Protocol for a database and an investigating into associations of interrupted time series in the healthcare setting

Protocol version: 2

Protocol Date: 29/01/2021

PhD Student

Jemma Hudson

29/01/2021

Signature Date

Supervisor

Craig Ramsay 29/01/2021

Signature Date

Supervisor

Shona Fielding 29/01/2021

Signature Date

Supervisor

Miriam Brazzelli 29/02/2021

Signature Date

Version history

Amendment No.	Protocol version No.	Description of changes	Date of protocol
	Version 1	New document	
1	Version 2	Seasonality, autocorrelation, and stationarity to be assessed pre-, post-intervention, and all of the data series and data to be extraction. Update of Hypothesis 2 to reflect this.	29/01/2021

Aim and Objectives

The ***aim*** of this work it is to help aid researchers in planning an interrupted time (ITS) study.

The two ***objectives*** are

1. To provide a database of ITS studies from healthcare settings (including design characteristics and associated parameter estimates).
2. To investigate which design characteristics are associated with size of autocorrelation and size of effect sizes.

Inclusion criteria

Interrupted time series studies will be eligible for inclusion if there is a minimum of two data points collected pre-intervention period and one post-intervention period. Outcomes will be restricted to continuous and interventions to be included will be associated with health and the healthcare of patients. For example, a program to prevent infection, policies on antibiotic use, clinical or educational. From the methodological systematic review of interrupted time series designs in healthcare[1], the majority of studies evaluated only one intervention (90/116 (78%)), therefore I have restricted inclusion to studies evaluating only one intervention. There is no restriction on the language of study and type of participants.

Identification of ITS studies

A maximum of 200 ITS studies will be identified. ITS studies will be obtained purposively from known sources of the research team including systematic reviews from the Cochrane Effective Practice and Organisation of Care group (EPOC) and a random sample from the studies identified from the methodological systematic review of interrupted time series designs in healthcare[1].

Analysis of data sets

To check whether the pre-intervention phase is linear, data will be plotted and checked visually as well as plotting the residuals versus fitted values. Presence of seasonality will be checked visually as well as using autocorrelation function (ACF) and the partial autocorrelation function (PACF). To determine whether autocorrelation is present across the whole time series data[2], the ACF, PACF and the Breusch-Godfrey test[3 4] will be used. Autocorrelation will be calculated from the ACF. To check whether the series is stationary the ACF, PACF and the augmented Dickey-Fuller test[4] will be used. If no autocorrelation is present and the series is stationary then an ordinary least squares segmented regression model

will be used and if there is seasonality, covariate adjustment will be used. If there is presence of autocorrelation and/or the series is non-stationary then an ARIMA model will be used, if seasonality is present then the SARIMA model will be used if there are 50 or more data points[5]. If there are less than 50 data points, then restricted maximum likelihood method will be used[6]. Analysis will be performed in Stata 16[7]. The Akaike information criterion as well as the Bayesian information criterion will be used to help determine the best fitting model. Sample code that will be used for the analysis to produce the estimates is provided at the end of this document.

Data to be extracted

The database will include characteristics of the individual studies. Effect estimates and autocorrelation results will be obtained from the analysis as described above.

Items to be extracted from included ITS studies:

- Study ID (last name and year of publication)
- Study reference
- Outcome (description of the outcome e.g. the number or antibiotic prescriptions prescribed per month)
- Type of intervention (program, policy, health system, guidelines, sales and dispensing and other)
- Level of analysis (e.g. country, hospital, region, hospital department)
- Frequency of data collection (monthly, quarterly, yearly, weekly and other)
- Type of outcome (continuous, count, rate)

Items to be obtained from analysis:

- Number of data points
    - All
    - Pre
    - Post
- Pre-intervention linear (yes/no)
- To be assessed and data extracted pre- and post-intervention as well as all of the data points included in the series
    - Seasonality (yes/no)

- Frequency of seasonality if present (monthly, quarterly, yearly and other)
- Autocorrelation estimate
- Autocorrelation present (determined by the Breusch-Godfrey test, ACF and PACF) (yes/no)
- Stationary (determined by the Dicky-fuller test, ACF and PACF) (yes/no)

For ITS studies that are linear the following data will be extracted:

- Baseline level or intercept (outcome at time 0)
  - Effect estimate
  - Standard error
  - P-value
- Pre-trend (or pre-slope) (corresponds to a unit increase (or decrease) in outcome for the pre-intervention phase)
  - Effect estimate
  - Standard error
  - P-value
- Post-trend (corresponds to a unit increase (or decrease) in outcome for the post-intervention phase)
  - Effect estimate
  - Standard error
  - P-value
- Change in trend (the difference between the pre-intervention trend and post-intervention trend)
  - Effect estimate
  - Standard error
  - P-value
- Immediate change in level (change in outcome at point of intervention)
  - Effect estimate
  - Standard error
  - P-value
  -

Outcomes and Hypotheses

The ***outcomes of interest*** of this study are autocorrelation, change in trend and the immediate change in level. The ***hypotheses*** are:

Hypothesis 1: There is an association between change in trend and level and the number of time points

Hypothesis 2. There is an association between the phase of the time points used to estimate the autocorrelation (pre data only; post data only; pre and post data combined) and size of autocorrelation.

Hypothesis 3: There is an association between length of time point collection e.g. monthly compared to yearly) and autocorrelation.

Hypothesis 4: There is an association between the level of intervention and autocorrelation. For example, individual level will have higher (or lower) autocorrelation compared to higher levels (hospital level).

Autocorrelation will be calculated from the method described above. For the change in trend and immediate change in level, these will be calculated using the methods described above but will be standardised by dividing by the standard error. Please refer to Appendix 1 for sample code that will be used for the analysis to produce the estimates.

Descriptive analysis

Basic summary statistics for the autocorrelation and effect estimates will be calculated for each hypothesis to be tested. This includes, means and standard deviations, median and ranges. Graphical representations such as histograms and boxplots of the distributions of autocorrelation will be used.

Statistical testing of hypotheses

Mean or median autocorrelations/effect estimates between the comparative groups will be initially tested using t-test or Mann-Whitney U-tests depending of the distribution of autocorrelation/effect estimates for two group comparisons. When there are more than two groups Kruskal-Wallis test will be used. For continuous variables, a Pearson's correlation will be used.

It could be possible that effects of one explanatory variable from the hypotheses defined above might influence the results of others. Therefore, a multivariable approach will be considered.

References

1. Hudson J, Fielding S, Ramsay CR. Methodology and reporting characteristics of studies using interrupted time series design in healthcare. *BMC Medical Research Methodology*. 2019;19(1):137. https://doi.org/10.1186/s12874-019-0777-x. doi: 10.1186/s12874-019-0777-x.

2. Cook TD, Campbell DT. *Quasi-experimentation: Design and analysis issues for field settings.* Rand McNally College Pub. Co; 1979. https://books.google.co.uk/books?id=6WxdAAAAIAAJ.

3. BREUSCH TS. Testing for autocorrelation in dynamic linear models*. *Aust Econ Pap*. 1978;17(31):334-355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x. doi: 10.1111/j.1467-8454.1978.tb00635.x.

4. Godfrey LG. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*. 1978;46(6):1293-1301. http://www.jstor.org/stable/1913829. doi: 10.2307/1913829.

5. Box GEP, Jenkins GM, Reinsel GC, Ljung GM, Ljung GM. *Time series analysis : Forecasting and control.* New York: John Wiley & Sons, Incorporated; 2015. http://ebookcentral.proquest.com/lib/abdn/detail.action?docID=2064681.

6. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.5525&rep=rep1&type=pdf

7. StataCorp. 2019. stata statistical software: Release 16. college station, TX: StataCorp LLC.
.

Example of analysis code and data

Stata Code

```
//generating intervention, trend variable and seasonal
variable
generate intervention = (_n >= 13)
generate trend = (time - 13)*intervention
generate seasonal = seq(), f(1) t(6) // 6 monthly seasonal
effect

** Segmented regression analysis **
regress outcome time intervention trend
regress outcome time intervention trend i.seasonal //
adjusting for seasonal effect

// baseline trend
lincomest _cons

// pre-trend
lincomest time

// change in trend
lincomest trend

// post-trend
lincomest time + trend

// change in level
lincomest intervention

** ARIMA code **

tsset time


arima outcome time intervention trend, arima(1,0,0) //
adjusting for autocorrelation of order 1
```

```
arima outcome time intervention trend, arima(1,0,0)
sarima(1,0,0,6) // adjusting for autocorrelation of order 1
and seasonality

// baseline trend
lincomest _cons

// pre-trend
lincomest time

// change in trend
lincomest trend

// post-trend
lincomest time + trend

// change in level
lincomest intervention
```

Data

Table 1. Hypothetical dataset

Time	Outcome	Intervention	Trend	Seasonal
1	10	0	0	1
2	11	0	0	2
3	15	0	0	3
4	17	0	0	4
5	20	0	0	5
6	21	0	0	6
7	25	0	0	1
8	28	0	0	2
9	29	0	0	3
10	29	0	0	4
11	30	0	0	5
12	33	0	0	6
13	28	1	1	1
14	26	1	2	2
15	24	1	3	3
16	22	1	4	4
17	23	1	5	5
18	22	1	6	6
19	21	1	7	1
20	20	1	8	2
21	18	1	9	3
22	18	1	10	4
23	16	1	11	5
24	17	1	12	6