

RESEARCH

Open Access



# Design and statistical analysis reporting among interrupted time series studies in drug utilization research: a cross-sectional survey

Yuanjin Zhang<sup>1,2,3,4†</sup>, Yan Ren<sup>1,2,3,4†</sup>, Yunxiang Huang<sup>1,2,3,4</sup>, Minghong Yao<sup>1,2,3,4</sup>, Yulong Jia<sup>1,2,3,4</sup>, Yuning Wang<sup>1,2,3,4</sup>, Fan Mei<sup>1,2,3,4</sup>, Kang Zou<sup>1,2,3,4</sup>, Jing Tan<sup>1,2,3,4\*†</sup> and Xin Sun<sup>1,2,3,4\*†</sup>

## Abstract

**Introduction** Interrupted time series (ITS) design is a commonly used method for evaluating large-scale interventions in clinical practice or public health. However, improperly using this method can lead to biased results.

**Objective** To investigate design and statistical analysis characteristics of drug utilization studies using ITS design, and give recommendations for improvements.

**Methods** A literature search was conducted based on PubMed from January 2021 to December 2021. We included original articles that used ITS design to investigate drug utilization without restriction on study population or outcome types. A structured, pilot-tested questionnaire was developed to extract information regarding study characteristics and details about design and statistical analysis.

**Results** We included 153 eligible studies. Among those, 28.1% (43/153) clearly explained the rationale for using the ITS design and 13.7% (21/153) clarified the rationale of using the specified ITS model structure. One hundred and forty-nine studies used aggregated data to do ITS analysis, and 20.8% (31/149) clarified the rationale for the number of time points. The consideration of autocorrelation, non-stationary and seasonality was often lacking among those studies, and only 14 studies mentioned all of three methodological issues. Missing data was mentioned in 31 studies. Only 39.22% (60/153) reported the regression models, while 15 studies gave the incorrect interpretation of level change due to time parameterization. Time-varying participant characteristics were considered in 24 studies. In 97 studies containing hierarchical data, 23 studies clarified the heterogeneity among clusters and used statistical methods to address this issue.

**Conclusion** The quality of design and statistical analyses in ITS studies for drug utilization remains unsatisfactory. Three emerging methodological issues warranted particular attention, including incorrect interpretation of level

<sup>†</sup>Yuanjin Zhang and Yan Ren have contributed equally to this work and share the first authorship.

<sup>†</sup>Jing Tan and Xin Sun are co-corresponding authors.

\*Correspondence:

Jing Tan  
tanjing84@outlook.com  
Xin Sun  
sunxin@wchscu.cn

Full list of author information is available at the end of the article



change due to time parameterization, time-varying participant characteristics and hierarchical data analysis. We offered specific recommendations about the design, analysis and reporting of the ITS study.

**Keywords** Drug utilization, Interrupted time series, Pharmacoepidemiology, Quasi-experimental design, Segmented regression

## Introduction

Drug utilization research has received substantial attention from health researchers and policymakers in recent years. Interventions in drug utilization research may range from clinical guideline publications to drug programmes or policies. The randomized controlled trial is considered as the gold standard design for evaluating the causal effect of an intervention. Nevertheless, it is not always feasible or ethical in the field, as these interventions are often targeted at population level [1–4]. As a strong quasi-experimental design, interrupted time series (ITS) design has increasingly been used for the evaluation of drug utilization interventions by comparing the level and trend of outcomes after intervention with the pre-intervention underlying level and trend [5–10].

Several important methodological issues need to be considered when conducting ITS studies, such as time period selection, sample size, missing data, autocorrelation, and non-stationary and seasonality, which have been described in previous tutorials [5, 11–15]. Three issues newly emerging in recent years also require additional methodological considerations. First, the correct setting and interpretation of the ITS regression model should be underlined. In a particular ITS model setting including parameters of level change and slope change,  $\beta_2$  represents the immediate level (or intercept) change immediately following the intervention [16]. However, some current peer-reviewed studies conducted the wrong ITS model but still described  $\beta_2$  as the level change at the time of interruption, which will lead to an erroneous and biased result for the main effect of the immediate level change of the time-series (see details in Appendix 1) [17]. Second, it is possible that the participants' characteristics are not constant at different time points. The ITS method might be affected by time-varying confounding, which may result in a misleading finding [5, 6]. Third, heterogeneity in clusters should be appropriately addressed in ITS studies if the dataset contains a hierarchical structure and has within and/or between cluster heterogeneity [18–20]. A study pointed out that authors need to consider this issue and use appropriate analysis methods such as mixed-effect model [19]. However, since not all articles contain multiple-level data, the proportion of studies that have not yet addressed this issue remains unclear.

Previous studies might not have comprehensively addressed these methodological issues [11, 21–25]. The

last survey on the ITS studies in drug utilization research was published in 2015 and did not cover the new methodological issues mentioned above [21]. Additionally, despite increasing tutorials for conducting ITS have been published in recent years, it is still unclear whether the quality of current ITS studies in drug utilization has improved. Thus, we conducted a cross-sectional survey among the published ITS studies in drug utilization, aiming to identify the potential methodological gaps and give suggestions for improvement.

## Methods

### Eligibility criteria

We included empirical studies that used ITS design and focused on intervention related to drug utilization, with no limitation to study population or types of outcomes. The definition of ITS study was followed as the previous methodological studies, as “a time series of a particular outcome of interest is used to establish an underlying trend, which is ‘interrupted’ by an intervention at a known point in time” [4, 5]. We focused on the ITS study about drug utilization, whose intervention was about various medical, social and economic aspects of drug use [26].

Studies meeting any of the following criteria were excluded: (1) letters, commentaries, study protocols, conference abstracts, systematic reviews, meta-analyses, randomized controlled studies; (2) not written in English; (3) a methodological paper with an ITS example; (4) ITS analysis was not the main result.

### Search strategy

We searched PubMed in January 2022 for ITS studies published in 2021. We used Mesh terms and text words correlated to interrupted time series to develop the search strategy, including “interrupted time series”, “change point”, “segmented regression”, “repeated measures study” and so on. The details of the search strategy are presented in Appendix 2.

### Study process

A structured, pilot-tested checklist was developed to screen titles, abstracts, and full texts for potentially eligible studies, using prespecified eligibility criteria. Two researchers (YZ and YH), who were trained in epidemiology and biostatistics with sufficient experiences in ITS

analysis, conducted the records screening independently. Any disagreements were resolved by the discussion and adjudication by a third reviewer (YR).

Before informally extracting the data, two researchers (YZ and YH) randomly selected 15 (10%) eligible studies and extracted the data independently. They checked for consistency, and any disagreements were adjudicated by a third reviewer (YR). In total, the agreement between the two researchers (YZ and YH) was above 95%. Then a single researcher (YZ) extracted the further 138 studies.

**Development of the data extraction form**

A structured questionnaire was developed to investigate the design and analysis characteristics of ITS studies on drug utilization research. Initially, we reviewed the published methodological literature and statements to design the initial data extraction form [5, 21–25]. Then, we invited four experts (XS, RY, JT and MY) in clinical epidemiology and biostatistics to review and discuss the data extraction form, assessing the relevance and applicability of candidate items. We randomly selected 30 studies as pilot extraction to check if there were any items inappropriate.

Finally, we identified three parts of the design and analysis characteristics of ITS studies, including (1) general

characteristics, (2) design, and (3) statistical analysis. The detailed items of the data extraction form are shown in Appendix 3.

**Data analysis**

All items in the data extraction form were summarized using descriptive statistics. For categorical variables, we presented frequencies and percentages; for continuous variables, we presented mean with standard deviation (SD) or median with interquartile range (IQR). All statistical analyses were conducted using Stata.15.1.

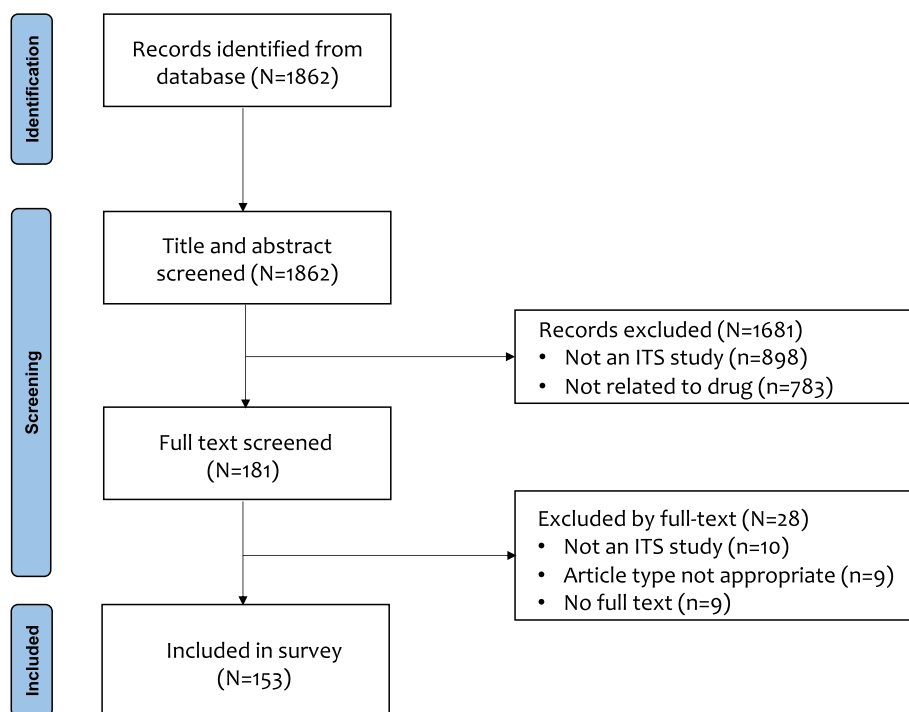
**Results**

Through the search from PubMed, we identified 1862 records. After reviewing titles, abstracts and full texts, 153 studies were finally included in our analysis (Fig. 1). Appendix 4 shows the details of all included studies.

**General characteristics**

Of the 153 included studies, 70.6% (108/153) were multi-site studies (Table 1). Hospital data (39.2%, 60/153), insurance databases (17.6%, 27/153) and other administrative databases (35.3%, 54/153) were the three most common data sources of the included ITS studies.

For the intervention characteristics, 82.4% (126/153) analyzed the single-stage intervention. The prescription



**Fig. 1** Flow diagram of the selection results

**Table 1** General characteristics of included ITS studies (N=153)

Characteristics	n	%
<b>Study sites</b>		
Multi-sites	108	70.6
Single-site	45	29.4
<b>Data source<sup>a</sup></b>		
Hospital data (from electronic medical record)	60	39.2
Other administrative databases	54	35.3
Insurance (claims) database	27	17.6
Others <sup>b</sup>	12	7.8
<b>Intervention stages</b>		
Single	126	82.4
Multiple	27	17.6
<b>Type of intervention</b>		
Prescription restriction	45	29.4
Drug price change (price, purchase or reimbursement policy)	27	17.6
Clinical guideline change (guidelines, statements or publications)	23	15.0
Education or training for physicians	22	14.4
Digital technology	13	8.5
Drug safety advisory	13	8.5
Drug approval or withdrawal	8	5.2
Others <sup>c</sup>	2	1.3
<b>Level of intervention</b>		
National	90	58.8
Hospital	30	19.6
State or province	21	13.7
Region	6	3.9
Others <sup>d</sup>	6	3.9
<b>Measure type of outcome</b>		
Drug utilization	125	81.7
Health outcomes	17	11.1
Expenditures	10	6.5
Others <sup>e</sup>	1	0.7
<b>Data type of outcome</b>		
Rate	98	64.1
Continuous	38	24.8
Count	15	9.8
Binary	2	1.3

<sup>a</sup> Administrative medical databases are massive repositories of data collected in healthcare for various purposes. Such databases are maintained in hospitals, health maintenance organizations and health insurance organizations. In this article, other administrative databases in this article included the administrative database which was not from hospital or health insurance organizations

<sup>b</sup> Others included registry (6 studies), survey (3 studies), statistical yearbook (2 studies) and cohort study (1 study)

<sup>c</sup> Others included drug shortage (1 study) and change drug packaging (1 study)

<sup>d</sup> Others included Europe (4 studies), city (1 study) and community (1 study)

<sup>e</sup> Others included cannabis-related criminal offences (1 study)

restriction was the most common intervention in the included studies (29.4%, 45/153). Drug price change and clinical guideline change were the next two most

common interventions (17.6%, 27/153 and 15.0%, 23/153). And the interventions were mainly implemented at the national level (58.8%, 90/153).

For the outcome characteristics, the most common outcomes were drug utilization (81.7%, 125/153), and some studies focused on health outcomes (11.1%, 17/153) and expenditures (6.5%, 10/153). Rate (64.1%, 98/153), continuous (24.8%, 38/153), and count (9.8%, 15/153) were the three most common data types of outcomes.

## Study design

### Rationale for ITS design

Among the included studies, only 28.1% (43/153) reported the rationale for using ITS design (Table 2). All studies gave a clear segment time of the intervention. There were 12.4% (19/153) studies that used ITS design with the control group to strengthen the validity of the study design.

### Data collection and aggregation

Of the 153 studies, most of the included studies (92.2%, 141/153) collected data retrospectively. In 12 studies used prospective data, and only 16.7% (2/12) had pre-specified study protocol. In total, 83.0% (127/153) contained individual-level data in the raw dataset, and only 17.0% (26/153) collected aggregated-level data. Most studies (97.4%, 149/153) used the aggregate unit as the ITS analysis unit, and only 2.6% (4/153) used the individual unit.

For the 149 studies with the aggregated unit, the most common time interval was monthly (73.8%, 110/149). The median (IQR) of total time points was 48 (30, 72). Only 20.8% (31/149) clarified the rationale for the number of time points (sample size calculation).

### The ITS model structure

Of 153 included studies, 90.8% (139/153) set the ITS model structure including both level change and slope change, 3.9% (6/153) included level change only and 3.9% (6/153) included slope change only. However, only 13.7% (21/153) gave the rationale for using this model structure. 27.5% (42/153) studies considered the potential delay effects, and 54.8% (23/153) of them reported the rationale for setting the transition period.

## Statistical analysis

### Basic statistical analysis characteristics

Various statistical methods were used to analyze the ITS studies (Table 3). Among the total 153 studies, OLS (30.7%, 47/153) and ARIMA (15.7%, 24/153) were the two most popular methods of the ITS regression model.

**Table 2** Design characteristics (rationale for ITS, data handling and model structure) of included studies

Characteristics	<i>n</i> or median	% or IQR
<b>Rationale for ITS</b>		
<b>Did the author give the reason/rationale for using ITS study design? (N = 153)</b>		
Yes	43	28.1
<b>Clearly segment time? (N = 153)</b>		
Yes	153	100.0
<b>Use of control group (N = 153)</b>		
Yes	19	12.4
<b>Type of control (N = 19)<sup>a</sup></b>		
Characteristic	11	57.9
Location	6	31.6
Historical	1	5.3
Outcome	1	5.3
<b>Data collection and aggregation</b>		
<b>Data collection (N = 153)</b>		
Retrospective	141	92.2
Prospective	12	7.8
<b>For studies used prospective data, if they have pre-specified protocol? (N = 12)</b>		
Yes	2	16.7
<b>Raw data (N = 153)</b>		
Contained individual-level data	127	83.0
Only contained aggregated-level data	26	17.0
<b>Analysis unit<sup>b</sup> (N = 153)</b>		
Aggregated unit	149	97.4
Individual unit	4	2.6
<b>Time intervals used for ITS with aggregated unit (N = 149)</b>		
Month	110	73.8
Quarter	17	11.4
Week	7	4.7
Year	6	4.0
Day	4	2.7
Six-month	2	1.3
Two-week	2	1.3
Two-month	1	0.7
<b>Number of time points (N = 149)<sup>c</sup></b>		
	48	30, 72
<b>Did the author give the rationale for the number of time points? (N = 149)</b>		
Yes	31	20.8
<b>The ITS model structure</b>		
<b>Type of ITS impact model (N = 153)</b>		
Level and slope change	139	90.8
Level change only	6	3.9
Slope change only	6	3.9
Unclear	2	1.3
<b>Did the author give the rationale for using this model? (N = 153)</b>		
Yes	21	13.7
<b>Did the author allow for a delay impact? (N = 153)</b>		
Yes	42	27.5
<b>Methods of dealing with delay (N = 42)<sup>d, e</sup></b>		
Sensitivity	19	45.2
Segment	11	26.2

**Table 2** (continued)

Characteristics	n or median	% or IQR
Excluded	10	23.8
Delay	6	14.3
<b>Did the author give the rationale for setting the transition period? (N= 42)</b>		
Yes	23	54.8

<sup>a</sup> Location: use a different area as control; Outcome: use an outcome not affected by the intervention as control; Characteristic: use a group not targeted by an intervention as control; Historical: compare a previous group to a current group

<sup>b</sup> Aggregated data refer to summary statistics (e.g., mean, percentage, median) calculated across individual data

<sup>c</sup> We reported the median and interquartile range (25% and 75%)

<sup>d</sup> Delay: where the delay was acknowledged and included in pre- or post-interruption segment; Excluded: where a separate segment was used for the delay time period, but this was excluded from the analysis; Segment: where a separate segment was used for the delay time period, and this was included in analysis; Sensitivity: where the delay was modelled as part of a sensitivity analysis, but ignored in main analysis)

<sup>e</sup> Some studies used more than one method to deal with delay

However, 15.0% (23/153) did not report the regression model.

Of these 153 studies, 31 studies addressed missing data and 49 studies did the sensitivity analysis. Most studies (86.3%, 132/153) reported the software for statistical analysis, in which SAS (35.3%, 54/153) and Stata (31.4%, 48/153) were the two most popular software for ITS analysis. Only 3.3% (5/153) studies uploaded the full code and 2.0% (3/153) shared the datasets.

#### **Basic methodological considerations (Autocorrelation, non-stationary and seasonality)**

Among the 149 studies with aggregated-level outcome and time series data, 14 studies considered all of the three methodological issues of time series data (Table 4). 117 studies considered at least one of three methodological issues. Specifically, autocorrelation was acknowledged in 108 studies, non-stationarity was acknowledged in 20 studies, and seasonality was acknowledged in 60 studies. Among the studies adjusted for autocorrelation, non-stationary and seasonality, 25.0% (27/108), 5% (1/20) and 16.7% (10/60) respectively failed to specify the methods they used.

We also compared the differences in studies that used the ARIMA model and the non-ARIMA model. The results showed that the studies that used the ARIMA model were more likely to consider the potential autocorrelation (83.3%, 20/24 vs 70.4%, 88/125), non-stationary (45.9%, 11/24 vs 7.2%, 9/125) and seasonality (62.5%, 15/24 vs 36.0%, 45/125).

#### **Additional methodological considerations**

##### **Incorrect interpretation of level change due to time parameterization**

Of the 153 studies, only 39.2% (60/153) reported the specific regression model and interpreted the coefficients in the article or supplementary files (Table 5).

Moreover, we found that 15 studies gave incorrect interpretations of level change due to time parameterisation. To be more specified, these studies reported the model as " $Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 T_t \bullet X_t$ ", which included parameters for level change and slope change. But they described  $\beta_2$  as "level change at the time of interruption. As we discussed in Appendix 1, this will lead to an incorrect result for the effect of the immediate level change if the study used this model for statistical analysis.

#### **Individual-level characteristics**

Among 153 included studies, 83.0% (127/153) contained individual-level data in the raw dataset. Of these 127 ITS studies, 18.9% (24/127) considered individual-level characteristics.

#### **Handling hierarchical data**

Of the 149 studies that conducted aggregated unit ITS analysis, 65.1% (97/149) studies included hierarchical data. However, only 23.7% (23/97) of them considered this hierarchical structure. In further analyzing these 23 studies, stratified analysis (56.5%, 13/23) was the most common method to address the hierarchical structure of data, mixed-effect model (26.1%, 6/23) and generalized estimating equation (8.7%, 2/23) were the next two common methods. Authors usually considered the hospital-level cluster effect (65.2%, 15/23), and 17.4% (4/23) studies considered the cluster effects in provinces. Additionally, 69.6% (16/23) estimated and reported the differences across sites.

## **Discussions**

### **Findings and interpretations**

This study provides updated evidence on the quality of ITS studies and found that most ITS studies in drug utilization fail to consider the methodological issues of design and statistical analysis comprehensively.

**Table 3** Basic statistical characteristics of included ITS studies

Characteristics	n	%
<b>Regression model<sup>a</sup> (N = 153)</b>		
OLS	47	30.7
ARIMA	24	15.7
GLS	13	8.5
OLS with Newey-West standard errors	12	7.8
Poisson	12	7.8
Mixed effect model	8	5.2
Logistic	4	2.6
Generalized estimating equation	4	2.6
Weighted least square regression	2	1.3
Others <sup>b</sup>	4	2.6
Unclear	23	15.0
<b>Other statistical analysis characteristics</b>		
<b>Did the author consider missing data? (N = 153)</b>		
Yes	31	20.3
<b>Sensitivity analysis (N = 153)</b>		
Yes	49	32.0
<b>Methods for sensitivity analysis (N = 49)<sup>c,d</sup></b>		
Transition period	19	38.8
Change measurement of outcomes	10	20.4
Change measurement of study population	11	22.5
Change ITS model setting	9	18.4
Add covariates	2	4.1
Other	11	22.4
<b>Statistical software (N = 153)</b>		
SAS	54	35.3
Stata	48	31.4
R	25	16.3
SPSS	5	3.3
Not report	21	13.7
<b>Data availability (N = 153)</b>		
Yes	3	2.0
<b>Code availability (N = 153)</b>		
Yes	5	3.3

<sup>a</sup> This item refers to the statistical method for the main results in a study

<sup>b</sup> Others included fixed effect model (1 study), negative binomial model (1 study), quasi-poisson model (1 study) and linear probability model (1 study)

<sup>c</sup> Transition period: change the interrupted time or time period in the regression model; Change ITS model setting: change the ITS impact model (e.g., from both level and slope change to only level change)

Some studies used more than one method for sensitivity analysis

Three main issues of ITS study design need to be considered. First, most studies did not give the rationale for using ITS design. Although it is an appropriate method when randomization is not feasible, the basic ITS design may be affected by confounding due to co-interventions or other events occurring around the study period [27, 28]. Thus, we recommended that the author should give the rationale for using ITS design, such as for ethical

consideration or no adequate control group. Second, most of the studies did not report the consideration of the study period, time interval and sample size. The selection of time period should be a balance between statistical requirements and research problem-driven decisions [29, 30]. A simulation study found that sample size per time point had a large impact on power in ITS study. Even though the studies meet the requirement of minimum time points, most analyses were underpowered if the sample size per time point was low [30]. Therefore, the author should balance the number of time points and the sample size per time point. Meanwhile, if the period is too short, there may be too little data to model the trend. However, if the period is too long, it may be affected by historical bias. Third, most of the studies used the ITS model structure with both level change and slope change. However, only a few studies analyzed the intervention (whether it will lead to immediate change or sustained change) and chose the ITS model structure to fit it well [5, 29]. When the model was misspecified, the results of ITS were not robust anymore [31].

Meanwhile, we found five issues that may affect the quality of statistical analysis. First, most of the studies did not mention the missing data. A study mentioned that most of the study used data aggregated at the population level, but it will lead to bias when data are missing at random at the individual level [11]. In a simulated scenario in this study, if the outcome is missing at random for male but is fully observed for female, the aggregated data will show a wrong seasonal pattern. Second, the considerations of autocorrelation, non-stationary and seasonality are still poor among current ITS studies. Ignoring the characteristics of time series data may not provide robust results [5]. Third, more than half studies did not report the regression model, which might lead to an unclear understanding of statistical methods for readers. Moreover, for the studies that reported the regression model, 15 studies used the setting “ $TX_t$ ” instead of “ $(T - T_0)X_t$ ” in the ITS regression model. But it will lead to an erroneous result for the main effects of the level change. Fourth, the consideration of time-varying confounding is lacking. Participants-level confounding should be considered and controlled if the population was changed at each time point [6, 19]. Fifth, most of the included studies ignored the hierarchical data structure and aggregated the outcome to the population level, even if they had the opportunity to aggregate the outcome at a lower level. As we discussed above, when the intervention is implemented regionwide or nationwide, the dataset may contain a hierarchical structure. If the outcome is aggregated at a higher level, which does not account for the heterogeneity among patients and across hospitals, it will lead to aggregation bias [18, 19, 25, 32].

**Table 4** Characteristics of the basic methodological considerations (autocorrelation, non-stationarity, seasonality) (Only for ITS with aggregated unit)

Characteristics	<i>n</i>	%
<b>Considered all of three methodological issues (autocorrelation, non-stationarity and seasonality) (N = 149)</b>		
Yes	14	9.4
<b>Considered at least one of three methodological issues (N = 149)</b>		
Yes	117	78.5
<b>Autocorrelation</b>		
<b>Autocorrelation acknowledged (N = 149)</b>		
Yes	108	72.5
<b>Autocorrelation acknowledged (ITS study used ARIMA model) (N = 24)</b>		
Yes	20	83.3
<b>Autocorrelation acknowledged (ITS study used Non-ARIMA model) (N = 125)</b>		
Yes	88	70.4
<b>Autocorrelation identified methods (N = 108)</b>		
Durbin Watson test	40	37.0
ACF	9	8.3
Cumby-Huizinga test	5	4.6
Ljung-Box2 test	3	2.8
Others <sup>a</sup>	4	3.7
Not reported	47	43.5
<b>Adjusted for autocorrelation (N = 108)</b>		
Yes	65	60.2
No adjustment for autocorrelation (after statistical test)	16	14.8
Unclear	27	25.0
<b>If yes, which method was used? (N = 65)</b>		
ARIMA	31	47.7
GLS	16	24.6
OLS with Newey-West standard errors	12	18.5
Add lag terms	5	7.7
Generalized Estimating Equation	1	1.5
<b>Non-stationarity</b>		
<b>Non-stationarity acknowledged (N = 149)</b>		
Yes	20	13.4
<b>Non-stationarity acknowledged (ITS study used ARIMA model) (N = 24)</b>		
Yes	11	45.9
<b>Non-stationarity acknowledged (ITS study used Non-ARIMA model) (N = 125)</b>		
Yes	9	7.2
<b>Non-stationarity identified methods (N = 20)</b>		
Augmented Dickey-Fuller test	8	40.0
Plot the raw data	1	5.0
Not reported	11	55.0
<b>Adjusted for non-stationarity (N = 20)</b>		
Yes	15	75.0
No adjustment for non-stationarity (after statistical test)	4	20.0
Unclear	1	5.0
<b>If yes, which method was used? (N = 15)</b>		
ARIMA	11	73.3
Others <sup>b</sup>	4	26.7
<b>Seasonality</b>		
<b>Seasonality acknowledged (N = 149)</b>		



**Table 4** (continued)

Characteristics	<i>n</i>	%
Yes	60	40.3
<b>Seasonality acknowledged (ITS study used ARIMA model) (N= 24)</b>		
Yes	15	62.5
<b>Seasonality acknowledged (ITS study used Non-ARIMA model) (N= 125)</b>		
Yes	45	36.0
<b>Seasonality identified methods (N= 60)</b>		
Augmented Dickey-Fuller test	4	6.7
Plot the raw data	3	5.0
Others <sup>c</sup>	6	10.0
Not reported	47	78.3
<b>Adjusted for seasonality (N= 60)</b>		
Yes	41	68.3
No adjustment for seasonality (after statistical test)	9	15.0
Unclear	10	16.7
<b>If yes, which method was used? (N= 41)</b>		
Add seasonality terms	18	43.9
ARIMA	13	31.7
Fourier function	8	19.5
Others <sup>d</sup>	2	4.9

<sup>a</sup> Others included residual plots (2 studies), Bartlett formula (1 study), Breusch-Godfrey test (1 study)

<sup>b</sup> Others included add dummy variable (3 studies) and first difference (1 study)

<sup>c</sup> Others included Cumby-Huizinga test (1 study), Kruskal–Wallis test (1 study), Webel-Ollech overall seasonality (1 study), Summary statistics (1 study), add seasonality terms (1 study), test lagged correlation (1 study)

<sup>d</sup> Others included Holt-Winters seasonal smoothing approach (1 study) and Lag period (1 study)

### Comparison with other studies

Several studies have systematically reviewed methodological issues regarding the design and statistical analysis of the ITS study [11, 21–25]. All of the previous reviews pointed out that the considerations of autocorrelation, non-stationary and seasonality were limited, which was aligned with our study. Five of them reported the sample size considerations which focused on the minimization of data points, while our study also pointed out that the maximum of data points should also be a consideration.

Some methodological issues have been improved among the ITS studies published in 2021. For example, for the item “clearly segment time”, the reported proportion has seen a notable increase, rising from 84.5% (as observed in Jandoc et al.’s review) to 100% in our study. However, some issues remain a concern (e.g., sample size, missing data, incorrect interpretation of level change due to time parameterization, time-varying participants-level confounding, and data hierarchical structure). A previous review that included a meta-analysis and re-analysis of ITS studies found that 5% (2/41) of studies did not report the statistical method used [33]. In our review, this proportion is 15.0% (23/153), indicating a higher proportion of inadequate reporting in original articles.

Our study gave a detailed analysis of three ever ignored but important methodological issues, including common errors in parameter interpretation of ITS models, limited consideration of individual-level characteristics and poor handling of heterogeneous data among clusters. Although a methodology study published in 2020 pointed out this problem and a corrigendum to the original tutorial had been made [17, 34], this was still a common mistake in ITS empirical studies published in 2021. Individual-level characteristic is also an important issue. If patient characteristics vary over time, it is essential to control for these changes using appropriate methods. For the potential cluster effects, our result showed that most of the studies had the opportunity to control the potential heterogeneity from different clusters, but few of them considered it.

### Strengths and limitations

This study gives a comprehensive survey of the methodological issues in the design and statistical analysis of ITS studies in drug utilization. To the best of our knowledge, this is the first cross-sectional survey that exclusively assesses the incorrect interpretation of level change due to time parameterization, time-varying individual-level covariates and handling of hierarchical data in current

**Table 5** Additional methodological considerations (parameters setting, individual-level covariates and hierarchical data structure) for ITS studies

Characteristics	<i>n</i>	%
<b>Incorrect interpretation of level change due to time parameterization <sup>a</sup></b>		
<b>Reported the regression model and interpreted the coefficients (N = 153)</b>		
Yes	60	39.2
<b>Where did the author report the regression model and the interpretation of coefficients? (N = 60)</b>		
In article	47	78.3
In supplementary material	13	21.7
<b>The interpretation of level change due to time parameterization was incorrect <sup>b</sup></b>		
Yes	15	-
<b>Individual-level characteristics</b>		
<b>Has individual-level data (N = 153)</b>		
Yes	127	83.0
<b>Consider individual-level characteristics (N = 127)</b>		
Yes	24	18.9
<b>How to control individual-level characteristics (N = 24) <sup>c</sup></b>		
Add covariates	21	87.5
Stratified Analysis	7	29.2
Other	4	16.7
<b>Hierarchical data structure</b>		
<b>Data structure for ITS analysis (N = 149) <sup>d</sup></b>		
Hierarchical data (more than one level) <sup>e</sup>	97	65.1
<b>Whether the author handled hierarchical data (N = 97)</b>		
Yes	23	23.7
<b>Methods for handling hierarchical data (N = 23)</b>		
Stratified by sites	13	56.5
Mixed effect model	6	26.1
Generalized estimating equation	2	8.7
Fixed-effect model	1	4.3
Two-stage analysis	1	4.3
<b>Considered cluster effects in which level? (N = 23)</b>		
Hospital/clinic/other healthcare provider	15	65.2
Province/State/Region	4	17.4
Nation	2	8.7
Unclear	2	8.7
<b>Whether the author reported the differences across sites (N = 23)</b>		
Yes	16	69.6
<b>If yes, how to present the differences across site? (N = 16)</b>		
Figure	7	43.8
Both table and figure	6	37.5
Table	3	18.8

<sup>a</sup> If the researchers set an ITS model with both level change and slope change, and used the product between their calendar time variable and the indicator variable indicating pre- versus post-intervention time periods to represent the post-intervention linear segment, then the interpretation was wrong (More details in Appendix 1)

<sup>b</sup> For this item, we did not calculate the proportion as the denominator is difficult to define. We believe that using either 60 (the number of studies reporting regression models) or 139 (the number of models including level change and slope change) as the denominator would be inappropriate

<sup>c</sup> Some studies used more than one method to control individual-level characteristics

<sup>d</sup> This part only included ITS studies with aggregated analysis units (*n* = 149) because the mishandling of data hierarchy only takes place in the ITS study with aggregated analysis unit

<sup>e</sup> For the studies that contained individual-level data, we calculated how many levels are there in the dataset excluded individual data (which cannot be repeated measured). For example, the raw data was a three-level hierarchy of patient, hospital and region and the repeated measured level were hospital and region. We defined this dataset as a two-level hierarchical data for ITS analysis. For the studies that only contained aggregated data, we calculated how many levels are there in the dataset directly

ITS studies, which have been highlighted in the methodological literature. Meanwhile, we updated the current practices of ITS in the field of drug utilization research. ITS is a frequently used method in evaluating a population-level intervention, and there is a series of literature on methodological considerations published over the past few years. It is worth analyzing and showing the limitations in methodological issues of ITS practices.

There are also three limitations in our studies. First, we only included the ITS studies published in 2021 and used a single database for searching. However, since PubMed contains nearly all healthcare science & service and public health research journals, we think that it can represent the current practices sufficiently. Another limitation is that we assess the design and statistical characteristics through the reporting of the article. If the reporting of these aspects is insufficient, we cannot determine the items and the results may be inaccurate. Third, some items may not be relevant to all studies. For example, in ITS studies using aggregated data, authors might not be able to assess the proportion of missing data at the individual level. Consequently, they may not report missing data in their articles.

## Conclusion

In summary, we identified a series of deficiencies in design and statistical analysis among current ITS studies, showing that the basic methodological issues are not improved, and some new issues are not widely considered (i.e., incorrect interpretation of level change due to time parameterization, time-varying individual characteristics and hierarchical data structure). Although a series of methodology reviews and tutorials mentioned the important issues in ITS design, there is still a significant gap between guidelines and practices of ITS studies in drug utilization research, accentuating that it is need to develop a clearer guide and checklist for conducting ITS study.

## Abbreviations

ITS	Interrupted time series
OLS	Ordinary least square
ARIMA	Autoregressive integrated moving average model
GLS	Generalized least square
IQR	Interquartile range

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-024-02184-8>.

### Supplementary Material 1.

## Authors' contributions

YZ, YR and XS conceived and designed the study. YZ and YR conducted the literature search. YZ and YH screened the articles and extracted the data. YZ, YR, YH, YJ, YW, FM and KZ conducted the analysis and interpreted the data. YZ drafted the manuscript. YR, MY, TJ and XS critically revised the manuscript.

## Funding

This study was supported by National Natural Science Foundation of China (Grant No. 72004149), China Postdoctoral Science Foundation (Grant No. 2022M720101), National Science Fund for Distinguished Young Scholars (Grant No. 82225049), Sichuan Provincial Central Government Guides Local Science and Technology Development Special Project (Grant No. 2022ZYD0127) and Xiamen Science and Technology Plan Project (3502Z20224004).

## Availability of data and materials

Details of the search strategy and included studies are in the appendix.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

## Author details

<sup>1</sup>Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, 37 Guo Xue Xiang, Chengdu 610041, Sichuan, China. <sup>2</sup>NMPA Key Laboratory for Real World Data Research and Evaluation in Hainan, Chengdu, China. <sup>3</sup>Sichuan Center of Technology Innovation for Real World Data, Chengdu, China. <sup>4</sup>Hainan Healthcare Security Administration Key Laboratory for Real World Data Research, Chengdu, China.

Received: 23 September 2023 Accepted: 20 February 2024

Published online: 09 March 2024

## References

- Fukuma S, Ikenoue T, Yamada Y, et al. Changes in Drug Utilization After Publication of Clinical Trials and Drug-Related Scandals in Japan: An Interrupted Time Series Analysis, 2005–2017. *J Epidemiol.* 2021;31(7):410–6. <https://doi.org/10.2188/jea.JE20200181>.
- Peek N, Gude WT, Keers RN, et al. Evaluation of a pharmacist-led actionable audit and feedback intervention for improving medication safety in UK primary care: An interrupted time series analysis. *PLoS Med.* 2020;17(10): e1003286. <https://doi.org/10.1371/journal.pmed.1003286>.
- Nichols J, Mamdani M, Gomes T, Shah BR, Casey CG, Yu CH. Impact of Clinical Practice Guidelines on Blood Glucose Test Strip Prescription Rates in Manitoba and Saskatchewan (Canada): An Interrupted Time-Series Analysis. *Canadian Journal of Diabetes.* 2021;45(6):557–565e2. <https://doi.org/10.1016/j.jcjd.2020.11.008>.
- Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther.* 2002;27(4):299–309. <https://doi.org/10.1046/j.1365-2710.2002.00430.x>.
- 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):348–55. <https://doi.org/10.1093/ije/dyw098>.
- Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *Bmj.* 2015;350:h2750. <https://doi.org/10.1136/bmj.h2750>.

7. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. *Acad Pediatr* Nov-Dec. 2013;13(6 Suppl):S38–44. <https://doi.org/10.1016/j.acap.2013.08.002>.
8. Bärnighausen T, Røttingen JA, Rockers P, Shemilt I, Tugwell P. Quasi-experimental study designs series-paper 1: introduction: two historical lineages. *J Clin Epidemiol*. 2017;89:4–11. <https://doi.org/10.1016/j.jclinepi.2017.02.020>.
9. Bärnighausen T, Tugwell P, Røttingen JA, et al. Quasi-experimental study designs series-paper 4: uses and value. *J Clin Epidemiol*. 2017;89:21–9. <https://doi.org/10.1016/j.jclinepi.2017.03.012>.
10. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *Int J Epidemiol*. 2013;42(4):1187–95. <https://doi.org/10.1093/ije/dyt092>.
11. Bazo-Alvarez JC, Morris TP, Carpenter JR, Petersen I. Current Practices in Missing Data Handling for Interrupted Time Series Studies Performed on Individual-Level Data: A Scoping Review in Health Research. *Clin Epidemiol*. 2021;13:603–13. <https://doi.org/10.2147/clep.S314020>.
12. Liu W, Ye S, Barton BA, et al. Simulation-based power and sample size calculation for designing interrupted time series analyses of count outcomes in evaluation of health policy interventions. *Contemp Clin Trials Commun*. 2020;17: 100474. <https://doi.org/10.1016/j.conctc.2019.100474>.
13. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. *J Clin Epidemiol*. 2011;64(11):1252–61. <https://doi.org/10.1016/j.jclinepi.2011.02.007>.
14. Turner SL, Karahalios A, Forbes AB, Taljaard M, Grimshaw JM, McKenzie JE. Comparison of six statistical methods for interrupted time series studies: empirical evaluation of 190 published series. *BMC Medical Research Methodology*. 2021;21(1):134. <https://doi.org/10.1186/s12874-021-01306-w>.
15. Schober P, Vetter TR. Segmented Regression in an Interrupted Time Series Study Design. *Anesth Analg*. 2021;132(3):696–7. <https://doi.org/10.1213/ane.0000000000005269>.
16. Simple Interrupted Time-Series Designs. *The Analysis of Covariance and Alternatives*. 2011:365–402.
17. Xiao H, Augusto O, Wagenaar BH. Reflection on modern methods: a common error in the segmented regression parameterization of interrupted time-series analyses. *Int J Epidemiol*. 2021;50(3):1011–5. <https://doi.org/10.1093/ije/dyaa148>.
18. Ewusie J, Beyene J, Thabane L, Straus SE, Hamid JS. An improved method for analysis of interrupted time series (ITS) data: accounting for patient heterogeneity using weighted analysis. *Int J Biostat*. 2022;18(2):521–35. <https://doi.org/10.1515/ijb-2020-0046>.
19. Ewusie JE, Thabane L, Beyene J, Straus SE, Hamid JS. MultiCenter Interrupted Time Series Analysis: Incorporating Within and Between-Center Heterogeneity. *Clin Epidemiol*. 2020;12:625–36. <https://doi.org/10.2147/clep.S231843>.
20. Zombré D, De Allegri M, Ridde V. Immediate and sustained effects of user fee exemption on healthcare utilization among children under five in Burkina Faso: A controlled interrupted time-series analysis. *Soc Sci Med*. 2017;179:27–35. <https://doi.org/10.1016/j.socscimed.2017.02.027>.
21. Jandoc R, Burden AM, Mamdani M, Lévesque LE, Cadarette SM. Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. *J Clin Epidemiol*. 2015;68(8):950–6. <https://doi.org/10.1016/j.jclinepi.2014.12.018>.
22. 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(1):137. <https://doi.org/10.1186/s12874-019-0777-x>.
23. 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 Global Health*. 2020;5:e003567.
24. 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. <https://doi.org/10.1016/j.jclinepi.2020.02.006>.
25. Ewusie JE, Soobiah C, Blondal E, Beyene J, Thabane L, Hamid JS. Methods, Applications and Challenges in the Analysis of Interrupted Time Series Data: A Scoping Review. *J Multidiscip Healthc*. 2020;13:411–23. <https://doi.org/10.2147/jmdh.S241085>.
26. Elseviers M, Wettermark B, Almarsdóttir AB, et al. *Drug utilization research: methods and applications*. John Wiley & Sons; 2016.
27. Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. *Int J Epidemiol*. 2018;47(6):2082–93. <https://doi.org/10.1093/ije/dyy135>.
28. Bottomley C, Scott J, Anthony G, Isham V. Analysing Interrupted Time Series with a Control. *Epidemiol Method*. 2019;8(1):20180010. <https://doi.org/10.1515/em-2018-0010>.
29. Lopez Bernal J, Soumerai S, Gasparrini A. A methodological framework for model selection in interrupted time series studies. *J Clin Epidemiol*. 2018;103:82–91. <https://doi.org/10.1016/j.jclinepi.2018.05.026>.
30. Hawley S, Ali MS, Berencsi K, Judge A, Prieto-Alhambra D. Sample size and power considerations for ordinary least squares interrupted time series analysis: a simulation study. *Clin Epidemiol*. 2019;11:197–205. <https://doi.org/10.2147/clep.S176723>.
31. Jiang H, Feng X, Lange S, Tran A, Manthey J, Rehm J. Estimating effects of health policy interventions using interrupted time-series analyses: a simulation study. *BMC Med Res Methodol*. 2022;22(1):235. <https://doi.org/10.1186/s12874-022-01716-4>.
32. Cruz M, Bender M, Ombao H. A robust interrupted time series model for analyzing complex health care intervention data. *Stat Med*. 2017;36(29):4660–76. <https://doi.org/10.1002/sim.7443>.
33. Korevaar E, Karahalios A, Turner SL, et al. Methodological systematic review recommends improvements to conduct and reporting when meta-analyzing interrupted time series studies. *J Clin Epidemiol*. 2022;145:55–69. <https://doi.org/10.1016/j.jclinepi.2022.01.010>.
34. 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;50(3):1045–1045. <https://doi.org/10.1093/ije/dyaa118>.

## Publisher's Note

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