



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Turner, SL;Korevaar, E;Karahalios, A;Forbes, AB;McKenzie, JE

Title:

Interrupted time series datasets from studies investigating the impact of interventions or exposures in public health and social science: a data note

Date:

2025-12-01

Citation:

Turner, S. L., Korevaar, E., Karahalios, A., Forbes, A. B. & McKenzie, J. E. (2025). Interrupted time series datasets from studies investigating the impact of interventions or exposures in public health and social science: a data note. BMC Research Notes, 18 (1), <https://doi.org/10.1186/s13104-024-07055-5>.

Persistent Link:

<https://hdl.handle.net/11343/359672>

License:

CC BY

DATA NOTE

Open Access



# Interrupted time series datasets from studies investigating the impact of interventions or exposures in public health and social science: a data note

Simon L Turner<sup>1</sup> , Elizabeth Korevaar<sup>1</sup> , Amalia Karahalios<sup>2</sup> , Andrew B Forbes<sup>1</sup>  and Joanne E McKenzie<sup>1\*</sup> 

## Abstract

**Objectives** The interrupted time series (ITS) design is commonly used to investigate the impact of an intervention or exposure in public health. There are many statistical methods that can be used to analyse ITS data and to meta-analyse their results. We undertook two empirical studies to investigate: (i) how effect estimates (and associated statistics) compared when six statistical methods were applied to 190 real-world datasets; and (ii) how meta-analysis effect estimates (and associated statistics) compared when the combinations of two ITS analysis methods and five meta-analysis methods were applied to 17 real-world meta-analyses including 283 ITS datasets. Here we present a curated repository of a subset of ITS datasets from these studies.

**Data description** The repository includes 430 ITS datasets curated from the two empirical studies. The datasets are diverse in the populations, interruptions and outcomes examined, and are methodologically diverse in the outcome types, aggregation time intervals, number of timepoints and segments. Most of the datasets are from public health. For each dataset, we provide the outcome value at each timepoint and the segment (indicating different interruptions), along with characteristics of the dataset. This repository may be of value for future research of ITS studies, and as a source of examples of ITS for use in teaching.

**Keywords** Interrupted time series, Meta-analysis, Public health, Empirical evaluation

## Objective

The interrupted time series (ITS) design is commonly used to investigate the impact of population-level interventions (e.g., government policy introduction [1–3]) or exposures (e.g., natural disaster [4, 5]). In an ITS design, data are collected at multiple time points before and after an interruption, and are often aggregated using a summary statistic (e.g., mean, proportion) within a time interval (e.g., weekly, monthly) [6]. A pre-post interruption comparison is then made, while controlling for the pre-interruption trend. The design is often used when randomisation is difficult or impossible, and is less

\*Correspondence:

Joanne E McKenzie

joanne.mckenzie@monash.edu

<sup>1</sup>School of Public Health and Preventive Medicine, Monash University, Level 4, 553 St Kilda Road, Melbourne, VIC 3004, Australia

<sup>2</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, VIC 3010, Australia



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Table 1** Overview of data files

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Datafile 1: Data dictionary	data_dictionary.xls	MS Excel file (.xlsx)	Monash University ( <a href="https://doi.org/10.26180/24287338">https://doi.org/10.26180/24287338</a> ) (22)
Datafile 2: Summary study description	study_information.xls	MS Excel file (.xlsx)	Monash University ( <a href="https://doi.org/10.26180/24287338">https://doi.org/10.26180/24287338</a> ) (22)
Datafile 3: Time series data	time_series_data.xls	MS Excel file (.xlsx)	Monash University ( <a href="https://doi.org/10.26180/24287338">https://doi.org/10.26180/24287338</a> ) (22)
Datafile 4: Version description	version_history.xls	MS Excel file (.xlsx)	Monash University ( <a href="https://doi.org/10.26180/24287338">https://doi.org/10.26180/24287338</a> ) (22)

susceptible to bias compared with other non-experimental designs [6–9]. These features have led to the design's inclusion in systematic reviews, and meta-analyses, to estimate the effects of population-level interruptions [10, 11].

Many statistical methods can be used to analyse ITS data (with the methods differing in how they account for underlying patterns in the time series data) and many methods are available to meta-analyse the resulting effect estimates. To understand whether the choice of ITS analysis method, and when meta-analysing results from ITS studies, whether the combination of ITS analysis and meta-analysis methods matters when applied to real-world datasets, we undertook two empirical evaluations [12, 13]. These empirical evaluations have companion numerical simulation studies that examine the performance of the same ITS analysis and meta-analysis methods under controlled scenarios [14, 15]. The ITS datasets were identified from two methodological reviews [10, 16]. A subset of these ITS datasets have been used to examine the accuracy of results calculated from data digitally extracted from ITS graphs [17].

We have curated the datasets from the empirical studies to form a repository including 430 ITS datasets. The datasets are diverse in the research questions addressed and in their methodological characteristics.

## Data description

### Description of how the ITS datasets were sourced

This repository of 430 datasets has been curated from ITS datasets included in two statistical empirical studies [12, 13]. Details of the methods used to obtain the datasets from each empirical study are now outlined.

Empirical study 1 – Turner et al. [12]: We sourced 190 ITS datasets from 200 ITS studies that investigated the impact of interruptions on public health related outcomes. Full details of the methods to select studies, and the series within, are available in Turner et al. [16, 18]. We sourced ITS datasets using three methods: (i) data provided with the study; (ii) data obtained from authors via email; and, (iii) digital extraction of data from graphs provided in the published manuscripts. Digital extraction was undertaken using the software WebPlotDigitizer, which has been shown to be an accurate tool for data extraction [17, 19, 20]. This repository includes a subset

of the ITS datasets sourced from methods (i) (8 datasets) and (iii) (176 datasets) above.

Empirical study 2 – Korevaar et al. [13]: We sourced 283 ITS datasets from 17 meta-analyses (included in 17 reviews investigating the impacts of, primarily, public health interruptions) that included results from at least two ITS studies. Full details of the methods used to select reviews, meta-analyses, and studies are available in Korevaar et al. [10, 21]. We sourced the ITS datasets using the same methods as for Turner et al. [16]. This repository includes a subset of the ITS datasets obtained from methods (i) (16 datasets), (ii) (220 datasets) and (iii) (10 datasets) above.

### Description of data files

The repository includes four Excel spreadsheet files (Table 1) accessible at <https://doi.org/10.26180/24287338> [22]. A brief description of the files follows:

- Datafile 1 - (data\_dictionary.xls) – is a data dictionary that describes the variables in *Datafile 2* and *Datafile 3*.
- *Datafile 2* - (study\_information.xls) – is the study information file containing an indicator of the repository source (Turner et al. [12] or Korevaar et al. [13]); study citation details; study and series identification numbers that link to the ITS datasets (in *Datafile 3*); description and type of the intervention; description, direction of benefit and type of the outcome (e.g., rate); length of the time series and time interval (e.g., 12 datapoints of monthly data); and, which method was used to source the ITS dataset.
- *Datafile 3* - (time\_series\_data.xls) - is the time series data file containing an indicator of the repository source (Turner et al. [12] or Korevaar et al. [13]); study and series identification numbers (i.e. linking variables with *Datafile 2*); time variable; numerical value of the outcome; segment (1 for the first segment, 2 for the second, etc.); and, an indicator of whether the datapoints were outliers or part of a transition period.
- *Datafile 4* - (version\_history.xls) – this file has been added in anticipation of further data being added to the repository and will include a description of

changes made between versions of the repository as new datasets are added.

If you wish to contribute ITS datasets to the repository, please contact the corresponding author.

### Limitations

We could not obtain data for all the eligible time series included in the reviews [10, 16]. There are additional limitations which vary by the methods used to obtain the data. For data sourced as part of the Turner et al. [12, 16] studies, limitations include:

- The majority of the data series were obtained using digital data extraction, and could contain errors (though these are unlikely to result in any substantive errors [17]).
- Due to difficulty digitally extracting data from particularly long data series not all of the longer series identified in the review [16] are part of this repository.

For data sourced as part of the Korevaar et al. [10, 13] studies, limitations include:

- The descriptions of the outcomes and interventions were extracted from the reviews, and not the primary studies. Therefore, any inaccuracy in the description of these variables in the review, will also be inaccurate in our repository.

Furthermore, our repository does not capture all details of the ITS datasets. For example, we have only captured the description of the first interruption, but not captured any descriptions of the pre-interruption, or potentially, other interruption segments; nor have we captured details of when and where the studies were undertaken. However, for those requiring further detail, citations are provided to the primary studies so that further information can be extracted if required.

### Abbreviations

ITS Interrupted Time Series

### Acknowledgements

We would like to acknowledge the review authors of the Korevaar et al. empirical study who provided the ITS datasets for inclusion in the repository.

### Author contributions

JEM conceived the study and all authors contributed to its design. SLT, EK, JEM, AK and ABF contributed to the data collection of the review. SLT and EK digitally extracted the data. SLT wrote the first draft of the manuscript, with contributions from EK and JEM. SLT, EK, JEM, AK, and ABF contributed to revisions of the manuscript and take public responsibility for its content.

### Funding

This work was supported by the Australian National Health and Medical Research Council (NHMRC) Project Grant (1145273). SLT and EK were

supported through an Australian Government Research Training Program (RTP) Scholarship (administered by Monash University, Australia) and by the Research Support Package of Joanne E McKenzie's NHMRC Investigator Grant (GNT2009612). JEM was supported by an NHMRC Career Development Fellowship (GNT1143429) and an NHMRC Investigator Grant (GNT2009612). The funders had no role in study design, decision to publish, or preparation of the manuscript.

### Data availability

The data described in this data note can be freely and openly accessed on the Monash University repository known as Bridges, <https://doi.org/10.26180/24287338>. Please see Table 1 and reference (22) for details and links to the data.

### Declarations

#### Ethics approval and consent to participate

For the Korevaar et al. [13] empirical study, ethics approval was obtained from the Monash University Human Research Ethics Committee (Project ID 30078). We sought consent from the corresponding author of the review for sharing the provided ITS datasets in the online repository.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 26 November 2023 / Accepted: 19 December 2024

Published online: 22 January 2025

### References

1. Baskerville NB, Brown KS, Nguyen NC, Hayward L, Kennedy RD, Hammond D, et al. Impact of Canadian tobacco packaging policy on use of a toll-free quit-smoking line: an interrupted time-series analysis. *CMAJ Open*. 2016;4(1):E59–65.
2. Baccini M, Carreras G. Analyzing and comparing the association between control policy measures and alcohol consumption in Europe. *Subst Use Misuse*. 2014;49(12):1684–91.
3. Hardie I, Stevely AK, Sasso A, Meier PS, Holmes J. The impact of changes in COVID-19 lockdown restrictions on alcohol consumption and drinking occasion characteristics in Scotland and England in 2020: an interrupted time-series analysis. *Addiction*. 2022;117(6):1622–39.
4. Zhang N, Song D, Zhang J, Liao W, Miao K, Zhong S, et al. The impact of the 2016 flood event in Anhui Province, China on infectious diarrhea disease: an interrupted time-series study. *Environ Int*. 2019;127:801–9.
5. Phung D, Chu C, Rutherford S, Nguyen HLT, Luong MA, Do CM, et al. Heavy rainfall and risk of infectious intestinal diseases in the most populous city in Vietnam. *Sci Total Environ*. 2017;580:805–12.
6. 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.
7. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2016. [dyw098](https://doi.org/10.1093/ije/dyw098).
8. 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: Br Med J*. 2015;350:h2750.
9. Penfold RB, Zhang F. Use of interrupted Time Series Analysis in evaluating Health Care Quality improvements. *Acad Pediatr*. 2013;13(6):S38–44.
10. Korevaar E, Karahalios A, Turner SL, Forbes AB, Taljaard M, Cheng AC, 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.
11. Vidanapathirana J, Abramson MJ, Forbes A, Fairley C. Mass media interventions for promoting HIV testing. *Cochrane Database Syst Rev*. 2005(3):Cd004775.
12. 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 Med Res Methodol.* 2021;21(1):134.
13. Korevaar E, Turner SL, Forbes AB, Karahalios A, Taljaard M, McKenzie JE. Comparison of statistical methods used to meta-analyse results from interrupted time series studies: an empirical study. *BMC Med Res Methodol.* 2024;24(1):31.
  14. Korevaar E, Turner SL, Forbes AB, Karahalios A, Taljaard M, McKenzie JE. Evaluation of statistical methods used to meta-analyse results from interrupted time series studies: a simulation study. *Res Synth Methods.* 2023;14(6):882–902.
  15. Turner SL, Forbes AB, Karahalios A, Taljaard M, McKenzie JE. Evaluation of statistical methods used in the analysis of interrupted time series studies: a simulation study. *BMC Med Res Methodol.* 2021;21(1):1–181.
  16. Turner SL, Karahalios A, Forbes AB, Taljaard M, Grimshaw JM, Cheng AC, et al. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review. *J Clin Epidemiol.* 2020;122:1–11.
  17. Turner SL, Korevaar E, Cumpston MS, Kanukula R, Forbes AB, McKenzie JE. Effect estimates can be accurately calculated with data digitally extracted from interrupted time series graphs. *Res Synthesis Methods.* 2023;14(4):622–38.
  18. Turner SL, Karahalios A, Forbes AB, Taljaard M, Grimshaw JM, Cheng AC, et al. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: protocol for a review. *BMJ Open.* 2019;9(1):e024096.
  19. Rohatgi A. *WebPlotDigitizer.* 4.2 ed. San Francisco, California, USA 2019.
  20. Drevon D, Fursa SR, Malcolm AL. Intercoder Reliability and Validity of Web-PlotDigitizer in extracting Graphed Data. *Behav Modif.* 2017;41(2):323–39.
  21. Korevaar E, Karahalios A, Forbes AB, Turner SL, McDonald S, Taljaard M, et al. Methods used to meta-analyse results from interrupted time series studies: a methodological systematic review protocol. *F1000Res.* 2020;9:110.
  22. Turner S, Korevaar E, Forbes A, Karahalios A, McKenzie J. Interrupted time series datasets from studies investigating the impact of interventions or exposures in. *Public health and social science: a data note.* Monash University; 2023. <https://doi.org/10.26180/24287338>.

### Publisher's note

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