

Methods, Applications and Challenges in the Analysis of Interrupted Time Series Data: A Scoping Review

This article was published in the following Dove Press journal:
Journal of Multidisciplinary Healthcare

Joycelyne E Ewusie¹
Charlene Soobiah^{2,3}
Erik Blondal^{2,3}
Joseph Beyene¹
Lehana Thabane^{1,4}
Jemila S Hamid^{1,5}

¹Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada; ²Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, Ontario, Canada; ³Institute of Health Policy Management and Evaluation (IHPME), University of Toronto, Toronto, ON, Canada; ⁴Biostatistics Unit, Father Sean O'Sullivan Research Centre, St Joseph's Healthcare, Hamilton, Ontario, Canada; ⁵Clinical Research Unit, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada

Objective: Interrupted time series (ITS) designs are robust quasi-experimental designs commonly used to evaluate the impact of interventions and programs implemented in healthcare settings. This scoping review aims to 1) identify and summarize existing methods used in the analysis of ITS studies conducted in health research, 2) elucidate their strengths and limitations, 3) describe their applications in health research and 4) identify any methodological gaps and challenges.

Design: Scoping review.

Data Sources: Searches were conducted in MEDLINE, JSTOR, PUBMED, EMBASE, CINAHL, Web of Science and the Cochrane Library from inception until September 2017.

Study Selection: Studies in health research involving ITS methods or reporting on the application of ITS designs.

Data Extraction: Screening of studies was completed independently and in duplicate by two reviewers. One reviewer extracted the data from relevant studies in consultations with a second reviewer. Results of the review were presented with respect to methodological and application areas, and data were summarized using descriptive statistics.

Results: A total of 1389 articles were included, of which 98.27% (N=1365) were application papers. Segmented linear regression was the most commonly used method (26%, N=360). A small percentage (1.73%, N=24) were methods papers, of which 11 described either the development of novel methods or improvement of existing methods, 7 adapted methods from other areas of statistics, while 6 provided comparative assessment of conventional ITS methods.

Conclusion: A significantly increasing trend in ITS use over time is observed, where its application in health research almost tripled within the last decade. Several statistical methods are available for analyzing ITS data. Researchers should consider the types of data and validate the required assumptions for the various methods. There is a significant methodological gap in ITS analysis involving aggregated data, where analyses involving such data did not account for heterogeneity across patients and hospital settings.

Keywords: interrupted time series, segmented linear regression, ARIMA, limitations, methods, scoping review

Introduction

Quasi-experimental designs (QEDs) refer to non-randomized designs that are used to evaluate the effect of interventions and programs.¹ Interrupted Time Series (ITS) design is considered the strongest among QEDs and is a powerful tool used for evaluating the impact of interventions and programs implemented in healthcare settings.^{2,3} With this design, outcomes are measured at different time points before

Correspondence: Jemila S Hamid
Children's Hospital of Eastern Ontario,
401 Smyth Road, Ottawa, ON K1H 8L1,
Canada
Tel +1 (613) 737-7600 x 4194
Email jhamid@uottawa.ca

and after implementing an intervention, allowing the change in level and trend of outcomes to be compared, to evaluate intervention effects.

ITS designs are applied in a wide range of applications and healthcare settings. ITS designs are commonly used to evaluate quality improvement initiatives and infection control programs in hospitals.^{3–8} A recently published systematic review also shows that ITS designs are being increasingly used in drug utilization research.⁹ With the increasing research focus on knowledge translation (KT) and evidence-based medicine (EBM), and the growing importance of the uptake of best evidence into clinical practice, ITS analysis has become a common tool used in assessing the effect of clinical practice guidelines and recommendations.^{10–14} For instance, ITS design was used to look at de-adoption of tight glycemic control in critically ill patients across 113 intensive care units (ICUs) in the USA following publication of a randomized trial showing that tight glycemic control can increase mortality.¹⁵ In another study, ITS design was used in evaluating the impact of Otolaryngology Head and Neck Surgery (OHNS) guidelines on perioperative care process and patient outcomes in children undergoing tonsillectomy.¹⁶ In recent years, ITS studies have also been included in Cochrane systematic reviews done by the Effective Practice and Organization of Care Review group,¹⁷ indicating their usefulness in studies of organizational and practice change interventions.

ITS designs are generally highly regarded for their rigor (compared to the traditional before and after studies) and are arguably considered the optimal approach in evaluating the impact of hospital-wide interventions and new policies implemented nationwide. ITS designs allow us to statistically test potential biases such as, autocorrelation, seasonality, non-stationarity, heteroskedasticity, history, maturation and random fluctuations.¹⁸ Moreover, to increase the internal validity of a study, ITS designs can be modified to include a nonequivalent dependent variable or a control outcome unaffected by the intervention to control for possible concurrent events.¹⁹ A control series which has not been exposed to the intervention can also be added to the ITS design to increase the internal validity of the study.²⁰

There are several statistical methods that can be used in analyzing data from an ITS design. The decision on what method to use is often based on several factors including type of outcome (eg, continuous, binary or count), distribution (eg, Gaussian, Skewed),²¹ assumptions such as autocorrelation or seasonality,^{22,23} the number of groups/sites included

in the ITS design (single site vs multi-site analysis),²⁴ or the inclusion of a control group.²⁵ Despite the availability of various statistical methods for ITS analysis, limitations exist in practical applications, where researchers frequently ignore checking the assumptions and the various factors influencing the optimality of the different methods.²⁶ As such, some ITS data are currently being analyzed using inadequate methods, and hence leading to inaccurate results, erroneous or misleading conclusions, and potentially affecting patient care.

Therefore, in order to reduce bias, increase precision and enhance statistical power it is imperative that researchers use appropriate methods for their analysis.¹⁸ There is also a need for the different statistical methods for ITS analysis to be identified and compared to inform researchers of the available methods and inform future research regarding the strengths and limitations of the methods as well as identify methodological gaps which will pave the way for improvements in ITS design and analysis.

In this study, we conducted a scoping review with the aim of 1) identifying and systematically summarizing available methods used in the analysis of ITS studies, 2) elucidating the strengths and limitations of existing methods, 3) identifying potential methodological gaps and 4) providing an extensive review of the applications of ITS designs and analysis in health research.

Methods

We performed the scoping review using the methods outlined by Arksey and O'Malley and the Joanna Briggs Handbook for conducting systematic scoping reviews.^{27,28} Scoping review methods were used, since our aim was to identify the methods that have been used to analyze ITS data, provide an overview of their strengths and limitations and evaluate the frequency of their application without an in depth assessment of the application papers with respect to their clinical content, risk of bias as well as quality of the design and analysis.²⁷ We developed a protocol for this scoping review based on the PRISMA-P guidelines.²⁹ The protocol is published in a peer-reviewed journal²⁶ and a brief description of our scoping review methods is outlined here.

Search Strategy

We searched electronic databases, MEDLINE, JSTOR, PUBMED, EMBASE, CINAHL, Web of Science and the Cochrane Library from inception until September 2017 for relevant articles and conference abstracts. An experienced

information specialist worked with JEE and JSH to further refine the search strategy presented in the study protocol. We also contacted methodological experts in the field of ITS for any difficult-to-locate and unpublished materials. Additionally, we scanned the references of included articles for other potentially relevant articles. We restricted our search to studies that were reported in English.

Eligibility Criteria

Our eligibility criteria were based on the recommendations by the Effective Practice and Organization of Care (EPOC) Cochrane Group.¹⁷ All health-related studies that reported on the development, adaptation, comparison or application of ITS design or methods were included. Studies with at least 3 time points before and after the intervention and had a clearly defined time point or period within which the intervention was implemented were included. Furthermore, studies were included if the outcome was objectively measured. We excluded ITS studies that had less than 3 time points per period to be consistent with the recommendations found in the literature.¹⁷

Study Selection and Data Collection

The search results and potentially relevant full-text articles were screened independently by two reviewers and in duplicate (JEE, CS and EB). To ensure reliability, a calibration exercise was performed prior to screening of titles and abstracts as well as full-text articles. Data extraction was completed by JEE in consultations with JSH. All conflicts that arose were resolved by discussion or consultation through a third reviewer. Since this is a scoping review, we did not assess the risk of bias of included studies.^{27,28}

Data Extraction and Synthesis

Data were extracted from all included studies. The studies were classified as methodological papers if they reported on the development, adaptation, important additions to common statistical methods, description or comparison of statistical methods for ITS design. Studies were classified as application papers if they used an ITS design or analysis for assessing intervention effect. Data extracted for methodological papers included; article characteristics (eg, title, author, year of publication), type of method, description of method, type of outcome the method is developed for, assumptions involved as well as strengths and limitations of the method as described by the authors. For application papers, the data extracted include; article characteristics, field of application (eg, clinical, pharmaceutical, etc.), setting (eg, single site, multi-site), statistical

method used as reported by the authors, number of time points, and the assumptions (eg, presence of autocorrelation) checked or tested.

Results from the review were summarized in two categories: methodological articles and articles on applications of ITS analysis. For methodological papers, we used narrative review to describe the methodological processes or analysis strategies. We also discussed the methods with respect to their strengths and weaknesses, the type of outcome assessed, the assumptions made, the type of software used and the limitations, if stated. Data from application papers were summarized using frequencies and percentages and we also provided tables and figures to depict the distribution.

Results

A total of 9278 articles were returned from the database search, of which 6890 were identified as unique articles. After the initial screening of titles and abstracts, 2262 articles were found to be relevant. After the full-text screening of these articles, 1389 articles were included for the review. A detailed illustration of the study flow is provided in [Figure 1](#).

Study Characteristics

Of the 1389 papers, about 2% (N=24) were classified as methods papers, and the majority (98%, N=1365) were classified as application papers. An overall description of papers included in the scoping review, using broad categories, is provided in [Table 1](#). As shown in the table, the methods papers were presented in three categories, with most of the articles presenting development of novel methods (46%, N=11). For application papers, ITS design has been used most frequently in clinical research (N=621, 46%) and in population and public health research (N=437, 32%). It is also applied frequently in multiple baseline or multiple site studies (N=392, 29%).

Review of Statistical Methods for ITS Analysis

Among the 11 methods papers proposing novel statistical methods for ITS analysis, a considerable percentage (55%, N=6) were developed for either controlled ITS designs^{24,30,31} or multiple baseline ITS designs,^{6,32,33} where multiple baseline indicates that the intervention is introduced in different localities (eg, sites, units) at different times. A significant percentage (25%, N=6) of the methods papers provided

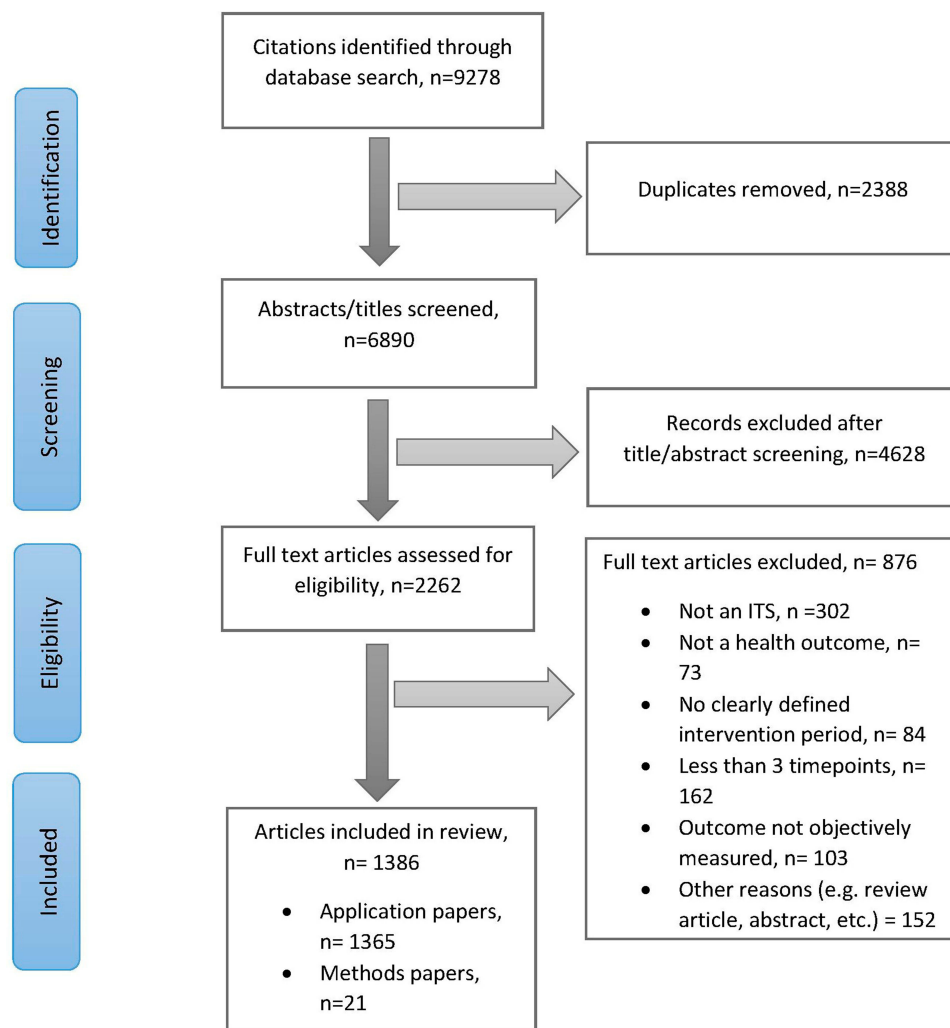


Figure 1 Flow chart outlining the search and review process, the records identified, included and excluded as well as the reasons for exclusion.

a comparative analysis of existing ITS methods where empirical analysis and/or simulated data were used to compare performance of the different methods.^{21,34-38}

Papers on Novel ITS Methods

A detailed summary of the 11 articles with respect to the descriptions of methods' process, assumptions made, statistical software used, strengths and limitations have been provided (see [Table S1](#), which describes the methods). The methods include the extended time series model introduced by Sun et al (2012) which allows the identification of immediate and gradual changes in an outcome of interest. This time series model comprises a stochastic component and a structural or intervention component. The stochastic component is modelled using ARIMA and the intervention component is divided into 4 subcomponents to detect: 1) pre-intervention marginal change in

outcome, 2) post-intervention marginal change in outcome, 3) short-term change in outcome and 4) additional impact of intervention over the observed period.³⁹ Their method also allows the assessment of different post-intervention time points.

However, it is not possible to assess the effect of concomitant interventions that may affect the outcome.

Duncan & Duncan (2004) applied the latent growth curve modelling approach to pooled interrupted time-series data. In their approach, they added a growth curve model that captures both the intercept and slope differences over the baseline and intervention periods. The model gives linear and average level change for both periods. This model can be applied to designs with multiple units and different time points, time spacing and growth functions.⁴⁰ It is indicated that their method can be used for a short time series (>2 time points per period)

Table 1 Description of Studies Included in the Review with Respect to Methods Used in the Analysis of ITS Data

Characteristic	Number of Studies Included in the Review, N, (%) *N= 1389
Methods papers	24 (1.73)
Novel methods	11 (45.8)
Method adaptation and important contribution	7 (29.2)
Method comparison	6 (25.0)
Application papers	1365 (98.27)
Field of Application	
Clinical	621 (45.5)
Pharmaceutical	238 (17.4)
Guideline implementation	69 (5.1)
Public health/policy	437 (32.0)
Setting/Design	
Single site	353 (25.9)
Multiple baseline/multi-site	392 (28.7)
Controlled ITS	237 (17.4)
National (population study)	383 (28.1)
Statistical Methods Used in Articles	
Segmented Regression	
Segmented regression using linear models	360 (26.4)
Segmented regression using GLM, GEE or GAM**	261 (19.1)
Segmented regression using ARIMA	268 (19.6)
Non-segmented regression	110 (19.6)
Non-regression methods eg t-test	82 (6.0)
Difference in differences	17 (1.2)
Unspecified	267 (19.6)
Type of Outcome of Interest	
Continuous	131 (9.6)
Count	1029 (75.4)
Binary	205 (15.0)
Number of Time Points	
Less than 16 (or < 8 per period)	141 (10.3)
At least 16 (or ≥ 8 per period)	634 (46.5)
At least 50	590 (43.2)
Autocorrelation Checked	
Yes	812 (59.5)
Other Biases Checked	
Yes	607 (44.5)
Specific Biases***	
Seasonality	407 (67.1)
Non-stationarity	290 (47.8)
Heteroskedasticity	123 (20.3)
Confounding	203 (33.4)
Clustering	68 (11.2)
Presentation of ITS Results	
Figures	414 (30.3)

(Continued)

Table 1 (Continued).

Characteristic	Number of Studies Included in the Review, N, (%) *N= 1389
Tables	105 (7.7)
Both	804 (58.9)
None	42 (3.1)

Notes: *All percentages are out of the total number of corresponding papers. **ITS- interrupted time series; GLM – Generalized Linear Models; GAM – Generalized Additive Models; GEE – Generalized Estimating Equation. ***The frequencies and percentages are out of the total of 607.

and allows the evaluation of treatment effects in the presence of different covariates. However, they also mentioned that their method is not optimum when the time series is lengthened since the linear growth form may not adequately define the series.

The robust interrupted time series method, using a two-stage approach, was proposed by Cruz et al (2017). The first stage involves identifying the change point and the second stage involves performing formal tests for differences in the correlation structure and variability between pre- and post-intervention.⁴¹ For this method, the time of intervention is assumed to not necessarily be the same as the time at which the effect of intervention initiate (ie, the change point). The authors demonstrated that their method performed better than traditional segmented regression with regards to mean squared error.

Kong et al (2012) developed an extended logistic regression method for count and proportion data. The authors showed how the incidence rate or proportion of an outcome of interest can be estimated while accounting for seasonality and serial correlation. The method incorporates harmonic functions to account for seasonality and a first-order autoregressive model to account for serial autocorrelation.⁴² The authors also described the different estimation procedures for estimating the parameters and their associated variances. The authors indicated that their method performed better than the conventional segmented linear regression model.

For studies involving multiple baseline ITS designs, Gbeski et al (2012) introduced the pooled and stacked ITS method of data analysis, where the data are from several units within one site (eg, hospital) or across multiple sites. The data are thus aggregated across the units (or sites) to obtain the overall intervention effect.⁶ For the pooled analysis, they

fitted a separate segmented regression model for each unit and then calculated the weighted average of the individual estimates of the parameters where the weights were the inverse variances of the estimates obtained from the individual fitted models. The stacked analysis approach involved fitting a single segmented regression model but accounting for the unit effect by incorporating a variable to represent the units using one unit as reference. Although their methods account for unit effect and heterogeneity among units, having large number of units may lead to over dispersion or an increase in Type 1 error due to the increase in number of required parameters.

Huitema et al (2014) implemented an extended time series regression model which involved a within unit and between unit analysis. For the within unit analysis which is performed for designs with multiple units and interventions that are rolled out over time and at different time points per unit, they incorporated a continuous variable which they called a penetration variable to evaluate the extent of penetration of the intervention. The penetration variable was calculated as the ratio of number of units with intervention to overall number of units, and it ranged from 0 to 1.³² Their method is appropriate for interventions that are not introduced in full after baseline. Moreover, the authors indicated that if an effect is identified, the function estimated from the regression can be used to predict the outcome from the degree of intervention penetration. This method, however, does not account for unit effect and heterogeneity across units.

Similarly, Velicer (1994) introduced the pooled time series analysis approach to analyze multiple baseline time-series data, where data are combined from different units and comparison between the units is allowed. The method involved using a patterned transformation matrix and a design matrix. The transformation matrix transforms serially dependent variables to independent variables. The design matrix is defined based on the parameters of interest (level and trend) and differences between units.³³ The author stated that the pooled time series method has the advantage of being easily implemented in existing computer programs with little modification; it can also be adapted to other modelling approaches. For instance, instead of the general transformation matrix used in the proposed method, the ARIMA (1,0,0) transformation matrix can be used for most cases, and thus the method can be implemented in R or SAS. This method, however, requires substantial understanding of transformation, which is a barrier to uptake of the proposed method. The

method does not also account for heterogeneity across units.

The propensity score-based weighted interrupted time series analysis was introduced by Linden & Adams (2011) for controlled series. The method allows the pre-intervention characteristics (level and slope) for the treatment and control groups to be comparable. This method uses standard regression techniques, and hence easy to implement.²⁴ However, the control groups must have significant overlap with treatment group, in terms of basic characteristics, for the weighting to be effective. Fretheim et al (2005) presented a method which is similar to the difference-in-differences approach. As with all controlled ITS studies, this analysis method allows for the detection of anomalous effects and co-interventions.³⁰ The authors indicated that this method requires more than 6 time points per period for reliable estimation of the regression coefficients.

The state-space method was introduced by Pechlivanoglou et al (2015) for controlled interrupted time series. The method comprises an observation equation and a state equation. The state equation contains the level and trend parameters that are affected by the intervention and allowed to vary over time.³¹ The main advantage of this method is the ability to capture effects of co-interventions by the addition of other variables. This method is, however, unable to test the assumption of comparability of the control group and thus overall results might be suboptimal if groups are not comparable. Park (2012) implemented the intervened time series central mean subspace, which is a non-parametric approach to analyzing ITS data. The method is an extension of the central mean subspace in time series to a nonparametric intervention analysis. The dynamic reduction technique is used to analyze the time series using interventions as a covariate.⁴³ The authors showed that the method is a viable alternative to ARIMA and does not require model specification. The authors stated that a large number of observations is required to ensure model accuracy, however, they did not explicitly specify how large this is required to be.

Papers on Improved or Adapted Methods

A total of 7 methods papers (29%) presented a description of existing statistical methods, used in other areas, that are adapted to ITS analysis^{3,5,44,45} or presented important contributions to the adapted methods.⁴⁶⁻⁴⁸ These articles include a publication by Gillings et al (1981) who described the implementation of ITS design and the application of segmented linear regression (SLR) methodology in health services research to assess mortality trends,

following the implementation of a regionalized perinatal care program.⁴⁴ The authors fitted both SLR and a single non-linear regression to the data; and they showed that SLR explains a greater proportion of variation in the data compared to fitting a single non-linear regression. Moreover, it is much easier to interpret. They also concluded that the SLR is relevant for analyzing ITS data, when the errors are independent.

Wagner et al (2002) presented the SLR approach, which is also known in statistical literature as piece-wise regression. The authors described how this method can be applied to evaluate the effect of a policy or educational intervention implemented to improve quality of medication use.³ The authors highlighted various factors that should be considered, when using SLR. These include detecting and controlling as well as testing for autocorrelation, seasonality as well as identifying outliers and their effects on the results. The authors indicated that SLR model is a robust modelling technique that allows the estimation of dynamic changes in outcomes following interventions intended to change the use of medication.

Yau et al (2004) described the extension of the zero-inflated Poisson (ZIP) regression model to handle time series of count data with excess zeros in application to occupational health.⁴⁵ In their paper, the authors discussed how the method enables the evaluation of an occupational intervention using population-level aggregated count data containing extra zeros. The authors stated the assumptions under which the ZIP model can be used, for instance, the existence of a perfect state (where there are zero counts because events are nearly impossible) and imperfect state (where there are zero counts although events are likely to occur without being avoidable). They also proposed the ZIP mixed autoregression model which is an extension of the ZIP model to account for the dependency between serially observed counts.

Studies presenting important contributions to existing methods include Zhang et al (2009). The authors considered SLR estimators for absolute and relative changes in outcomes and provided methods for obtaining confidence intervals (CIs) of the estimators.⁴⁸ Similarly, Zhang et al (2011) considered design aspects of SLR and provided simulation-based power calculations. They indicated that SLR models with 24 or more time points have more than 80% power to detect an effect size of 1 or greater, depending on the degree of autocorrelation and number of parameters to be estimated.⁴⁷ Huitema & McKean (2007) proposed a new method for dealing with autocorrelation

in ITS analysis, in the context of SLR. Their method is a form of portmanteau test that identifies autocorrelated errors generated from processes of higher-order autoregressive models. Unlike the conventional portmanteau tests, their method provides more satisfactory small sample properties and better inferential properties by accounting for the biases associated with the autocorrelation estimator and the error variance estimators.⁴⁶

Papers on Methodological Comparisons

Six of the studies focused on comparative evaluation of statistical methods used in the analysis of ITS studies.^{21,34-38}

Shardell et al (2007) performed empirical evaluation to compare the 2 group tests, SLR analysis and time series (ARIMA and ARMA) analysis in terms of the characteristics, assumptions, strengths and limitations of the methods. The authors used data from a study conducted with the objective of evaluating the impact of a hospital-based intervention to reduce antimicrobial infection rates and overall length of stay. The conclusions drawn from this study indicated that all three methods can be effectively used for ITS data analysis.²¹ For statistical validity, however, researchers must consider the research question, the data requirements and modeling assumptions to obtain high quality and unbiased results. The authors also provided guidance on examining the requirements for each of the methods.

Nunes et al (2011) compared two statistical methods, seasonal autoregressive integrated moving average (ARIMA) and cyclical regression models, used to analyze ITS data on mortality. Based on their results, the seasonal ARIMA was found to produce non-autocorrelated residuals in the weeks where there were excess deaths due to influenza epidemic. The seasonal ARIMA model had a lower residual mean square than the cyclical regression model. These results led to a lower 95% confidence limit for the estimates obtained from the seasonal ARIMA model, making it more efficient.³⁸

Hagiwara et al (2016) compared SLR analysis and statistical process control (SPC) for evaluating the longitudinal effects of quality improvement interventions. SPC method has similar data requirements as the SLR and uses a run chart which is a graphical display of data to plot the data. With the SPC method, however, mathematical calculations are not needed to interpret the data. The authors discussed the implementation processes of both methods, their differences, strengths and limitations. Based on their analysis of the empirical data, the authors concluded that SLR analysis was more statistically robust than SPC analysis in comparing

the effectiveness of different interventions, while the SPC was more appropriate for controlling a process since it is relatively easy to conduct and interpret.³⁴ Kontopantelis et al (2015) compared different regression modeling approaches for ITS analysis, based on the level of complexity in their implementation, which they labeled as basic, advanced and expert levels. The authors provided a description of the method processes, the assumptions made, and the technical details of each modeling technique using secondary data obtained from a quality and outcomes framework performance research.³⁷ In the same way, Burke et al (2016) compared three methods; SLR, ARIMA, and standardized incidence ratio (SIR) and compared the methods' ability to determine whether Amerithrax influenced patient utilization. SIR method is usually used for evaluating disease clusters by comparing observed number of cases to expected number of cases to determine if the incidence of the disease in the identified population is similar to that observed in the standard population.³⁵ The results obtained based on the three methods were discussed and their differences were reported. The authors, however, stated that, they were not able to successfully evaluate the impact of the intervention using the three methods due to the limited level of detail in the timeframes.³⁵ They therefore concluded that granularity of timeframes is as important as number of data points in a time series analysis.

Review of the Application Studies Involving ITS Design and Analysis

A large percentage (98%, N=1365) of the studies in our review involves application papers, where studies reported use of ITS design and/or analysis. The most common application area identified was clinical research (46%, N=621) followed by public health or health policy applications (32%, N=437) and pharmaceutical research (17%, N= 238) (Table 1). In clinical research, ITS designs are mostly utilized in studies involving interventions to reduce anti-microbial resistance and in other quality improvement studies (40%, N=246). Common public health and policy research areas, where ITS designs are used, include assessing the effect of mass media campaign programs and national or regional regulations such as smoking ban and traffic laws (27%, N=120). Pharmaceutical research areas, where ITS designs, are frequently used include studies on drug dose evaluation and drug regulatory approval.

Our review reveals that there is an overall increasing trend in ITS use (Figure 2). Most of the application papers

using ITS are published in recent years, where 197 studies using ITS analysis were published in 2016 alone compared to only 69 in 2011 (approximately 185% point increase in the last 5 years); and just 27 in 2006 indicating more than 600% point increase in publications within the last decade.

Evaluation of the application papers with respect to the methods utilized in the statistical analysis shows that most of the studies (65%, N=889) reported segmented regression (SR) techniques. Specifically, SR using ordinary linear regression modelling (SLR) was the most commonly applied method (26%, N=360), followed by ARIMA modeling (20%, N=268).

SR using other modelling approaches such as generalized linear model (eg, Poisson and Binomial models) was implemented in 19% (N=261) of the studies. Although most of the studies used segmented regression approaches (eg, SLR) to analyze data, there was considerable shortfall in the application of these methods. For instance, among the studies that used ARIMA models, approximately 23% (N= 62) did not meet its at least 50 time points requirement for adequate analysis.⁴⁹ Similarly, 10% (N=65) of the studies that used SLR had less than 8 time points per period. Of the studies that reported using SR modelling approaches, other than ARIMA, 66% (N=407) stated that they tested and where necessary accounted for the presence of autocorrelation in their data.

Overall, about 60% (N=812) of the included studies checked or accounted for autocorrelation while approximately 45% (N=607) checked or accounted for other biases, such as non-stationarity (48%, N=290), heteroskedasticity (20%, N=123), confounding variables such as age and sex (30%, N=203), clustering effect (11%, N=68) and seasonality (67%, N=407) when suitable. Furthermore, a considerable percentage (21%, N=132) of the studies inappropriately used SLR or ARIMA to analyze ITS data with count outcome summarized as frequency. This was often done without checking for linearity or validating assumptions such as the normality assumption. For considerable percentage of the application studies (48%, N=658), the outcome of interest was summarized as rate.

A sizeable number of studies (6%, N=82), that implemented an ITS design, inappropriately used statistical methods that do not account for time. The methods used include ANOVA, t-tests and Chi-square tests. Approximately 20% (N=267) of the studies included in our review reported using ITS analysis without specifying the modelling approach they implemented in performing analysis.

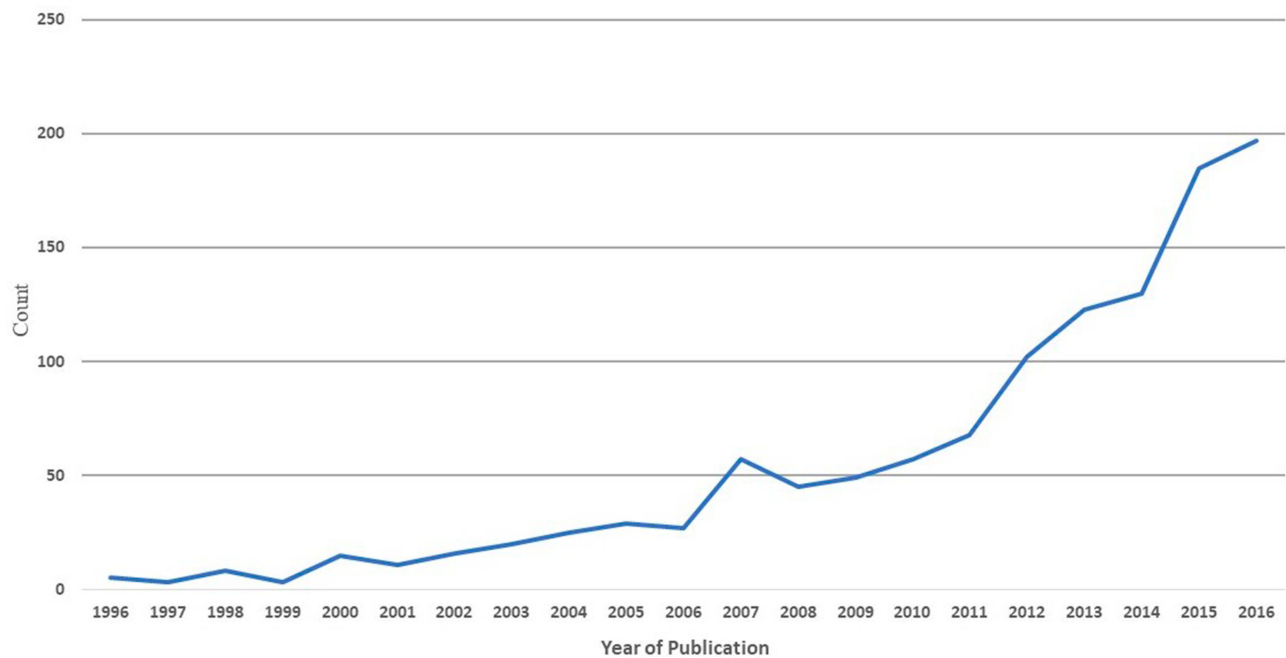


Figure 2 Trend of interrupted time series application papers over the last two decades.

Substantial number of the studies included in our review implemented a multiple baseline (multi-site) design (29%, N=392) and a considerable number (17%, N=237) implemented controlled ITS designs. For studies that implemented controlled ITS designs, a very small percentage (7%, N=16) reported using methods such as difference in differences in analyzing their data, which is a method arguably considered as non-ITS. For studies that reported the statistical software used (70%, N=958), a considerable number performed their analysis using either SAS (37%, N=353) or STATA (36%, N=348) statistical software. PROC AUTOREG in SAS is the most commonly reported software package (15%, N=52) among the SAS users.

Discussion

Our review demonstrated the increasing trend in ITS use in health research, where publications using ITS designs and analysis have almost tripled within the last decade. This increase might be attributed to advancements in the field of implementation science in recent years. Our findings also show that one of the factors contributing to this increasing trend is the use of routinely collected administrative data to answer important healthcare-related questions. These administrative data provide relatively inexpensive options compared to using resource-intensive prospective data. ITS designs are also being increasingly used in evidence synthesis and uptake,

where ITS studies are used to evaluate the impact of evidence-based recommendations and policies, clinical practice guidelines (CPGs) as well as publication of other important evidence. With the growing knowledge and awareness of health ethics combined with other advances in patient care delivery, feasibility and ethical considerations might be contributing factors to the increasing use of ITS designs.

Our results show that ITS designs are being implemented mostly in clinical research and public health and policy applications. The observed increase in ITS applications is consistent with findings in a recent systematic review focused on ITS use in drug utilization research.⁹ Our review also revealed that most of the studies used either a multiple baseline design or a controlled ITS design. A few of the studies also used a control outcome or non-equivalent dependent variable. These design adaptations have been recommended in literature to increase the internal validity of ITS designs by controlling for threats such as time-varying confounders.^{3,19} Moreover, even without a control group, ITS designs can address important biases such as history and maturation by having multiple observations before and after the intervention.³ We identified several methods that have been utilized in ITS data analysis. The most common methods were SR analysis and ARIMA. We also identified some new and improved methods that have been developed in recent years such as the robust ITS, the extended ITS and the pooled ITS analysis.

This scoping review highlights some major limitations in the implementation of ITS designs and their analysis. First, a considerable number of studies analyzed their data inappropriately. For instance, despite having sufficient time points per phase (study period), some studies used methods such as ANOVA and t-tests when performing analyses. These methods do not account for any underlying secular trend or autocorrelations often present in the data, thus making such methods suboptimal.¹⁸ Similarly, although the recommendation by the Effective Practice and Organization of Care (EPOC) Cochrane Group¹⁷ suggests that ITS designs should have at least 3 time points per period for inclusion in their reviews, methodological literature examined in this scoping review show that a minimum of 8 time points per period (or 50 time points for ARIMA models) is required to gain sufficient power in estimating the regression coefficients.² Nevertheless, our findings revealed that some of the studies, despite fewer time points, used time series regression techniques such as ARIMA to analyze their data thus making their results underpowered. Second, for studies that used appropriate time series regression approaches, over 40% of them did not test or account for potential biases that might be present in the data due to, for instance, autocorrelation, seasonality, or heteroskedasticity. The effects of correlated errors or serial dependency in time series data have been highlighted in the literature.^{18,22,23,50} Testing or accounting for such potential biases in time series data is imperative to ensure that the standard errors are not biased, and that the significance of the intervention is not overestimated.

There are some methodological gaps in current and frequently used ITS methods that were identified in our review. Although a few of the gaps have been acknowledged in previous studies,^{6,18,48} the limited methodological advances in the area remain a problem. One major issue, for instance, is the use of SR to analyze aggregated data.³ For almost all identified ITS studies, data at a given time point are often aggregated or summarized across different patients or participants. Further, with the multiple baseline (or multi-site) designs, the final data are pooled across the sites or units, resulting in another level of data aggregation. Thus, for both single and multi-site designs, data at each time point are estimated and hence associated with an imprecision due to variability of patient outcomes or variability across sites at a given time point. Nevertheless, our results show that a substantial number of these studies used SLR or ARIMA models, which do not account for this imprecision and hence may lead to biased (aggregation

bias) or suboptimal results. The method proposed by Gebiski and colleagues,⁶ accounts for the heterogeneity across sites using meta-analytic approaches. However, the method does not account for imprecision introduced due to aggregation across patients within the same site. Moreover, pooling the intercept and slope estimates lead to loss of information (and hence loss of power), since summarized data rather than individual patient data are used. Similarly, other methods that have been developed for multi-site studies do not account for the heterogeneity across the patients or across the different sites/units.^{32,33} Hence, there is a need for methods that account for the variability at both the patient and site levels.

Another issue highlighted in our review is the lack of guidance in design aspect of ITS studies and lack of clarity on the adequate sample size per time point. Although a sample size of 100 at each time point is stated as desirable for acceptable level of variability of estimates,³ we found no information in the literature concerning the adequate sample size required at each time point to provide optimal results in analyzing ITS data. Since sample size calculation is essential in estimating effect sizes, it will be imperative for future research to consider studies that will help determine the minimum sample size required at each time point for adequate evaluation of intervention effects.

Yet another issue is the slow uptake and implementation of the methods that have been developed in recent years. While only a few recently developed methods were identified in our review (N=11), we noticed that none of the recently developed methods have been applied by other researchers in the analysis of ITS data, even if they had similar data structures. This issue is, however, not unique to ITS methods; the problem of dissemination and uptake of new statistical methods into health research has been discussed extensively in the literature.⁵¹ The reasons for the limited uptake of new methods include: 1) the lack of statistical expertise to implement the method and understand the findings especially for methods such as extended time series model³² and pooled time series model³³ that require substantial mathematical or statistical knowledge 2) lack of software packages to implement the method and 3) lack of awareness of the existence of the method. This knowledge translation gap emphasizes the objective of this scoping review of available methods used in ITS data analysis.

Based on these observed shortfalls in the analysis of ITS data and the gaps present in current methods used in ITS analysis, we make the following recommendations: 1) future

researchers and reviewers of ITS studies must carefully consider the assumptions and requirements of the various statistical methods to ensure that conclusions about intervention effects are not spurious, 2) future research comparing available methods should be based on simulation studies to assess the bias, mean square error, and power. This will enable readers to confidently use methods based on their preferences, 3) ITS method developers should provide user-friendly software packages as was done by Cruz et al (2017), to ensure effective and efficient uptake of their new methods in biomedical research, and 4) there is the need for the provision of some form of guideline to help researchers and novice users decide on the appropriate use of ITS methods with respect to both design and analysis. Currently, two articles have provided a checklist for researchers and reviewers of ITS designs.^{9,18} In addition, reporting guidelines will also be helpful in ensuring consistency in ITS articles. These guidelines will have the potential to reduce heterogeneity as well as standardize analysis, interpretation and reporting of ITS studies. This standardization will not only enhance the methodological rigor, but also facilitate the more appropriate assessment of study quality, evidence gathering and evidence synthesis through meta-analyses of ITS studies.

In our study, we considered specific databases and studies published in English to limit the scope of our research, thus we may not have retrieved all studies particularly those on the application of ITS designs. We acknowledge that this may be a possible limitation of our study, however, we believe that our sample is representative of ITS methods utilized in biomedical research since we have captured all the methods papers that meet our eligibility criteria. Furthermore, the application papers were needed to describe the overall trend in ITS use, thus the number of papers included in this study is sufficient to provide a reliable estimate and achieve our aim. Moreover, we expect the magnitude of increase in ITS to be higher than what is reported in this paper if non-English studies and applications outside health research are included in our review. This further underscores the importance of ITS methods. Finally, we acknowledge the lack of detail in the presentation of the application papers as another limitation of this review. However, we believe that the level of detail presented is enough to answer our objective. Nonetheless, we highlight that the studies found in our literature search can be used as bases for future systematic review with an in-depth evaluation of application articles.

The findings from this study accentuate the need for improved methods for design and analysis of ITS studies such as when the data are aggregated per time point. This

review serves as a first step towards developing standard guidelines for ITS studies and for filling the methodological gaps identified in current literature.

Data Sharing Statement

All data generated or analyzed during this study are included in this published article [and its [supplementary Table S1](#)].

Acknowledgments

The authors thank Mr. Andrew Colgoni for helping to develop the search strategy for the literature search and Dr Monica Taaljad for providing additional data. The authors would also like to thank the reviewers for their detailed and careful review of our paper and providing suggestions that have significantly improved the presentation of our paper.

Disclosure

The authors report no funding and no conflicts of interest for this work.

References

1. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Cengage learning; 2002.
2. Penfold RB, Fang Z. *Use of Interrupted Time Series Analysis in Evaluating Health Care Quality Improvements*. Academic Pediatrics; 2013:S38–S44.
3. 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(4):299–309. doi:10.1046/j.1365-2710.2002.00430.x
4. Harris AD, McGregor JC, Perencevich EN, et al. The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inf Assoc*. 2006;13(1):16–23. doi:10.1197/jamia.M1749
5. 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(1):77. doi:10.1186/1748-5908-9-77
6. 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(12):2131–2141. doi:10.1017/S0950268812000179
7. Kastner M, Sawka AM, Hamid J, et al. A knowledge translation tool improved osteoporosis disease management in primary care: an interrupted time series analysis. *Implement Sci*. 2014;9(1):109. doi:10.1186/s13012-014-0109-9
8. Liu B, Moore JE, Almaawiy U, et al. Outcomes of Mobilisation of Vulnerable Elders in Ontario (MOVE ON): a multisite interrupted time series evaluation of an implementation intervention to increase patient mobilisation. *Age Ageing*. 2017;47(1):112–119. doi:10.1093/ageing/afx128
9. 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(8):950–956. doi:10.1016/j.jclinepi.2014.12.018

10. Cortoos P-J, Gilissen C, Mol PGM, et al. Empirical management of community-acquired pneumonia: impact of concurrent A/H1N1 influenza pandemic on guideline implementation. *J Antimicrob Chemother.* 2011;66(12):2864–2871. doi:10.1093/jac/dkr366
11. Dayer MJ, Jones S, Prendergast B, et al. Incidence of infective endocarditis in England, 2000–13: a secular trend, interrupted time-series analysis. *Lancet.* 2015;385(9974):1219–1228. doi:10.1016/S0140-6736(14)62007-9
12. Dowell D, Tian LH, Stover JA, et al. Changes in fluoroquinolone use for gonorrhoea following publication of revised treatment guidelines. *Am J Public Health.* 2012;102(1):148–155. doi:10.2105/AJPH.2011.300283
13. Etchepare F, Pambrun E, Verdoux H, et al. Trends in patterns of antidepressant use in older general population between 2006 and 2012 following publication of practice guidelines. *Int J Geriatr Psychiatry.* 2017;32(8):849–859. doi:10.1002/gps.4536
14. Matowe L, Ramsay CR, Grimshaw JM, et al. Effects of mailed dissemination of the Royal College of Radiologists' guidelines on general practitioner referrals for radiography: a time series analysis. *Clin Radiol.* 2002;57(7):575–578. doi:10.1053/crad.2001.0894
15. Niven DJ, Rubenfeld GD, Kramer AA, et al. Effect of published scientific evidence on glycemic control in adult intensive care units. *JAMA Intern Med.* 2015;175(5):801–809. doi:10.1001/jamainternmed.2015.0157
16. Mahant S, Hall M, Ishman SL, et al. Association of national guidelines with tonsillectomy perioperative care and outcomes. *Pediatrics.* 2015;136(1):53–60. doi:10.1542/peds.2015-0127
17. Bero L, Grilli R, Grimshaw JM, et al. Cochrane Effective Professional and Organisation of Care Group. Cochrane Collaboration. *Cochrane Lib*;2002:2.
18. 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(4):613–623. doi:10.1017/S0266462303000576
19. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol.* 2017;46(1):348–355. doi:10.1093/ije/dyw098
20. St. Clair T, Hallberg K, Cook TD. The validity and precision of the comparative interrupted time-series design: three within-study comparisons. *J Educ Behav Stat.* 2016;41(3):269–299. doi:10.3102/1076998616636854
21. Shardell M, Shardell M, Harris AD, et al. Statistical analysis and application of quasi experiments to antimicrobial resistance intervention studies. *Clin Infect Dis.* 2007;45(7):901–907. doi:10.1086/521255
22. Carroll N. Application of segmented regression analysis to the Kaiser Permanente Colorado critical drug interaction program. In: Proceedings of the Western Users of SAS Software 2008 Conference; 2008. Available from: <https://www.lexjansen.com/wuss/2008/an/an08.pdf>. Accessed May 4, 2020.
23. Nelson BK. Statistical methodology: v. Time series analysis using autoregressive integrated moving average (ARIMA) models. *Acad Emerg Med.* 1998;5(7):739–744. doi:10.1111/j.1553-2712.1998.tb02493.x
24. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. *J Eval Clin Pract.* 2011;17(6):1231–1238. doi:10.1111/j.1365-2753.2010.01504.x
25. Boel J, Andreasen V, Jarlöv JO, et al. Impact of antibiotic restriction on resistance levels of *Escherichia coli*: a controlled interrupted time series study of a hospital-wide antibiotic stewardship programme. *J Antimicrob Chemother.* 2016;71(7):2047–2051. doi:10.1093/jac/dkw055
26. Ewusie JE, Blondal E, Soobiah C, et al. Methods, applications, interpretations and challenges of interrupted time series (ITS) data: protocol for a scoping review. *BMJ Open.* 2017;7(6):e016018. doi:10.1136/bmjopen-2017-016018
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* 2005;8(1):19–32. doi:10.1080/1364557032000119616
28. Peters MD, Godfrey CM, Khalil H, et al. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc.* 2015;13(3):141–146. doi:10.1097/XEB.000000000000050
29. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4(1):1. doi:10.1186/2046-4053-4-1
30. Fretheim A, Zhang F, Ross-Degnan D, 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):324–333. doi:10.1016/j.jclinepi.2014.10.003
31. Pechlivanoglou P, Wieringa JE, de Jager T, Postma MJ. The effect of financial and educational incentives on rational prescribing. *State-Space Approach Health Econ.* 2015;24(4):439–453.
32. Huitema BE, Van Houten R, Manal H. Time-series intervention analysis of pedestrian countdown timer effects. *Accid Anal Prev.* 2014;72:23–31. doi:10.1016/j.aap.2014.05.025
33. Velicer WF. Time series models of individual substance abusers. *NIDA Res Monogr.* 1994;142:264–301.
34. Andersson Hagiwara M, Andersson Gare B, Elg M. Interrupted time series versus statistical process control in quality improvement projects. *J Nurs Care Qual.* 2016;31(1):E1–E8. doi:10.1097/NCQ.0000000000000130
35. Burke LK, Brown CP, Johnson TM. Historical data analysis of hospital discharges related to the Amerithrax attack in florida historical data analysis of hospital discharges related to the Amerithrax attack in Florida. *Perspect Health Inf Manage.* 2016;13:1–16.
36. Harrington M, Velicer WF. Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behav Res.* 2015;50(2):162–183. doi:10.1080/00273171.2014.973989
37. Kontopantelis E, Doran T, Springate DA, et al. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ.* 2015;350(jun09 5):h2750. doi:10.1136/bmj.h2750
38. Nunes B, Natario I, Carvalho ML. Time series methods for obtaining excess mortality attributable to influenza epidemics. *Stat Methods Med Res.* 2011;20(4):331–345. doi:10.1177/0962280209340201
39. Sun P, Chang J, Zhang J, et al. Evolutionary cost analysis of valsartan initiation among patients with hypertension: a time series approach. *J Med Econ.* 2012;15(1):8–18. doi:10.3111/13696998.2011.626097
40. Duncan TE, Duncan SC. A latent growth curve modeling approach to pooled interrupted time series analyses. *J Psychopathol Behav Assess.* 2004;26(4):271–278. doi:10.1023/B:JOBA.0000045342.32739.2f
41. Cruz M, Bender M, Ombao H. A robust interrupted time series model for analyzing complex health care intervention data. *Stat Med.* 2017;29:29.
42. Kong M, Cambon A, Smith MJ. Extended logistic regression model for studies with interrupted events, seasonal trend, and serial correlation. *Commun Stat Theory Methods.* 2012;41(19):3528–3543. doi:10.1080/03610926.2011.563020
43. Park JH. Nonparametric approach to intervention time series modeling. *J Appl Stat.* 2012;39(7):1397–1408. doi:10.1080/02664763.2011.650684
44. Gillings D, Makuc D, Siegel E. Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care. *Am J Public Health.* 1981;71(1):38–46. doi:10.2105/AJPH.71.1.38
45. Yau KK, Lee AH, Carrivick PJ. Modeling zero-inflated count series with application to occupational health. *Comput Methods Programs Biomed.* 2004;74(1):47–52. doi:10.1016/S0169-2607(03)00070-1
46. Huitema BE, McKean JW. An improved portmanteau test for auto-correlated errors in interrupted time-series regression models. *Behav Res Methods.* 2007;39(3):343–349. doi:10.3758/BF03193002

47. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. *J Clin Epidemiol.* 2011;64(11):1252–1261. doi:10.1016/j.jclinepi.2011.02.007
48. Zhang F, Wagner AK, Soumerai SB, et al. Methods for estimating confidence intervals in interrupted time series analyses of health interventions. *J Clin Epidemiol.* 2009;62(2):143–148. doi:10.1016/j.jclinepi.2008.08.007
49. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci.* 2000;1(1):31–49. doi:10.1023/A:1010024016308
50. Hartmann DP, Gottman JM, Jones RR, et al. Interrupted time-series analysis and its application to behavioral data. *J Appl Behav Anal.* 1980;13(4):543–559. doi:10.1901/jaba.1980.13-543
51. Pullenayegum EM, Platt RW, Barwick M, et al. Knowledge translation in biostatistics: a survey of current practices, preferences, and barriers to the dissemination and uptake of new statistical methods. *Stat Med.* 2016;35(6):805–818. doi:10.1002/sim.6633

Journal of Multidisciplinary Healthcare

Dovepress

Publish your work in this journal

The Journal of Multidisciplinary Healthcare is an international, peer-reviewed open-access journal that aims to represent and publish research in healthcare areas delivered by practitioners of different disciplines. This includes studies and reviews conducted by multidisciplinary teams as well as research which evaluates the results or conduct of such teams or healthcare processes in general. The journal

covers a very wide range of areas and welcomes submissions from practitioners at all levels, from all over the world. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-inflammation-research-journal>