A. Hyvarinen, M.R. Hirvonen, and A. Nevalainen *Effects of moisture-damage repairs on microbial exposure and symptoms in schoolchildren*. Indoor Air, 2005. **15 Suppl 10**, 40-7 DOI: 10.1111/j.1600-0668.2005.00357.x.
11. Sauni, R., J.H. Verbeek, J. Uitti, M. Jauhiainen, K. Kreiss, and T. Sigsgaard *Remediating buildings damaged by dampness and mould for preventing or reducing respiratory tract symptoms, infections and asthma*. Cochrane Database Syst Rev, 2015. **2**, CD007897 DOI: 10.1002/14651858.CD007897.pub3.
12. Shortt, N. and J. Rugkasa *"The walls were so damp and cold" fuel poverty and ill health in Northern Ireland: results from a housing intervention*. Health Place, 2007. **13**, 99-110 DOI: 10.1016/j.healthplace.2005.10.004.
13. Levine, D.I., M.W. Toffel, and M.S. Johnson *Randomized government safety inspections reduce worker injuries with no detectable job loss*. Science, 2012. **336**, 907-11 DOI: 10.1126/science.1215191.
14. Mischke, C., J.H. Verbeek, J. Job, T.C. Morata, A. Alvesalo-Kuusi, K. Neuvonen, et al. *Occupational safety and health enforcement tools for preventing occupational diseases and injuries*. Cochrane Database Syst Rev, 2013. **8**, CD010183 DOI: 10.1002/14651858.CD010183.pub2.
15. Nelson, N.A., J. Kaufman, J. Kalat, and B. Silverstein *Falls in construction: injury rates for OSHA-inspected employers before and after citation for violating the Washington State Fall Protection Standard*. Am J Ind Med, 1997. **31**, 296-302.

16. Tucker, J.S., A.E. Fitzmaurice, M. Imamura, S. Penfold, G.C. Penney, E. Teijlingen, et al. *The effect of the national demonstration project Healthy Respect on teenage sexual health behaviour*. Eur J Public Health, 2007. **17**, 33-41 DOI: 10.1093/eurpub/ckl044.
17. Hayes, S.L., M.K. Mann, F.M. Morgan, H. Kitcher, M.J. Kelly, and A.L. Weightman *Collaboration between local health and local government agencies for health improvement*. Cochrane Database Syst Rev, 2011. CD007825 DOI: 10.1002/14651858.CD007825.pub5.
18. Hultberg, E.L., K. Lonnroth, and P. Allebeck *Interdisciplinary collaboration between primary care, social insurance and social services in the rehabilitation of people with musculoskeletal disorder: effects on self-rated health and physical performance*. J Interprof Care, 2005. **19**, 115-24 DOI: 10.1080/13561820400024134.
19. Kaushal, R., D.W. Bates, E.L. Abramson, J.R. Soukup, and D.A. Goldmann *Unit-based clinical pharmacists' prevention of serious medication errors in pediatric inpatients*. Am J Health Syst Pharm, 2008. **65**, 1254-60 DOI: 10.2146/ajhp070522.
20. Maaskant, J.M., H. Vermeulen, B. Apampa, B. Fernando, M.A. Ghaleb, A. Neubert, et al. *Interventions for reducing medication errors in children in hospital*. Cochrane Database Syst Rev, 2015. **3**, CD006208 DOI: 10.1002/14651858.CD006208.pub3.
21. Morriss, F.H., Jr., P.W. Abramowitz, S.P. Nelson, G. Milavetz, S.L. Michael, S.N. Gordon, et al. *Effectiveness of a barcode medication administration system in reducing preventable adverse drug events in a neonatal intensive care unit: a prospective cohort study*. J Pediatr, 2009. **154**, 363-8, 368 e1 DOI: 10.1016/j.jpeds.2008.08.025.
22. Coyne, T., M. Dowling, and D. Condon-Paoloni, *Evaluation of preschool meals programmes on the nutritional health of Aboriginal children.* Med J Aust, 1980. **2**(7): p. 369-75.
23. Kristjansson, E., D.K. Francis, S. Liberato, M. Benkhalti Jandu, V. Welch, M. Batal, et al. *Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years*. Cochrane Database Syst Rev, 2015. **3**, CD009924 DOI: 10.1002/14651858.CD009924.pub2.
24. Santos, I.S., D.P. Gigante, D.C. Coitinho, H. Haisma, N.C. Valle, and G. Valente *Evaluation of the impact of a nutritional program for undernourished children in Brazil*. Cad Saude Publica, 2005. **21**, 776-85 DOI: /S0102-311X2005000300011.
25. Maizlish, N., L. Rudolph, and K. Dervin *The surveillance of work-related pesticide illness: an application of the Sentinel Event Notification System for Occupational Risks (SENSOR)*. Am J Public Health, 1995. **85**, 806-11.
26. Curti, S., R. Sauni, D. Spreeuwers, A. De Schryver, M. Valenty, S. Riviere, et al. *Interventions to increase the reporting of occupational diseases by physicians*. Cochrane Database Syst Rev, 2015. **3**, CD010305 DOI: 10.1002/14651858.CD010305.pub2.
27. Smits, P.B., A.G. de Boer, P.P. Kuijer, I. Braam, D. Spreeuwers, A.F. Lenderink, et al. *The effectiveness of an educational programme on occupational disease reporting*. Occup Med (Lond), 2008. **58**, 373-5 DOI: 10.1093/occmed/kqn061.
28. Meyer, G.D. and D.B. Wirth *An evaluation of the U.S. Air Force's detailed follow-up audiometric examination program*. Mil Med, 1993. **158**, 603-5.
29. Verbeek, J.H., E. Kateman, T.C. Morata, W.A. Dreschler, and C. Mischke *Interventions to prevent occupational noise-induced hearing loss*. Cochrane Database Syst Rev, 2012. **10**, CD006396 DOI: 10.1002/14651858.CD006396.pub3.
30. Nilsson, R. and F. Lindgren, *The effect of long term use of hearing protectors in industrial noise.* Scand Audiol Suppl, 1980(Suppl 12): p. 204-11.
31. Jordan, S., D. Snow, C. Hayes, and A. Williams *Introducing a nutrition screening tool: an exploratory study in a district general hospital*. J Adv Nurs, 2003. **44**, 12-23.
32. Omidvari, A.H., Y. Vali, S.M. Murray, D. Wonderling, and A. Rashidian *Nutritional screening for improving professional practice for patient outcomes in hospital and primary care settings*. Cochrane Database Syst Rev, 2013. **6**, CD005539 DOI: 10.1002/14651858.CD005539.pub2.

Supplemental Table 3 ITS study characteristics

The grey shaded area on the left includes the information depicted in Table 5 in the publication. The right part (in white) presents the additional information.

	Intervention assessed	Clear intervention time point	Data points before	Data points after	Time period data point	Control group	Analysis characteristics included	Reported analysis method	Review ID	Type of intervention	Cochrane definition	Temporality	No. of Observations/data point
Grooten-dorst 2005	Reference pricing of nonsteroidal anti-inflammatory drugs	Yes	13 (for intervention #2 31)	86 (for int. #2 68)	Monthly	No	Trend, autocorrelation	Linear regression	Acosta 2014	Health policy	Yes	Retr.	>100
Puig 2007	Reference pricing for generics	Yes (but differing between sites & drugs)	16 (varies depending on site & drug)	30(varies depending on site & drug)	Monthly	No	Trend, autocorrelation	Generalised Least-Squares regression	Acosta 2014	Health policy	Yes	Retr.	<100
Smart 1976	Ban on alcohol advertisements	Yes	> 12	>12	Monthly	Yes	Trend	Calculation of geometric mean consumption, t-test for comparison	Siegfried 2014	Health policy	Yes	Retr.	>100
Makowksy 1991	Lifting of an advertising ban on alcohol	Yes	32	46	Monthly	Yes	Trend, autocorrelation	ARIMA (Box and Jenkins method)	Siegfried 2014	Health policy	Yes	Retr.	>100
Khan 2003	Change in antibiotic policy and use of antibiotics	Yes (diff. Time points for 2 separate interventions)	6 (1st intervention)	12 (1st intervention)	Quarterly	No	-	No statistical analysis	Davey 2013	Health policy	Yes	Retr.	>100
Mercer 1999	Antibiotic control policy	Yes	12	12	Monthly	No	-	No statistical analysis	Davey 2013	Health policy	Yes	Retr.	>100
Goldwater 1989	Introduction of recapping device for needles	Yes	9	36	Monthly	No	-	No statistical analysis	Lavoie 2014	Health system	Yes	Both	<100
Sossai 2010	Sharps awareness campaign and needlestick prevention devices	Yes (year)	0	5	Yearly	No	-	No statistical analysis	Lavoie 2014	Occupational	No	Retr.	>100
Carpenter 2011	Anti-drug media campaign	Yes	0	3	Na	No		Multivariate logistic regression of post-intervention time	Ferri 2013	Behavioural	No	Retr.	>100
Idaho Meth 2011	Messaging campaign on drug use	Yes	1	3	Yearly	No	-	No statistical analysis	Ferri 2013	Behavioural	Yes	-	>100
Jackevicius 2001	Publication of scientific evidence on	Yes	32	28	Monthly	No	Trend, autocor-	Segmented regression analysis,	Giguère 2012	Other (publishin	Yes	Retr.	>100

	medical practice						relation	linear regression		g of trial			
Lam 2009	Publishing of large RCT about statins in nephrology	Yes	33	7	Other	No	Autocor-relation	Linear regression to estimate annual increase in statin use and subsequent F-test to assess slope difference	Giguère 2012	Other (publishing of trial)	Yes	Retr.	>100
Beal 2007	Regulation on architectural design for construction sites	Yes	14	10	Yearly	No	-	No statistical analysis	Van der Molen 2012	Other (policy)	Yes	Retr.	>100
Lipscomb 2003	Washington State fall standard for the construction industry	Yes	8	31	Quarterly	No	Trend	Poisson regression	Van der Molen 2012	Other (policy)	Yes	Retr.	Unclear
Joy 2007	Permissible exposure level (PEL) for noise exposure in coal mining	Yes	12	5	Yearly	No	-	Linear regression	Verbeek 2012	Occupational	Yes	Retr.	>100
Rabinowitz 2011	Mandatory hearing protection programme	Yes (year)	5	4	Yearly	Yes	-	Difference-in differences analysis based on individual-specific regression coefficients before and after the intervention	Verbeek 2012	Occupational	Yes	Retr.	<100

Supplemental Table 3 References

Acosta, A., Ciapponi, A., Aaserud, M., Vietto, V., Austvoll-Dahlgren, A., Kosters, J. P., . . . Oxman, A. D. (2014). Pharmaceutical policies: effects of reference pricing, other pricing, and purchasing policies. *Cochrane Database Syst Rev, 10*, CD005979. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25318966 doi:10.1002/14651858.CD005979.pub2

Beal, A. (2007). *CDM Regulations: 12 years of pain but little gain*. Retrieved from Leeds, UK:

Carpenter, C. S., & Pechmann, C. (2011). Exposure to the Above the Influence antidrug advertisements and adolescent marijuana use in the United States, 2006-2008. *Am J Public Health, 101*(5), 948-954. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21421952 doi:10.2105/AJPH.2010.300040

Davey, P., Brown, E., Charani, E., Fenelon, L., Gould, I. M., Holmes, A., . . . Wilcox, M. (2013). Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev, 4*, CD003543. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23633313 doi:10.1002/14651858.CD003543.pub3

Ferri, M., Allara, E., Bo, A., Gasparrini, A., & Faggiano, F. (2013). Media campaigns for the prevention of illicit drug use in young people. *Cochrane Database Syst Rev, 6*, CD009287. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23740538 doi:10.1002/14651858.CD009287.pub2

GfK Roper Public Affairs & Corporate Communications. (2011). *Idaho Meth Use & Attitudes Survey 2010. Statewide survey measuring attitudes and behaviors towards methamphetamine in Idaho*. Retrieved from New York, NY:

Giguere, A., Legare, F., Grimshaw, J., Turcotte, S., Fiander, M., Grudniewicz, A., . . . Gagnon, M. P. (2012). Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev, 10*, CD004398. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23076904 doi:10.1002/14651858.CD004398.pub3

Goldwater, P. N., Law, R., Nixon, A. D., Officer, J. A., & Cleland, J. F. (1989). Impact of a recapping device on venepuncture-related needlestick injury. *Infect Control Hosp Epidemiol, 10*(1), 21-25.

Grootendorst, P. V., Marshall, J. K., Holbrook, A. M., Dolovich, L. R., O'Brien, B. J., & Levy, A. R. (2005). The impact of reference pricing of nonsteroidal anti-inflammatory agents on the use and costs of analgesic drugs. *Health Serv Res, 40*(5 Pt 1), 1297-1317. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16174135 doi:10.1111/j.1475-6773.2005.00420.x

Jackevicius, C. A., Anderson, G. M., Leiter, L., & Tu, J. V. (2001). Use of the statins in patients after acute myocardial infarction: does evidence change practice? *Arch Intern Med, 161*(2), 183-188. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11176731

Joy, G. J., & Middendorf, P. J. (2007). Noise exposure and hearing conservation in U.S. coal mines--a surveillance report. *J Occup Environ Hyg, 4*(1), 26-35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17162478 doi:10.1080/15459620601067209

Khan, R., & Cheesbrough, J. (2003). Impact of changes in antibiotic policy on Clostridium difficile-associated diarrhoea (CDAD) over a five-year period in a district general hospital. *J Hosp Infect, 54*(2), 104-108. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12818582

Lam, N. N., Jain, A. K., Hackam, D. G., Cuerden, M. S., Suri, R. S., Huo, C. Y., . . . Garg, A. X. (2009). Results of a randomized controlled trial on statin use in dialysis patients had no influence on statin prescription. *Kidney Int, 76*(11), 1172-1179. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19776719 doi:10.1038/ki.2009.323

Lavoie, M. C., Verbeek, J. H., & Pahwa, M. (2014). Devices for preventing percutaneous exposure injuries caused by needles in healthcare personnel. *Cochrane Database Syst Rev, 3*, CD009740. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24610008 doi:10.1002/14651858.CD009740.pub2

Lipscomb, H. J., Li, L., & Dement, J. (2003). Work-related falls among union carpenters in Washington State before and after the Vertical Fall Arrest Standard. *Am J Ind Med, 44*(2), 157-165. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12874848 doi:10.1002/ajim.10254

Makowsky, C. R., & Whitehead, P. C. (1991). Advertising and alcohol sales: a legal impact study. *J Stud Alcohol, 52*(6), 555-567. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1758183

Mercer, K. A., Chintalapudi, S. R., & Visconti, E. B. (1999). Impact of targeted antibiotic restriction on usage and cost in a community hospital. *Journal of Pharmacy Technology, 15*(3), 79-84. Retrieved from

Puig-Junoy, J. (2007). The impact of generic reference pricing interventions in the statin market. *Health Policy, 84*(1), 14-29. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17368619 doi:10.1016/j.healthpol.2007.02.010

Rabinowitz, P. M., Galusha, D., Kirsche, S. R., Cullen, M. R., Slade, M. D., & Dixon-Ernst, C. (2011). Effect of daily noise exposure monitoring on annual rates of hearing loss in industrial workers. *Occup Environ Med, 68*(6), 414-418. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21193566 doi:10.1136/oem.2010.055905

Siegfried, N., Pienaar, D. C., Ataguba, J. E., Volmink, J., Kredo, T., Jere, M., & Parry, C. D. (2014). Restricting or banning alcohol advertising to reduce alcohol consumption in adults and adolescents. *Cochrane Database Syst Rev, 11*, CD010704. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25369459 doi:10.1002/14651858.CD010704.pub2

Smart, R. G., & Cutler, R. E. (1976). The alcohol advertising ban in British Columbia: problems and effects on beverage consumption. *Br J Addict Alcohol Other Drugs, 71*(1), 13-21.

Sossai, D., Puro, V., Chiappatoli, L., Dagnino, G., Odone, B., Polimeri, A., . . . Scognamiglio, P. (2010). Using an intravenous catheter system to prevent needlestick injury. *Nurs Stand, 24*(29), 42-46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20426370 doi:10.7748/ns2010.03.24.29.42.c7628

van der Molen, H. F., Lehtola, M. M., Lappalainen, J., Hoonakker, P. L., Hsiao, H., Haslam, R., . . . Verbeek, J. H. (2012). Interventions to prevent injuries in construction workers. *Cochrane Database Syst Rev, 12*, CD006251. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23235627 doi:10.1002/14651858.CD006251.pub3

Verbeek, J. H., Kateman, E., Morata, T. C., Dreschler, W. A., & Mischke, C. (2012). Interventions to prevent occupational noise-induced hearing loss. *Cochrane Database Syst Rev, 10*, CD006396. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23076923 doi:10.1002/14651858.CD006396.pub3

Appendix A

Ten intervention types:

1. Behavioural/educational: any intervention to change people's behaviour, e.g. information campaign to reduce smoking
2. Clinical: any medical diagnostic or treatment procedure going beyond administration of pharmaceuticals only, e.g. heart surgery
3. Environmental: targeting the physical environment, e.g. constructions of latrines to reduce diarrhoea
4. Health policy: laws, regulations or policies with the primary goal to improve health, e.g. policy to control costs of pharmaceuticals
5. Health system: targeting health institutions or the organisation of the health system, e.g. training of supervisory health staff to improve quality of care
6. Nutrition: modifying people's nutrition directly, e.g. zinc supplementation among children, or indirectly, e.g. education programmes or policies
7. Occupational: improving health at or through the workplace, e.g. hearing protection for construction workers
8. Pharmaceutical: e.g. aspirin treatment to reduce hypertension
9. Screening: e.g. screening for colon cancer in men
10. Vaccination: e.g. vaccination against human papillomavirus to reduce cervical cancer rates

Appendix B

References of studies including CBA and ITS studies

Acosta, A., Ciapponi, A., Aaserud, M., Vietto, V., Austvoll-Dahlgren, A., Kosters, J. P., . . . Oxman, A. D. (2014). Pharmaceutical policies: effects of reference pricing, other pricing, and purchasing policies. *Cochrane Database Syst Rev, 10*, CD005979. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25318966 doi:10.1002/14651858.CD005979.pub2

Akl, E. A., Kairouz, V. F., Sackett, K. M., Erdley, W. S., Mustafa, R. A., Fiander, M., . . . Schunemann, H. (2013). Educational games for health professionals. *Cochrane Database Syst Rev*(3), CD006411. doi:10.1002/14651858.CD006411.pub4

Algie, C. M., Mahar, R. K., Wasiak, J., Batty, L., Gruen, R. L., & Mahar, P. D. (2015). Interventions for reducing wrong-site surgery and invasive clinical procedures. *Cochrane Database Syst Rev,* (3), CD009404. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25821069 doi:10.1002/14651858.CD009404.pub3

Atherton, H., Sawmynaden, P., Meyer, B., & Car, J. (2012). Email for the coordination of healthcare appointments and attendance reminders. *Cochrane Database Syst Rev*(8), CD007981. doi:10.1002/14651858.CD007981.pub2

Atherton, H., Sawmynaden, P., Sheikh, A., Majeed, A., & Car, J. (2012). Email for clinical communication between patients/caregivers and healthcare professionals. *Cochrane Database Syst Rev, 11*, CD007978. doi:10.1002/14651858.CD007978.pub2

Aubin, M., Giguere, A., Martin, M., Verreault, R., Fitch, M. I., Kazanjian, A., & Carmichael, P. H. (2012). Interventions to improve continuity of care in the follow-up of patients with cancer. *Cochrane Database Syst Rev*(7), CD007672. doi:10.1002/14651858.CD007672.pub2

Baker, P. R., Francis, D. P., Soares, J., Weightman, A. L., & Foster, C. (2015). Community wide interventions for increasing physical activity. *Cochrane Database Syst Rev, 1*, CD008366. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25556970 doi:10.1002/14651858.CD008366.pub3

Ballini, L., Negro, A., Maltoni, S., Vignatelli, L., Flodgren, G., Simera, I., . . . Grilli, R. (2015). Interventions to reduce waiting times for elective procedures. *Cochrane Database Syst Rev,* (2), CD005610. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25706039 doi:10.1002/14651858.CD005610.pub2

Beller, E. M., van Driel, M. L., McGregor, L., Truong, S., & Mitchell, G. (2015). Palliative pharmacological sedation for terminally ill adults. *Cochrane Database Syst Rev, 1*, CD010206. doi:10.1002/14651858.CD010206.pub2

Brocklehurst, P., Price, J., Glenny, A. M., Tickle, M., Birch, S., Mertz, E., & Grytten, J. (2013). The effect of different methods of remuneration on the behaviour of primary care dentists. *Cochrane Database Syst Rev*(11), CD009853. doi:10.1002/14651858.CD009853.pub2

Brusamento, S., Ghanotakis, E., Tudor Car, L., van-Velthoven, M. H., Majeed, A., & Car, J. (2012). Male involvement for increasing the effectiveness of prevention of mother-to-child HIV transmission (PMTCT) programmes. *Cochrane Database Syst Rev, 10*, CD009468. doi:10.1002/14651858.CD009468.pub2

Coppo, A., Galanti, M. R., Giordano, L., Buscemi, D., Bremberg, S., & Faggiano, F. (2014). School policies for preventing smoking among young people. *Cochrane Database Syst Rev*(10), CD009990. doi:10.1002/14651858.CD009990.pub2

Curti, S., Sauni, R., Spreeuwers, D., De Schryver, A., Valenty, M., Riviere, S., & Mattioli, S. (2015). Interventions to increase the reporting of occupational diseases by physicians. *Cochrane Database Syst Rev, 3*, CD010305. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25805310 doi:10.1002/14651858.CD010305.pub2

Dangour, A. D., Watson, L., Cumming, O., Boisson, S., Che, Y., Velleman, Y., . . . Uauy, R. (2013). Interventions to improve water quality and supply, sanitation and hygiene practices, and

their effects on the nutritional status of children. *Cochrane Database Syst Rev*(8), CD009382. doi:10.1002/14651858.CD009382.pub2
Davey, P., Brown, E., Charani, E., Fenelon, L., Gould, I. M., Holmes, A., . . . Wilcox, M. (2013). Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev, 4*, CD003543. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23633313 doi:10.1002/14651858.CD003543.pub3
de Jongh, T., Gurol-Urganci, I., Vodopivec-Jamsek, V., Car, J., & Atun, R. (2012). Mobile phone messaging for facilitating self-management of long-term illnesses. *Cochrane Database Syst Rev, 12*, CD007459. doi:10.1002/14651858.CD007459.pub2
Dyer, T. A., Brocklehurst, P., Glenny, A. M., Davies, L., Tickle, M., Issac, A., & Robinson, P. G. (2014). Dental auxiliaries for dental care traditionally provided by dentists. *Cochrane Database Syst Rev*(8), CD010076. doi:10.1002/14651858.CD010076.pub2
Ferri, M., Allara, E., Bo, A., Gasparrini, A., & Faggiano, F. (2013). Media campaigns for the prevention of illicit drug use in young people. *Cochrane Database Syst Rev, 6*, CD009287. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23740538 doi:10.1002/14651858.CD009287.pub2
Fiander, M., McGowan, J., Grad, R., Pluye, P., Hannes, K., Labrecque, M., . . . Tugwell, P. (2015). Interventions to increase the use of electronic health information by healthcare practitioners to improve clinical practice and patient outcomes. *Cochrane Database Syst Rev*(3), CD004749. doi:10.1002/14651858.CD004749.pub3
Flodgren, G., Conterno, L. O., Mayhew, A., Omar, O., Pereira, C. R., & Shepperd, S. (2013). Interventions to improve professional adherence to guidelines for prevention of device-related infections. *Cochrane Database Syst Rev,* (3), CD006559. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23543545 doi:10.1002/14651858.CD006559.pub2
Galvagno, S. M., Jr., Thomas, S., Stephens, C., Haut, E. R., Hirshon, J. M., Floccare, D., & Pronovost, P. (2013). Helicopter emergency medical services for adults with major trauma. *Cochrane Database Syst Rev*(3), CD009228. doi:10.1002/14651858.CD009228.pub2
Gentry, S., van-Velthoven, M. H., Tudor Car, L., & Car, J. (2013). Telephone delivered interventions for reducing morbidity and mortality in people with HIV infection. *Cochrane Database Syst Rev*(5), CD009189. doi:10.1002/14651858.CD009189.pub2
Giguere, A., Legare, F., Grimshaw, J., Turcotte, S., Fiander, M., Grudniewicz, A., . . . Gagnon, M. P. (2012). Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev, 10*, CD004398. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23076904 doi:10.1002/14651858.CD004398.pub3
Gillaizeau, F., Chan, E., Trinquart, L., Colombet, I., Walton, R. T., Rege-Walther, M., . . . Durieux, P. (2013). Computerized advice on drug dosage to improve prescribing practice. *Cochrane Database Syst Rev*(11), CD002894. doi:10.1002/14651858.CD002894.pub3
Gomes, B., Calanzani, N., Curiale, V., McCrone, P., & Higginson, I. J. (2013). Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers. *Cochrane Database Syst Rev,* (6), CD007760. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23744578 doi:10.1002/14651858.CD007760.pub2
Goyder, C., Atherton, H., Car, M., Heneghan, C. J., & Car, J. (2015). Email for clinical communication between healthcare professionals. *Cochrane Database Syst Rev*(2), CD007979. doi:10.1002/14651858.CD007979.pub3
Gupta, S., Carmichael, C., Simpson, C., Clarke, M. J., Allen, C., Gao, Y., . . . Murray, V. (2012). Electric fans for reducing adverse health impacts in heatwaves. *Cochrane Database Syst Rev*(7), CD009888. doi:10.1002/14651858.CD009888.pub2
Gurol-Urganci, I., de Jongh, T., Vodopivec-Jamsek, V., Car, J., & Atun, R. (2012). Mobile phone messaging for communicating results of medical investigations. *Cochrane Database Syst Rev*(6), CD007456. doi:10.1002/14651858.CD007456.pub2

Hardt, J., Meerpohl, J. J., Metzendorf, M. I., Kienle, P., Post, S., & Herrle, F. (2013). Lateral pararectal versus transrectal stoma placement for prevention of parastomal herniation. *Cochrane Database Syst Rev*(11), CD009487. doi:10.1002/14651858.CD009487.pub2

Harrod, C. S., Goss, C. W., Stallones, L., & DiGuiseppi, C. (2014). Interventions for primary prevention of suicide in university and other post-secondary educational settings. *Cochrane Database Syst Rev, 10*, CD009439. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25353703 doi:10.1002/14651858.CD009439.pub2

Hayes, S. L., Mann, M. K., Morgan, F. M., Kitcher, H., Kelly, M. J., & Weightman, A. L. (2011). Collaboration between local health and local government agencies for health improvement. *Cochrane Database Syst Rev,* (6), CD007825. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21678371 doi:10.1002/14651858.CD007825.pub5

Hughes, C., Tunney, M., & Bradley, M. C. (2013). Infection control strategies for preventing the transmission of meticillin-resistant Staphylococcus aureus (MRSA) in nursing homes for older people. *Cochrane Database Syst Rev*(11), CD006354. doi:10.1002/14651858.CD006354.pub4

Jayaraman, S., Sethi, D., & Wong, R. (2014). Advanced training in trauma life support for ambulance crews. *Cochrane Database Syst Rev,* (8), CD003109. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25144654 doi:10.1002/14651858.CD003109.pub3

Jia, L., Yuan, B., Huang, F., Lu, Y., Garner, P., & Meng, Q. (2014). Strategies for expanding health insurance coverage in vulnerable populations. *Cochrane Database Syst Rev*(11), CD008194. doi:10.1002/14651858.CD008194.pub3

Khangura, J. K., Flodgren, G., Perera, R., Rowe, B. H., & Shepperd, S. (2012). Primary care professionals providing non-urgent care in hospital emergency departments. *Cochrane Database Syst Rev, 11*, CD002097. doi:10.1002/14651858.CD002097.pub3

Kristjansson, E., Francis, D. K., Liberato, S., Benkhalti Jandu, M., Welch, V., Batal, M., . . . Petticrew, M. (2015). Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years. *Cochrane Database Syst Rev, 3*, CD009924. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25739460 doi:10.1002/14651858.CD009924.pub2

Lavoie, M. C., Verbeek, J. H., & Pahwa, M. (2014). Devices for preventing percutaneous exposure injuries caused by needles in healthcare personnel. *Cochrane Database Syst Rev, 3*, CD009740. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24610008 doi:10.1002/14651858.CD009740.pub2

Lazzerini, M., Rubert, L., & Pani, P. (2013). Specially formulated foods for treating children with moderate acute malnutrition in low- and middle-income countries. *Cochrane Database Syst Rev*(6), CD009584. doi:10.1002/14651858.CD009584.pub2

Legare, F., Stacey, D., Turcotte, S., Cossi, M. J., Kryworuchko, J., Graham, I. D., . . . Donner-Banzhoff, N. (2014). Interventions for improving the adoption of shared decision making by healthcare professionals. *Cochrane Database Syst Rev*(9), CD006732. doi:10.1002/14651858.CD006732.pub3

Lewis, S. R., Nicholson, A., Smith, A. F., & Alderson, P. (2014). Physician anaesthetists versus non-physician providers of anaesthesia for surgical patients. *Cochrane Database Syst Rev*(7), CD010357. doi:10.1002/14651858.CD010357.pub2

Maaskant, J. M., Vermeulen, H., Apampa, B., Fernando, B., Ghaleb, M. A., Neubert, A., . . . Soe, A. (2015). Interventions for reducing medication errors in children in hospital. *Cochrane Database Syst Rev, 3*, CD006208. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25756542 doi:10.1002/14651858.CD006208.pub3

Meyer, B., Atherton, H., Sawmynaden, P., & Car, J. (2012). Email for communicating results of diagnostic medical investigations to patients. *Cochrane Database Syst Rev*(8), CD007980. doi:10.1002/14651858.CD007980.pub2

Mischke, C., Verbeek, J. H., Job, J., Morata, T. C., Alvesalo-Kuusi, A., Neuvonen, K., . . . Pedlow, R. I. (2013). Occupational safety and health enforcement tools for preventing occupational diseases and injuries. *Cochrane Database Syst Rev, 8*, CD010183. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23996220 doi:10.1002/14651858.CD010183.pub2

Moran, P. S., Flattery, M. J., Teljeur, C., Ryan, M., & Smith, S. M. (2013). Effectiveness of systematic screening for the detection of atrial fibrillation. *Cochrane Database Syst Rev*(4), CD009586. doi:10.1002/14651858.CD009586.pub2

Muckle, W., Muckle, J., Welch, V., & Tugwell, P. (2012). Managed alcohol as a harm reduction intervention for alcohol addiction in populations at high risk for substance abuse. *Cochrane Database Syst Rev, 12*, CD006747. doi:10.1002/14651858.CD006747.pub2

Murthy, L., Shepperd, S., Clarke, M. J., Garner, S. E., Lavis, J. N., Perrier, L., . . . Straus, S. E. (2012). Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians. *Cochrane Database Syst Rev,* (9), CD009401. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22972142 doi:10.1002/14651858.CD009401.pub2

Omidvari, A. H., Vali, Y., Murray, S. M., Wonderling, D., & Rashidian, A. (2013). Nutritional screening for improving professional practice for patient outcomes in hospital and primary care settings. *Cochrane Database Syst Rev, 6*, CD005539. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23744516 doi:10.1002/14651858.CD005539.pub2

Pande, S., Hiller, J. E., Nkansah, N., & Bero, L. (2013). The effect of pharmacist-provided non-dispensing services on patient outcomes, health service utilisation and costs in low- and middle-income countries. *Cochrane Database Syst Rev*(2), CD010398. doi:10.1002/14651858.CD010398

Pani, P. P., Trogu, E., Maremmani, I., & Pacini, M. (2013). QTc interval screening for cardiac risk in methadone treatment of opioid dependence. *Cochrane Database Syst Rev*(6), CD008939. doi:10.1002/14651858.CD008939.pub2

Parmelli, E., Flodgren, G., Fraser, S. G., Williams, N., Rubin, G., & Eccles, M. P. (2012). Interventions to increase clinical incident reporting in health care. *Cochrane Database Syst Rev,* (8), CD005609. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22895951 doi:10.1002/14651858.CD005609.pub2

Pega, F., Carter, K., Blakely, T., & Lucas, P. J. (2013). In-work tax credits for families and their impact on health status in adults. *Cochrane Database Syst Rev, 8*, CD009963. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23921458 doi:10.1002/14651858.CD009963.pub2

Reeves, S., Perrier, L., Goldman, J., Freeth, D., & Zwarenstein, M. (2013). Interprofessional education: effects on professional practice and healthcare outcomes (update). *Cochrane Database Syst Rev,* (3), CD002213. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23543515 doi:10.1002/14651858.CD002213.pub3

Rockers, P. C., & Barnighausen, T. (2013). Interventions for hiring, retaining and training district health systems managers in low- and middle-income countries. *Cochrane Database Syst Rev,* (4), CD009035. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23633365 doi:10.1002/14651858.CD009035.pub2

Rolfe, A., Cash-Gibson, L., Car, J., Sheikh, A., & McKinstry, B. (2014). Interventions for improving patients' trust in doctors and groups of doctors. *Cochrane Database Syst Rev*(3), CD004134. doi:10.1002/14651858.CD004134.pub3

Ruotsalainen, J. H., Verbeek, J. H., Marine, A., & Serra, C. (2014). Preventing occupational stress in healthcare workers. *Cochrane Database Syst Rev,* (12), CD002892. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25482522 doi:10.1002/14651858.CD002892.pub4

Rutebemberwa, E., Kinengyere, A. A., Ssengooba, F., Pariyo, G. W., & Kiwanuka, S. N. (2014). Financial interventions and movement restrictions for managing the movement of health

workers between public and private organizations in low- and middle-income countries. *Cochrane Database Syst Rev*(2), CD009845. doi:10.1002/14651858.CD009845.pub2

Saeterdal, I., Lewin, S., Austvoll-Dahlgren, A., Glenton, C., & Munabi-Babigumira, S. (2014). Interventions aimed at communities to inform and/or educate about early childhood vaccination. *Cochrane Database Syst Rev*(11), CD010232. doi:10.1002/14651858.CD010232.pub2

Sauni, R., Verbeek, J. H., Uitti, J., Jauhiainen, M., Kreiss, K., & Sigsgaard, T. (2015). Remediating buildings damaged by dampness and mould for preventing or reducing respiratory tract symptoms, infections and asthma. *Cochrane Database Syst Rev, 2*, CD007897. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25715323 doi:10.1002/14651858.CD007897.pub3

Sawmynaden, P., Atherton, H., Majeed, A., & Car, J. (2012). Email for the provision of information on disease prevention and health promotion. *Cochrane Database Syst Rev, 11*, CD007982. doi:10.1002/14651858.CD007982.pub2

Siegfried, N., Pienaar, D. C., Ataguba, J. E., Volmink, J., Kredo, T., Jere, M., & Parry, C. D. (2014). Restricting or banning alcohol advertising to reduce alcohol consumption in adults and adolescents. *Cochrane Database Syst Rev, 11*, CD010704. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25369459 doi:10.1002/14651858.CD010704.pub2

Tudor Car, L., Gentry, S., van-Velthoven, M. H., & Car, J. (2013). Telephone communication of HIV testing results for improving knowledge of HIV infection status. *Cochrane Database Syst Rev*(1), CD009192. doi:10.1002/14651858.CD009192.pub2

Turley, R., Saith, R., Bhan, N., Rehfuess, E., & Carter, B. (2013). Slum upgrading strategies involving physical environment and infrastructure interventions and their effects on health and socio-economic outcomes. *Cochrane Database Syst Rev, 1*, CD010067. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23440845 doi:10.1002/14651858.CD010067.pub2

van-Velthoven, M. H., Tudor Car, L., Gentry, S., & Car, J. (2013). Telephone delivered interventions for preventing HIV infection in HIV-negative persons. *Cochrane Database Syst Rev*(5), CD009190. doi:10.1002/14651858.CD009190.pub2

van der Molen, H. F., Lehtola, M. M., Lappalainen, J., Hoonakker, P. L., Hsiao, H., Haslam, R., . . . Verbeek, J. H. (2012). Interventions to prevent injuries in construction workers. *Cochrane Database Syst Rev, 12*, CD006251. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23235627 doi:10.1002/14651858.CD006251.pub3

van Ginneken, N., Tharyan, P., Lewin, S., Rao, G. N., Meera, S. M., Pian, J., . . . Patel, V. (2013). Non-specialist health worker interventions for the care of mental, neurological and substance-abuse disorders in low- and middle-income countries. *Cochrane Database Syst Rev,* (11), CD009149. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24249541 doi:10.1002/14651858.CD009149.pub2

Verbeek, J. H., Kateman, E., Morata, T. C., Dreschler, W. A., & Mischke, C. (2012). Interventions to prevent occupational noise-induced hearing loss. *Cochrane Database Syst Rev, 10*, CD006396. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23076923 doi:10.1002/14651858.CD006396.pub3

Vodopivec-Jamsek, V., de Jongh, T., Gurol-Urganci, I., Atun, R., & Car, J. (2012). Mobile phone messaging for preventive health care. *Cochrane Database Syst Rev, 12*, CD007457. doi:10.1002/14651858.CD007457.pub2

Walshe, D. P., Garner, P., Abdel-Hameed Adeel, A. A., Pyke, G. H., & Burkot, T. (2013). Larvivorous fish for preventing malaria transmission. *Cochrane Database Syst Rev*(12), CD008090. doi:10.1002/14651858.CD008090.pub2

Webster, J., & Flint, A. (2014). Exit interviews to reduce turnover amongst healthcare professionals. *Cochrane Database Syst Rev*(8), CD006620. doi:10.1002/14651858.CD006620.pub5

Publication II

scientific reports

Check for updates

OPEN

Interrupted time series study found mixed effects of the impact of the Bavarian smoke-free legislation on pregnancy outcomes

Stephanie Polus[1,2]✉, Jacob Burns[1,2], Sabine Hoffmann[1,2], Tim Mathes[3], Ulrich Mansmann[1,2], Jasper V. Been[4], Nicholas Lack[5], Daniela Koller[1,2], Werner Maier[6] & Eva A. Rehfuess[1,2]

In 2007 the German government passed smoke-free legislation, leaving the details of implementation to the individual federal states. In January 2008 Bavaria implemented one of the strictest laws in Germany. We investigated its impact on pregnancy outcomes and applied an interrupted time series (ITS) study design to assess any changes in preterm birth, small for gestational age (primary outcomes), and low birth weight, stillbirth and very preterm birth. We included 1,236,992 singleton births, comprising 83,691 preterm births and 112,143 small for gestational age newborns. For most outcomes we observed unclear effects. For very preterm births, we found an immediate drop of 10.4% (95%CI − 15.8, − 4.6%; p = 0.0006) and a gradual decrease of 0.5% (95%CI − 0.7, − 0.2%, p = 0.0010) after implementation of the legislation. The majority of subgroup and sensitivity analyses confirm these results. Although we found no statistically significant effect of the Bavarian smoke-free legislation on most pregnancy outcomes, a substantial decrease in very preterm births was observed. We cannot rule out that despite our rigorous methods and robustness checks, design-inherent limitations of the ITS study as well as country-specific factors, such as the ambivalent German policy context have influenced our estimation of the effects of the legislation.

Over the past two decades, a range of policies and programmes at global, national and regional levels have been designed to reduce the detrimental harms associated with tobacco use[1,2]. There is strong evidence that smoke-free legislation improves adult health outcomes, such as cardiovascular health and mortality from smoking-related illnesses[3]. There is also evidence that smoke-free legislation improves pregnancy outcomes and child health[3,4]. For example, a recent systematic review found reductions in preterm birth rates, perinatal mortality and hospital attendance rates for asthma following smoke-free legislation[4]. However, findings are not fully consistent and based on less rigorous study methods[3,4], and more rigorous studies are needed to strengthen the evidence base[5].

Since ratification of the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) in 2004, Germany is obliged by international law to implement appropriate measures to reduce and prevent tobacco consumption and second-hand smoke (SHS) exposure[6]. Germany prohibited smoking in the workplace in 2004[7]. In 2007 a national law was passed to protect non-smokers from the harmful consequences of SHS and required the implementation of federal state level legislation to prohibit smoking in public places[8]. Sargent et al.[9] investigated the short-term effects of the smoke-free legislation on the national level and found a significant decrease in hospital admissions due to acute coronary events after implementation of the smoke-free legislation. No study, however, has assessed the effects of the smoke-free legislation on pregnancy outcomes in the German context.

[1]Institute for Medical Information Processing, Biometry, and Epidemiology – IBE, LMU Munich, Munich, Germany. [2]Pettenkofer School of Public Health, Munich, Germany. [3]Institute for Research in Operative Medicine, Faculty of Health, School of Medicine, Witten/Herdecke University, Cologne, Germany. [4]Division of Neonatology, Department of Paediatrics, Department of Obstetrics and Gynaecology, Department of Public Health, Erasmus MC – Sophia Children's Hospital, Rotterdam, The Netherlands. [5]German Bavarian Quality Assurance Institute for Medical Care, Munich, Germany. [6]Institute of Health Economics and Health Care Management, Helmholtz Zentrum München – German Research Center for Environmental Health (GmbH), Neuherberg, Germany. ✉email: polus@ibe.med.uni-muenchen.de

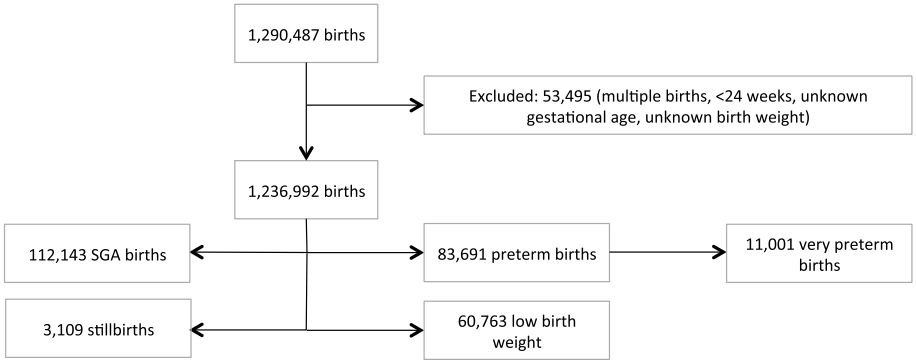


Figure 1. Flow chart of study population and primary and secondary outcomes.

In this study, we assess the impact of the smoke-free legislation on pregnancy outcomes implemented on 1 January 2008 in Bavaria, Germany's largest and second most populous state with more than 12.5 million inhabitants and approximately one-sixth of all births in Germany[10].

Results

There were 1,290,487 deliveries between 1 January 2005 and 31 December 2016. Due to the standardized data collection process related to Bavarian hospital births, there are no missing data of hospital births. We excluded 53,495 births (4.15%) because inclusion criteria were not met (see Fig. 1). The analysed time series thus included 1,236,992 singleton deliveries, which represents a monthly mean of 8,950 births (range 7,266–10,825). The number of births per month increased steadily over the last 5 years of the study period (see Supplementary Fig. S1 online).

Overall, there were 83,691 preterm births (< 37 gestational weeks) during the study period. During the same time span there were 112,143 babies born small for gestational age (SGA) (< 10th percentile) and 60,763 born with low birth weight (< 2500 g); 11,001 were very preterm births (< 32 gestational weeks) and 3,109 deliveries were stillbirths (intrauterine death > 500 g). Maternal, newborn and subgroup characteristics are specified in Table 1. Throughout the study period, the outcome rates of all primary and secondary outcomes stayed relatively constant (see Fig. 2). We did not observe a significant underlying trend in primary or secondary outcomes over the study period in the regression models of our main analyses.

Results of main analysis. For the two primary and two out of three secondary outcomes, i.e. preterm birth, SGA, low birth weight and stillbirth, any effect on level changes or slope changes following implementation of the legislation in January 2008 was not statistically significant. We observed small effects with 95% confidence intervals (CIs) that include a decrease as well as an increase of the outcome rates. Results are presented in Table 2 (rate ratios retrieved from exponential beta coefficients) and Fig. 2. These effect estimates can be interpreted as illustrated through the following example: For preterm births, we observed a rate ratio of 1.0163 (95%CI 0.9762, 1.0580) for level change. This represents an immediate relative increase of 1.63% (95%CI − 2.38, 5.80%) in the preterm birth rate (i.e. percentage of total births), which corresponds to a predicted increase in the preterm birth rate from 6.95 to 7.06% from December 2007 to January 2008. The calculated rate ratio of 0.9995 (95%CI 0.9976, 1.0013) for the slope change represents a gradual decrease of 0.05% (95%CI − 0.24, 0.13%) in the preterm birth rate. This corresponds, for example, to a change in the predicted monthly preterm birth rate from 7.031% in May 2008 to a rate of 7.025% in June 2008 and represents an average of two preterm births less every month after implementation of the intervention.

For the secondary outcome very preterm births we did observe a rate ratio of 0.8960 (95%CI 0.8413, 0.9542) for level change. This represents an immediate relative decrease of very preterm births by 10.40% (95%CI − 15.87, − 4.58%, p = 0.0006), corresponding to a level change from a very preterm birth rate of 0.98% in December 2007 to 0.89% in January 2008. We also observed a rate ratio of 0.9954 (95% 0.9928, 0.9982) for a slope change in very preterm births. This represents a decrease of 0.46% (95%CI − 0.72, − 0.18%, p = 0.0010) in the very preterm birth rate, corresponding to an additional relative decrease from e.g. 0.983% in July 2009 to 0.979% in August 2009, representing four very preterm births less each month.

Subgroup and sensitivity analyses. Consistent with the main analysis, most of the subgroup and sensitivity analyses showed small effects with confidence intervals suggesting that the effect could be in either direction (see Figs. 3 and 4 and Supplementary Fig. S2-7 online). The detailed results of the sensitivity and subgroup analyses are shown in Supplementary Table S1 online. Active smoking during pregnancy decreased throughout the study period (see Supplementary figure S8 online). For 284,421 deliveries (23%) smoking status information was missing and therefore not included in the subgroup analysis. The mean preterm birth rate was 2.48% (95%CI 2.25–2.70%) and the mean SGA rate 10.52% (95%CI 10.18–10.86%) higher for smokers than for non-smokers (see Supplementary Fig. S9 and S10 online). We did not observe a statistically significant level or slope change in the number of smoking mothers following implementation of the intervention (level = − 1.53, 95%CI − 6.83,

Maternal characteristics	Live and non-live births (%) (n = 1,236,992)	Preterm births (%) (n = 83,691)	SGA (%) (n = 112,143)	LBW (%) (n = 60,763)	Stillbirths (%) (n = 3,109)	Very preterm births (%) (n = 11,001)
Maternal age (years)						
< 20	19,045 (1.54)	1,594 (1.90)	2,544 (2.27)	1,363 (2.24)	59 (1.90)	249 (2.26)
20–24	129,279 (10.45)	9,311 (11.13)	14,940 (13.32)	7,425 (12.22)	384 (12.35)	1,299 (11.80)
25–29	329,984 (26.67)	21,916 (26.19)	30,763 (27.43)	15,714 (25.86)	745 (23.96)	2,711 (24.64)
30–34	436,317 (35.27)	27,910 (33.35)	36,667 (32.70)	19,499 (32.09)	984 (31.65)	3,486 (31.69)
35–39	258,574 (20.90)	17,695 (21.14)	21,320 (19.01)	12,784 (21.04)	714 (22.97)	2,517 (22.88)
≥ 40	63,793 (5.16)	5,265 (6.29)	5,909 (5.27)	3,978 (6.55)	223 (7.17)	738 (6.71)
Missing	0	0	0	0	0	1 (0.01)
Parity						
0	518,743 (41.94)	38,173 (45.61)	60,078 (53.57)	30,119 (45.57)	1,304 (41.94)	4,863 (44.21)
1	408,042 (32.99)	22,734 (27.16)	30,385 (27.09)	15,620 (25.71)	909 (29.24)	2,882 (26.20)
2	185,488 (15.00)	11,873 (14.19)	12,788 (11.40)	7,938 (13.06)	489 (15.73)	1,610 (14.64)
≤ 3	124,691 (10.10)	10,909 (13.03)	8,890 (7.93)	7,083 (11.66)	407 (13.09)	1,645 (14.95)
Missing	28 (0.00)	2 (0.00)	2 (0.00)	3 (0.00)	0 (0.00)	1 (0.01)
Smoking status						
Smokers	69,156 (5.59)	6,240 (7.46)	12,974 (11.57)	6,846 (11.27)	248 (7.98)	901 (8.19)
Non-smokers	883,415 (71.42)	57,795 (69.10)	73,264 (65.53)	39,481 (64.98)	2,061 (66.29)	7311 (66.46)
missing	284,421 (22.99)	19,656 (23.49)	25,902 (23.10)	14,434 (18.82)	800 (25.73)	2,789 (25.35)
SES according to BIMD quintiles						
BIMD 1 (least deprived)	163,522 (13.22)	10,677 (12.76)	13,569 (12.10)	7,402 (12.18)	404 (12.99)	1,263 (11.48)
BIMD2	147,174 (11.90)	9,823 (11.74)	12,876 (11.48)	6,921 (11.39)	338 (10.87)	1,185 (10.77)
BIMD3	168,880 (13.65)	11,366 (13.58)	15,095 (13.46)	8,222 (13.53)	399 (12.83)	1,413 (12.84)
BIMD4	362,012 (29.27)	23,660 (28.27)	32,558 (29.03)	17,333 (28.53)	895 (28.79)	3,145 (28.59)
BIMD5 (most deprived)	314,219 (25.40)	22,318 (26.67)	30,679 (27.36)	16,647 (27.40)	853 (27.44)	3,098 (28.16)
Missing	81,185 (6.56)	5,847 (6.97)	7,556 (6.74)	4,138 (6.81)	220 (7.08)	897 (8.15)
Nationality						
German	999,383 (80.79)	67,699 (80.89)	90,647 (80.83)	49,011 (80.66)	2,367 (76.13)	8,403 (76.39)
Other nationality/ missing	237,609 (19.21)	15,992 (19.11)	21,496 (19.17)	11,752 (19.34)	742 (23.87)	2,598 (23.62)
Infant characteristics						
Sex						
Male	634,111 (51.26)	46,290 (55.31)	57,606 (51.37)	28,863 (47.50)	1,633 (52.52)	6,037 (54.88)
Female	602,881 (48.74)	37,401 (44.69)	54,537 (48.63)	31,900 (52.50)	1,476 (47.48)	4,963 (45.11)
Missing	0	0	0	0	0	1

Table 1. Maternal, newborn and subgroup characteristics by outcome; the numbers represent the number of newborns/mothers with the respective characteristic and the percentage of the births within the outcome group; SES = socio-economic status, BIMD = Bavarian Index of Multiple Deprivation, ranging from least deprived quintile (BIMD1) to most deprived quintile (BIMD5); SGA = small for gestational age, LBW = low birth weight; percent values are rounded to the second decimal place.

4.06; slope = − 0.18, 95%CI − 0.45, 0.01). We further performed post-hoc sensitivity analyses for our secondary outcome very preterm birth where we could replicate our findings however without reaching statistical significance (see Supplementary Table S2 online).

Discussion

This study is the first to assess the impact of smoke-free legislation on pregnancy outcomes conducted in Germany. We did not observe clear immediate (level change) or gradual (slope change) effects of the smoke-free legislation on preterm birth, SGA, low birth weight and stillbirth; 95%-confidence intervals surrounding effect estimates suggest that the effect could be in either direction. Our findings were consistent across the majority of sensitivity and subgroup analyses.

We observed statistically significant immediate and gradual reductions for very preterm births. We could replicate these findings in the sensitivity and subgroup analyses especially for smokers; however, here the effects were not statistically significant, probably due to a lack of power considering the much smaller population. Although the preterm birth and SGA rates were higher for the most deprived quintile (BIMD 5) as compared to the least deprived (BIMD 1), consistent with the literature[4,11], we observed no clear differential effect of the legislation according to SES.

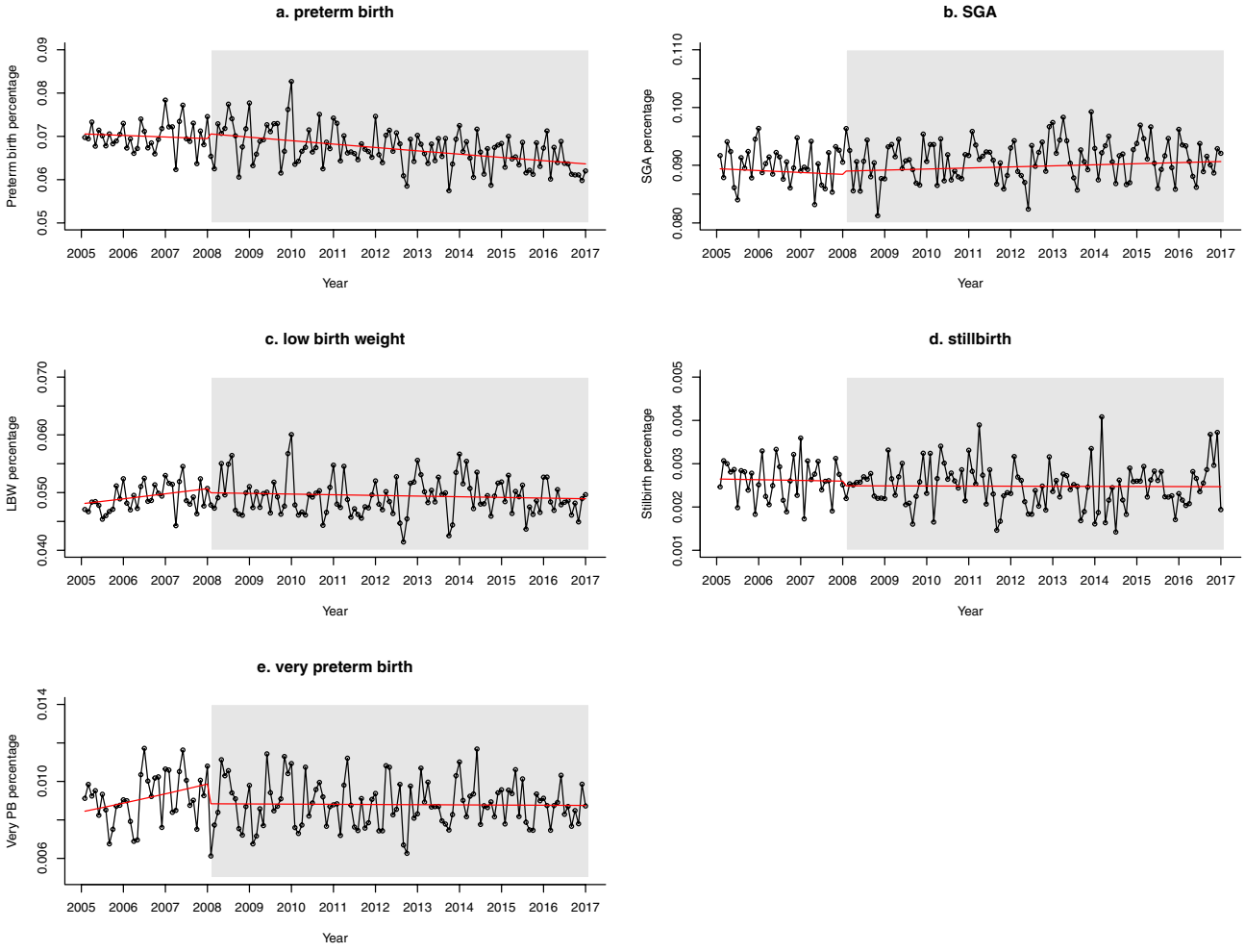


Figure 2. Panel figure presenting time series of pregnancy outcomes in panels (**a**–**e**); dots represent the monthly percentages; white = pre-intervention period, grey = post-intervention period, red line = regression line; SGA = small for gestational age (R version 3.5.1, https://www.r-project.org/).

Outcome	Rate ratio (95%CI)[1] level change	Rate ratio (95%CI) slope change	Model type
Preterm birth	1.0163 (0.9762, 1.0580)	0.9995 (0.9976, 1.0013)	Negative binomial model with autocorrelation terms
SGA	1.0063 (0.9839, 1.0292)	1.0005 (0.9995, 1.0014)	Poisson model with seasonal dummies and autocorrelation terms
LBW	0.9861 (0.9484, 1.0254)	0.9983 (0.9966, 1.0000)	Negative Binomial with seasonal dummies
Very preterm birth	0.8960 (0.8413, 0.9542), p = 0.0006	0.9954 (0.9928,0.9982), p = 0.0010	Negative binomial model with autocorrelation terms
Stillbirth	0.9583 (0.8165,1.1247)	1.0004 (0.9936, 1.0073)	Poisson model

Table 2. Estimates of level and slope changes in main analysis of primary and secondary outcomes; SGA = small for gestational age, LBW = low birth weight; All values are rounded to the fourth decimal place.

The literature shows mixed effects for the impact of smoke-free legislation on pregnancy outcomes, with a tendency towards a protective effect. Specifically, a recent systematic review and meta-analysis of mostly interrupted time series (ITS) studies performed by Faber et al.[4] observed an immediate drop (level change) in preterm birth, low birth weight, SGA and very preterm birth, but a gradual decline (slope change) only in SGA and very preterm birth. The effect estimates (risk differences) reported by the included studies were rather small for the outcomes preterm birth, SGA and low birth weight and greater for very preterm birth. In our study we observed, consistent with the meta-analysis in terms of relative effect size, unclear changes in all pregnancy outcomes following the smoke-free legislation except for very preterm birth.

In the absence of studies assessing pregnancy outcomes in Germany, up to now, only one published study employed a time series design to investigate the smoke-free legislation and found a short-term immediate decrease in hospital admissions for acute coronary events[9]. A difference-in-difference study examined short-term

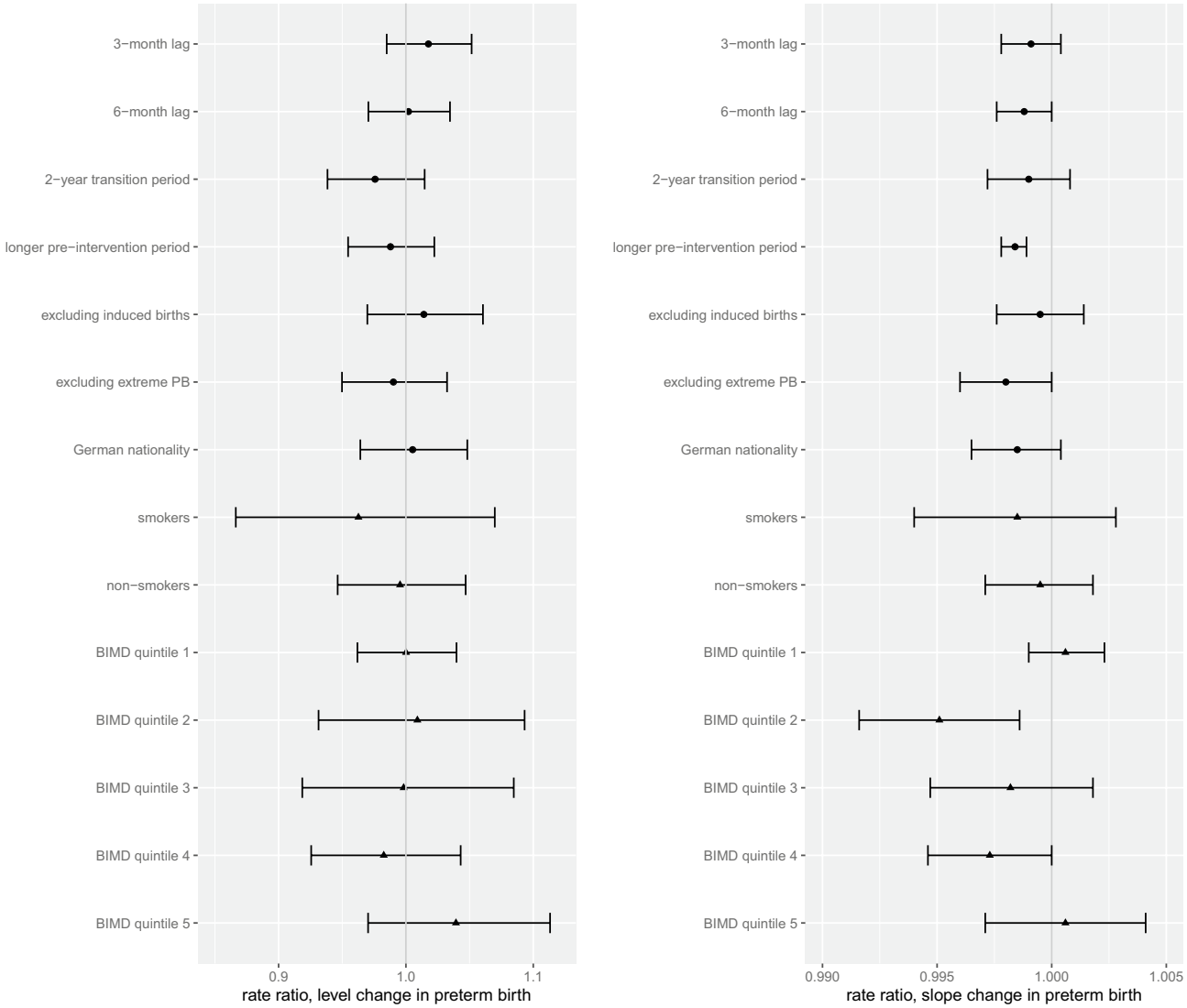


Figure 3. Rate ratios with 95%CIs for level and slope changes of preterm birth, sensitivity (●) and subgroup (▲) analyses (R version 3.5.1, https://www.r-project.org/).

effects of the legislation on smoking behaviour using German socio-economic panel (SOEP) data. It found no impact of the legislation in the general population but only for those frequenting bars and restaurants more often[12].

A previous study conducted in Canada[13] identified several factors, apart from methodological differences, that may explain the heterogeneous findings across studies assessing the effect of smoke-free legislation on pregnancy outcomes, namely: (i) different policy environments in terms of smoking prevalence and smoking norms, (ii) the presence of existing legislation prior to the smoke-free legislation under investigation, and (iii) differences in policy implementation and enforcement. These factors may contribute to understanding why we did not observe a clearer effect related to the Bavarian smoke-free legislation.

With regards to different policy environments, the smoking prevalence in other countries was similar to the prevalence in Germany, ranging from 18 to 27% among the female population[14–20]. However, Germany, with a prevalence of 27% in 2008–2011 lies at the upper end of this range[21] and Germany's efforts in tobacco control have been poor compared to other countries[22–24]. According to the Tobacco Control Ranking Scale 2019[25], Germany occupies the last rank for successful implementation of tobacco control among 36 mostly European countries. For example, Germany currently is the only EU country, which still allows tobacco advertising on billboards[1]. Tobacco smoking is a well-established risk factor for fetal growth restriction and preterm birth. Indeed, we found large differences between smoking and non-smoking mothers regarding rates of preterm birth and especially SGA. However, we did not observe any drop or gradual decrease of preterm birth or SGA rates in either subgroup following the implementation of the smoke-free legislation. Furthermore, we did not observe changes in smoking rates related to the implementation of the legislation as assumed in our logic model and shown in other studies[14,17,18]. However, caution is merited, as a 5.9% smoking prevalence in our data is very likely an underestimation. On the other hand, the lack of a clear effect could, for example, also be explained by

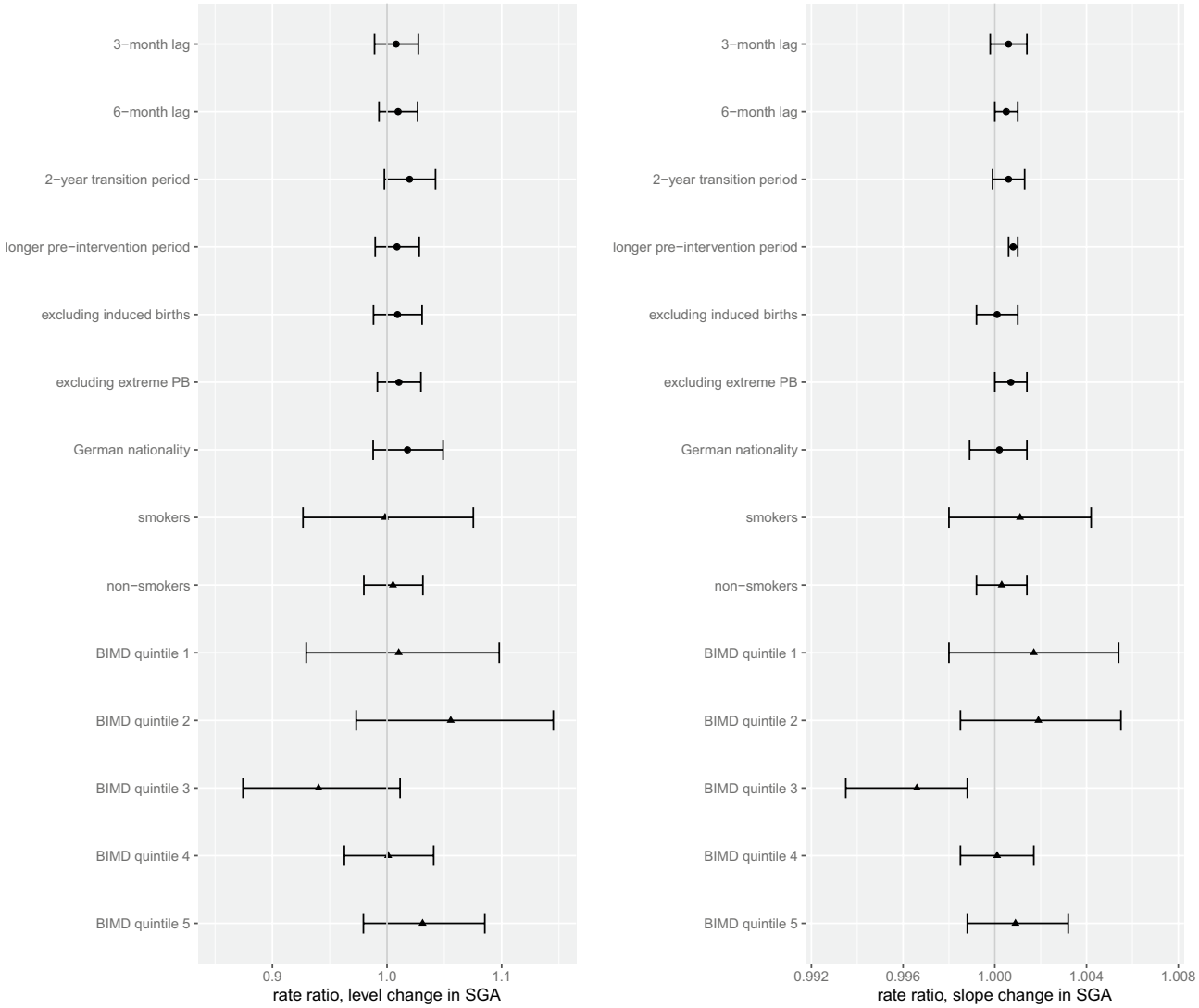


Figure 4. Rate ratios with 95%CIs for level and slope changes of small for gestational age (SGA), sensitivity (●) and subgroup (▲) analyses (R version 3.5.1, https://www.r-project.org/).

more health-conscious behaviour among pregnant women, who, prior to implementation of the smoke-free legislation, already avoided exposure to second-hand smoke.

Considering pre-existing legislation, as laid out in our logic model, the smoke-free legislation is not the only factor influencing pregnancy outcomes, and we did identify co-interventions and existing legislation prior to implementation of the legislation under investigation that may have subdued a more prevalent effect.

Finally, regarding the role of policy implementation and enforcement, recent studies have shown that the health impact is larger when the smoke-free legislation is more comprehensive[4,26]. While Bavaria has one of the strictest smoke-free legislations within Germany, the legislation lacked supporting interventions, such as an accompanying media campaign or tax increase. Studies in Spain, England and Scotland, where improvements in pregnancy outcomes associated with legislation were observed, also found high compliance with the legislation[17,18,27]. While we do not have data in Bavaria regarding actual enforcement or compliance, it is possible that the unclear effects could be explained by a lack of enforcement or compliance considering Germany's lacking efforts in tobacco control[12,22,24].

The ITS study design is prone to certain methodological limitations. The lack of randomization and thus potential confounding make it difficult to definitely attribute causality to the intervention-outcome relationship[5,28]. Beyond this design-inherent limitation, the use of rigorous a priori methods is important.

We used a high quality, large dataset and followed the steps outlined in the tutorial developed by Lopez Bernal et al.[29] to account for common methodological and conceptual flaws in assessing population level interventions with time series designs[30]. In particular, we took a complex systems approach and used a logic model to conceptualize our study and identified co-interventions and other risk factors prior to conducting our study[31]. We registered a detailed study protocol, in which we defined an impact model, the main statistical analysis, as well as the sensitivity and subgroup analyses. Had we used, for example, our originally planned impact model

including the full range of data from 2000 to 2016, we would have come to different conclusions. However, in such a model the pre-intervention slope (and therefore level and slope change) was defined primarily by a major breakpoint in 2004 caused by several co-interventions in this year. This emphasizes the importance of choosing an appropriate impact model, as described by Lopez Bernal et al.[32].

Our choice of the correct impact model, however, was also associated with several uncertainties. Many aspects of biological processes of pregnancy and especially the exact window of susceptibility of pregnancy to smoking are insufficiently understood[33,34]. Also, the interrupted implementation of the legislation in Bavaria from 2008 to 2010 makes it difficult to identify the exact time at which we can expect to see an effect. Some studies have investigated an immediate onset[18,19], or even an anticipatory intervention time point[17] whereas others have used an intervention time point nine months after the actual implementation of the intervention[13]. A study which investigated the smoke-free legislation in the different cantons of Switzerland, found that the more time a mother spent under the smoke-free legislation the fewer were the risks for preterm birth and early-term births[20].

A further design-inherent weakness of the single-arm ITS study design is the lack of a concomitant, geographical control group[35]. Indeed, a recent study has shown the limitations of single group ITS studies assessing the impact of smoke-free legislation on mortality in Spain where initial protective intervention effects from a single group ITS study were not confirmed after the addition of a comparable geographical control site[36]. Despite careful preparation and consideration, we may have failed to identify important confounders or co-interventions, considering the complexity of the intervention as well as of the system in which it was implemented. Additionally, little concrete guidance on choice of statistical model exists, and determining the 'best-fit' model among a range of alternatives remains at least partially arbitrary. Gasparrini et al.[37] already concluded that the model specifications, among other factors, have a strong impact on the effect estimate when assessing smoke-free legislation on acute myocardial infarction. We aimed, however, to define statistical parameters a priori, where possible, and to comprehensively report modelling choices by publishing our code alongside the manuscript.

Methods

We applied an ITS study design to assess the association between implementation of the smoke-free legislation and pregnancy outcomes using monthly data from all births in Bavaria between 2005 and 2016. The ITS study design is considered to be one of the best alternatives to assess intervention effectiveness of population-level interventions where randomization is considered infeasible[29,38]. It is increasingly used in the field of healthcare and public health[39,40]. This study design usually draws on routine data collected over time to identify any underlying time trends, and can thereby observe changes after the implementation of an intervention compared to a counterfactual scenario (i.e. a hypothetical scenario in which the intervention was not implemented[41]). We pre-specified the study methods including the main impact model, main analysis, and subgroup and sensitivity analyses in a study protocol (available at www.drks.de, study ID: DRKS00014805).

Data source. We included aggregated data from a high-quality routine dataset of maternal and neonatal health indicators, collected and managed by the Bavarian Institute for Quality Assurance in hospital care (BAQ). This dataset contains all Bavarian in-hospital births, which constitute about 99% of all births in Bavaria[42]. It provides extensive information retrieved from all hospitals in Bavaria regarding maternal and neonatal demographic and health-related characteristics, clinical management, and pregnancy complications. The data are subjected to a series of formal and contextual plausibility checks[43].

Outcomes. Our primary outcomes included preterm birth (< 37 gestational weeks), and SGA (< the 10th percentile, adjusted for gestational age and sex based on Voigt et al.[44]), both measured as the percentage of these outcomes among all births that occurred during a given month. Secondary outcomes included monthly percentages of low birth weight (< 2500 g), very preterm birth (< 32 gestational weeks), and stillbirth (intrauterine death > 500 g).

Inclusion and exclusion criteria. We included all live and non-live singleton births from 24 until 42 completed weeks of gestation that occurred between January 2005 and December 2016. We excluded pregnancies with multiple births due to their increased risk of preterm birth, low birth weight and other pregnancy complications. We also excluded pregnancies with unknown gestational length and children with unknown birth weight.

Logic model. We developed a logic model that describes how the intervention and other factors influence pregnancy outcomes (see Fig. 5) to conceptualise the study and decide on the impact model and statistical analysis. The logic model, derived from literature searches, within-team discussions and expert consultations, provides a structure to help authors address complexity and thus better understand the interactions between the intervention, its implementation and multiple outcomes among a population and context[45,46].

We identified several co-interventions (i.e. other interventions, measures or policies that occur during the same time period) and other risk factors influencing pregnancy outcomes over the study period, which we describe in detail in our study protocol.

Impact model. The 16 German federal states are responsible for the implementation of the national law for the protection of non-smokers, and individual state legislation varies in strength (e.g. partial smoke-free laws), as well as in timing of implementation (ranging from August 2007 to July 2008). Bavaria implemented the smoke-free legislation on 1 January 2008. Thereafter, smoking was prohibited in all public buildings and institutions, such as universities, hospitals, retirement and nursing homes, and restaurants and bars[47]. Due to political

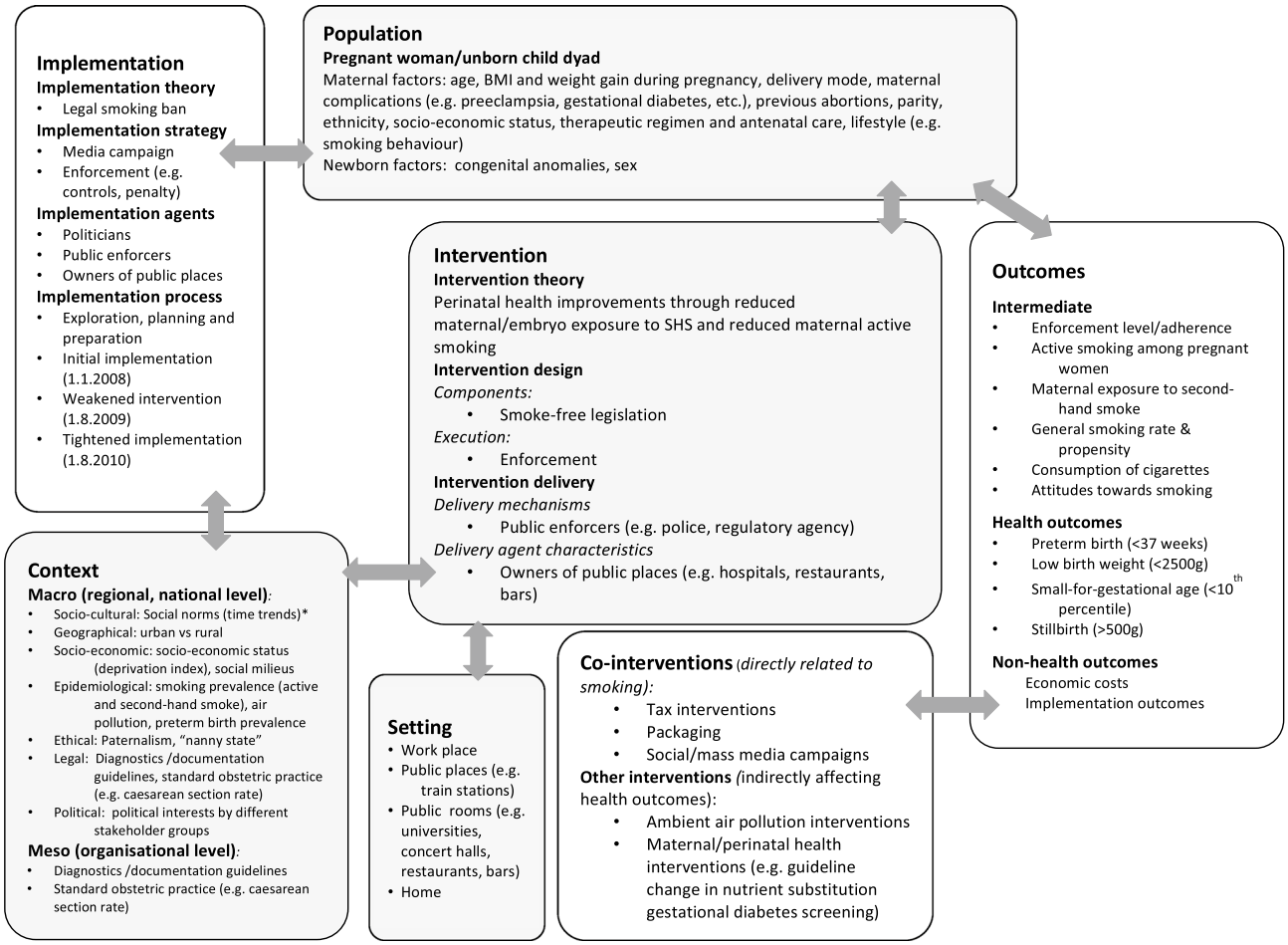


Figure 5. Logic model of the Bavarian smoke-free legislation.

arguments, however, the Bavarian legislation was loosened on 1 August 2009, and smoking was permitted again in a subset of restaurants, e.g. in restaurants larger than 75 m^2 mainly serving drinks. Following a referendum, which allowed Bavarian citizens to vote directly for or against more restrictive smoke-free legislation, the legislation was tightened again on 1 August 2010. This iteration additionally banned smoking in beer and event tents[48,49], making the Bavarian smoke-free legislation one of the strictest in Germany[48]. Violations of smoke-free legislation for smokers as well as event organisers include fines between 5 and 1,000 Euros, but information on enforcement and compliance is lacking.

We hypothesized that the effects of the smoke-free legislation may be detected as an immediate drop (level change) and gradual decline (slope change) in pregnancy outcomes at the first introduction of the smoke-free legislation on 1 January 2008. The changes could be the result of an immediate reduction in maternal exposure to SHS, and/or an immediate reduction in active maternal smoking (in public places and potentially elsewhere). They could also be impacted by longer-term influences on sociocultural norms, affecting smoking behaviours in different settings[50–52].

Originally, we planned to use data from 2000 to 2016 with equal time periods pre- and post-intervention. Upon visual inspection of the data we identified, however, a series of pronounced changes in outcome rates in the year 2004, during the pre-intervention period, that we were not able to sufficiently account for through adjustments in the analysis. These changes were potentially triggered by the smoking ban at work in August 2004[7], a major cigarette price increase in September 2004[53], as well as a documentation change initiated in January 2004. Therefore, we shortened the pre-intervention time period, using data from 2005 to 2016. The main impact model is therefore based on a pre-intervention period from January 2005 to 31 December 2007 and a post-intervention period from 1 January 2008 until 31 December 2016. We report, however, additional sensitivity analyses using data from 2000 to 2016.

Statistical analyses. We performed a segmented regression analysis using a generalized linear model with log-link for all analyses[29,36]. As the analysed monthly data were not independent from one another, we adjusted for seasonality through the inclusion of monthly dummy variables and/or for autocorrelation through the inclusion of auto-regressive structures in the model. We performed goodness-of-fit tests to decide whether to use a Poisson or a more flexible Negative Binomial model. We scrutinized auto-correlation function (ACF) and partial auto-correlation function (PACF) plots visually and compared Akaike's information criteria (AIC) to see whether the model performed better after adjustment for seasonality and/or remaining autocorrelation.

Sensitivity analyses
Varying time lags We tested different intervention time points with three and six month time lags to assess if our assumption that smoking can affect pregnancy outcomes at any stage during pregnancy was correct
Excluding transition period We excluded the data from 1 January 2008 (when the first smoking ban was implemented) until 1 August 2010 (when the smoking ban was reinstated) to compare the time period prior to the first ban to the period after implementation of the tightened ban
Including a longer pre-intervention period (2000–2016) (post-hoc) We analysed the originally planned impact model with a study period from 2000 to 2016
Excluding induced births We analysed spontaneous preterm births only, as smoking is associated with the spontaneous preterm onset of labour due to the inflammatory responses it triggers and higher risk of intrauterine uterine infections (Goldenberg, Culhane, Iams, & Romero, 2008). We could thereby also account for the potential effects of the introduction of gestational diabetes screening in 2011[56], which is associated with induced preterm births
Excluding preterm infants at the border of viability We excluded infants born between 24 and 27 completed gestational weeks and only assessed infants born between 28 and 36 gestational weeks to rule out any effect of changes in data documentation practices after implementation of the guidelines on premature infants on the border of viability[57]
Including only mothers of German nationality We excluded all mothers born outside of Germany to rule out any effect of the recent increase in refugees starting in 2014 in Germany
Subgroup analyses
Smoking status We tested if outcomes rates differed between actively smoking and non-smoking mothers (see Mackay et al.[17]). Smoking, as reported by the mother, is recorded when registering in hospital. We further assessed post-hoc whether the smoking rates differed after implementation of the smoke-free legislation
Socio-economic status (SES) We wanted to assess the impact of tobacco control policies on marginalised populations and assessed whether the legislation had a different effect on different socio-economic groups. Individual-level SES data were not available, and we thus used the area-level Bavarian Index of Multiple Deprivation (BIMD) as a proxy for individual SES[58,59]. We therefore assigned each mother a BIMD quintile based on the postal code of her residential address. We then performed subgroup analyses according to each BIMD quintile

Table 3. Sensitivity and subgroup analysis descriptions.

The final models were generalized linear models (Poisson or Negative Binomial depending on the outcome) including seasonal dummy variables and/or a random effect term comprising the appropriate autoregressive terms. As we were dealing with count data, we were using the population as an offset variable in order to transform back to rates. The main statistical formula is depicted below:

$$Number\ of\ PB_t \sim Poisson(\boldsymbol{\lambda}_t) \text{ or } Number\ of\ PB_t \sim Negative\ Binomial(\boldsymbol{\lambda}_t)$$

$$\boldsymbol{\lambda}_t = \left(\log\left(Total\ Number\ of\ Birth_t\right)\right) + \beta_0 + \beta_1\, time_t + \beta_2\, level_j + \beta_3\, slope_{jt} + \sum_{k=1}^{12} I_{\{month(t)=k\}}$$

where $\boldsymbol{\lambda}_t$ is the log of monthly outcome rates measured at each month of observation t, and $time_t$ is a continuous variable modelling each month since January 2005 (1,2,3…–145), $level_j$ a binary predictor for the legislation, which is modelled as 0 in the pre-legislation time period (January 2005–December 2007) and 1 in the post-legislation period (January 2008–December 2016). $slope_{jt}$ is an interaction term of the legislation with time. In this model β_0 represents the baseline outcome rate, β_1 the change in outcome rate per one unit increase in time (month) (i.e. the underlying pre-legislation trend), β_2 the level change in outcome following the legislation and β_3 the slope change in outcome following the legislation.

We performed data management with SAS software version 9.4[54], and the analyses using R[55]. The complete R code for the main statistical analysis is available in the Supplementary Information online.

Subgroup and sensitivity analyses. We specified a series of sensitivity and subgroup analyses a priori for the primary outcomes preterm birth and SGA (see Table 3). We further performed post-hoc sensitivity analyses for the secondary outcome very preterm birth to verify the findings in the main analysis.

Ethics statement. As we use anonymous, routinely collected data, separate ethics approval was not required for this study, as confirmed by a waiver obtained from the ethics commission of the LMU Munich. The Bavarian Institute for Quality Assurance in hospital care (BAQ) approved the use of the data.

Data availability

The data that support the findings of this study are available from the Bavarian Institute for Quality Assurance in hospital care (BAQ) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of BAQ.

Received: 24 April 2020; Accepted: 1 February 2021
Published online: 18 February 2021

References

1. WHO. *Report on the Global Tobacco Epidemic* (World Health Organization, Geneva, 2019).
2. WHO. *Tobacco Free Initiative. WHO report on the global tobacco epidemic, 2019*. MPOWER (2020).
3. Frazer, K. *et al.* Legislative smoking bans for reducing harms from secondhand smoke exposure, smoking prevalence and tobacco consumption. *Cochrane Database Syst. Rev.* **2**, CD005992 (2016).
4. Faber, T. *et al.* Effect of tobacco control policies on perinatal and child health: a systematic review and meta-analysis. *Lancet Public Health* **2**(9), e420–e437 (2017).
5. Been, J. V. & Sheikh, A. Investigating the link between smoke-free legislation and stillbirths. *Expert Rev. Respir. Med.* **10**(2), 109–112 (2016).
6. Kahnert, S., et al. *Perspektiven für Deutschland: Das Rahmenübereinkommen der WHO zur Eindämmung des Tabakgebrauchs. WHO Framework Convention on Tobacco Control (FCTC)*. Heidelberg (2011).
7. Bundesministerium der Justiz und für Verbraucherschutz, *Verordnung über Arbeitsstätten (Arbeitsstättenverordnung - ArbStättV)* (2004).
8. Verbraucherschutz, B.d.J.u.f. *Gesetz zur Einführung eines Rauchverbotes in Einrichtungen des Bundes und öffentlichen Verkehrsmitteln (Bundesnichtraucherschutzgesetz - BNichtrSchG)* (2007). Last accessed January 2021. https://www.gesetze-im-internet.de/bnichtrschg/BJNR159510007.html.
9. Sargent, J. D. *et al.* Smoking restrictions and hospitalization for acute coronary events in Germany. *Clin. Res. Cardiol.* **101**(3), 227–235 (2012).
10. Statistisches Bundesamt (Destatis). *Lebendgeborene: Bundesländer, Jahre, Geschlecht* (2017). Last accessed January 2021. https://www-genesis.destatis.de/genesis//online?operation=table&code=12612-0100&bypass=true&levelindex=0&levelid=1609930120465-abreadcrumb.
11. Hill, S. *et al.* Impact of tobacco control interventions on socioeconomic inequalities in smoking: review of the evidence. *Tob. Control* **23**(e2), e89-97 (2014).
12. Anger, S., Kvasnicka, M. & Siedler, T. One last puff? Public smoking bans and smoking behavior. *J. Health Econ.* **30**(3), 591–601 (2011).
13. McKinnon, B., Auger, N. & Kaufman, J. S. The impact of smoke-free legislation on educational differences in birth outcomes. *J. Epidemiol. Community Health* **69**(10), 937–943 (2015).
14. Bharadwaj, P., Johnsen, J.V. & Loken, K.V. Smoking bans, maternal smoking and birth outcomes. In *IZA Discussion Papers*. 2012, Institute for the Study of Labor, Bonn.
15. Cox, B. *et al.* Impact of a stepwise introduction of smoke-free legislation on the rate of preterm births: analysis of routinely collected birth data. *BMJ* **346**, f441 (2013).
16. Kabir, Z. *et al.* Smoking ban and small-for-gestational age births in Ireland. *PLoS ONE* **8**(3), e57441 (2013).
17. Mackay, D. F. *et al.* Impact of Scotland's smoke-free legislation on pregnancy complications: retrospective cohort study. *PLoS Med* **9**(3), e1001175 (2012).
18. Been, J. V. *et al.* Impact of smoke-free legislation on perinatal and infant mortality: a national quasi-experimental study. *Sci. Rep.* **5**, 13020 (2015).
19. Peelen, M. J. *et al.* Tobacco control policies and perinatal health: a national quasi-experimental study. *Sci. Rep.* **6**, 23907 (2016).
20. Vicedo-Cabrera, A. M. *et al.* Benefits of smoking bans on preterm and early-term births: a natural experimental design in Switzerland. *Tob. Control* **25**(e2), e135–e141 (2016).
21. Lampert, T., Lippe, E.v.d. & Müters, S. Prevalence of smoking in the adult population of Germany (Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung, 2013)
22. Boeckmann, M., et al. German public support for tobacco control policy measures: results from the german study on tobacco use (DEBRA), a representative national survey. *Int. J. Environ. Res. Public Health* 15(4) (2018).
23. Gruning, T. *et al.* Tobacco industry attempts to influence and use the German government to undermine the WHO Framework Convention on Tobacco Control. *Tob. Control* **21**(1), 30–38 (2012).
24. Kuijpers, T. G., Kunst, A. E. & Willemsen, M. C. Who calls the shots in tobacco control policy? Policy monopolies of pro and anti-tobacco interest groups across six European countries. *BMC Public Health* **19**(1), 800 (2019).
25. Joossens, L., Feliu, A. & Fernandez, E. The Tobacco Control Scale 2019 in Europe. (Association of European Cancer Leagues, Catalan Institute of Oncology, Brussels, 2020)
26. Hone, T. *et al.* Smoke-free legislation and neonatal and infant mortality in Brazil: longitudinal quasi-experimental study. *Tob. Control* **29**(3), 312–319 (2020).
27. Simón, L., et al. Smoke-free legislation in Spain and prematurity. Pediatrics 139(6) (2017).
28. Barnighausen, T. *et al.* Quasi-experimental study designs series - Paper 7: assessing the assumptions. *J. Clin. Epidemiol.* **89**, 53–66 (2017).
29. Lopez Bernal, J., Cummins, S. & Gasparrini, A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int. J. Epidemiol.* **46**(1), 348–355 (2016).
30. Huesch, M. D., Ostbye, T. & Ong, M. K. Measuring the effect of policy interventions at the population level: some methodological concerns. *Health Econ.* **21**(10), 1234–1249 (2012).
31. Rohwer, A. *et al.* Logic models help make sense of complexity in systematic reviews and health technology assessments. *J. Clin. Epidemiol.* **83**, 37–47 (2016).
32. Lopez Bernal, J., Soumerai, S. & Gasparrini, A. A methodological framework for model selection in interrupted time series studies. *J. Clin. Epidemiol.* **103**, 82–91 (2018).
33. Prabhu, N. *et al.* First trimester maternal tobacco smoking habits and fetal growth. *Thorax* **65**(3), 235–240 (2010).
34. Shakeri, B., Mongelli, M. & Condous, G. First trimester growth: is it influenced by cigarette smoking, and other substances?. *Australas J. Ultrasound Med.* **16**(2), 42–43 (2013).
35. Lopez Bernal, J., Cummins, S. & Gasparrini, A. The use of controls in interrupted time series studies of public health interventions. *Int. J. Epidemiol.* **47**(6), 2082–2093 (2018).
36. Barrio, G. *et al.* The limits of single-group interrupted time series analysis in assessing the impact of smoke-free laws on short-term mortality. *Int. J. Drug Policy* **73**, 112–120 (2019).
37. Gasparrini, A., Gorini, G. & Barchielli, A. On the relationship between smoking bans and incidence of acute myocardial infarction. *Eur. J. Epidemiol.* **24**(10), 597–602 (2009).
38. Kontopantelis, E. *et al.* Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ* **350**, h2750 (2015).
39. Ijaz, S. *et al.* Inclusion of nonrandomized studies in Cochrane systematic reviews was found to be in need of improvement. *J. Clin. Epidemiol.* **67**(6), 645–653 (2014).
40. Polus, S. *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.* **91**, 56–69 (2017).

41. Shadish, W. R., Cook, T. D. & Cook, D. T. *Experimental and quasiexperimental designs for generalized causal inference* (Houghton Mifflin, Boston, 2002).
42. Gesellschaft für Qualität in der außerklinischen Geburtshilfe e.V., *HG und HgE Geburten im Jahr 2015 nach Bundesländern*. Storkow (2015).
43. BAQ. *Neonatologie* (2017). Last accessed January 2021. http://www.baq-bayern.de/leistungsbereiche/gynaekologiegeburtshilfeneonatologie/neo-neonatologie/neonatologie.
44. Voigt, M., Schneider, K. T. & Jahrig, K. Analysis of a 1992 birth sample in Germany 1: new percentile values of the body weight of newborn infants. *Geburtshilfe Frauenheilkd* **56**, 550–558 (1996).
45. Rehfuess, E. A. *et al.* Towards a taxonomy of logic models in systematic reviews and health technology assessments: a priori, staged, and iterative approaches. *Res Synth Methods* **9**(1), 13–24 (2018).
46. Rohwer, A. *et al.* Series: clinical epidemiology in South Africa. Paper 3: logic models help make sense of complexity in systematic reviews and health technology assessments. *J. Clin. Epidemiol.* **83**, 37–47 (2017).
47. Bundesverfassungsgericht. *L e i t s ä t z e zum Urteil des Ersten Senats vom 30. Juli 2008* (2008) Last accessed January 2021. http://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2008/07/rs20080730_1bvr326207.html.
48. Bayerische Staatskanzlei. *Gesetz zum Schutz der Gesundheit (Gesundheitsschutzgesetz – GSG) Vom 23. Juli 2010 (GVBl S. 314) BayRS 2126-3-G* (2010). Last accessed January 2021. https://www.gesetze-bayern.de/Content/Document/BayGSG.
49. Bayerische Staatsregierung, *VOLKSENTSCHEID zum Nichtraucherschutz in Bayern Bekanntmachung der Bayerischen Staatsregierung vom 20. April 2010, Az.: II 2-G 58/09* 2010.
50. Akhtar, P. C. *et al.* Smoking restrictions in the home and secondhand smoke exposure among primary schoolchildren before and after introduction of the Scottish smoke-free legislation. *Tob. Control* **18**(5), 409–415 (2009).
51. Mons, U. *et al.* Impact of national smoke-free legislation on home smoking bans: findings from the International Tobacco Control Policy Evaluation Project Europe Surveys. *Tob. Control* **22**(e1), e2-9 (2013).
52. Lee, J. T., Glantz, S. A. & Millett, C. Effect of smoke-free legislation on adult smoking behaviour in England in the 18 months following implementation. *PLoS ONE* **6**(6), e20933 (2011).
53. Statistisches Bundesamt (Destatis), *Fachserie. 14, Finanzen und Steuern. 9, Verbrauchsteuern. 1, Tabaksteuer. 1, Absatz von Tabakwaren*. 2017, Statistisches Bundesamt (Destatis), Wiesbaden.
54. SAS Institute Inc. Cary, NC, USA.
55. R Core Team, *R: A Language and Environment for Statistical Computing*. 2018, R Foundation for Statistical Computing: Vienna, Austria. URL https://www.R-project.org/.
56. Gemeinsamer Bundesausschuss, *Bekanntmachung eines Beschlusses des Gemeinsamen Bundesausschusses über eine Änderung der Richtlinien über die ärztliche Betreuung während der Schwangerschaft und nach der Entbindung (Mutterschafts-Richtlinien): Einführung eines Screenings auf Gestationsdiabetes* (Bundesministerium für Gesundheit, Berlin, 2011)
57. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, *S2k Leitlinie: Frühgeborene an der Grenze der Lebensfähigkeit*. (2014).
58. Fairburn, J., Maier, W. & Braubach, M. Incorporating environmental justice into second generation indices of multiple deprivation: lessons from the UK and progress internationally. *Int. J. Environ. Res. Public Health* **13**(8), 750 (2016).
59. Maier, W., Fairburn, J. & Mielck, A. Regional deprivation and mortality in Bavaria. Development of a community-based index of multiple deprivation. *Gesundheitswesen* **74**(7), 416–425 (2012).

Author contributions

S.P. conceived the study, developed the methods, conducted and interpreted the statistical analysis, and wrote the manuscript. J.B. co-conceived the study, developed the methods, supported and interpreted the statistical analysis, and commented on the manuscript. S.H. and T.M. advised on and supported the analyses and their interpretation, and commented on the manuscript. U.M. supported the statistical analyses, and commented on the manuscript. J.V.B. supported developing the methods and commented on the manuscript. N.L. helped retrieve the data and supported analyses, and commented on the manuscript. D.K. and W.M. developed the methods for socio-economic subgroup analyses and commented on the manuscript. E.A.R. co-conceived the study and developed the methods, supervised and interpreted the analyses, and commented on the manuscript. All authors contributed to the editing and final approval of the manuscript.

Funding

SP is additionally funded by a doctoral scholarship through the Heinrich-Boell-Foundation (www.boell.de). JVB is funded by a personal fellowship from the Netherlands Lung Foundation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-83774-0.

Correspondence and requests for materials should be addressed to S.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Interrupted time series study found mixed effects of the impact of the Bavarian smoke-free legislation on pregnancy outcomes

Stephanie Polus [*1,2], Jacob Burns[1,2], Sabine Hoffmann [1,2], Tim Mathes[3], Ulrich Mansmann[1,2], Jasper V Been[4], Nicholas Lack[5], Daniela Koller[1,2], Werner Maier[6], Eva A Rehfuess[1,2]

*corresponding author
[1] Institute for Medical Information Processing, Biometry, and Epidemiology – IBE, LMU Munich, Marchioninistr. 17, 81377 Munich, Germany
[2] Pettenkofer School of Public Health, Munich, Germany
[3] Institute for Research in Operative Medicine, Faculty of Health, School of Medicine, Witten/Herdecke University, Cologne, Germany
[4] Division of Neonatology, Department of Paediatrics, Department of Obstetrics and Gynaecology, Department of Public Health, Erasmus MC – Sophia Children's Hospital, Rotterdam, Netherlands
[5] German Bavarian Quality Assurance Institute for Medical Care, Munich, Germany
[6] Institute of Health Economics and Health Care Management, Helmholtz Zentrum München – German Research Center for Environmental Health (GmbH), Neuherberg, Germany

Table of Contents

Figure S1 Number of births 2005-2016

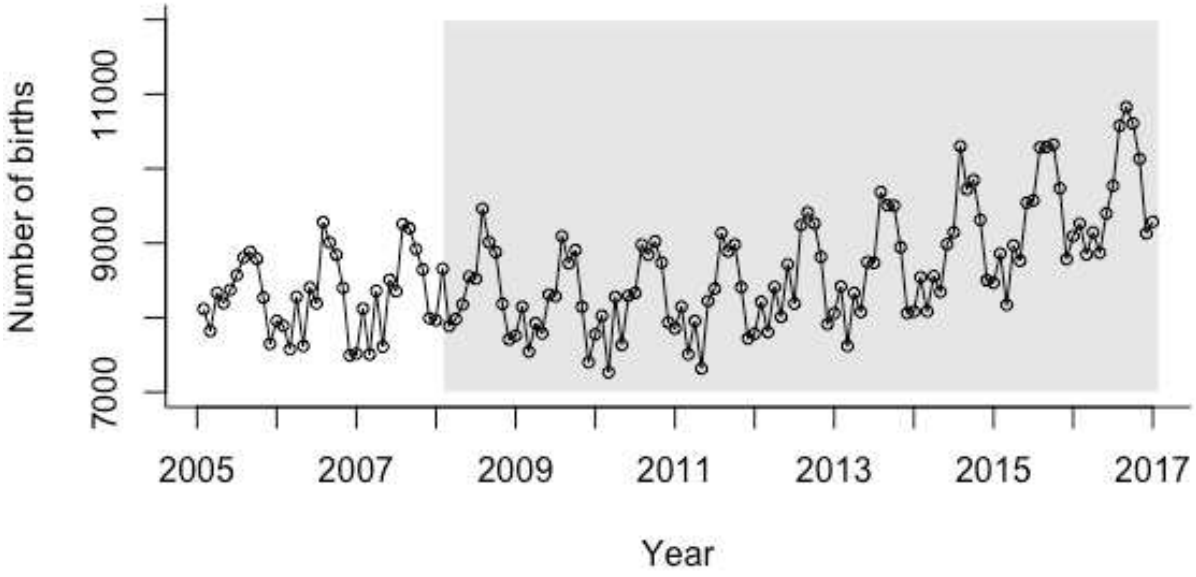


Figure S2 Preterm birth rate 2000-2016 including regression line

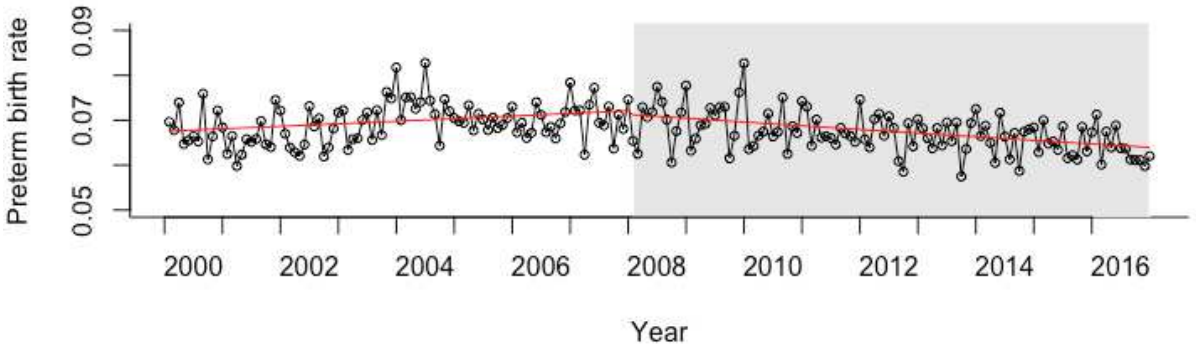


Figure S3 SGA rate 2000-2016 including regression line

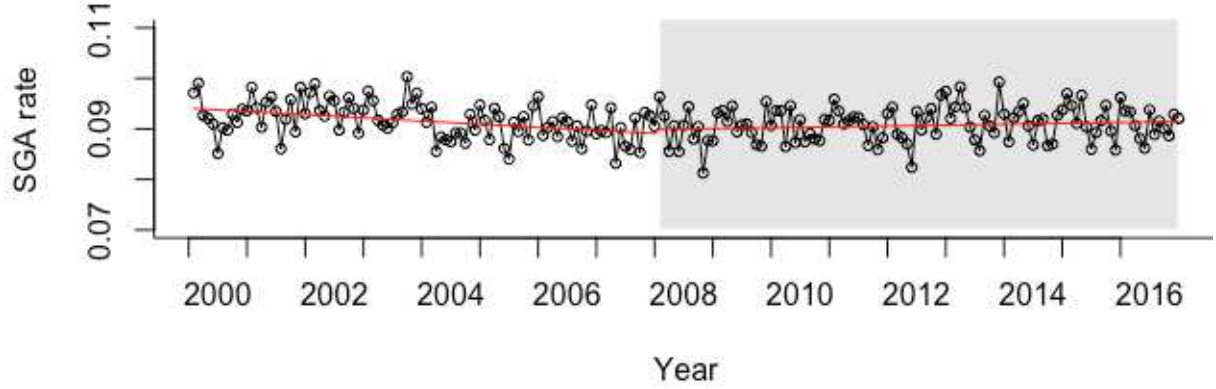


Figure S4 Preterm birth rate by highest (BIMD 1) and lowest (BIMD 5) SES, 2005-2016

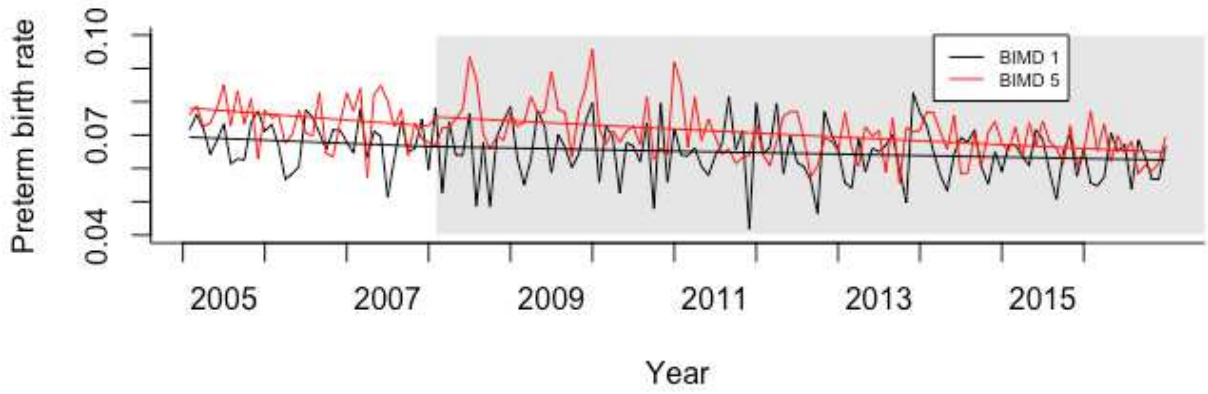


Figure S5 SGA rate by highest (BIMD 1) and lowest (BIMD 5) SES, 2005-2016

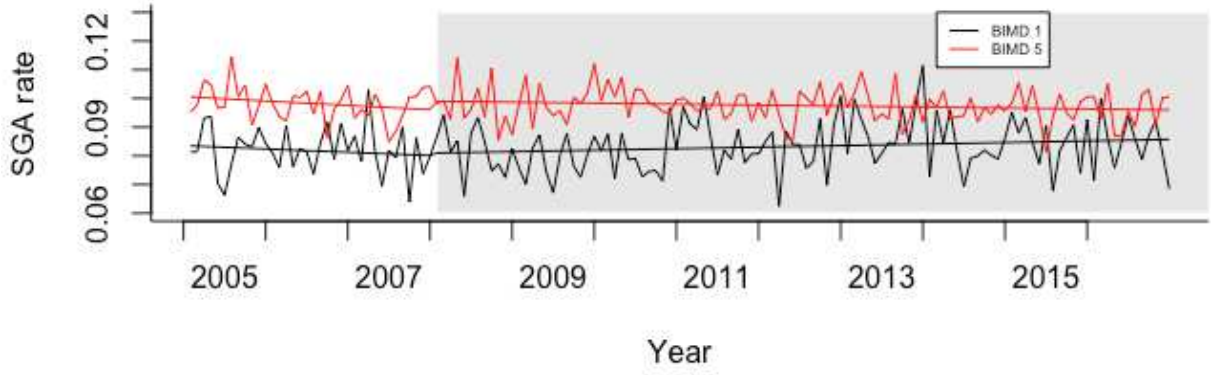


Figure S6 Regression lines preterm birth rates according to SES (BIMD quintile)

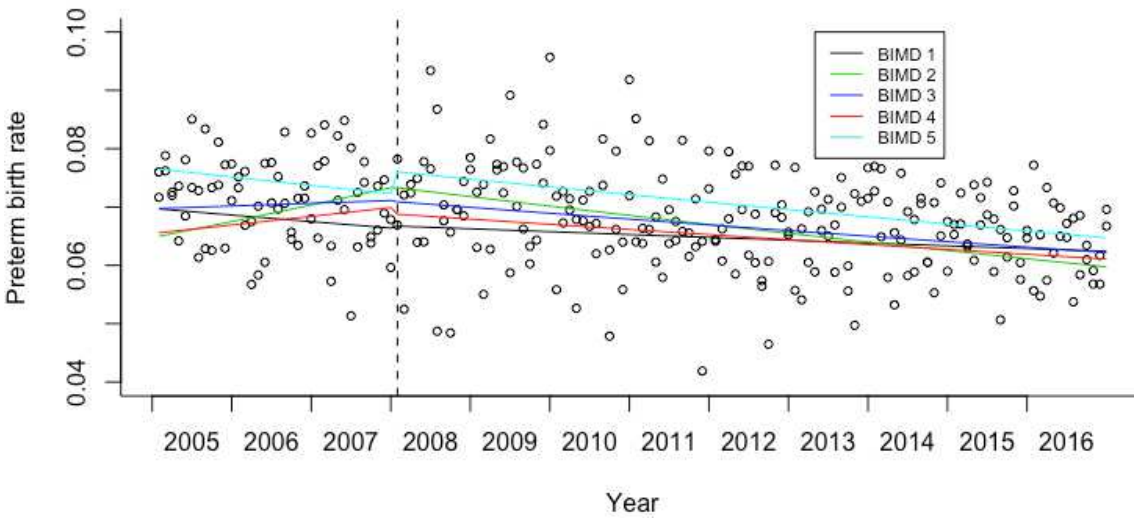


Figure S7 Regression lines SGA rates according to SES (BIMD quintile)

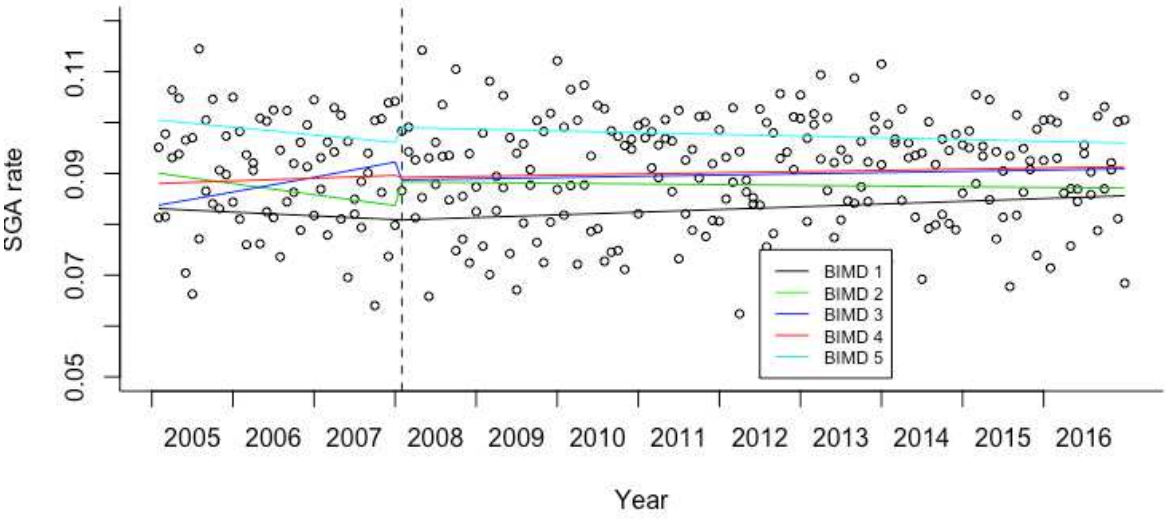


Figure S8 Number of smoking mothers 2005-2016

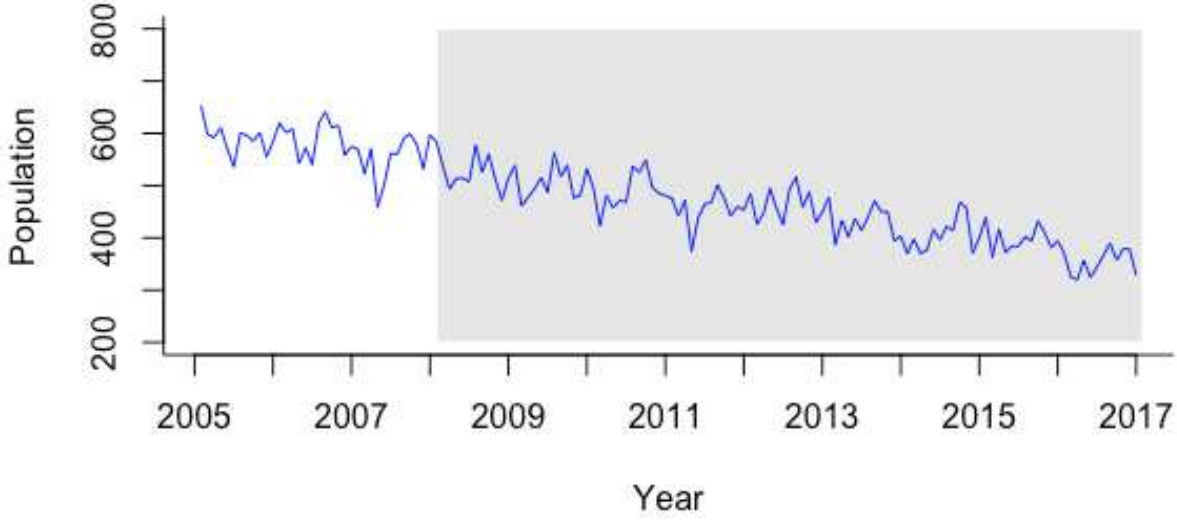


Figure S9 Preterm birth rates among smoking and non-smoking mothers, 2005-2016

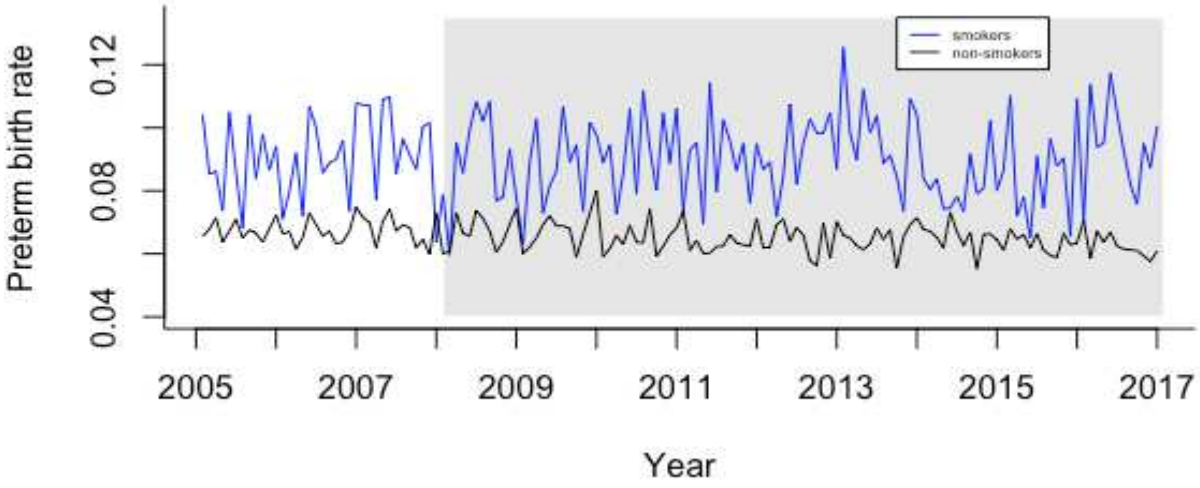


Figure S10 SGA rates smoking and non-smoking mothers, 2005-2016

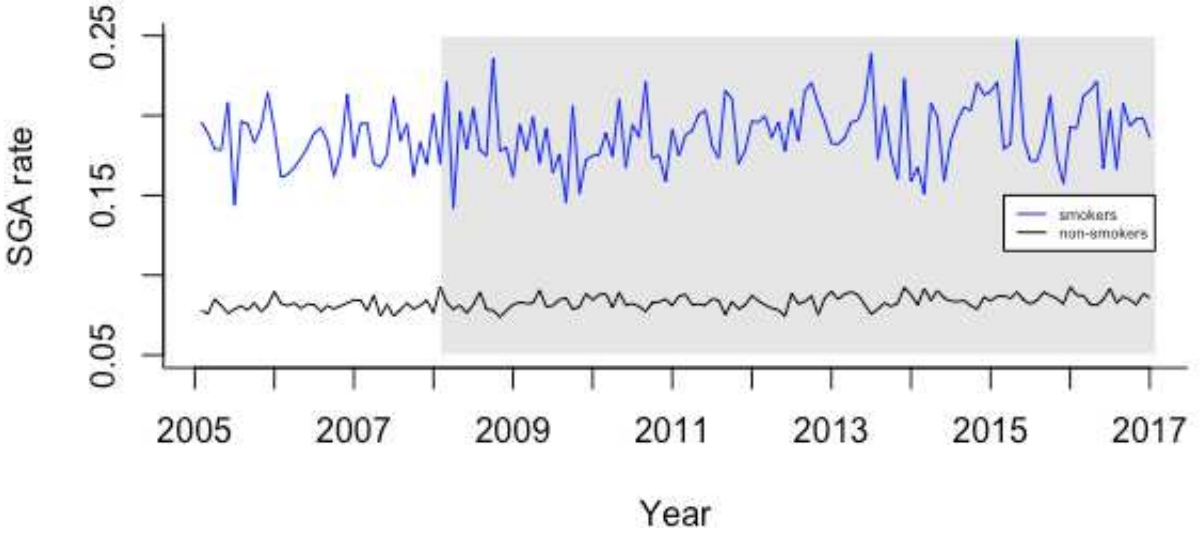


Table S1 Estimates of changes in level and slope in subgroup and sensitivity analyses of primary outcomes.

Outcome	Exponential level coefficient (95%CI)	Exponential slope coefficient (95%CI)	Model type (and R package used)
Sensitivity analyses			
3-month lag after legislation implementation			
Preterm birth	1.0177 (0.9849,1.0516)	0.9991 (0.9978,1.0004)	Seasonally adjusted Negative Binomial model (glm)
SGA	1.0080 (0.9891,1.0273)	1.0006 (0.9998, 1.0014)	Poisson model (glarma) with seasonal dummies and autocorrelation terms
6-month lag after legislation implementation			
Preterm birth	1.002 (0.9705, 1.0346)	0.9988 (0.9976,1.0000), p=0.03947	Seasonally adjusted Negative Binomial model (glm)
SGA	1.0097 (0.9930, 1.0267)	1.0005 (1.0000, 1.0010)	Poisson model (glarma) with seasonal dummies and autocorrelation terms
Excluding data of a 2-year transition period after legislation implementation			
Preterm birth	0.9758 (0.9384, 1.0147)	0.9990 (0.9972,1.0008)	Negative binomial model (glarma) with autocorrelation terms
SGA	1.0197 (0.9977,1.0423)	1.0006 (0.9999, 1.0013)	Poisson model (glarma) with autocorrelation terms
Including a longer pre-intervention period (2000-2016) (post-hoc)			
Preterm birth	0.9879 (0.9546, 1.0223)	0.9984 (0.9978, 0.9989), p=0.0000	Negative Binomial model (glarma) with seasonal dummies and autocorrelation terms
SGA	1.0087 (0.9897,1.0281)	1.0008 (1.0006, 1.0010), p=0.0000	Poisson model (glarma) with seasonal dummies & autocorrelation terms
Excluding induced births			
Preterm birth	1.0141 (0.9697, 1.0605)	0.9995 (0.9976, 1.0014)	Negative binomial model (glm) with seasonal dummies
SGA	1.0092 (0.9882,1.0306)	1.0001 (0.9992,1.0010)	Poisson model (glarma) with autocorrelation terms
Excluding preterm infants at the border of viability			
Moderate preterm birth	0.9902 (0.9498, 1.0323)	0.9980 (0.9960,1.0000), p=0.0384	Negative binomial model (glarma) with autocorrelation terms
SGA	1.0104 (0.9916, 1.0296)	1.0007 (1.0000,1.0014)	Poisson model (glarma) with autocorrelation terms
Including only mothers of German nationality			
Preterm birth	1.0053 (0.9642, 1.0482)	0.9985 (0.9965, 1.0004)	Negative binomial model (glarma) with autocorrelation terms
SGA	1.0179 (0.9879, 1.0489)	1.0002 (0.9989, 1.0014)	Unadjusted Poisson model (glm)
Subgroup analyses			

Smoking status			
Preterm birth	Smokers: 0.9627 (0.8664,1.0698)	0.9985 (0.9940,1.0028)	Unadjusted Poisson model (glm)
	Non-smokers: 0.9954 (0.9464,1.0469)	0.9995 (0.9971,1.0018)	Negative binomial model (glarma) with autocorrelation terms
SGA	Smokers: 0.9983 (0.9267, 1.0753)	1.0011 (0.9980,1.0042)	Unadjusted Poisson model (glm)
	Non-smokers: 1.0052 (0.9798, 1.0313)	1.0003 (0.9992, 1.0014)	Poisson model (glarma) with autocorrelation terms
Socio-economic status – Preterm birth			
BIMD quintile 1 (highest SES)	1.0001 (0.9619, 1.0397)	1.0006 (0.9990, 1.0023)	Poisson model (glarma) with autocorrelation terms
BIMD quintile 2	1.0090 (0.9314, 1.0932)	0.9951 (0.9916, 0.9986), p=0.0061	Poisson model (glarma) with autocorrelation terms
BIMD quintile 3	0.9982 (0.9186, 1.0847)	0.9982 (0.9947, 1.0018)	Unadjusted Poisson model
BIMD quintile 4	0.9826 (0.9256, 1.0430)	0.9973 (0.9946, 1.0000), p=0.04	Poisson model (glarma) with autocorrelation term
BIMD quintile 5 (lowest SES)	1.0393 (0.9703, 1.1132)	1.0006 (0.9971, 1.0041)	Negative binomial model (glarma) with autocorrelation terms
Socio-economic status – SGA			
BIMD quintile 1 (highest SES)	1.0102 (0.9295, 1.0979)	1.0017 (0.9980, 1.0054)	Poisson model (glarma) with autocorrelation terms
BIMD quintile 2	1.0556 (0.9731, 1.1450)	1.0019 (0.9985, 1.0055)	Unadjusted Poisson model
BIMD quintile 3	0.9403 (0.8743, 1.0114)	*0.9966 (0.9935, 0.9988), p=0.03*	Poisson model (glarma) with autocorrelation terms
BIMD quintile 4	1.0010 (0.9628, 1.0407)	1.0001 (0.9985, 1.0017)	Poisson model (glarma) with autocorrelation terms
BIMD quintile 5 (lowest SES)	1.0310 (0.9794, 1.0853)	1.0009 (0.9988, 1.0032)	Unadjusted Poisson model

Table S2 Post-hoc sensitivity analyses very preterm birth

Sensitivity analyses	**Exponential level coefficient (95%CI)***	**Exponential slope coefficient (95%CI)**	**Model type (and R package used)**
3 month lag	0.9077 (0.8291, 0.9938), p= 0.0364	0.9962 (0.9922, 1.0002)	Negative binomial model (glarma) with autocorrelation terms
6 month lag	0.9080 (0.8287,0.9949), p=0.0384	0.9962 (0.9922, 1.0002)	Negative binomial model (glarma) with autocorrelation terms
2-year transition period	0.9745 (0.9357, 1.0149)	0.9985 (0.9961,1.0001)	Negative binomial model (glarma) with autocorrelation terms
Longer time period 2000-2016	0.9967 (0.9051, 1.0329)	0.9993 (0.9982, 1.0004)	Negative binomial model (glarma) with autocorrelation terms
Excluding induced births	0.8667 (0.7297, 1.0308)	0.9954 (0.9880,1.0028)	Negative binomial model (GLM) with seasonal adjustment

Including only mothers of German nationality	0.9268 (0.8250,1.0412)	0.9957 (0.9906,1.0007)	Negative binomial model (glarma) with autocorrelation terms
Smokers	0.8161 (0.6318, 1.0540)	0.9959 (0.9851,1.0007)	Poisson model (glarma) with autocorrelation terms
Non-smokers	0.9093 (0.8134,1.0164)	0.9956 (0.9908,1.0004)	Unadjusted negative binomial model (glm)
BIMD 1	0.9365 (0.7316,1.1987)	0.9963 (0.9856,1.0071)	Unadjusted Poisson model
BIMD 2	1.0361 (0.7958,1.349)	0.9938 (0.9822,1.0055)	Unadjusted Poisson model
BIMD 3	1.0595 (0.8311,1.3508)	1.0037 (0.9934,1.0141)	Unadjusted Poisson model
BIMD 4	0.8845 (0.7690,1.0174)	0.9935 (0.8637,1.1473)	Negative binomial model (glarma) with autocorrelation terms
BIMD 5	0.8992 (0.7869,1.0275)	0.9979 (0.9909,1.0049)	Negative binomial model (glarma) with autocorrelation terms

R Script of main analysis

```
#load necessary packages
library(foreign)
library(dplyr)
library(MASS)
library(glarma)
source("/Users/likTestsNEW.R") #in the glarma package this was necessary to be able to use the package properly
source("/Users/summary.glarmaNEW.R") #sources kindly provided by William Duinsmuir

#retrieve data
smoke <- read.xport("/smoke.xpt")


#rename variables
smoke$year <- smoke$YEAR
smoke$YEAR <- NULL
smoke$month <- smoke$MONTH
smoke$MONTH <- NULL
smoke$pb <- smoke$PB
smoke$PB <- NULL
smoke$vpb <- smoke$VPB
smoke$VPB <- NULL
smoke$sga <- smoke$SGA
smoke$SGA <- NULL
smoke$lbw <- smoke$LBW
smoke$LBW <- NULL
smoke$still <- smoke$STILL
smoke$STILL <- NULL
```

```
smoke$n <- smoke$N
smoke$N <- NULL
smoke$t <- smoke$T
smoke$T <- NULL

#We delete the data smaller 2005 and bigger than 2016
smoke <- smoke %>% filter(year >=2005 & year != 2017)

#We set time variabe t at new dataset start
smoke$t <- 1:nrow(smoke)

###############
#create time variables
################
cp <- 37 #set intervention time point on 1 January 2008
smoke$level <- ifelse(smoke$t<cp, 0, 1)

smoke$slope <- c(rep(0, cp-1), 1:(145-cp))


#plot  whole time series
# start the plot, excluding the points and the x-axis
plot(smoke$pb,type="n",ylim=c(0.05,0.09),xlab="Year", ylab="Preterm birth rate",
    bty="l",xaxt="n")
# Add line indicating the policy changes
abline(v=37,lty=2)
# plot the observed rate for intervention period
points(smoke$pb,cex=0.7, col="red")  #oder mit points
lines(smoke$pb,cex=0.7, col="red")
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("Preterm rate, 2005-2016")



#plot numbers
plot(smoke$pbn,type="n",ylim=c(450,700),xlab="Year", ylab="Preterm birth rate",
    bty="l",xaxt="n")
# Add line indicating the policy changes
abline(v=37,lty=2)
# plot the observed rate for intervention period
points(smoke$pbn,cex=0.7)  #oder mit points
lines(smoke$pbn,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
```

```
# add a title
title("Preterm rate, 2005-2016")

#create offset for model
smoke$pbn <- smoke$pb * smoke$n
smoke$logn <- log(smoke$n)
head(smoke)

#first try of glm model
model1 <- glm(pbn ~ offset(logn) + t + level + slope, data=smoke, family = poisson())

summary(model1)

# goodness of fit test: if test is significant, model fit is not good
1 - pchisq(summary(model1)$deviance,
        summary(model1)$df.residual)


#negative binomial model
model2 <- glm.nb(pbn ~ offset(logn) + t + level + slope, data=smoke, link=log)
summary(model2)

1 - pchisq(summary(model2)$deviance,
        summary(model2)$df.residual) #better fit

par(mfrow=c(1, 2))
acf(residuals(model2))
acf(residuals(model2),type='partial') #looks like months/seasonality play more of a role

#plot residuals
plot(smoke$t, residuals(model2), type='o', pch=16)

#adjust with monthly dummies
smoke$jan <- ifelse(smoke$month==1, 1,0)
smoke$feb <- ifelse(smoke$month==2, 1,0)
smoke$mar <- ifelse(smoke$month==3, 1,0)
smoke$apr <- ifelse(smoke$month==4, 1,0)
smoke$may <- ifelse(smoke$month==5, 1,0)
smoke$jun <- ifelse(smoke$month==6, 1,0)
smoke$jul <- ifelse(smoke$month==7, 1,0)
smoke$aug <- ifelse(smoke$month==8, 1,0)
smoke$sep <- ifelse(smoke$month==9, 1,0)
smoke$oct <- ifelse(smoke$month==10, 1,0)
smoke$nov <- ifelse(smoke$month==11, 1,0)
smoke$dec <- ifelse(smoke$month==12, 1,0)


model3 <- glm.nb(pbn ~ offset(logn) + t + level + slope + jan + feb + mar + apr + may +
```

```
            jun + jul + aug  + oct + nov + dec, data=smoke, link=log, maxit=300)

summary(model3)

#plot residuals
plot(smoke$t, residuals(model3), type='o', pch=16)
#goodness of fit test
1 - pchisq(summary(model3)$deviance,
        summary(model3)$df.residual) #looks good


#look for remaining autocorrelation
par(mfrow=c(1, 2))
acf(residuals(model3))
acf(residuals(model3),type='partial') #still autocorrelation present

#glarma model
library(glarma)
source("/Users/stephie/LRZ Sync+Share/ITS PhD/likTestsNEW.R")
source("/Users/stephie/LRZ Sync+Share/ITS PhD/summary.glarmaNEW.R")

predictors <- cbind(intercept <-(rep(1, dim(smoke)[1])),
              slope <-smoke$slope, level <- smoke$level, t <-smoke$t)

colnames(predictors) <- c("intercept", "slope", "level", "t")

model5 <- glarma(smoke$pbn, predictors, offset = smoke$logn, type = "NegBin",
            phiLags = c(9,12,21)) #final model

par(mfrow=c(1, 2))
acf(residuals(model5))
acf(residuals(model5),type='partial')#looks good

#plot regression line
dev.off()
# start the plot, excluding the points and the x-axis
plot(smoke$pb,type="n",ylim=c(0.05,0.09),xlab="Year", ylab="Preterm birth percentage",
   bty="l",xaxt="n")
# Add line indicating the policy changes
rect(37,0.05,145,0.09,col=grey(0.9),border=F)
# plot the observed rate for intervention period
points(smoke$pb,cex=0.7)
lines(smoke$pb,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("a. Preterm birth")
```

```
#plot regression line with glarma model
coef(model5)
predict <- exp(smoke$logn + coef(model5)$beta[1]
           + smoke$t*coef(model5)$beta[4]
           + smoke$level*coef(model5)$beta[3]
           + smoke$slope*coef(model5)$beta[2])


lines(predict/smoke$n,col=2)


#########################################################
#SGA
#####
# start the plot, excluding the points and the x-axis
plot(smoke$sga,type="n",ylim=c(0.08,0.11),xlab="Year", ylab="SGA rate",
    bty="l",xaxt="n")
# Add line indicating the policy changes
abline(v=37,lty=2)
# plot the observed rate for intervention period
points(smoke$sga,cex=0.7, col="red")  #oder mit points
lines(smoke$sga,cex=0.7, col="red")
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("SGA rate, 2005-2016")


#create offset for model
smoke$sgan <- smoke$sga * smoke$n
head(smoke)
model1 <- glm(sgan ~ offset(logn) + t + level + slope, data=smoke, family = poisson())

summary(model1)

# If test is significant, model fit is not good
1 - pchisq(summary(model1)$deviance,
        summary(model1)$df.residual) #good fit

par(mfrow=c(1, 2))
acf(residuals(model1))
acf(residuals(model1),type='partial') #again seasonality?

#plot residuals
plot(smoke$t, residuals(model1), type='o', pch=16)
```

```
#model with monthly dummies
model2 <- glm(sgan ~ offset(logn) + t + level + slope + jan + feb + mar + apr + may
          + jun + jul + aug  + oct + nov + dec,
          data=smoke, family=poisson)
summary(model2)

par(mfrow=c(1, 2))
acf(residuals(model4))
acf(residuals(model4),type='partial')


#glarma
predictors <- cbind(intercept <-(rep(1, dim(smoke)[1])),
                slope <-smoke$slope, level <- smoke$level, t <-smoke$t,
                jan <- smoke$jan, feb <- smoke$feb,mar <- smoke$mar,
                apr <- smoke$apr, may <- smoke$may, jun <- smoke$jun,
                jul <- smoke$jul, aug <- smoke$aug, oct <- smoke$oct,
                nov <- smoke$nov, dec <- smoke$dec)

colnames(predictors) <- c("intercept", "slope", "level", "t", "jan", "feb", "mar",
                    "apr", "may", "jun", "jul", "aug", "oct", "nov", "dec")


model4 <- glarma(smoke$sgan, predictors, offset = smoke$logn, type = "Poi") #seasonal
model does not look like a good fit

summary(model4)


par(mfrow=c(1, 2))
acf(residuals(model4))
acf(residuals(model4),type='partial')



#try glarma with autocorrelation
model4_10 <- glarma(smoke$sgan, predictors, offset = smoke$logn, type = "Poi", phiLags =
c(4,10)) #final model
summary(model4_10)

acf(residuals(model4_10))
acf(residuals(model4_10),type='partial') #looks ok with monthly dummies and AR structure

#plot regression line
plot(smoke$sga,type="n",ylim=c(0.08,0.11),xlab="Year", ylab="SGA percentage",
    bty="l",xaxt="n")
# Add line indicating the policy changes
rect(37,0.08,145,0.11,col=grey(0.9),border=F)
```

```
#abline(v=37,lty=2)
# plot the observed rate for intervention period
points(smoke$sga,cex=0.7)
lines(smoke$sga,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("b. SGA")

#print regression line
coef(model4_10)

predict <- exp(smoke$logn + coef(model4_10)$beta[1]
          + smoke$t*coef(model4_10)$beta[4]
          + smoke$level*coef(model4_10)$beta[3]
          + smoke$slope*coef(model4_10)$beta[2])


lines(predict/smoke$n,col=2)



########################################
#lbw
plot(smoke$lbw,type="n",ylim=c(0.04,0.07),xlab="Year", ylab="LBW percentage",
   bty="l",xaxt="n")
# shade the post intervention period grey
rect(37,0.04,145,0.07,col=grey(0.9),border=F)
# plot the observed rate for intervention period
points(smoke$lbw,cex=0.7)
lines(smoke$lbw,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("c. LBW")


head(smoke)
#create offset for model
smoke$lbwn <- smoke$lbw * smoke$n
head(smoke)


model2 <- glm.nb(lbwn ~ offset(logn) + t + level + slope, data=smoke, link=log)
summary(model2)
```

```
# goodness of fit test
1 - pchisq(summary(model2)$deviance,
        summary(model2)$df.residual) #neg bin model is the better fit

par(mfrow=c(1, 2))
acf(residuals(model2))
acf(residuals(model2),type='partial')

#try with seasonal adjustment
model3 <- glm.nb(lbwn ~ offset(logn) + t + level + slope + jan + feb + mar + apr + may
            + jun + jul + aug  + oct + nov + dec, data=smoke, link=log, maxit = 300)

summary(model3)


acf(residuals(model3))
acf(residuals(model3),type='partial') #looks good with seasonal adjustment

#plot regression line
#pred <- predict(model3,type="response")
plot(smoke$lbw,type="n",ylim=c(0.04,0.07),xlab="Year", ylab="LBW rate",
    bty="l",xaxt="n")
# Add line indicating the policy changes
#abline(v=37,lty=2)
# shade the post intervention period grey
rect(37,0.04,145,0.07,col=grey(0.9),border=F)
# plot the observed rate for intervention period
points(smoke$lbw,cex=0.7)
lines(smoke$lbw,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("LBW rate")



#without seasonality visible:

newdata  <- smoke
head(newdata)
newdata[(14:25)] <- 0
#plot new data without seasonality
pred3 <- predict(model3,newdata=newdata)
coef(model3)
mean_months <- mean(coef(model3)[5:15])

lines(exp(pred3 + mean_months)/smoke$n,col="red")
```

```
#####################################
#stillbirth
##################

#plot stillbirth rate
plot(smoke$still,type="n",ylim=c(0.001,0.005),xlab="Year", ylab="stilbirth percentage",
    bty="l",xaxt="n")
# Add line indicating the policy changes
#abline(v=37,lty=2)
rect(37,0.001,145,0.005,col=grey(0.9),border=F)
# plot the observed rate for intervention period
points(smoke$still,cex=0.7)
lines(smoke$still,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("d. stillbirth")


#create offset for model
smoke$stilln <- smoke$still * smoke$n
head(smoke)
model1 <- glm(stilln ~ offset(logn) + t + level + slope, data=smoke, family = poisson())

summary(model1)

# If test is significant, model fit is not good
1 - pchisq(summary(model1)$deviance,
        summary(model1)$df.residual) #fits


par(mfrow=c(1, 2))
acf(residuals(model1))
acf(residuals(model1),type='partial') #looks already good

#plot regression line

plot(smoke$still,type="n",ylim=c(0.001,0.005),xlab="Year", ylab="stillbirth rate",
    bty="l",xaxt="n")
# Add line indicating the policy changes
#abline(v=37,lty=2)
rect(37,0.001,145,0.006,col=grey(0.9),border=F)
# plot the observed rate for intervention period
points(smoke$still,cex=0.7)
```

```
lines(smoke$still,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("Stillbirth rate, 2005-2016")

#plot regression line
pred <- predict(model1,type="response")
lines(pred/smoke$n,col=2)


#############
#VPB
#############

plot(smoke$vpb,type="n",ylim=c(0.005,0.014),xlab="Year", ylab="Very PB percentage",
    bty="l",xaxt="n")
# Add line indicating the policy changes
#abline(v=37,lty=2)
rect(37,0.005,145,0.014,col=grey(0.9),border=F)
# plot the observed rate for intervention period
points(smoke$vpb,cex=0.7)  #oder mit points
lines(smoke$vpb,cex=0.7)
#specify the x-axis (i.e. time units)
axis(1,at=0:12*12,labels=F)
axis(1,at=0:12*12,tick=F,labels=2005:2017)
# add a title
title("e. very preterm birth")

smoke$vpbn <- smoke$vpb * smoke$n


library(MASS)
model2 <- glm.nb(vpbn ~ offset(logn) + t + level + slope, data=smoke, link=log)
summary(model2)

1 - pchisq(summary(model2)$deviance,
        summary(model2)$df.residual) #better fit

par(mfrow=c(1, 2))
acf(residuals(model2))
acf(residuals(model2),type='partial') #autocorrelation still present

#model with autocorrelation adjustment
predictors <- cbind(intercept <-(rep(1, dim(smoke)[1])),
                slope <-smoke$slope, level <- smoke$level, t <-smoke$t)
```

```
colnames(predictors) <- c("intercept", "slope", "level", "t")

model4 <- glarma(smoke$vpbn, predictors, offset = smoke$logn, type = "NegBin", phiLags =
c(3,5,11,12))
#final model

summary(model4)

acf(residuals(model4))
acf(residuals(model4),type='partial')

#plot regression line
coef(model4)

predict <- exp(smoke$logn + coef(model4)$beta[1]
          + smoke$t*coef(model4)$beta[4]
          + smoke$level*coef(model4)$beta[3]
          + smoke$slope*coef(model4)$beta[2])

lines(predict/smoke$n,col="red")
```

References

1. 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**: p. 56-69.
2. Polus S, Burns J, Hoffmann S, Mathes T, Mansmann U, Been JV, et al. *Interrupted time series study found mixed effects of the impact of the Bavarian smoke-free legislation on pregnancy outcomes.* Sci Rep, 2021. **11**(1): p. 4209.
3. Polus S, Burns J, Pedron S, Paudel D, and Rehfuess E. *The choice of analysis in controlled before-after studies and its impact on effect size and statistical precision*. Unpublished Manuscript.
4. Laky S. *Public health*. Fact Sheets on the European Union 2020; Available from: https://www.europarl.europa.eu/factsheets/en/sheet/49/public-health.
5. World Health Organization. *Facing the future: opportunities and challenges for 21st-century public health in implementing the Sustainable Development Goals and the Health 2020 policy framework*. 2018, World Health Organization Regional Office for Europe: Copenhagen, Denmark.
6. Dalstra JA, Kunst AE, Borrell C, Breeze E, Cambois E, Costa G, et al. *Socioeconomic differences in the prevalence of common chronic diseases: an overview of eight European countries.* Int J Epidemiol, 2005. **34**(2): p. 316-26.
7. World Health Organization. *Tobacco*. 2020 [cited 2021; Available from: https://www.who.int/news-room/fact-sheets/detail/tobacco.
8. GBD Tobacco Collaborators. *Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015.* Lancet, 2017. **389**(10082): p. 1885-1906.
9. World Health Organization. *Tobacco Free Initiative. WHO report on the global tobacco epidemic, 2019*. MPOWER 2020 [cited 2020 26 February 2020].
10. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. *Developing and evaluating complex interventions: the new Medical Research Council guidance*. BMJ, 2008. **337**, a1655 DOI: 10.1136/bmj.a1655.
11. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. *Developing and evaluating complex interventions: the new Medical Research Council guidance.* BMJ, 2008. **337**: p. a1655.
12. Rehfuess EA and Akl EA. *Current experience with applying the GRADE approach to public health interventions: an empirical study*. BMC Public Health, 2013. **13**, 9 DOI: 10.1186/1471-2458-13-9.
13. Rychetnik L, Frommer M, Hawe P, and Shiell A. *Criteria for evaluating evidence on public health interventions*. J Epidemiol Community Health, 2002. **56**, 119-27.
14. Rehfuess EA and Bartram J. *Beyond direct impact: evidence synthesis towards a better understanding of effectiveness of environmental health interventions*. Int J Hyg Environ Health, 2014. **217**, 155-9 DOI: 10.1016/j.ijheh.2013.07.011.
15. Victora CG, Habicht JP, and Bryce J. *Evidence-based public health: moving beyond randomized trials.* Am J Public Health, 2004. **94**(3): p. 400-5.
16. Lavis JN, Posada FB, Haines A, and Osei E. *Use of research to inform public policymaking.* Lancet, 2004. **364**(9445): p. 1615-21.
17. Cochrane. *About Cochrane Reviews*. 2020; Available from: https://www.cochranelibrary.com/about/about-cochrane-reviews.
18. Noyes J, Booth A, Cargo M, Flemming K, Harden A, Harris J, et al. *Chapter 21: Qualitative evidence*, in *Cochrane Handbook for Systematic Reviews of Interventions*

version 6.0 (updated July 2019), TJ Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors), Editor. 2019, Cochrane, .
19. Khan KS, Daya S, and Jadad A. *The importance of quality of primary studies in producing unbiased systematic reviews.* Arch Intern Med, 1996. **156**(6): p. 661-6.
20. Shadish W, Cook T, and Campbell D. *Experimental and quasiexperimental designs for generalized causal inference*. 2002, Boston: Houghton Mifflin.
21. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. *Using natural experiments to evaluate population health interventions: new Medical Research Council guidance.* J Epidemiol Community Health, 2012. **66**(12): p. 1182-6.
22. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. *Evaluating non-randomised intervention studies*. Health Technol Assess, 2003. **7**, iii-x, 1-173.
23. Kontopantelis E, Doran T, Springate DA, Buchan I, and Reeves D. *Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis.* BMJ, 2015. **350**: p. h2750.
24. Sanson-Fisher RW, D'Este CA, Carey ML, Noble N, and Paul CL. *Evaluation of systems-oriented public health interventions: alternative research designs*. Annu Rev Public Health, 2014. **35**, 9-27 DOI: 10.1146/annurev-publhealth-032013-182445.
25. Norris SL and Atkins D. *Challenges in using nonrandomized studies in systematic reviews of treatment interventions*. Ann Intern Med, 2005. **142**, 1112-9.
26. Reeves BC DJ, Higgins JPT, Shea B, Tugwell P, Wells GA. *Chapter 24: Including non-randomized studies on intervention effects*, in *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)*, TJ Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA Editor. 2019, Cochrane.
27. Higgins JPT, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC, et al. *Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions*. Research Synthesis Methods, 2012. 12-25 DOI: 10.1002/jrsm.1056.
28. Reeves BC, Higgins JP, Ramsay C, Shea B, Tugwell P, and Wells GA. *An introduction to methodological issues when including non-randomised studies in systematic reviews on the effects of interventions.* Res Synth Methods, 2013. **4**(1): p. 1-11.
29. Craig P, Katikireddi SV, Leyland A, and Popham F. *Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research.* Annu Rev Public Health, 2017. **38**: p. 39-56.
30. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions*. BMJ, 2016. **355**, i4919 DOI: 10.1136/bmj.i4919.
31. Medical Research Council. *Using natural experiments to evaluate population health interventions: guidance for producers and users of evidence*. 2011.
32. Glenton C, Lewin S, Mayhew A, Scheel I, and Odgaard-Jensen J. *Nonrandomized studies are not always found even when selection criteria for health systems intervention reviews include them: a methodological study.* J Clin Epidemiol, 2013. **66**(4): p. 367-70.
33. Peinemann F, Tushabe DA, and Kleijnen J. *Using multiple types of studies in systematic reviews of health care interventions--a systematic review*. PLoS One, 2013. **8**, e85035 DOI: 10.1371/journal.pone.0085035.
34. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. *Increasing value and reducing waste in research design, conduct, and analysis.* Lancet, 2014. **383**(9912): p. 166-75.
35. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, and 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**, 613-23.

36. Valentine JC and Thompson SG. *Issues relating to confounding and meta-analysis when including non-randomoized studies in systematic reviews on the effects of interventions.* Research Synthesis Methods, 2013. **4**(1): p. 26-35.
37. Lopez Bernal J, Cummins S, and Gasparrini A. *Interrupted time series regression for the evaluation of public health interventions: a tutorial.* Int J Epidemiol, 2016.
38. Ijaz S, Verbeek JH, Mischke C, and Ruotsalainen J. *Inclusion of nonrandomized studies in Cochrane systematic reviews was found to be in need of improvement.* J Clin Epidemiol, 2014. **67**(6): p. 645-53.
39. Hartling L, Bond K, Santaguida PL, Viswanathan M, and Dryden DM. *Testing a tool for the classification of study designs in systematic reviews of interventions and exposures showed moderate reliability and low accuracy*. J Clin Epidemiol, 2011. **64**, 861-71 DOI: 10.1016/j.jclinepi.2011.01.010.
40. Wagner AK, Soumerai SB, Zhang F, and Ross-Degnan D. *Segmented regression analysis of interrupted time series studies in medication use research*. J Clin Pharm Ther, 2002. **27**, 299-309.
41. Zhang F, Wagner AK, and Ross-Degnan D. *Simulation-based power calculation for designing interrupted time series analyses of health policy interventions*. J Clin Epidemiol, 2011. **64**, 1252-61 DOI: 10.1016/j.jclinepi.2011.02.007.
42. Effective Practice and Organisation of Care Group (EPOC). *What study designs should be included in an EPOC review and what should they be called ?* 2013; Available from: http://epoc.cochrane.org/epoc-specific-resources-review-authors.
43. Penfold RB and Zhang F. *Use of interrupted time series analysis in evaluating health care quality improvements.* Acad Pediatr, 2013. **13**(6 Suppl): p. S38-44.
44. Grimshaw J, Wilson B, Campbell M, Eccles M, and Ramsay C. *Epidemiological methods*, in *Studying the Organisation and Delivery of Health Services Research Methods*, NB Pauline Allen, Aileen Clarke, Naomi Fulop, Stuart Anderson, Editor. 2001, Routledge: New York, USA.
45. Chaiton MO, Schwartz R, Tremblay G, and Nugent R. *Association of flavoured cigar regulations with wholesale tobacco volumes in Canada: an interrupted time series analysis.* Tob Control, 2019. **28**(4): p. 457-461.
46. Biglan A, Ary D, and Wagenaar AC. *The value of interrupted time-series experiments for community intervention research.* Prev Sci, 2000. **1**(1): p. 31-49.
47. Svoronos T and Fretheim A. *Clarifying the interrupted time series study design*. BMJ Qual Saf, 2015. **24**, 475 DOI: 10.1136/bmjqs-2015-004122.
48. Lopez Bernal J, Cummins S, and Gasparrini A. *The use of controls in interrupted time series studies of public health interventions.* Int J Epidemiol, 2018. **47**(6): p. 2082-2093.
49. Lopez Bernal J, Soumerai S, and Gasparrini A. *A methodological framework for model selection in interrupted time series studies.* J Clin Epidemiol, 2018. **103**: p. 82-91.
50. Fretheim A, Soumerai SB, Zhang F, Oxman AD, and 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): p. 883-7.
51. Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, et al. *A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation.* J Clin Epidemiol, 2015. **68**(3): p. 324-33.
52. St.Clair T, Cook T, and 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. **35**(3): p. 311-327.
53. Eccles M, Grimshaw J, Campbell M, and Ramsay C. *Research designs for studies evaluating the effectiveness of change and improvement strategies.* Qual Saf Health Care, 2003. **12**(1): p. 47-52.

54. Grimshaw J, Campbell M, Eccles M, and Steen N. *Experimental and quasi-experimental designs for evaluating guideline implementation strategies*. Fam Pract, 2000. **17 Suppl 1**, S11-6.
55. Reeves BC, Wells GA, and Waddington H. *Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels.* J Clin Epidemiol, 2017. **89**: p. 30-42.
56. Gulayin PE, Irazola V, Rubinstein A, Bruno R, Rossi Diaz A, Gulayin M, et al. *Smoke-Free Adolescents. Effectiveness of an educational intervention. Controlled, before and after study.* Arch Argent Pediatr, 2018. **116**(3): p. e392-e400.
57. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. *Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals*. Cochrane Database Syst Rev, 2012. **11**, MR000030 DOI: 10.1002/14651858.MR000030.pub2.
58. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. *Grading quality of evidence and strength of recommendations*. BMJ, 2004. **328**, 1490 DOI: 10.1136/bmj.328.7454.1490.
59. Sterne JAC, Hernán MA, McAleenan A, Reeves BC, and JPT H. *Chapter 25: Assessing risk of bias in a non-randomized study.*, in *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)*, TJ Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, Editor. 2019, Cochrane.
60. Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, et al. *Interventions to reduce ambient particulate matter air pollution and their effect on health.* Cochrane Database Syst Rev, 2019. **5**: p. CD010919.
61. von Philipsborn P, Stratil JM, Burns J, Busert LK, Pfadenhauer LM, Polus S, et al. *Environmental interventions to reduce the consumption of sugar-sweetened beverages and their effects on health.* Cochrane Database Syst Rev, 2019. **6**: p. CD012292.
62. Chow JC. *Measurement methods to determine compliance with ambient air quality standards for suspended particles.* J Air Waste Manag Assoc, 1995. **45**(5): p. 320-82.
63. World Health Organization Europe. *Review of evidence on health aspects of air pollution – REVIHAAP Project: Technical Report*. 2013, WHO Regional Office for Europe: Copenhagen.
64. GBD Risk Factor Collaborators. *Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet, 2018. **392**(10159): p. 1923-1994.
65. Been JV, Nurmatov UB, Cox B, Nawrot TS, van Schayck CP, and Sheikh A. *Effect of smoke-free legislation on perinatal and child health: a systematic review and meta-analysis.* Lancet, 2014. **383**(9928): p. 1549-60.
66. Faber T, Kumar A, Mackenbach JP, Millett C, Basu S, Sheikh A, et al. *Effect of tobacco control policies on perinatal and child health: a systematic review and meta-analysis.* Lancet Public Health, 2017. **2**(9): p. e420-e437.
67. Frazer K, Callinan JE, McHugh J, van Baarsel S, Clarke A, Doherty K, et al. *Legislative smoking bans for reducing harms from secondhand smoke exposure, smoking prevalence and tobacco consumption.* Cochrane Database Syst Rev, 2016. **2**: p. CD005992.
68. Hudson J, Fielding S, and Ramsay CR. *Methodology and reporting characteristics of studies using interrupted time series design in healthcare.* BMC Med Res Methodol, 2019. **19**(1): p. 137.
69. Gasparrini A, Gorini G, and Barchielli A. *On the relationship between smoking bans and incidence of acute myocardial infarction.* Eur J Epidemiol, 2009. **24**(10): p. 597-602.

70. Ewusie J. *Improved methods for interrupted time series analysis useful when outcomes are aggregated: accounting for heterogeneity across patients and healthcare settings*, in *Health Research Methods, Evidence, and Impact*. 2018, McMaster University: Hamilton, Ontario, Canada. p. 135.
71. Paudel D, Shrestha IB, Siebeck M, and Rehfuess E. *Impact of the community-based newborn care package in Nepal: a quasi-experimental evaluation.* BMJ Open, 2017. **7**(10): p. e015285.
72. MOH, NewERA, and ICF;. *Nepal demographic and health survey 2016 key indicator report.* . 2017: Kathmandu, Nepal.
73. Burns J, Hoffmann S, Kurz C, Laxy M, Polus S, and Rehfuess E. *COVID-19 mitigation measures and nitrogen dioxide - A quasi-experimental study of air quality in Munich, Germany.* Atmos Environ (1994), 2021. **246**: p. 118089.
74. Turner SL, Karahalios A, Forbes AB, Taljaard M, Grimshaw JM, Cheng AC, et al. *Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: protocol for a review.* BMJ Open, 2019. **9**(1): p. e024096.
75. Jandoc R, Burden AM, Mamdani M, Levesque LE, and Cadarette SM. *Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations*. J Clin Epidemiol, 2015. **68**, 950-6 DOI: 10.1016/j.jclinepi.2014.12.018.
76. EQUATOR Network. *Reporting guidelines under development for observational studies or STROBE extensions*. 2020 [cited 2020; Available from: https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-observational-studies/#92.
77. Lopez Bernal J, Cummins S, and Gasparrini A. *Difference in difference, controlled interrupted time series and synthetic controls.* Int J Epidemiol, 2019. **48**(6): p. 2062-2063.
78. Benmarhnia T and Rudolph KE. *A rose by any other name still needs to be identified (with plausible assumptions).* Int J Epidemiol, 2019. **48**(6): p. 2061-2062.
79. Soumerai SB, Starr D, and Majumdar SR. *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**: p. E101.
80. Rockers PC, Feigl AB, Rottingen JA, Fretheim A, de Ferranti D, Lavis JN, et al. *Study-design selection criteria in systematic reviews of effectiveness of health systems interventions and reforms: A meta-review*. Health Policy, 2012. **104**, 206-14 DOI: 10.1016/j.healthpol.2011.12.007.
81. Linden A and Yarnold PR. *Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis.* J Eval Clin Pract, 2018. **24**(4): p. 740-744.
82. Ryan AM, Kontopantelis E, Linden A, and Burgess JF, Jr. *Now trending: Coping with non-parallel trends in difference-in-differences analysis.* Stat Methods Med Res, 2019. **28**(12): p. 3697-3711.

Appendix A: Additional publication III

The choice of analysis in controlled before-after studies and its impact on effect size and statistical precision

Stephanie Polus, Institute for Medical Information Processing, Biometry, and Epidemiology – IBE, LMU Munich, Munich, Germany, Pettenkofer School of Public Health, Munich, Germany, polus@ibe.med.uni-muenchen.de, corresponding author,

Jacob Burns, Institute for Medical Information Processing, Biometry, and Epidemiology – IBE, LMU Munich, Munich, Germany, Pettenkofer School of Public Health, Munich, Germany, burns@ibe.med.uni-muenchen.de

Sara Pedron, Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Munich, Germany, sara.pedron@helmholtz-muenchen.de

Deepak Paudel, Save the Children, Kathmandu, Nepal, deepak.paudel@savethechildren.org

Eva A Rehfuess, Institute for Medical Information Processing, Biometry, and Epidemiology – IBE, LMU Munich, Munich, Germany, Pettenkofer School of Public Health, Munich, Germany, rehfuess@ibe.med.uni-muenchen.de

Abstract

Objective:

In this research note, we assess the influence that different methods to analyse controlled before-after (CBA) studies have on the size of the effect estimate and the associated precision using a worked example.

Results:

Methods commonly applied in CBA studies can lead to misleading conclusions if the statistical analysis fails to fully exploit the structure of the collected data, or to take clustering into account. With increasing use of CBA studies, researchers should make best use of difference-in-differences methods and appropriately adjust for clustering. Systematic reviewers may need to re-analyse incorrectly or insufficiently analysed CBA studies.

Keywords

Systematic review, quasi-experimental studies, controlled before-after studies, statistics, epidemiology, public health

Introduction

The controlled before-after (CBA) study was found, along with the interrupted time series (ITS) study, to be the nonrandomised study design most often included in Cochrane reviews (1, 2). The CBA study uses pre- and post-intervention measurements in an intervention and a control

group to assess whether an effect of an intervention is observed in the intervention group relative to the control group (3). CBA studies often use a clustered design where sites, such as healthcare centres, schools or regions within a country, are allocated to an intervention or a control group (4). They are commonly used in health services and public health research, where randomisation may be unfeasible, unethical or simply difficult to implement (5).

In a recent study assessing the methodological characteristics of CBA and ITS studies included in Cochrane reviews, we found that many CBA studies applied suboptimal or even inappropriate analysis methods (2). This may lead to distorted conclusions about the intervention effect. In randomised controlled trials (RCTs), researchers have long focused on analysis methods (6). Bland and Altman (7), for example, described in detail how within group comparisons in RCTs may lead to highly misleading interpretations. In CBA studies where randomisation is lacking, it is all the more important to understand the implications of different analysis methods, both with respect to the findings of individual studies as well as with respect to the conclusions of systematic reviews drawing on such studies.

We assessed the influence that different analysis methods commonly applied in CBA studies have on the size of the effect estimate and the associated precision using a worked example.

Main text

Methods

We re-analysed a previously conducted CBA study that assessed the impacts of the so-called community-based newborn care package (CBNCP) in Nepal (8). The CBNCP, which comprises seven community- and home-based interventions, was implemented by the Ministry of Health in ten out of 75 districts in Nepal in 2009 and 2010 to tackle major causes of neonatal

mortality. The study examined the impacts of this large-scale government programme on women's behaviours influencing neonatal health in ten intervention and ten control districts, using data from the Nepal Demographic and Health Survey (DHS) among other sources (9). The ten control districts were chosen using propensity score matching (8). We used the DHS data analysed in this study for our worked example, focusing on one of the selected behaviours among women, i.e. *recommended antenatal care seeking*.

The controlled before-after study design

The CBA study is broadly described as a study design in which observations are made before and after the implementation of an intervention, both in a group that receives the intervention and in a control group that does not. Figure 1 is a simplified illustration of the design showing the intervention group (grey square) and the control group (black dot). They may represent one, or a mean of several measurements taken over time before and after the intervention, respectively. In the following we introduce commonly used analysis methods identified in a previous study (2). Figure 2 comprises panels a-f representing the different analysis choices

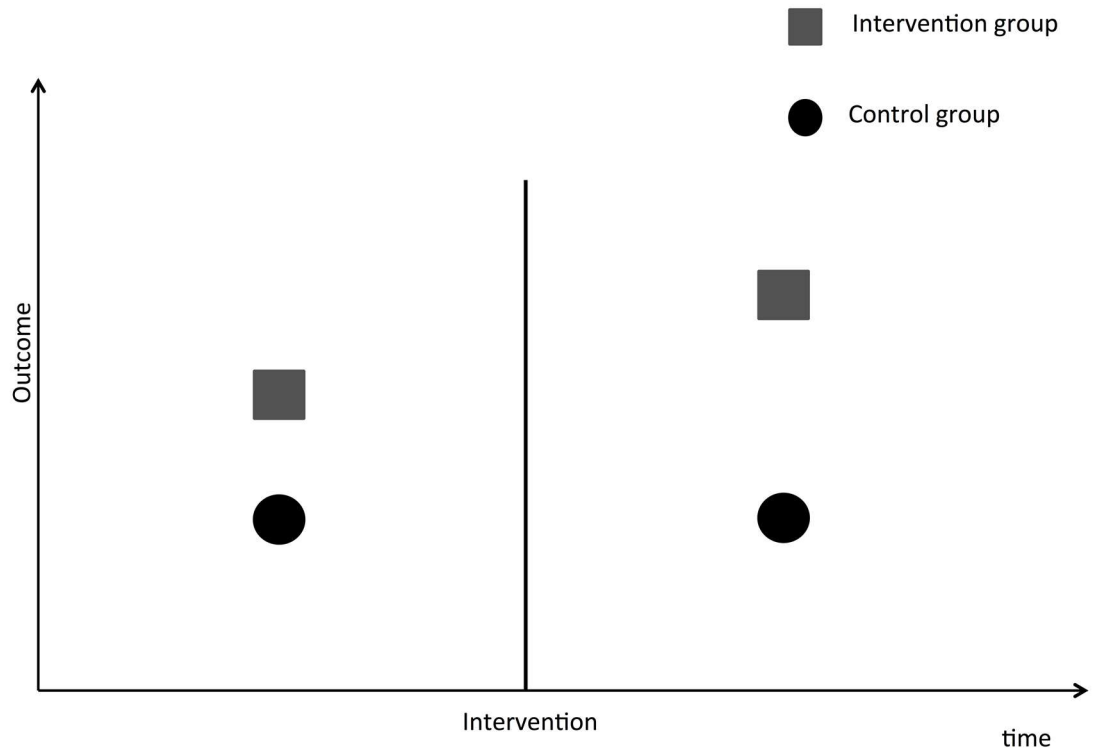


Figure 1 Simplified controlled before-after study design, adapted from Grimshaw, Campbell (10)

Analysis A: Comparison of post-intervention means between intervention and control groups

The study may utilize only post-intervention data in the analysis, i.e. the post-intervention mean in the intervention group is compared with the post-intervention mean of the control group (see Fig 2a). This type of analysis has already been criticized for RCTs (7). In CBA studies, where balanced baseline groups are much more difficult to obtain if at all possible, ignoring baseline differences between groups in the statistical analysis is likely to bias the effect estimate.

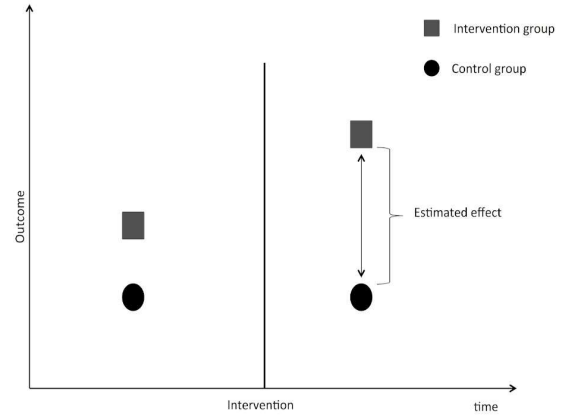


Figure 2a Comparison of post-intervention means between intervention and control group (analysis A)

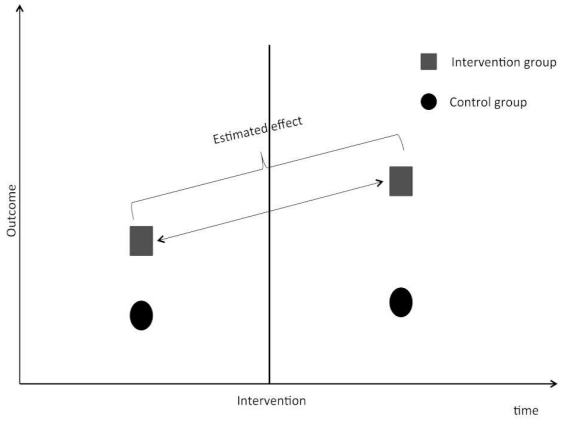


Figure 2b Comparison of pre- and post-intervention means of the intervention group only (analysis B)

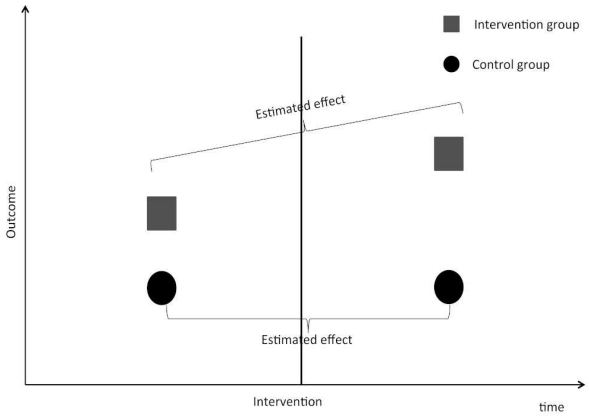


Figure 2c Comparison of pre-post differences separately (analysis C)

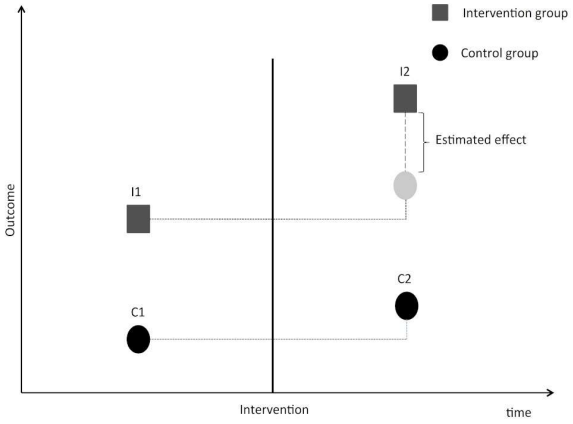


Figure 2d Difference-in-differences analysis (analysis D)

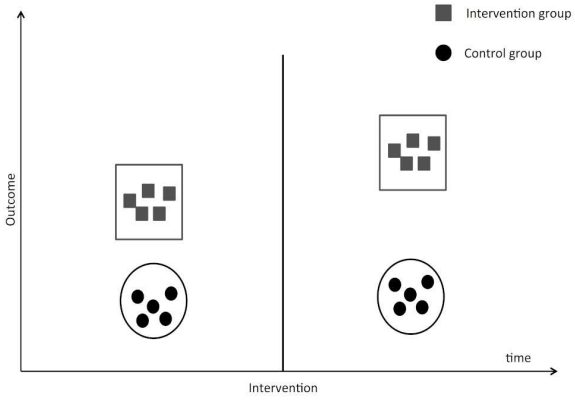


Figure 2e Controlled before-after study with two level-clustering

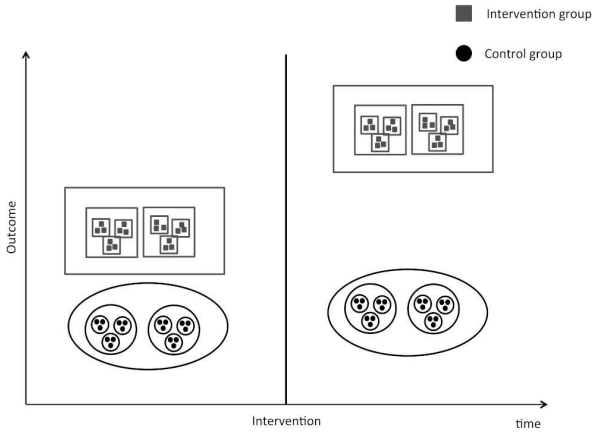


Figure 2f Controlled before-after study with three level-nesting

93

94 Figure 2 comprising panels a-f representing the different analysis choices

95

Analysis B: Comparison of pre- and post-intervention means of the intervention group only

The study, despite measuring the outcome in a control group, may not include these measurements in a statistical analysis but may calculate the difference between pre- and post-measurements in the intervention group only (see Fig 2b). Changes in the outcome, however, might also reflect general developments in the population, independent from the intervention (e.g. economic trends, implementation of other large-scale programs, new regulations, etc.). By focusing only on the pre-post differences in the intervention group, the effect of the intervention will not be isolated, providing biased effect estimates.

Analysis C: Separate comparison of pre-post differences in intervention and control group

The study may analyse within-group differences but not compare the differences between the groups statistically (see Fig 2c). Again, the effect of treatment will not be isolated from the effect of other potential influences, resulting in potentially biased estimates.

Analysis D: Difference-in-differences analysis, adjusting for clustering

The difference-in-differences (DiD) analysis takes into account all measurements in both groups and estimates the differences between the within-group differences. It therefore accounts for time-dependent trends in the outcome unrelated to the intervention, assuming parallel trends in the groups. The estimated effect is calculated by (I2−I1) - (C2−C1) (11) (see Fig 2d). In studies where data are aggregated from different sites or groups, it is important that the analysis takes this clustering into account (see detailed explanation below in analysis E).

Analysis E: Difference-in-differences analysis, not adjusting for clustering

In CBA studies there may be two (e.g. individual, family) (see Fig 2e) or more levels (e.g. individual, family, community) (see Fig 2f) that should be accounted for in the analysis. This is called clustering (in the case of two levels) or nesting (if three or more levels are present) (5). If the analysis simply lumps the means (and standard deviations) of each of the sites together into a single site, ignoring the individual-level data, the study will make inefficient use of the data and the standard error will be much larger than needed. On the other hand, if the analysis aggregates the data of all sites together and then treats it on an individual level, it will create a serious unit of analysis error: the standard errors will be much smaller than they actually are (12).

Application to worked example

For our worked example we calculated the proportion of women adhering to *recommended antenatal care seeking*, measured as completing four or more antenatal care visits. For the above-described analyses A-C we applied a simple linear regression model:

(A) $\Upsilon_i = \beta_0 + \beta_1 intervention_i + \epsilon_{it}$

(B) $\Upsilon_{it} = \beta_0 + \beta_1 post_t + \epsilon_{it}$

(C) $\Upsilon_{int_{it}} = \beta_0 + \beta_1 post_t + \epsilon_{it}$

$\Upsilon_{cont_{it}} = \beta_0 + \beta_1 post_t + \epsilon_{it}$

where, for person i at time t, Υ represents the outcome, proportion of women adhering to recommended antenatal care seeking; $intervention$ is an indicator variable that defines whether a person belongs to the intervention or control group, taking the value 0 for the control group and 1 for the intervention group; $post$ is an indicator variable that defines which

time period the measure belongs to, taking the value 0 before and 1 after the intervention. ε represents the idiosyncratic error of the regression.

In analysis C, the estimation is conducted separately for the intervention group and the control group, however no statistical comparison takes place.

For analyses D and E, we performed a DiD analysis:

$$\text{(D) } Y_{it} = \beta_0 + \beta_1 intervention_{it} + \beta_2 post_t + \beta_3 post_t * intervention_{it} + \epsilon_{it}$$

In analysis D, we adjusted for clustering at district level. In analysis E, the standard errors of the coefficients do not account for clustering at district level.

We applied sample weights and complex survey design methods in all analyses as provided by the DHS (8). All analyses were conducted in Stata 14 (13) (see additional file 1 for the code of the statistical analysis).

Results

Table 1 displays the analysis results and an interpretation of the findings according to analysis method.

Table 1 Analysis results and interpretation of findings according to analysis method

Discussion

We apply in this worked example the most commonly used analysis methods and illustrate how the findings and their interpretation vary accordingly. For example, a statistical comparison of the pre-and post-intervention rates of the intervention group only shows a

clear, statistically significant increase in recommended antenatal care seeking (i.e. analysis B). A simple observation of before-after differences within the intervention and control groups (i.e. analysis C) suggests a much larger increase in recommended antenatal care seeking in the intervention group compared to a much smaller and statistically non-significant improvement in the control group. However it is unclear how relevant this difference is, given that we have no direct statistical comparison. The findings of the DiD analysis (i.e. analysis D), however, imply a relative, yet statistically non-significant, 9% improvement of recommended antenatal care seeking in the intervention group compared to the control group. In this worked example we see that the choice of different statistical analysis approaches can lead to very different interpretations in terms of effect size and statistical precision and thus in terms of drawing conclusions about intervention effectiveness.

Limitations

Little methodological guidance exists for the correct implementation and analysis of the CBA study design, despite its importance for public health and health services research and its increased application in primary research as well as inclusion in systematic reviews (2). In this paper, we provide only a simple comparison of the most commonly used analysis methods identified in a previous study. We recommend that detailed guidance be developed, also including a detailed discussion of different more advanced statistical analysis methods (14).

Abbreviations

CBA study = Controlled before-after study

CBNCP = Community-based newborn care package

DHS = Demographic and Health Survey

DiD analysis = Difference-in-differences analysis

ITS study = Interrupted time series study

RCT = Randomised controlled trial

Declarations

Ethics approval

Not applicable

Consent for publication

Not applicable

Availability of data and material

The data used are available from the public domain of the Demographic Health Survey (DHS) Program (www.dhsprogram.com). The DHS program requires all researchers wishing to use data for the purpose of research or study, to register. The authors did not have any special access privileges that others would not have. Information regarding registration can be found here: https://dhsprogram.com/data/new-user-registration.cfm.

Competing interests

The authors declare that they have no competing of interests.

Funding

SP is funded through a doctoral scholarship by the Heinrich Boell Foundation (Germany).

Authors contributions

SP conceived the study, developed the methods, conducted and interpreted the statistical analysis, and wrote the manuscript. JB co-conceived the study, interpreted the analysis, and commented on

206 the manuscript. SPe conducted the analysis, interpreted the analysis, and commented on the
207 manuscript. DP aided with the statistical analysis and commented on the manuscript. EAR co-
208 conceived the study and commented on the manuscript.

209 Acknowledgements

210 Not applicable

211 Legends

212 Figure 1 Simplified controlled before-after study design, adapted from Grimshaw, Campbell
213 (10)

214

215 Figure 2 Panels a-f representing the different analysis choices

216

217 Table 1 Proportion of women displaying recommended antenatal care seeking (4 or more
218 antenatal care visits) and interpretation of findings, by analysis method; estimates of
219 ordinary least squares regression with relative standard error (SE), 95% confidence intervals
220 (95% CI) and p-value (p)

221

222 Additional file 1 Code of the statistical analysis (Stata do file.pdf)

223 References

224 1. Ijaz S, Verbeek JH, Mischke C, Ruotsalainen J. Inclusion of nonrandomized studies in
225 Cochrane systematic reviews was found to be in need of improvement. Journal of clinical
226 epidemiology [Internet]. 2014 Jun; 67(6):[645-53 pp.]. Available from:
227 http://www.ncbi.nlm.nih.gov/pubmed/24725644.
228 2. Polus S, Pieper D, Burns J, Fretheim A, Ramsay C, Higgins JP, et al. Heterogeneity in
229 application, design and analysis characteristics was found for controlled before-after (CBA) and
230 interrupted time series (ITS) studies included in Cochrane reviews. Journal of clinical epidemiology.
231 2017.

3. Shadish W, Cook T, Campbell M. Quasi-experimental designs that use both control groups and pretests. In: Prancan K, editor. Experimental and Quasi-Experimental Designs for Generalized Causal Inference: Wadsworth Cengage Learning; 2002.
4. Reeves BC, Higgins JP, Ramsay C, Shea B, Tugwell P, Wells GA. An introduction to methodological issues when including non-randomised studies in systematic reviews on the effects of interventions. Res Synth Methods. 2013;4(1):1-11.
5. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med. 2004;2(3):204-8.
6. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. Lancet. 1990;335(8682):149-53.
7. Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. Trials. 2011;12:264.
8. Paudel D, Shrestha IB, Siebeck M, Rehfuess E. Impact of the community-based newborn care package in Nepal: a quasi-experimental evaluation. BMJ Open. 2017;7(10):e015285.
9. Ministry of Health and Population - MOHP/Nepal, New ERA/Nepal, ICF International. Nepal Demographic and Health Survey 2011. Kathmandu, Nepal: MOHP/Nepal, New ERA, and ICF International; 2012.
10. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Family practice [Internet]. 2000 Feb; 17 Suppl 1:[S11-6 pp.]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10735262.
11. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. JAMA. 2014;312(22):2401-2.
12. Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. Bmj. 1998;317(7167):1171-2.
13. StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LLC. 2015.
14. Senn S. Change from baseline and analysis of covariance revisited. Stat Med. 2006;25(24):4334-44.

Tables

Table 1 Analysis results and interpretation of findings according to analysis method

	A. Comparison of post-intervention means	B. Comparison of pre- post-intervention rates of the intervention group only	C. Separate comparisons of pre-post differences in intervention and control group	D. Difference-in-differences, adjusting for clustering at district level[1]	E. Difference-in-differences, not adjusted for clustering at district level
Recommended antenatal care seeking	0.165 (0.117), 95%CI (-0.070, 0.399) p=0.16	0.160 (0.069), 95%CI (0.021, 0.300) p=0.025	Intervention: 0.160 (0.069), 95%CI (0.021, 0.300), p=0.025 Control: 0.068 (0.086), 95%CI (-0.110, 0.246), p=0.438	0.092 (0.105), 95%CI (-0.117, 0.302), p=0.383	0.092 (0.087), 95%CI (-0.078, 0.262) p=0.287

Possible inter-pretation	*In the intervention group 16.5% (p=0.165) more women were seeking recommended antenatal care compared to the control group, however, this difference was not statistically significant.*	*There was a statistically significant increase of 16.0% (p=0.025) of women in the intervention group seeking recommended antenatal care after the intervention.*	*Whereas in the intervention group an increase of 16.0% (p=0.025) in recommended antenatal care seeking was found, a smaller, statistically not significant increase of 6.8% (p=0.086) was observed in the comparison group.*	*We found a statistically non-significant increase of 9.2% (p=0.383) in recommended antenatal care seeking among women in the intervention compared with the comparison group.*	*We found a statistically non-significant increase of 9.2% (p=0.287) for recommended antenatal care seeking in the intervention group compared with the control group after the intervention.*

[1]analysis used in original study by Paudel, Shrestha (8)

Legend: Proportion of women displaying recommended antenatal care seeking (4 or more antenatal care visits) and interpretation of findings, by analysis method; estimates of ordinary least squares regression with relative standard error (SE), 95% confidence intervals (95% CI) and p-value (p)

Do_File_Publication.do - Printed on 09/07/2020 16:56:57

```
** ANALYSIS OF IMPACT BASED ON DHS SURVEY, shortened for outcome antenatal care seeking

 *use "H:\XX.dta", clear
 use "\\XX ", clear

 generate weight = v005/1000000
egen strata = group(v022), label
svyset [pweight=weight], psu(v021) strata(strata) singleunit(center)
  ** categorize districts for intervention and control areas
 generate float district = v001/100
 recode district (1/1.99=1)(2/2.99=2)(3/3.99=3)(4/4.99=4)(5/5.99=5)(6/6.99=6) ///
 (7/7.99=7)(8/8.99=8)(9/9.99=9)(10/10.99=10)(11/11.99=11)(12/12.99=12) ///
 (13/13.99=13)(14/14.99=14)(15/15.99=15)(16/16.99=16)(17/17.99=17)(18/18.99=18) ///
 (19/19.99=19)(20/20.99=20)(21/21.99=21)(22/22.99=22)(23/23.99=23)(24/24.99=24) ///
 (25/25.99=25)(26/26.99=26)(27/27.99=27)(28/28.99=28)(29/29.99=29)(30/30.99=30) ///
 (31/31.99=31)(32/32.99=32)(33/33.99=33)(34/34.99=34)(35/35.99=35)(36/36.99=36) ///
 (37/37.99=37)(38/38.99=38)(39/39.99=39)(40/40.99=40)(41/41.99=41)(42/42.99=42) ///
 (43/43.99=43)(44/44.99=44)(45/45.99=45)(46/46.99=46)(47/47.99=47)(48/48.99=48) ///
 (49/49.99=49)(50/50.99=50)(51/51.99=51)(52/52.99=52)(53/53.99=53)(54/54.99=54) ///
 (55/55.99=55)(56/56.99=56)(57/57.99=57)(58/58.99=58)(59/59.99=59)(60/60.99=60) ///
 (61/61.99=61)(62/62.99=62)(63/63.99=63)(64/64.99=64)(65/65.99=65)(66/66.99=66) ///
 (67/67.99=67)(68/68.99=68)(69/69.99=69)(70/70.99=70)(71/71.99=71)(72/72.99=72) ///
 (73/73.99=73)(74/74.99=74)(75/75.99=75)
 recode district (5 6 7 24 34 35 47 56 58 70=1 "Intervention") ///
 (4 14 17 20 25 31 39 45 59 72=0 "Control") /// (nonmiss=3 "Other"), gen(dist)
generate intv=.
replace intv = b3 if district==70|district==58|district==56|district==47| /// district==35|
district==24|district==34|district==6|district==5|district==7 replace intv = b3 if
district==4|district==14|district==17|district==20| /// district==25|
district==31|district==39|district==45|district==59|district==72
 recode intv (1955/1994=0) (2008/2015=1) (nonmiss=3) if district==70|district==20
 recode intv (1955/1993=0) (2000/2015=1) (nonmiss=3) if district==58|district==59
 recode intv (1955/1999=0) (2005/2015=1) (nonmiss=3) if district==56|district==59
 recode intv (1955/2004=0) (2008/2015=1) (nonmiss=3) if district==47|district==25
 recode intv (1955/2001=0) (2007/2015=1) (nonmiss=3) if district==35|district==72
 recode intv (1955/2001=0) (2008/2015=1) (nonmiss=3) if district==24|district==14
 recode intv (1955/1995=0) (2008/2015=1) (nonmiss=3) if district==34|district==45
 recode intv (1955/1992=0) (2003/2015=1) (nonmiss=3) if district==6|district==31
 recode intv (1955/2004=0) (2008/2015=1) (nonmiss=3) if district==5|district==17
 recode intv (1955/2004=0) (2008/2015=1) (nonmiss=3) if district==7|district==39
 recode intv (.=2)
recode intv (0=0 "Before") (1=1 "After")(3=3 "During")(2=2 "Other"), gen(interv)
 *grouping
 gen grp=" "
 replace grp="Intv_Before" if dist==1 & intv==0
 replace grp="Intv_After" if dist==1 & intv==1
```

```
 replace grp="Cont_Before" if dist==0 & intv==0
 replace grp="Cont_After" if dist==0 & intv==1
 replace grp="Other" if dist==3 | intv==2 | intv==3 keep if dist<=1 & intv<=1 & bidx==1 &
(v008-b3)<60

* Overall ANC care seeking
generate float tt2 = 0 if m1==0|m1==1|m1==8|m1a==1|m1a==2|m1a==3|m1a==4|m1a==8
replace tt2=1 if m1==2|m1==3|m1==4|m1==5|m1a==5|m1a==6|m1a==7
label variable tt2 "TT lifttime protection"
recode m46 (90/300=1 "Iron 90+ tab")(0/89 998=0 "Iron - no or <90")(else=.), gen
replace m46a=0 if m45==0

 Do_File_Publication.do - Printed on 09/07/2020 16:56:58
generate anc_seek=0
replace anc_seek=1 if (anc_n4==1&tt2==1&m46a==1)


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

*ANALYSIS
*A. comparison of post intervention rates
svy: reg anc_seek dist if interv==1
*B. comparison of pre- vs. post-intervention rates of the intervention group only
svy: reg anc_seek interv if dist==1
*C. before after comparing means
svy: reg anc_seek interv if dist==0
*D: full did with clustering at district level
svyset [pweight=weight], psu(v021) strata(strata) singleunit(center)
svy: reg anc_seek dist##interv
*E: full did WITHOUT clustering at district level
 svyset [pweight=weight], strata(strata) singleunit(center)
 svy: reg anc_seek dist##interv
```

Acknowledgements

First I want to thank my supervisor Eva Rehfuess for her continuous and dependable outstanding supervision and support throughout the years. A big thanks also to Ulrich Mansmann and Atle Fretheim for their engagement being on my PhD advisory board. I would also like to thank the Heinrich-Boell Foundation for enabling this work through a promotion grant.

I am most grateful for my friend and secret supervisor Jake Burns without whom I would not have been able to do this. Thanks to him and Lisa Pfadenhauer, who as the real freefolk made work so much fun. I am grateful to my husband Lukas for his support and my extended family, the big Assmann Clan, and my mother Heidi for keeping my back with hours of babysitting, providing a working space or food. I am so grateful for my two little daughters Leonie and Clara who actually kept me from finishing this work smoothly, which was the best thing that ever happened to me.

List of all scientific publications

Brereton L, Wahlster P, Mozygemba K, Lysdahl KB, Burns J, Polus S, et al., *Stakeholder Involvement Throughout Health Technology Assessment: An Example from Palliative Care.* Int J Technol Assess Health Care, 2017. **33**(5): p. 552-561.

Burns J, Hoffmann S, Kurz C, Laxy M, Polus S, and Rehfuess E, *COVID-19 mitigation measures and nitrogen dioxide - A quasi-experimental study of air quality in Munich, Germany.* Atmos Environ (1994), 2021. **246**: p. 118089.

Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, et al., *Interventions to reduce ambient air pollution and their effects on health: An abridged Cochrane systematic review.* Environ Int, 2020. **135**: p. 105400.

Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, et al., *Interventions to reduce ambient particulate matter air pollution and their effect on health.* Cochrane Database Syst Rev, 2019. **5**: p. CD010919.

Burns J, Polus S, Brereton L, Chilcott J, Ward SE, Pfadenhauer LM, et al., *Looking beyond the forest: Using harvest plots, gap analysis, and expert consultations to assess effectiveness, engage stakeholders, and inform policy.* Res Synth Methods, 2018. **9**(1): p. 132-140.

Konsgen N, Polus S, Rombey T, and Pieper D, *Clowning in children undergoing potentially anxiety-provoking procedures: a systematic review and meta-analysis.* Syst Rev, 2019. **8**(1): p. 178.

Mathes T, Pieper D, Morche J, Polus S, Jaschinski T, and Eikermann M, *Pay for performance for hospitals.* Cochrane Database Syst Rev, 2019. **7**: p. CD011156.

Mathes T, Willms G, Polus S, Stegbauer C, Messer M, Klingler C, et al., *Health technology assessment of public health interventions: an analysis of characteristics and comparison of methods-study protocol.* Syst Rev, 2018. **7**(1): p. 79.

Mathes T, Antoine SL, Prengel P, Buhn S, Polus S, and Pieper D, *Health Technology Assessment of Public Health Interventions: A Synthesis of Methodological Guidance.* Int J Technol Assess Health Care, 2017. **33**(2): p. 135-146.

Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, et al., *Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework.* Implement Sci, 2017. **12**(1): p. 21.

Polus S, Mathes T, Klingler C, Messer M, Gerhardus A, Stegbauer C, et al., *Health Technology Assessment of Public Health Interventions Published 2012 to 2016: An Analysis of Characteristics and Comparison of Methods.* Int J Technol Assess Health Care, 2019. **35**(4): p. 280-290.

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**: p. 56-69.

Polus S, Pfadenhauer L, Brereton L, Leppert W, Wahlster P, Gerhardus A, et al., *A Consultation Guide for Assessing the Applicability of Health Technologies:A Case Study.* Int J Technol Assess Health Care, 2017. **33**(5): p. 577-585.

Polus S, Lewin S, Glenton C, Lerberg PM, Rehfuess E, and Gulmezoglu AM, *Optimizing the delivery of contraceptives in low- and middle-income countries through task shifting: a systematic review of effectiveness and safety.* Reprod Health, 2015. **12**: p. 27.

Polus S, Lerberg P, Vogel J, Watananirun K, Souza JP, Mathai M, et al., *Appraisal of WHO guidelines in maternal health using the AGREE II assessment tool.* PLoS One, 2012. **7**(8): p. e38891.

Prediger B, Mathes T, Polus S, Glatt A, Buhn S, Schiermeier S, et al., *A systematic review and time-response meta-analysis of the optimal timing of elective caesarean sections for best maternal and neonatal health outcomes.* BMC Pregnancy Childbirth, 2020. **20**(1): p. 395.

Prediger B, Polus S, Mathes T, Buhn S, Louwen F, Neugebauer EAM, et al., *(Update of a) systematic review on the impact of elective early term (< 39th gestational week) caesarean sections on maternal and neonatal health - a protocol.* Syst Rev, 2018. **7**(1): p. 119.

Runkel B, Bein G, Sieben W, Sow D, Polus S, and Fleer D, *Targeted antenatal anti-D prophylaxis for RhD-negative pregnant women: a systematic review.* BMC Pregnancy Childbirth, 2020. **20**(1): p. 83.

von Philipsborn P, Stratil JM, Burns J, Busert LK, Pfadenhauer LM, Polus S, et al., *Environmental Interventions to Reduce the Consumption of Sugar-Sweetened Beverages: Abridged Cochrane Systematic Review.* Obes Facts, 2020. **13**(4): p. 397-417.

von Philipsborn P, Stratil JM, Burns J, Busert LK, Pfadenhauer LM, Polus S, et al., *Environmental interventions to reduce the consumption of sugar-sweetened beverages and their effects on health.* Cochrane Database Syst Rev, 2019. **6**: p. CD012292.

Wahlster P, Brereton L, Burns J, Hofmann B, Mozygemba K, Oortwijn W, et al., *An Integrated Perspective on the Assessment of Technologies: Integrate-Hta.* Int J Technol Assess Health Care, 2017. **33**(5): p. 544-551.