Author/s:
Korevaar, E;Karahalios, A;Turner, SL;Forbes, AB;Taljaard, M;Cheng, AC;Grimshaw, JM;Bero, L;McKenzie, JE

Title:
Methodological systematic review recommends improvements to conduct and reporting when meta-analyzing interrupted time series studies

Date:
2022-05-01

Citation:
Korevaar, E., Karahalios, A., Turner, S. L., Forbes, A. B., Taljaard, M., Cheng, A. C., Grimshaw, J. M., Bero, L. & McKenzie, J. E. (2022). Methodological systematic review recommends improvements to conduct and reporting when meta-analyzing interrupted time series studies. Journal of Clinical Epidemiology, 145, pp.55-69. https://doi.org/10.1016/j.jclinepi.2022.01.010.

Persistent Link:
https://hdl.handle.net/11343/302626

Journal of Clinical Epidemiology 145 (2022) 55–69

Journal of Clinical Epidemiology

REVIEW

Methodological systematic review recommends improvements to conduct and reporting when meta-analyzing interrupted time series studies

Elizabeth Korevaar[a], Amalia Karahalios[b], Simon L Turner[a], Andrew B Forbes[a], Monica Taljaard[c,d], Allen C Cheng[a,e], Jeremy M Grimshaw[c,d,f], Lisa Bero[g], Joanne E McKenzie[a,*]

[a] *School of Public Health and Preventive Medicine, Monash University, Melbourne, 3004, Victoria, Australia*
[b] *Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, 3010, Victoria, Australia*
[c] *Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, K1Y 4E9, Canada*
[d] *School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada*
[e] *Infection Prevention and Healthcare Epidemiology Unit, Alfred Health, Melbourne, 3004, Victoria, Australia*
[f] *Department of Medicine, University of Ottawa, Ottawa, Ontario, K1H 8M5, Canada*
[g] *School of Medicine and Colorado School of Public Health, Center for Bioethics and Humanities, University of Colorado Anschutz Medical Campus, 13080 E. 19th Ave, Aurora, CO 80045 | Mail Stop B137, Denver, Colorado*

Accepted 13 January 2022; Available online 16 January 2022

Abstract

Objectives: Interrupted Time Series (ITS) are a type of nonrandomized design commonly used to evaluate public health policy interventions, and the impact of exposures, at the population level. Meta-analysis may be used to combine results from ITS across studies (in the context of systematic reviews) or across sites within the same study. We aimed to examine the statistical approaches, methods, and completeness of reporting in reviews that meta-analyze results from ITS.

Study Design and Settings: Eight electronic databases were searched to identify reviews (published 2000–2019) that meta-analyzed at least two ITS. Characteristics of the included reviews, the statistical methods used to analyze the ITS and meta-analyze their results, effect measures, and risk of bias assessment tools were extracted.

Results: Of the 4213 identified records, 54 reviews were included. Nearly all reviews (94%) used two-stage meta-analysis, most commonly fitting a random effects model (69%). Among the 41 reviews that re-analyzed the ITS, linear regression (39%) and ARIMA (20%) were most commonly used; 38% adjusted for autocorrelation. The most common effect measure meta-analyzed was an immediate level-change (46/54). Reporting of the statistical methods and ITS characteristics was often incomplete.

Conclusion: Improvement is needed in the conduct and reporting of reviews that meta-analyze results from ITS.

Keywords: Systematic review; Meta-analysis; Interrupted time series; Segmented regression; Statistical methods; Reporting quality

Conflict of interest: All authors have no competing interests to declare.

Authors contributions: J.E.M. conceived the study, and all authors contributed to its design. E.K., A.K., A.B.F., S.L.T. and J.E.M. contributed to the data collection. E.K. wrote the first draft of the manuscript, with contributions from J.E.M. A.K., A.B.F., S.L.T., M.T., A.C.C., J.M.G., L.B. and J.E.M. contributed to revisions of the manuscript.

Funding: E.K. is supported through an Australian Government Research Training Program (RTP) Scholarship administered by Monash University, Australia. J.E.M. supported by an NHMRC Career Development Fellowship (1143429). The project is funded by the Australian National Health and Medical Research Council (NHMRC) project grant GNT1145273, "How should we analyse, synthesize, and interpret evidence from interrupted time series studies? Making the best use of available evidence", McKenzie JE, Forbes A, Taljaard M, Cheng A, Grimshaw J, Bero L, Karahalios A. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

* Corresponding author.
E-mail address: joanne.mckenzie@monash.edu (J.E. McKenzie).

https://doi.org/10.1016/j.jclinepi.2022.01.010

What is new?

Key findings

In our methodological review of 54 reviews that included a meta-analysis of results from interrupted time series (ITS), we found that:

- determining the key characteristics of included ITS was frequently not possible (eg, time interval of the datapoints, average number of datapoints per series);
- ITS were frequently re-analyzed for inclusion in meta-analyses, and the majority of re-analysis models included level- and slope-change parameters (i.e. segmented regression model). However, the entire model structure, or if autocorrelation was adjusted for, could not commonly be determined; and,
- the two-stage meta-analysis was the most common approach to combine results from multiple ITS, however the heterogeneity variance estimator and confidence interval method were frequently not reported. Furthermore, an immediate level-change was the most commonly meta-analyzed effect measure and, for the majority, was the only effect measure used to evaluate the effects of the interruption.

What this adds to what is known

- There have been a number of methodological reviews examining meta-analysis methods for including randomized and observation study designs, as well as examining the characteristics of individual ITS and their analysis. This review provides novel methodological insight into meta-analyses including ITS.

What is the implication, and what should change now

- Improvements are needed in the reporting of whether and how autocorrelation is accounted for in the individual series, use of multiple effect measures (reflecting both immediate and gradual changes), and in the complete and accurate reporting of the analysis methods.

1. Introduction

Interrupted Time Series (ITS) are a type of nonrandomized design commonly used to evaluate public health policy interventions, and the impact of exposures (eg, natural disasters), at the population level; henceforth referred to as an "interruption." In an ITS, measurements on groups of individuals are collected repeatedly both before and after an interruption [1]. The period before the interruption can be used to estimate the underlying time trend, which when modeled correctly and projected into the postinterruption period, provides a counterfactual for what would have occurred in the absence of the interruption [2] (Fig. 1(A)). A range of effect measures can then be calculated to characterize both the short- and long-term effects of the interruption (eg, level-change and slope-change). Care is required in selecting statistical methods to analyze time series data that appropriately deal with features such as autocorrelation (i.e. the propensity for data points closer in time to be more similar than points further apart) and seasonality [1,3–5].

Meta-analysis can be used to estimate a pooled effect from separate ITS in the context of systematic reviews (henceforth referred to as a "review of studies") or "multiple series" within the same study (eg, regions within a country) (henceforth referred to as a "primary study with multiple series"). Two approaches for meta-analyzing results from ITS have been proposed [6]. These include the standard two-stage approach, where effect estimates from each ITS are first computed and then meta-analyzed (see Fig. 1(B) for an example), and the one-stage approach, where a single model including all series is fitted to simultaneously obtain the combined effect estimates [6].

The specific analytical decisions will be dictated by a range of factors such as: whether raw time series data are available (as would be the case in a "primary study with multiple series," or if the data could be extracted from graphs or tables of the included ITS studies in a "review of studies" [6–9]); the desired meta-analysis approach (one- or two-stage); and, whether there is a need to re-analyze the primary ITS studies. Re-analysis of ITS could be necessary when, for example, autocorrelation has not been accounted for in the original analysis (potentially leading to estimates of standard errors that are too small) [10–13], when the model structure in the original analysis is inappropriate [14], or when inconsistent effect measures have been reported across the studies [14–16]. These decisions can lead to different analytical pathways as depicted in Fig. 2(A).

In addition to the analytical decisions, an assessment of the strengths and weaknesses in the design and conduct of studies is an essential step for meta-analyses in the context of systematic reviews [17]. Weaknesses in a study may introduce bias, and the presence of such bias threatens the validity of the meta-analysis results [18]. Tools and criteria for assessing the risk of bias or methodological quality of ITS studies are less well established than for other study designs (eg, randomized trials).

To our knowledge, there have been no methodological reviews examining the approaches and methods used to meta-analyze effect estimates from ITS. In this review, we therefore aimed to: (1) investigate whether reviewers re-analyze the included ITS, and if so, what re-analysis methods are used; (2) what meta-analysis methods are used; (3) what effect measures are used, and how completely

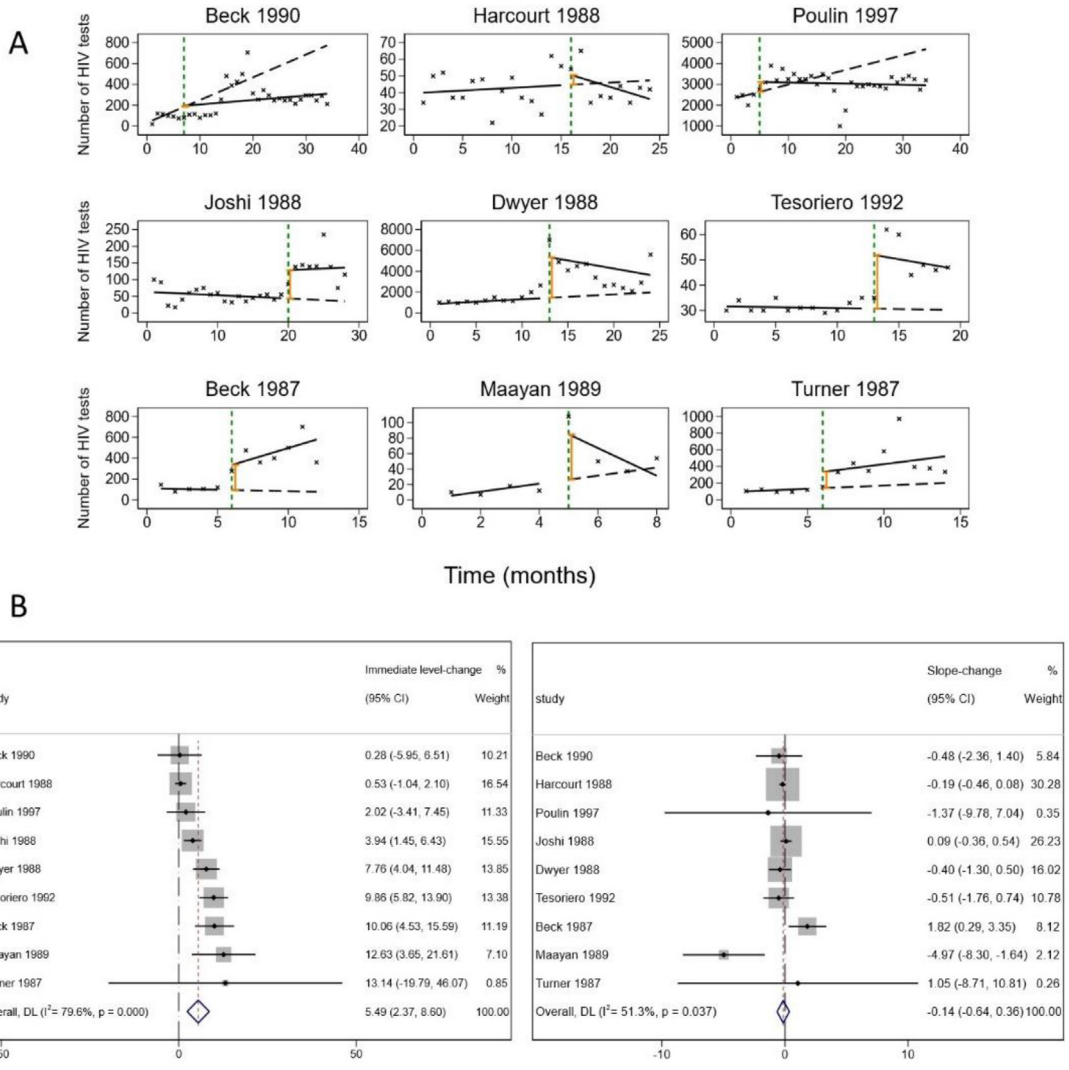

Fig. 1. (A) Nine interrupted time series (ITS) examining the effect of mass media campaigns on the ordering of HIV tests [19]. The crosses represent data points, the solid lines represent the pre- and post-interruption trend lines and the dashed line represents the counterfactual trend line. The green dashed line indicates the time of the interruption, while the capped orange lines indicate the estimated immediate level-change. (B) Forest plots depicting study-level and meta-analysis estimates of immediate level-change (left) and slope change (right).

the estimated meta-analyzed effects are reported; and (4) in "reviews of studies," what tools and domains are used to assess the risks of bias or methodological quality of the included ITS studies.

2. Review methods

The paper is structured as follows. In this section, we briefly describe the review methods, with further details available in the protocol [15] (deviations from our planned methods are presented in Appendix A). In Section 3, we present the results, structured according to Fig. 2(B), and finally, in Section 4, we present the key findings and implications for practice.

2.1. Eligibility criteria

The complete eligibility criteria are available in the protocol [15]. The inclusion criteria were:

1) a "review" that included at least two ITS which met the review authors' definition of an ITS design; and
2) included at least one meta-analysis of ITS.

Our definition of a "review" was very broad. It included "systematic reviews" and "reviews of selected studies" (both referred to as a "review of studies"), and primary studies that combined ITS across sites within the same study (ie, "primary study with multiple series"). We opted for broad inclusion since our primary interest was in the meta-analysis methods, which apply regardless of the particular study design. For ease of description, we use the term "review" in this paper to refer to both a "review of studies" and a "primary study with multiple series", although we recognize that the latter are not technically considered reviews.

In determining which included studies in a "review of studies" were ITS, we used the review authors' judgments when these were available, irrespective of the design

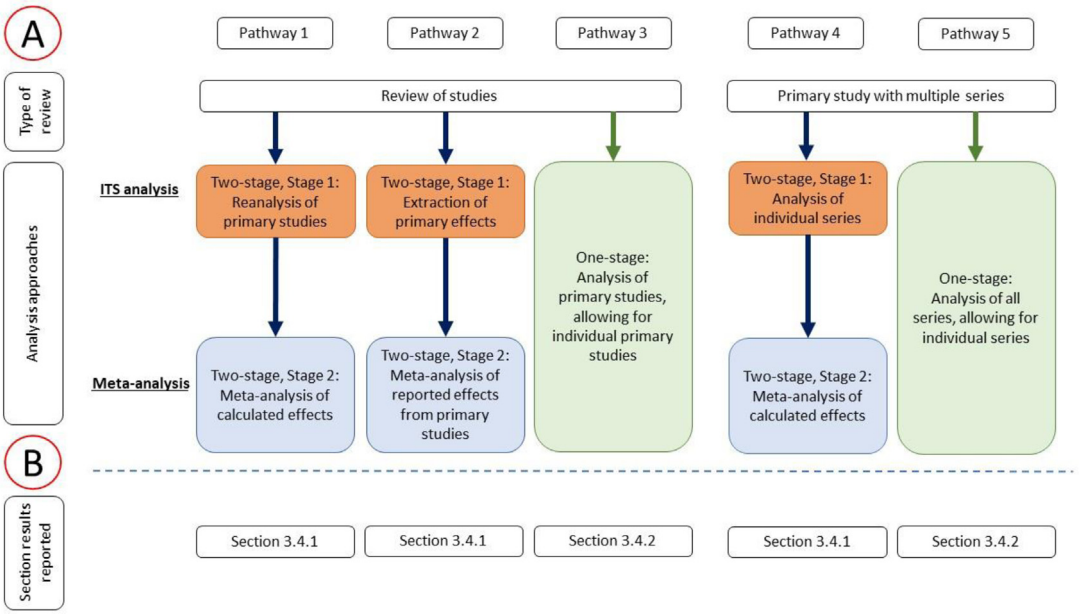

Fig. 2. (A) Depicts possible pathways arising from different analytical decisions. Green boxes represent the one-stage approach, orange and blue boxes represent the two-stage meta-analysis approach. Pathways 1 to 3 represent "reviews of studies". Pathways 4 and 5 represent "primary studies with multiple series." (B) Identifies the manuscript sections where the results for the pathways are reported.

features they had used to define an ITS (eg, number of timepoints pre and post). In reviews that used broader study labels, such as "quasi-experimental," we applied the following criteria to define a study as an ITS:

1) a clearly defined timepoint when the interruption occurred; and,
2) at least three datapoints before and after the interruption.

2.2. Literature search

Our search strategy was informed by previous publications that have reviewed ITS studies [11,12,20], and may be found in the supplementary material of the protocol [15]. Several databases were searched to capture the broad range of disciplines that use the ITS design. To capture reviews in health, we searched MEDLINE (Ovid), EMBASE (Ovid), Campbell Systematic Reviews, the Cochrane Database of Systematic Reviews (CDSR) and 3ie. To capture economic reviews, we searched EconLit (EBSCOhost) and for psychology and education disciplines, we searched PsycINFO (Ovid) and ERIC (ProQuest). The search was limited to the period January 1, 2000 to October 11, 2019 for all databases except CDSR which was limited to January 1, 2000 to August 9, 2019.

2.3. Study selection

Citations identified from the searches were imported into Endnote X8 [21] to remove duplicates, sorted by most recent first, and screened against the eligibility criteria. Abstracts and titles were screened independently and in duplicate by two members of the review team (E.K., A.K., S.L.T., A.B.F., and J.E.M.). Full-text articles of eligible abstracts were retrieved, sorted by most recent first and screened (independently and in duplicate by two members of the review team) until the target sample size of 100 reviews [15] was reached. Conflicts in screening decisions at both stages were resolved via discussion among the review team.

2.4. Comparison-outcome selection

Reviews may include several meta-analyses of ITS for *different* comparisons and outcomes (eg, the impact of two different policies to reduce drink-driving [eg, media campaign or random roadside breath testing] on injury and fatality rates). We examined the meta-analysis methods for only *one* comparison-outcome per review (for which there could be multiple meta-analyses of different effect measures) using the following hierarchy of rules:

The comparison-outcome with the largest number of effect measures (eg, the comparison-outcome with meta-analyses of level-change and slope-change estimates would be selected ahead of a comparison-outcome with only a meta-analysis of level-change estimates);

1) The comparison-outcome with the largest number of ITS; or
2) The comparison-outcome that was first reported in the abstract, then the methods section, then the results section of the manuscript.

Justification for this hierarchy is provided in the protocol [15]. The selected comparison-outcome was chosen by a single reviewer (E.K.); any uncertainty in the selection of the comparison-outcome was resolved through discussion with the review team.

2.5. Data extraction

Data extraction was undertaken using Research Electronic Data Capture [22,23], with the list of items available in Korevaar et al [15,24]. The review team independently piloted data extraction on 10 reviews to uncover ambiguity, identify missing items and test the logic of the form. As a result, several items were added, removed or clarified (see Appendix A). Data extraction was performed, independently, for 41% (22/54) of reviews by two or more authors (E.K. and A.K., S.L.T., A.B.F. or J.E.M.) and by one author (E.K.) for the remaining 59% (32/54) of reviews. Inconsistencies or uncertainties in data extraction were resolved via discussion between the data extractors or through consultation with the review team.

2.6. Analysis

We summarized the characteristics of included reviews with descriptive statistics using Stata 15.0 [25].

3. Results

3.1. Search results

The search identified 4,213 records, of which 339 were eligible for full text screening. Of these, 285 records were excluded, resulting in 54 included reviews (Fig. A1).

3.2. Characteristics of the included reviews

The 54 included reviews were published between 2005 and 2019 (Fig. A2), with publications steadily increasing from 2014. The majority of reviews explored public health interruptions (87%, 47/54), followed by crime (15%, 8/54) [Table 1]. The interruptions were predominantly targeted at the population level (63%, 34/54) [eg, state-wide legislation] or organizational level (28%, 15/54) [eg, police strategy]. The majority of included reviews were "reviews of studies" (63%, 34/54) [Fig. 2(A), pathways 1–3], while the remaining were "primary studies with multiple series" (37%, 20/54) [pathways 4–5].

3.3. Characteristics of included meta-analyses and their ITS

The median number of ITS for the selected comparison-outcomes in all reviews was 12 (interquartile range [IQR]: 6–26) [Table 2]. In "reviews of studies", the median number of studies was 5 (IQR: 3–7.5), with some studies contributing multiple time series: median number of series 7 (IQR: 6–10. A range of outcome types were analyzed, with the most common being proportion (31%; 17/54) and count (30%; 16/54) outcomes. The most common time interval was monthly (31%; 17/54); but in 46% (25/54) of reviews, the interval could not be determined. The average number of datapoints in a series per meta-analysis had a median of 48 (IQR: 23–157; N=31) across the reviews, with a similar number of datapoints pre- and post-interruption. The characteristics of the included meta-analyses varied depending on the type of review, with "primary studies with multiple series" including a larger number of series, more commonly modeling count outcomes, and including longer series, than "reviews of studies" (Table 2).

In "reviews of studies", 65% (22/34) provided a definition of ITS; half of which used the criteria defined by the Cochrane Effective Practice and Organisation of Care (EPOC) Group [26]). In 44% (15/34) of "reviews of studies," study designs other than ITS (eg, randomized trials) were included in the meta-analysis.

3.4. Analysis methods

Of the 54 reviews, 51 (94%) performed a two-stage meta-analysis (Fig. 2(A): pathways 1, 2, 4), and three performed a one-stage meta-analysis (pathways 3, 5). These approaches were implemented using a range of statistical packages (Table 1). In this section, we first report analysis details separated by approach (two-stage [Section 3.4.1] and one-stage [Section 3.4.2]), followed by additional analysis issues (approaches for incorporating control series [Section 3.4.3], investigation of heterogeneity and sensitivity analysis [Section 3.4.4]), and methods for assessing risk of bias in ITS and bias due to missing results (Section 3.4.5).

3.4.1. Two-stage meta-analysis

Of the 51 reviews using a two-stage meta-analysis, in 41 (80%) the primary studies were (re-)analyzed (where (re-)analysis includes both analysis in the "primary studies with multiple series" and re-analysis of ITS studies in "reviews of studies" to calculate an effect estimate for each study; Fig. 2(A) pathways 1, 4), in 8 (15%) the available effect estimates from the primary studies were used (Fig. 2(A) pathway 2), and in 2 (5%), we could not determine which approach had been used.

Stage 1: (Re-)analysis of ITS or extraction of effect estimates from primary studies

In reviews performing their own analysis of primary data (n = 41), 23 were "reviews of studies" (Fig. 2(A) pathway 1) and 18 were "primary studies with multiple series" (Fig. 2(A) pathway 4). In 61% (14/23) of "reviews of studies", a reason for re-analyzing the primary ITS studies was given; the most frequently reported reason was to analyze the study as an interrupted time series (79%, 11/14) [eg, where the primary study may have been analyzed as a before-after study] [Table 3]. In 46% (19/41) of reviews,

Table 1. Characteristics of the included reviews

		"Reviews of studies" (N = 34) n (%)	"Primary studies with multiple series" (N = 20) n (%)	All reviews (N = 54) n (%)
Discipline/Topic[I]				
	Public health	30 (88)	17 (85)	47 (87)
	Crime	4 (12)	4 (20)	8 (15)
	Economics	0 (0)	3 (15)	3 (6)
	War	0 (0)	1 (5)	1 (2)
	Psychology	0 (0)	1 (5)	1 (2)
Interruption target				
	Population	15 (44)	19 (95)	34 (63)
	Organisation	14 (41)	1 (5)	15 (28)
	Individual[II]	3 (9)	0 (0)	3 (6)
	Combination	2 (6)	0 (0)	2 (4)
Interruption types[I]				
	Policy change	21 (62)	14 (70)	35 (65)
	Practice change	13 (38)	0 (0)	13 (25)
	Communication (campaign)	6 (18)	2 (10)	8 (15)
	Exposure	2 (6)	6 (30)	8 (15)
	Educational method	6 (18)	0 (0)	6 (11)
	Policing strategy	3 (8)	0 (0)	3 (6)
	Clinical intervention	3 (9)	0 (0)	3 (6)
Reported analysis software				
	RevMan	11 (41)	0 (0)	11 (20)
	Stata	3 (9)	6 (30)	9 (17)
	R	2 (6)	3 (15)	5 (9)
	SAS	0 (0)	2 (10)	2 (4)
	SPSS	1 (3)	0 (0)	1 (2)
	Other	3 (9)	0 (0)	3 (6)
	Mix[III]	13 (38)	4 (20)	17 (31)
	Can't determine	1 (3)	5 (25)	6 (11)

[I] Multiple response options are possible – percentages sum to greater than 100%.
[II] Interruptions classified as 'individual-level interruption' were treatments directed at an individual (eg, treatment with beta-blockers), however, the measurements were still aggregated over units of time (eg, number of cardiac events each year).
[III] Reviews were classified as 'mix' if they used more than one software package for their analysis.

the model structure was reported. Most models included a level-change (93%, 38/41), 51% included both a level and slope-change (21/41), and for 37% (15/41) we could not determine the entire model structure. The most common statistical methods were linear regression (52%, 16/31), Poisson regression (32%, 10/41) and ARIMA (20%, 8/41). Autocorrelation and seasonality were adjusted for in 51% (21/41) and 41% (17/41) of reviews respectively, and not mentioned at all in 29% (12/41) and 44% (18/41) [Table 4]. Only one review adjusted for nonstationarity (2%).

In the majority of reviews that extracted available effect estimates from the primary studies (Fig. 2(A): pathway 5), details of acknowledgement or adjustment for autocorrelation and seasonality were reported (5/8, 62%), however none mentioned non-stationarity (Table 4).

Stage 2: Meta-analysis

Seven types of effect measures were meta-analyzed (Table 5). An immediate level-change was most common, in 84% of reviews (43/51), followed by long-term level-change (24%, 12/51) and slope-change (19%, 10/51) [Table 5, Table A1]. In 75% (38/51) of reviews, only one type of effect measure was used, which was commonly the immediate level-change (82%, 31/38). In reviews that reported multiple effect measures, the most common combination was an immediate level-change and slope-change (62%, 8/13). Standardization of the effect estimates prior to meta-analysis was reported in 31% of reviews (16/51) [Table 5]. The standardization procedures varied, including, for example, rescaling each series to the same unit of time (eg, converting from weekly to monthly datapoints)

Table 2. Characteristics of meta-analyses and of the included studies/series.

	"Reviews of studies" (N = 34) n (%) or median (IQR)	"Primary studies with multiple series" (N = 20) n (%) or median (IQR)	All reviews (N = 54) n (%) or median (IQR)
Number of ITS studies	5 (3.0 – 7.5) [I]	N/A	
Number of ITS series	7 (6 – 10) [I]	17 (7 – 26.5)	12 (6 – 26) [I]
Considered control series[II]			
Yes	9 (26)	8 (40)	17 (31)
No	21 (62)	11 (55)	32 (59)
Can't determine	4 (12)	1 (5)	5 (9)
Outcome types [III]			
Proportion	15 (44)	2 (10)	17 (31)
Count	6 (18)	10 (50)	16 (30)
Rate	7 (21)	2 (10)	9 (17)
Continuous	2 (6)	4 (20)	6 (11)
Combination [IV]	2 (6)	0	2 (4)
Other	1 (3)	2 (10)	3 (6)
Can't determine	1 (3)	1 (5)	2(4)
Time interval for datapoints			
Days	0	3 (15)	3 (6)
Weeks	0	3 (15)	3 (6)
Months	8 (24)	9 (45)	17 (31)
Years	4 (12)	2 (10)	6 (11)
Can't determine	22 (64)	3 (15)	25 (46)
Average[V] **number of datapoints per series** [VI]			
Overall	**(N = 16)** 24 (14 – 33)	**(N = 15)** 157 (48 – 312)	**(N = 31)** 48 (23 – 157)
Preinterruption	**(N = 14)** 12 (7 – 14)	**(N = 14)** 70 (24 – 133)	**(N = 28)** 23 (11 – 90)
Postinterruption	**(N = 14)** 13 (9 – 24)	**(N = 14)** 38 (24 – 87)	**(N = 28)** 24 (13 – 48)
Definition of ITS			
Yes	22 (65)	N/A	N/A
Cochrane Effective Practice and Organisation of Care criteria [VII] [26]	11 (50)		
ITS have a clearly defined time point	6 (27)		
The reviewers defined the number of datapoints pre- and post-interruption required by an ITS	11 (50)		
No	12 (35)	N/A	N/A
Combined with other designs			
Yes	15 (44)	N/A	N/A
No	19 (56)	N/A	N/A

ITS, Interrupted time series. IQR, interquartile range. N/A, not applicable.

[I] The number of included ITS studies could not be determined for two "reviews of studies," thus N = 32 for "reviews of studies" and N = 52 for all reviews.

[II] Reviews were categorized as having 'considered control series' if the controls were mentioned in the methods or the results section, regardless of whether the controls were used in a statistical comparison, a narrative comparison, or if the nature of use could not be determined.

[III] Multiple response options possible therefore percentages sum to greater than 100%.

[IV] Combination indicates where reviewers combined multiple data types (eg, combining studies using proportion and rate outcomes).

[V] The average number of datapoints were calculated across the series included in each review. These averages were then summarized across the reviews using the median and IQR.

[VI] Series lengths (overall, pre- and post-interruption lengths) were not available in every review.

[VII] Cochrane Effective Practice and Organisation of Care criteria[26] defines an ITS study as one which has i) a clearly defined time point when the interruption occurred, and ii) at least three datapoints before and after the interruption.

Table 3. Characteristics of the model structure and statistical methods in reviews where primary studies were (re-) analyzed.

	"Reviews of studies" (N = 23) n (%)	"Primary studies with multiple series"(N = 18) n (%)	All reviews (N = 41) n (%)
Reported a reason for (re-)analysis[I]			
Yes	14 (61)	N/A	N/A
To analyse as an interrupted time series	11 (79)		
To calculate the same effect measures across studies	5 (36)		
To adjust for autocorrelation	1 (7)		
Other	10 (71)		
No	9 (39)	N/A	N/A
Model characteristics			
Provided model structure			
Yes	7 (30)	12 (67)	19 (46)
No	16 (70)	6 (33)	22 (54)
Included parameters			
Level-change only	4 (17)	1 (6)	5 (12)
Level-change and slope-change	14 (61)	7 (39)	21 (51)
Included level, inclusion of slope could not be determined	3 (13)	9 (50)	12 (29)
Neither level nor slope could be determined	2 (9)	1 (6)	3 (7)
Statistical method			
ARIMA	2 (9)	6 (33)	8 (20)
Non-ARIMA	19 (83)	12 (67)	31 (76)
Linear regression	13 (68)	3 (25)	16 (52)
Without adjustment for autocorrelation	0 (0)	1 (33)	1 (6)
With adjustment for autocorrelation	8 (61)	0 (0)	8 (50)
Prais-Winsten	3 (38)	0 (0)	3 (38)
Newey-West	1 (13)	0 (0)	1 (13)
Cannot determine how autocorrelation was adjusted for	4 (50)	0 (0)	4 (50)
Cannot determine if autocorrelation was adjusted for	5 (38)	2 (66)	7 (44)
Poisson	1 (4)	9 (75)	10 (32)
Without adjustment for autocorrelation	0 (0)	1 (11)	1 (10)
With adjustment for autocorrelation	1 (100)	3 (33)	4 (40)
Included autoregressive lagged terms in their model	0 (0)	3 (100)	3 (75)
Cannot determine how autocorrelation was adjusted for	1 (100)	0 (0)	1 (25)
Cannot determine if autocorrelation was adjusted for	0 (0)	5 (56)	5 (50)
Negative-binomial without adjustment for autocorrelation	1 (4)	0 (0)	1 (3)
Simple statistical test (i.e. t/z test) [II]	4 (17)	0 (0)	4 (13)
Can't determine	2 (9)	0 (0)	2 (5)

ARIMA: autoregressive integrated moving average, ITS: Interrupted time series

[I] Denominator includes "reviews of studies" performing a two-stage meta-analysis and a re-analysis of the included ITS studies (n = 23). Multiple response options possible– percentages sum to greater than 100%.

[II] Effect estimate (eg, odds ratio, risk ratio) calculated from comparing summary statistics in the post- to pre-interruption period (eg, odds in the post-period to the odds in the preperiod)

Table 4. Acknowledgement and adjustment for autocorrelation, seasonality and nonstationarity.

	Method of obtaining first stage analysis data	
Analysis feature *mentioned* by reviewers [I]	**(Re-)analysis of ITS data** [II] **(N = 41) n (%)**	**Extraction of effect estimates from primary studies** [III] **(N = 8) n (%)**
Autocorrelation		
Yes	28 (68)	5 (62)
No	12 (29)	3 (38)
Can't determine	1 (2)	0 (0)
Seasonality		
Yes	23 (56)	5 (62)
No	18 (44)	3 (38)
Non-stationarity		
Yes	1 (2)	0 (0)
No	40 (98)	8 (100)
Analysis feature *adjusted for* by reviewers [IV]		
Autocorrelation	**(N = 41)**	N/A
Yes	21 (51)	
No	3 (7)	
Can't determine	4 (10)	
Seasonality	**(N = 41)**	N/A
Yes	17 (41)	
No	4 (10)	
Can't determine	2 (5)	
Non-stationarity	**(N = 41)**	N/A
Yes	1 (2)	

[I] Reviews were categorised as having 'mentioned' the analysis feature if the feature was explicitly mentioned but may or may not have been adjusted for (eg, "We assessed autocorrelation for each model by examining the plot of residuals and the partial autocorrelation function.").
[II] (Re-)analysis of ITS data reflects pathways 1 and 4 (Fig. 2(A)).
[III] Extraction of effect estimates from primary studies reflects pathway 2 (Fig. 2(A)).
[IV] Reviews categorised as having 'adjusted' for the analysis feature, a subset of reviews that 'mentioned' the analysis feature, if they explicitly stated that the feature was incorporated into the analysis (eg, "To control for autocorrelation, the first-order weekly lagged term of residuals was included in the model.").

or accounting for population denominators (eg, incidence per 100,000).

Of the 51 reviews using two-stage meta-analysis, 50 used pairwise meta-analysis and one used network meta-analysis. In 42 reviews (82%), meta-analysis of effect estimates was used (with one fitted in a Bayesian framework), in seven, other methods were used (most commonly calculating the median effect size and range (5/51, 10%)), and in two we could not determine the method. Of the 42 reviews using meta-analysis of effect estimates, the most common meta-analysis model was the random effects (36/42, 86%) [Table 5], with the fixed / common effect model used in 14% (6/42) of reviews. Half (27/51, 53%) provided a justification for their chosen meta-analysis model. Justifications included using thresholds of heterogeneity/inconsistency to decide between fixed effect or random effects (eg, $I^2>50\%$ would lead to use of random effects meta-analysis), "due to heterogeneity", or if studies varied by context and approach.

In a random effects meta-analysis, the most commonly used estimator of the heterogeneity variance (τ^2) was the DerSimonian and Laird method of moments estimator (19/35, 54%), but in 40% (14/35) of the reviews, the estimator could not be determined. The Wald type confidence interval method was used most commonly (18/35, 51%), and in 15/35 (43%) the method used to estimate the confidence interval could not be determined.

The calculated pooled effects were commonly reported in text (38/51, 75%), forest plots (34/51, 67%) and tables (25/51, 49%) [Table 5]. Other displays included bar graphs and box plots. From the 51 included reviews, 82 meta-analyses of effect measures were extracted, because one comparison-outcome from a review could contribute multiple meta-analyses of different effect measures (eg, one

Table 5. Effect measures meta-analysed, meta-analysis methods used, and reporting of justifications and heterogeneity

Effect measure [I]	"Reviews of studies" (N = 33) n (%)	"Primary studies with multiple series" (N = 18) n (%)	All reviews (N = 51) n (%)
Level			
Immediate level-change [II]	26 (79)	17 (94)	43 (84)
Long-term level-change [III,IV]	9 (27)	3 (17)	12 (24)
Slope			
Change	8 (24)	2 (11)	10 (19)
Pre-interruption slope	1 (3)	3 (17)	4 (8)
Post-interruption slope	1 (3)	2 (11)	3 (6)
Area Under the Curve (AUC)	1 (3)	0 (0)	1 (2)
Standardization [V]			
Yes	8 (24)	8 (44)	16 (31)
No	20 (61)	10 (56)	30 (59)
Can't determine	5 (15)	0 (0)	5 (10)
Meta-analysis method			
Meta-analysis of effect estimates	28 (85)	14 (78)	42 (82)
Random effects	25 (89)	11 (79)	36 (86)
Common effect / fixed effect	3 (11)	3 (21)	6 (14)
Other methods	5 (15)	2 (11)	7 (14)
Can't determine	0 (0)	2 (11)	2 (4)
Provided a justification for chosen meta-analysis model			
Yes	17 (52)	10 (56)	27 (53)
No	16 (48)	8 (44)	24 (47)
Heterogeneity variance estimator [VI]	**(N = 24)**	**(N = 11)**	**(N = 35)**
DerSimonian and Laird	17 (71)	2 (18)	19 (54)
REML	0 (0)	2 (18)	2 (6)
Can't determine	7 (29)	7 (64)	14 (40)
Confidence interval method [VI]	**(N = 24)**	**(N = 11)**	**(N=35)**
Wald type	17 (71)	1 (9)	18 (51)
Hartung-Knapp	1 (4)	1 (9)	2 (6)
Can't determine	6 (25)	9 (82)	15 (43)
Method of reporting the effect estimates [I]	**(N = 33)**	**(N = 18)**	**(N = 51)**
Text	24 (73)	14 (78)	38 (75)
Forest plot	23 (70)	11 (61)	34 (67)
Tables	16 (48)	9 (50)	25 (49)
Other	4 (12)	2 (11)	6 (12)
Effect estimate confidence interval reported [VII]	**(N = 55)**	**(N = 27)**	**(N = 82)**
Yes	51 (93)	25 (93)	76 (93)
No	4 (7)	2 (7)	6 (7)
Report an estimate of heterogeneity/inconsistency (eg, τ^2, I^2) [VII]	**(N = 55)**	**(N = 27)**	**(N = 82)**
Yes	32 (58)	17 (63)	49 (60)
No	23 (42)	10 (37)	33 (40)

All reviews conducting a two-stage meta-analysis (N = 51).

[I] Multiple response options are possible – percentages sum to greater than 100%.

[II] Immediate level-change includes meta-analyses of level-changes with before-after analysis and meta-analyses of level-changes calculated at the time of the interruption with segmented regression analysis.

[III] Effect measures were categorized as long-term level-changes where the effect of the interruption was evaluated at some time after the interruption (ie, using a combination of the level-change and slope-change parameters).

[IV] Two "reviews of studies" included more than one meta-analysis of a long-term level-change (ie, one included three meta-analyses of long-term level changes, the other included eight), giving a total of 82 meta-analyses from the 51 outcome-comparison assessments.

[V] Standardization includes, for example, rescaling each series to the same unit of time (eg, converting from weekly datapoints to monthly datapoints), dividing by the standard error, or accounting for population denominators (eg, incidence per 100,000).

[VI] Denominator includes reviews conducting random effects meta-analysis and excludes the one review conducting a Bayesian random effects meta-analysis (N = 35).

[VII] Denominator includes all meta-analyses extracted from the 51 reviews (N = 82).

of level-change and one of slope-change). For these 82 meta-analyses, confidence intervals for the pooled effect were almost always reported (93%, 76/82), however only for 60% of meta-analyses was a measure of heterogeneity reported (49/82) [Table 5].

3.4.2. One-stage meta-analysis

In the three reviews that used one-stage meta-analysis, two were "primary studies with multiple series" (Fig. 2(A): pathway 1), and one was a "review of studies" (pathway 3). The reviews used different regression models (linear, negative binomial, logistic). All reviews included a parameter to capture the immediate effect of the interruption (ie, level-change); none included a slope-change parameter. Two reviews included a fixed effect for each series (to allow the intercept to vary across sites/studies), while one review used a random intercept for each series. One review mentioned and adjusted for autocorrelation in their analysis, and one other mentioned but did not adjust for autocorrelation. None of the three reviews mentioned seasonality or nonstationarity.

3.4.3. Incorporation of control series

Of the included reviews, 31% considered control series in their analysis (17/54), more than half did not (32/54, 59%) and for five reviews we could not determine whether control series were used (9%; Table 2). Of those reviews that considered controls (for at least some of their included series), the most common approach was to use a unified model (7/17, 41%; ie, a one-stage approach or to incorporate the control series in the first step of a two-stage meta-analysis [eg, calculating an effect estimate adjusted for any difference observed in a control series]).

3.4.4. Meta-regression, subgroup, and sensitivity analysis

In 57% (31/54) of reviews, an investigation of heterogeneity was conducted using meta-regression (22%, 12/54) or subgroup analysis (28/54, 52%); in most reviews, these investigations were noted as planned (Table 6). A sensitivity analysis was conducted in 56% of reviews (30/54). Examples included examination of whether the results were robust to the fitted ITS model (eg, fixing the slopes across series vs. allowing these to vary) and to the meta-analysis method (eg, fixed effect vs. random effects meta-analysis). Reported reasons for failing to conduct the planned sensitivity analyses included few studies, "heterogeneity of included studies" or "short nature of the time series."

3.4.5. Methods for assessing risk of bias and bias due to missing results

In "reviews of studies" (Fig. 2(A): pathways 1–3), nearly all assessed the risk of bias or methodological quality of the included ITS (33/34, 97%; Table A2). The most commonly used tool was the Cochrane EPOC tool [26], which was used in 55% (19/34) reviews. The EPOC tool addresses the domains: intervention independent of other changes, prespecification of the intervention shape (including defined time point for the intervention and shape of the intervention effect), likelihood of the intervention affecting data collection, knowledge of the allocated interventions adequately prevented during the study, incomplete outcome data, selective outcome reporting, and other factors (including, for example, whether seasonality may explain observed effects) [27].

In 65% (22/34) of reviews, methods used to assess the presence of reporting bias were planned or undertaken. The most commonly used methods included funnel plots (10/13, 77%) and tests for funnel plot asymmetry (8/13, 62%). The most common reason reported for not completing the planned assessment for reporting bias was that the number of included effect estimates in the meta-analysis were too few.

4. Discussion

Our methodological review provides insights into: the characteristics of 54 reviews that include ITS, their series and their meta-analyses; the statistical and review methods; and the completeness of reporting. We found that while ITS may be included to address particular review questions (eg, where randomization is not possible, such as when examining the effects of war), there was a large degree of diversity in the statistical methods used to (re-)analyze individual series, meta-analyze results across series, and in how completely the methods and results were reported. We now consider these aspects.

4.1. Analysis of ITS

In many systematic reviews, reviewers are restricted to the aggregate-level summary statistics or effect estimates reported in the study publications. In ITS studies, graphs of time series data are commonly presented [16], providing an opportunity for reviewers to extract the data and undertake a re-analysis. Re-analyses can be valuable for overcoming limitations in the original analysis (eg, not allowing for autocorrelation) and to allow for a consistent effect measure to be calculated for each study for meta-analysis [15,16]. In our review, we found that re-analysis was common in "reviews of studies," with the most commonly noted reason being to analyze the data as interrupted time series, when in the original study, this may not have been the case. In the majority of re-analyses, the fitted models included variables representing functions of time, with the most common parameterization including variables for level- and slope-change (ie, segmented regression model); although, we could not determine the model structure in more than a third of reviews.

A notable percentage of reviews did not report using a statistical method that adjusted for autocorrelation, or we could not determine if there was adjustment. Furthermore, in reviews where the effects were extracted from

Table 6. The number and percentage of reviews that planned and/or conducted meta-regression, subgroup and sensitivity analyses.

	"Reviews of studies" (N = 34)n (%)	"Primary studies with multiple series" (N = 20) n (%)	All reviews (N = 54) n (%)
Meta-regression analysis			
Planned[l]	5 (15)	9 (45)	14 (26)
Conducted	3 (9)	9 (45)	12 (22)
Not conducted	2 (6)	0 (0)	2 (4)
Not planned	29 (85)	11 (55)	40 (74)
Conducted	0 (0)	0 (0)	0 (0)
Not conducted	29 (85)	11 (55)	40 (74)
Subgroup analysis			
Planned[l]	22 (65)	13 (65)	35 (65)
Conducted	14 (41)	12 (60)	26 (48)
Not conducted	8 (24)	1 (5)	9 (17)
Not planned	12 (35)	7 (35)	19 (35)
Conducted	1 (3)	1 (5)	2 (4)
Not conducted	11 (32)	6 (30)	17 (31)
Sensitivity analysis			
Planned[l]	16 (47)	16 (80)	32 (59)
Conducted	11 (32)	16 (80)	27 (50)
Not conducted	5 (15)	0 (0)	5 (9)
Not planned	18 (53)	4 (20)	22 (41)
Conducted	3 (9)	0 (0)	3 (6)
Not conducted	15 (44)	4 (20)	19 (35)

[l] A meta-regression, subgroup or sensitivity analysis was considered to be 'planned' if the authors explicitly stated the analysis in their methods section, or indicated that they had intended to conduct the analysis.

the primary studies (Fig. 2(A): pathway 2), the issue of autocorrelation was mentioned in less than two thirds of the reviews. Failing to account for autocorrelation in the analysis of ITS has been previously identified in methodological reviews of ITS studies, which have identified that autocorrelation was not considered in between 34% and 45% of included ITS [9,11,12,28,29]. The implications of not accounting for autocorrelation may be amplified in the context of meta-analysis where multiple unadjusted series may contribute to the meta-analytic effect having a confidence interval that is too narrow. However, the impact is likely to depend on the length of the series of the included ITS. This is because statistical methods that attempt to adjust for autocorrelation have been shown to outperform ordinary least squares for series with more than 12 data points [5]. Furthermore, the impact may differ by review type, given the series lengths of the included ITS were generally longer in "primary studies with multiple series" than in "reviews of studies" (potentially due to administrative databases being used more frequently in the former review type).

4.2. Meta-analysis

When undertaking a two-stage meta-analysis of ITS, decisions must be made regarding the choice of effect measure(s). For ITS, multiple effect measures are often available, which characterize the effects of the interruption differently. We found that most of the included reviews using two-stage meta-analysis combined the immediate level-change, and for the majority, this was the only effect measure used (Table 3). While the immediate level-change is important for understanding the interruption's immediate impact, presented in isolation, it provides an incomplete picture. For example, the same immediate level-change could be observed in two ITS, but with very different long-term effects (as illustrated in Fig. A3), and thus with different consequences for policy decisions. Therefore, using multiple effect measures, or those that integrate level and slope changes (eg, long-term level-change), is recommended for providing a more complete understanding.

Reviews using two-stage meta-analysis, most commonly fit the random effects model. This was perhaps to be expected, particularly in "reviews of studies," where reviewers might expect diversity in characteristics (eg, populations) across studies, potentially leading to statistical heterogeneity. Consistent with other studies examining random effects methods, the DerSimonian and Laird heterogeneity variance estimator was dominant, along with the Wald-type confidence interval for the pooled effect [30–32]. In the presence of a small number of studies, as we observed in the "reviews of studies," these methods

often yield confidence interval coverage less than the nominal 95% [33–35].

4.3. Reporting

Complete reporting of characteristics of the primary ITS studies, the statistical methods used to (re-)analyze the ITS and meta-analyze their results, are critical for reproducibility [36,37], and enabling the assessment of whether appropriate and consistent analysis methods have been used. While the performance of statistical methods for analyzing individual ITS is importantly dependent on the series length [5,38], in more than half of the reviews the length of the included series could not be determined, aligning with previous research [9].

4.4. Strengths and limitations

This review follows a pre-specified protocol, detailing the systematic review methods [15]. The search of a range of disciplines enabled retrieval of reviews from health, economics, psychology and education. However, our search strategy has not been validated, so the sensitivity of the strategy is unknown. In particular, our search may have missed identifying studies that undertake a one-stage analysis, where the terms "pooled" or "meta-analysis" were not noted in the abstract. This may, in part, explain the small number of reviews of such studies.

We extracted information from the reviews, but not the primary publications of included ITS. Therefore, our ability to assess the methods used was hindered by incomplete reporting in the reviews. This may have led to an incomplete catalogue of the methods being used (particularly for the (re-)analysis of ITS), as well as under- or over-estimation of the percentage of reviews using particular methods.

4.5. What this review adds to what is already known

Methodological reviews examining and providing guidance for meta-analysis methods have focused on reviews including RCTs and observational study designs, but we are unaware of any that have specifically focused on reviews including ITS [7,30,39–42]. Therefore, our review provides novel insight into the meta-analysis methods, effect measures, and completeness of reporting for this design, as well as other review methods (eg, risk of bias).

Furthermore, our review provides characteristics of the included ITS, and the methods review authors have used to (re-)analyze them. There have been a number of methodological reviews examining the characteristics and analysis methods of primary ITS studies [9,11,12,28,29]. The characteristics of the ITS and methods of analysis in the present review were generally consistent with the findings from these other methodological reviews, including the series length, time interval between datapoints, use of linear regression as the most common statistical method for ITS analysis, model structure and acknowledgement of autocorrelation.

4.6. Implications of this research

For those conducting and reporting ITS studies, complete and accurate reporting of the analysis methods (including details of any adjustments [eg, for autocorrelation, seasonality], model structure, estimation method) and results (ie, clearly describing the effect measure, effect estimate and measure of precision) is critical, not only for readers to be able to fully interpret the findings, but also for the study findings to be incorporated into meta-analyses. Furthermore, provision of the time series data is recommended, as this provides greater opportunity for reviewers to include the study in a meta-analysis, for example, by applying consistent ITS analysis methods across the included series.

Similarly, for those undertaking reviews of ITS, complete and accurate reporting of the analysis methods (at both the ITS and meta-analysis levels) and results is necessary for readers to fully interpret the findings and assess their validity. In reviews that meta-analyze the reported effects from the primary studies, reporting the analysis methods used in the primary studies is important, as is noting when they cannot be determined. Furthermore, we recommend that reviewers use multiple effect measures, or effect measures that integrate changes in level and slope.

5. Conclusions

Reviews examining the effects of interruptions targeted at populations may meta-analyze results from interrupted time series. Our review provides novel insight into the statistical and review methods, and reporting in such reviews. We found a large diversity in the statistical methods used to (re-)analyze individual series and meta-analyze results across series, and in how completely the methods and results were reported. More attention should be given to adjusting for autocorrelation in the individual series, use of multiple effect measures, and in the complete and accurate reporting of the analysis methods.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2022.01.010.

References

[1] Wagner AK, Soumerai SB, Zhang F, et al. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002;27:299–309 2002/08/14.

[2] Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin; 2002.
[3] Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2017;46:348–55 2016/06/11. doi:10.1093/ije/dyw098.
[4] Bernal JL, Cummins S, Gasparrini A. Corrigendum to: Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2020;49:1414 2020/09/04. doi:10.1093/ije/dyaa118.
[5] Turner SL, Forbes AB, Karahalios A, et al. Evaluation of statistical methods used in the analysis of interrupted time series studies: a simulation study. BMC Med Res Methodol 2021;21:181 2021/08/30. doi:10.1186/s12874-021-01364-0.
[6] McKenzie JE, Beller EM, Forbes AB. Introduction to systematic reviews and meta-analysis. Respirology 2016;21:626–37 2016/04/22. doi:10.1111/resp.12783.
[7] Deeks J, Higgins J, Altman D, et al. Chapter 10: Analysing data and undertaking meta-analyses Cochrane Handbook for Systematic Reviews of Interventions. Higgins J, Thomas J, Chandler J, et al., editors. Cochrane; 2019.
[8] Gebski V, Ellingson K, Edwards J, et al. Modelling interrupted time series to evaluate prevention and control of infection in healthcare. Epidemiol Infect 2012;140:2131–41 2012/02/18. doi:10.1017/S0950268812000179.
[9] Turner SL, Karahalios A, Forbes AB, et al. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review. J Clin Epidemiol 2020;122:1–11 2020/02/29. doi:10.1016/j.jclinepi.2020.02.006.
[10] Ramsay CR, Matowe L, Grilli R, et al. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. Int J Technol Assess Health Care 2003;19:613–23 2004/04/21.
[11] Jandoc R, Burden AM, Mamdani M, et al. 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.
[12] Hudson J, Fielding S, Ramsay CR. Methodology and reporting characteristics of studies using interrupted time series design in healthcare. BMC Med Res Methodol 2019;19:137 2019/07/06. doi:10.1186/s12874-019-0777-x.
[13] Polus S, Pieper D, Burns J, 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;91:56–69 2017/07/29. doi:10.1016/j.jclinepi.2017.07.008.
[14] Taljaard M, McKenzie JE, Ramsay CR, et al. The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. Implement Sci 2014;9:77 2014/06/20. doi:10.1186/1748-5908-9-77.
[15] Korevaar E, Karahalios A, Forbes AB, et al. Methods used to meta-analyse results from interrupted time series studies: A methodological systematic review protocol. F1000Res 2020;9:110 2020/12/24. doi:10.12688/f1000research.22226.3.
[16] Turner SL, Karahalios A, Forbes AB, et al. Creating effective interrupted time series graphs: Review and recommendations. Res Synth Methods 2021;12:106–17 2020/07/14. doi:10.1002/jrsm.1435.
[17] Last J, EA I. A Dictionary of Epidemiology. New York: Oxford University Press; 2001.
[18] Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 2019/08/30. doi:10.1136/bmj.l4898.
[19] Vidanapathirana J, Abramson MJ, Forbes A, et al. Mass media interventions for promoting HIV testing. Cochrane Database Systemat Rev 2005:Cd004775 2005/07/22.. doi:10.1002/14651858.CD004775.pub2.
[20] Turner SL, Karahalios A, Forbes AB, 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:e024096 2019/01/31. doi:10.1136/bmjopen-2018-024096.
[21] The EndNote Team EndNote. EndNote X9 ed. Philadelphia, PA: Clarivate; 2013.
[22] Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208 2019/05/13. doi:10.1016/j.jbi.2019.103208.
[23] Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–81 2008/10/22. doi:10.1016/j.jbi.2008.08.010.
[24] Korevaar E, Karahalios A, Forbes A, et al. Methods used to meta-analyse results from interrupted time series studies: A methodological systematic review protocol - Appendix 2 Data extraction form (version 1). 2020.
[25] StataCorp. Stata statistical software: release 15. Tx: College Station StataCorp LLC; 2017.
[26] Cochrane Effective Practice and Organisation of Care (EPOC). What study designs can be considered for inclusion in an EPOC review and what should they be called? 2017. http://epoc.cochrane.org/resources/epoc-resources-review-authors accessed 09 February 2021.
[27] Cochrane Effective Practice and Organisation of Care (EPOC). Suggested risk of bias criteria for EPOC reviews 2017. http://epoc.cochrane.org/resources/epoc-resources-review-authors accessed 21 December 2021.
[28] Hategeka C, Ruton H, Karamouzian M, et al. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: a methodological systematic review. BMJ Glob Health 2020:5 2020/10/16. doi:10.1136/bmjgh-2020-003567.
[29] Ewusie JE, Soobiah C, Blondal E, et al. Methods, Applications and Challenges in the Analysis of Interrupted Time Series Data: A Scoping Review. J Multidiscip Healthc 2020;13:411–23 2020/06/05. doi:10.2147/JMDH.S241085.
[30] Page MJ, Altman DG, McKenzie JE, et al. Flaws in the application and interpretation of statistical analyses in systematic reviews of therapeutic interventions were common: a cross-sectional analysis. J Clin Epidemiol 2018;95:7–18 2017/12/06. doi:10.1016/j.jclinepi.2017.11.022.
[31] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88 1986/09/01. doi:10.1016/0197-2456(86)90046-2.
[32] Collaboration C. Review Manager(RevMan)[Computer Program] Version 5.2. 3. Copenhagen: The Nordic Cochrane Centre; 2012. Health Psychol Rev 2014;17.
[33] Veroniki AA, Jackson D, Bender R, et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. Res Synth Methods 2019;10:23–43 2018/08/22. doi:10.1002/jrsm.1319.
[34] Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods 2016;7:55–79 2015/09/04. doi:10.1002/jrsm.1164.
[35] Cornell JE, Mulrow CD, Localio R, et al. Random-effects meta-analysis of inconsistent effects: a time for change. Ann Intern Med 2014;160:267–70 2014/04/15. doi:10.7326/M13-2886.
[36] Page MJ, Altman DG, Shamseer L, et al. Reproducible research practices are underused in systematic reviews of biomedical interventions. J Clin Epidemiol 2018;94:8–18 2017/11/09. doi:10.1016/j.jclinepi.2017.10.017.

[37] Lopez-Nicolas R, Lopez-Lopez JA, Rubio-Aparicio M, et al. A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000-2020). Behav Res Methods 2021 2021/06/27. doi:10.3758/s13428-021-01644-z.

[38] Nelson BK. Statistical methodology: V. Time series analysis using autoregressive integrated moving average (ARIMA) models. Acad Emerg Med 1998;5:739–44. doi:10.1111/j.1553-2712.1998.tb02493.x.

[39] Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. Syst Rev 2017;6:263 2017/12/21. doi:10.1186/s13643-017-0663-8.

[40] McKenzie JE, Herbison GP, Deeks JJ. Impact of analysing continuous outcomes using final values, change scores and analysis of covariance on the performance of meta-analytic methods: a simulation study. Res Synth Methods 2016;7:371–86 2015/12/31. doi:10.1002/jrsm.1196.

[41] Higgins J, Eldridge S, Li T, et al. Chapter 23: Including variants on randomized trials Cochrane Handbook for Systematic Reviews of Interventions. Higgins J, Thomas J, Chandler J, et al., editors. Cochrane; 2021.

[42] Reeves B, Deeks J, Higgins J, et al. Chapter 24: Including non-randomized studies on intervention effects Cochrane Handbook for Systematic Reviews of Interventions. Higgins J, Thomas J, Chandler J, et al., editors. Cochrane; 2021.