PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Methods, Applications, Interpretations and Challenges of Interrupted Time Series (ITS) Data: Protocol for a Scoping Review
AUTHORS	Ewusie, Joycelyne; Blondal, Erik; Soobiah, Charlene; Beyene, Joseph; Thabane, Lehana; Straus, Sharon; Hamid, Jemila

VERSION 1 - REVIEW

REVIEWER	Robert Penfold Kaiser Permanente of Washington Research Institute
REVIEW RETURNED	02-Mar-2017

GENERAL COMMENTS	This scoping review proposes an approach to develop guidance on methods for interrupted time series analysis. A manuscript or book providing such guidance would be useful. However, the content of this proposal is unclear about how this will be accomplished. Page 3: ITS designs are useful for examining changes in outcomes relative to an underlying secular trend as well as competing non-contemporaneous interventions. ITS designs do not test the impact of implementation strategies per se. Segmented regression and ARIMA models are not mutually exclusive approaches. The introduction and approach overall must discuss the quasi-experimental design as well as the statistical approach. Any longitudinal regression model can be turned into an ITS. However, whether one uses a single time series, difference-in-differences, or triple differences approach is critical information needed to assess the quality of the research conducted and appropriateness of the statistical approach. Page 6 I disagree with the eligibility criteria. Specifically, I would require that there be at least 8 time periods pre- and post- index date. Second, "soft" index dates (where an intervention is rolled out over a period of time) are the norm rather than the rule in health care. There are several methods for handling these roll-out periods analytically and these should be discussed in the review.

REVIEWER	Javier Virues-Ortega
	The University of Auckland, New Zealand
REVIEW RETURNED	14-Mar-2017

GENERAL COMMENTS	Below I describe some reasons for concern. I hope they could help to make your study even better.
	1. Confining target studies to those published in English is a source of bias acknowledged by many (see for example McDonagh et al 2013, Avoiding Bias in Selecting Studies)
	2. 3-13 What is meant by "robust quasi-experimental design"?
	3. 3-35 How often are they included as part of Cochrane systematic review and meta-analyses?
	4. 4-8 Single-subject experimental designs (SSED) have been characterized by some as a form of ITS. Please, pay more attention to these designs (see for example Kazdin, 2011 or Morgan & Morgan, 2008) and to the journals that publish them (e.g., Journal of Applied Behavior Analysis). Please, acknowledge also the statistical models developed for SSED specifically (see for example the recent work by William Shadish).
	5. 5-7 Consider incorporating meta-analytical methods in order to evaluate potential systematic biases present in ITS studies relative to comparable RCTs. This would require focusing on an area for which sufficient studies from both designs are available. Ioannidis et al. (2001) provide an example of a visual display to conduct such a comparison (they compared RCT vs non-randomized studies).
	6. 5-43 It would be important to include Psychinfo given the prevalence of ITS in the psychological literature. On another note, if the intent of the authors is to characterize current trends it may be sensible to include a search start date.
	7 7-31 "Data Extraction" may be a more standard term in this context.
	8. 8-22 The data synthesis section is poorly described. Please provide greater detail on the expected outcomes and how would they be summarized and presented. Quantitative synthesis strategies, in particular, would strengthen the review.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Robert Penfold
Institution and Country: Kaiser Permanente of Washington Research Institute
Please state any competing interests or state 'None declared': none declared

Please leave your comments for the authors below

General Comment
We thank the reviewer for helpful comments and suggestions. We have revised our manuscript accordingly and provided a point by point response to the comments below. We have also provided a

list of some references at the end of this comment section.

"This scoping review proposes an approach to develop guidance on methods for interrupted time series analysis. A manuscript or book providing such guidance would be useful. However, the content of this proposal is unclear about how this will be accomplished".

We agree with the reviewer that a manuscript or book providing guidance on methods for interrupted time series analysis would be useful. That is in fact, what motivated this scoping review. However, the objective of this study is to systematically review literature with the aim of identifying available methods that have been used in the analysis of interrupted time series data. We also aim to evaluate the application of the methods and identify methodological gaps, with the purpose of addressing the methodological gaps. We would also like to highlight that the review is motivated by an interdisciplinary project, conducted by our group, in evaluating mobility of elderly before and after a knowledge translation intervention, for which we encountered lack of optimal statistical method that incorporates heterogeneity across patients within a hospital as well as heterogeneity among the hospitals (settings). As a first step towards filling this methodological gap, we decided to conduct a review of available methods and their applications.

We therefore believe that this review will provide the ground work towards developing a guidance manuscript or book on methods for analyzing ITS data as well as help identify potential gaps (in addition to the one mentioned above) in existing methodology. Having said that, we have modified the manuscript accordingly to clearly define the objectives of our review. (Page 4, Paragraph 3; Page 10, Paragraph 2).

Page 3:
"ITS designs are useful for examining changes in outcomes relative to an underlying secular trend as well as competing non-contemporaneous interventions. ITS designs do not test the impact of implementation strategies per se."

We thank the reviewer in raising an important question regarding "impact" of intervention. By impact, we mean "effect" of intervention on outcomes, however, we understand that our use of "impact" might be misleading since the term is often used to mean "influence of health care practice or policy". In this regard, we agree with the reviewer that ITS designs although useful for examining changes in outcomes does not test the impact of implementation strategies. We have therefore revised the manuscript accordingly. (Page 3, Paragraph 2)

"Segmented regression and ARIMA models are not mutually exclusive approaches".

We agree that in a broad sense, these two methods may not be mutually exclusive. In fact, both can be described as linear models, where the difference can be described in a way they handle serial correlation and account for other factors such as seasonality. Having said that, segmented regression and ARIMA are considered different statistical modeling strategies in current ITS literature (Zhang et al., 2009; Shardell et al., 2007). Some differences between the methods are described in literature. For instance, ARIMA model requires sample sizes of at least 50 consecutive time points or 20 time points pre-intervention and assume a complex error correlation structure while segmented regression tolerate fewer time points and a simpler error correlation structure.

"The introduction and approach overall must discuss the quasi-experimental design as well as the statistical approach. Any longitudinal regression model can be turned into an ITS. However, whether one uses a single time series, difference-in-differences, or triple differences approach is critical information needed to assess the quality of the research conducted and appropriateness of the statistical approach".

We thank the reviewer for highlighting these important methodological issues. We agree that the type of design influences choice of modeling and statistical analysis. As stated above, our overall objective in this study is to systematically search literature with the aim identify the various statistical methods that have been utilized in analyzing interrupted time series data. Therefore, every statistical approach that has been used to analyze ITS data will be identified and described, including some of the important issues highlighted by the reviewer in this comment.

The difference in differences or triple difference designs are usually classified as different quasi-experimental designs from an interrupted time series design (Shadish et al., 2002; Harris et al., 2006), and will not be included in this review. However, we will discuss the various quasi experimental designs in our main manuscript and how the type of design influences the statistical method used. We have also provided more information on the different quasi experimental designs in the protocol accordingly. (Page 3, Paragraph 1).

Page 6
"I disagree with the eligibility criteria. Specifically, I would require that there be at least 8 time periods pre- and post- index date. Second, "soft" index dates (where an intervention is rolled out over a period of time) are the norm rather than the rule in health care. There are several methods for handling these roll-out periods analytically and these should be discussed in the review".

We are aware that at least 8 time points have been recommended as ideal by some researchers (Penfold & Zhang, 2013) and our eligibility criteria does not in any way support use of 3 time points in ITS analysis. In fact, we strongly agree with the reviewer that at least 8 time points pre- and post (if not more depending on the variability of data) should be used and we will highlight this in the review paper. Nevertheless, the EPOC criteria [2002] suggested that at least 3 time points will be enough to qualify as a short time series and thus an ITS study, which in turn means studies involving evidence synthesis use 3 time points as eligibility criteria.

In our review, we decided to use the least recommended number of time points to include all articles that reported the use of ITS. We plan to discuss the characteristics of included studies including number of time points pre- and post- intervention. We will also discuss the limitations of the methods (e.g. 3 time points vs 8 time points) and the appropriateness of the statistical approaches utilized with respect to design.

We also agree with the reviewer and are aware of the soft index dates being the norm in clinical practice. In fact, in some of the knowledge translation projects we have worked on, we mostly have "implementation" period in addition of pre- and post- intervention periods. This is mainly because, as the reviewer indicated, interventions are rolled out over a period. We plan to discuss this in detail in our review as this influences the design as well as analysis of the data resulting from such ITS study. We thank the reviewer, once again, for raising this and we have modified the protocol accordingly to reflect this by providing more details about the type of intervention. Specifically, we added this statement, "… or a definition of the time within which the intervention was rolled out since most interventions are rolled out over a period." (Page 6 Paragraph 2).

References
Zhang F, Wagner AK, Soumerai SB, Ross-Degnan D. Methods for estimating confidence intervals in interrupted time series analyses of health interventions. Journal of Clinical Epidemiology 2009,62:143-148
Shardell M, Harris AD, El-Kamary S, Furuno JP, Miller R, Perencevich EN. Statistical analysis and application of quasi experiments to antimicrobial resistance intervention studies. Antimicrobial resistance 2007, 45:901:907.
William R. Shadish, Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Wadsworth Cengage learning.

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical informatics. Journal of the American Medical Informatics Association, 13(1), 16-23.

Penfold, R. B., & Zhang, F. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. Academic pediatrics, 13(6), S38-S44.

Bero L, Grilli R, Grimshaw JM, et al. The Cochrane Effective Practice and Organisation of care Group (EPOC) Module. In: The Cochrane Library. Oxford: Update Software, 2002 Issue 1.

Reviewer: 2
Reviewer Name: Javier Virues-Ortega
Institution and Country: The University of Auckland, New Zealand
Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below
Dear authors,
Below I describe some reasons for concern. I hope they could help to make your study even better.

We thank the reviewer for helpful comments and suggestions. We have revised the manuscript accordingly and provided point by point responses below.

1. "Confining target studies to those published in English is a source of bias acknowledged by many (see for example McDonagh et al 2013, Avoiding Bias in Selecting Studies)"

We agree with the reviewer that in most systematic reviews, confining the target studies within the English language leads to bias. However, we would like to highlight that this is methods review and the main objective is to identify existing methods in ITS analysis and to understand how the methods, overall, are being used in current practice. We believe this can be achieved by using studies published in English, which comprises majority of the available literature. Having said that, we acknowledge this as a limitation and we have reported it accordingly in the protocol. (Page 2 line 47).

2. 3-13 "What is meant by "robust quasi-experimental design"?"

There are different types of quasi-experimental designs. Robust quasi-experimental design refers to the strongest quasi-experimental design with respect to internal validity and inferential power (Shadish et al., 2002). Interrupted time series has been identified as the strongest quasi-experimental design with its internal validity. Shadish et al. [2002], have provide detailed information on the different quasi-experimental designs and how they differ in terms of the number, transparency and testability of their assumptions as well as their performance with regards to inferential power and internal validity. Other researchers such as Harris et al., [2006], have confirmed this assertion and thus our use of the phrase "robust quasi-experimental design" for ITS was influenced by this.

3. 3-35 "How often are they included as part of Cochrane systematic review and meta-analyses?"

There has not been any study, to our knowledge, that looks at how often ITS studies are included in Cochrane systematic reviews and meta-analyses. Therefore, we are currently unable to provide a definite answer to this question. However, according to the Effective Practice and Organization of Care (EPOC) Cochrane Group [2002], due to the impracticability of using randomized trials to test the effectiveness of all interventions within the scope of EPOC, guidelines were developed to include studies that utilized ITS designs. Hence, since there are guidelines available for including ITS studies for such reviews, we believe that they are being included in Cochrane systematic reviews quite often.

Ramsey et al. [2003] also confirms this assertion in their paper, where they examined the methodological quality of ITS studies included in systematic reviews.

Having said that, it would be useful in the future to conduct a review of systematic reviews and meta-analyses to investigate how often ITS studies are being included in evidence synthesis as well as to examine if they are indeed being included in meta-analyses using appropriate methodology.

4. 4-8 "Single-subject experimental designs (SSED) have been characterized by some as a form of ITS. Please, pay more attention to these designs (see for example Kazdin, 2011 or Morgan & Morgan, 2008) and to the journals that publish them (e.g., Journal of Applied Behavior Analysis). Please, acknowledge also the statistical models developed for SSED specifically (see for example the recent work by William Shadish)."

We agree that Single Subject Experimental Designs (SSEDs) may have been characterized as a form of ITS by some researchers and thus will be captured in our review, since they meet our eligibility criteria. We anticipate that we will be able to comment on designs that have been characterized as a form of ITS including SSEDs and the corresponding approaches to modelling such data. Nevertheless, there is no statistical inference involved in single-subject experimental data, and hence SSEDs will not be the focus of our review.

5. 5-7 "Consider incorporating meta-analytical methods in order to evaluate potential systematic biases present in ITS studies relative to comparable RCTs. This would require focusing on an area for which sufficient studies from both designs are available. Ioannidis et al. (2001) provide an example of a visual display to conduct such a comparison (they compared RCT vs non-randomized studies)."

We thank the reviewer for raising this interesting issue and we agree that meta-analytical methods including evaluating potential systematic biases are crucial when conducting systematic reviews involving outcomes of some sort. However, our study is a review of methods, where the main objective is to identify available methodologies, investigate methodological gaps as well as describe how the available methods are being used in health research. Thus, no patient outcome will be collected from any of the studies and hence meta-analysis does not apply to our study. However, we anticipate that we will encounter studies involving meta-analysis in our review and we plan to explore how ITS studies are being incorporated in meta-analysis and how authors handled biases.

6. 5-43 "It would be important to include Psychinfo given the prevalence of ITS in the psychological literature. On another note, if the intent of the authors is to characterize current trends it may be sensible to include a search start date. "

We thank the reviewer for the suggestion. We will include PsycINFO database in our search and our protocol has been revised accordingly. (Page 5 line 51).

The search is done from inception of the databases to ensure the review is comprehensive. We are concerned that if we limit the search date, we may miss useful studies and key methodologies for ITS analysis. However, in our review we will also discuss emerging methods as well as current trends in ITS application.

7. 7-31 " "Data Extraction" may be a more standard term in this context. "

We thank the reviewer for this suggestion. We used the term data abstraction to be consistent with the terms used in systematic or scoping review literature. We would, therefore, like to keep the term. However, if the reviewer strongly feels that data extraction is a more standard term we will modify our manuscript accordingly.

8. 8-22 "The data synthesis section is poorly described. Please provide greater detail on the expected

outcomes and how would they be summarized and presented. Quantitative synthesis strategies, in particular, would strengthen the review."

We realize based on this comment that our use of the term "data synthesis" is quite misleading, so we have changed the sub-title to "data summary" to reflect what the expected results will be. (Page 8 line 35).

This is because, our study is a methodological review where the main objective, as stated above, is to identify available methods used in the analysis of ITS studies, identify potential methodological gaps in current literature, as well as to investigate their application in health research. Due to the nature of this review, no outcomes will be extracted and hence data synthesis (as in meta-analysis) will not be performed. In terms of results, we will provide a summarized description of the methods identified including the assumptions made and their strengths and limitations as well as the frequency of application in health research. We will also identify gaps in current methods and discuss their appropriateness in terms of different designs, data types (continuous, binary, count etc.) and distributions. For the application papers, summarized results will include the frequency of use of the methods over time, where they are applied in terms of setting and kind of data. We believe the title "Data summary" is, therefore, more appropriate than "Data Synthesis".

References

William R. Shadish, Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Wadsworth Cengage learning.

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical informatics. Journal of the American Medical Informatics Association, 13(1), 16-23.

Bero L, Grilli R, Grimshaw JM, et al. The Cochrane Effective Practice and Organisation of care Group (EPOC) Module. In: The Cochrane Library. Oxford: Update Software, 2002 Issue 1.

Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behaviour change strategies. International Journal of Health Technology Assessment in Health Care 2003; 19:613-23.

VERSION 2 – REVIEW

REVIEWER	Robert Penfold Kaiser Permanente Washington Health Research Institute
REVIEW RETURNED	20-Apr-2017

GENERAL COMMENTS	I strongly disagree with the scope of the proposed manuscript. This literature review will have little to no impact on the conduct of interrupted time series analysis. I leave you with a quote from Judy Singer: "You cannot fix with statistics that which you bungle by design".