Draft
EPOC Methods Paper
Including Interrupted Time Series (ITS) Designs in a EPOC Review

Although well-conducted randomised trials provide the most reliable evidence on the effectiveness of interventions, these are not always feasible for interventions within the EPOC scope. Examples include the diffusion of guidelines through mass media or the implementation of an organisational change to a health care delivery system. The following guidelines have been developed to assist reviewers in making decisions about when to include studies that use interrupted time series (ITS) designs and how to assess their methodological quality. The next step in the development of these guidelines will be to provide further details on the statistical analysis from the review of the impact of mass media on health services utilisation (Grilli et al 1997).

What is an ITS design?

Simply put, ITS designs are multiple observations **over time** that are 'interrupted' usually by an intervention or treatment. The investigators must indicate a specific point in time when the intervention occurred. A control group may or may not be present. Cook and Campbell describe different types of effects of the intervention: indicators of the effect (level, slope, variance and pattern of seasonality), permanence of the effect (continuous or discontinuous) and the type of impact (immediate or delayed) .

Firstly, there may be a discontinuity or change in level at the point where the intervention occurred. The pre and post intervention slopes would have different intercepts. The second change occurs when there is a difference in slopes. Other types of changes include post intervention changes in variances around each mean and changes related to seasonality. Effects can also be characterised as continuous (no decay over time) or discontinuous (decay or improvement over time). Effects can also be instantaneous or delayed following implementation.

ITS designs are subject to threats to internal validity that are related to history (such as seasonality) that influence the dependent variable, maturation bias where there is a pattern of improvement in the experimental group prior to the intervention, instrumentation bias, for example changes in the way records are kept or the way the outcomes are measured, and selection bias which could cause a differential drop out in the experimental group.

In order to provide some protection against these threats to internal validity, Cook and Campbell suggest that about 50 observations may be required to estimate the correlated error but an adequate analysis may require less observations depending on the expected impact of the intervention. For the purposes of analysis, ITS designs will be considered as long or short series.

Statistical analysis of ITS designs

Statistical methods based on ordinary least squares methods are inappropriate for analysing ITS designs partly because these methods assume independence of errors. When events or behaviours are measured over time, they are usually correlated with each other resulting in biased standard deviations of the parameter estimates.

Long time series

Cook and Campbell suggest the use of autoregressive integrated moving average (ARIMA) models developed by Box and Jenkins (1976) which are designed to provide unbiased estimates of the error in a series. First the 'noise' in a series is modelled, then the invention component is added to the model. The critical issue is to establish whether the intervention adds significantly to predicting the behaviour of a time series over and above the prediction derived from understanding the regular and seasonal components of the noise. ARIMA models require at least 20 observation points **pre-intervention**.

Short time series
This type of series also consists of pre and post intervention phases. This type of series need to have at least three observation points in the pre-intervention phase and three in the post intervention phase. The series may be modelled using multiple t-tests, analysis of variance and repeated measures analysis.

Including studies with ITS designs

The following two criteria must be met for a study with an ITS design to be included in an EPOC review:

a) Clearly defined point in time when the intervention occurred

Score DONE if the investigators report that intervention occurred at a clearly defined point in time
Score NOT CLEAR if not reported (will be treated as NOT DONE if information cannot be obtained from the authors)
Score NOT DONE if reported that intervention did not occur at a clearly defined point in time

b) At least three data points before and three after the intervention

Score DONE if 3 or more data points recorded before and 3 or more data points recorded after the intervention.
Score NOT CLEAR if not specified in paper, e.g. number of discrete data points not mentioned in text or tables (will be treated as NOT DONE if information cannot be obtained from the authors)
Score NOT DONE if less than 3 data points recorded before and 3 data points recorded after intervention

If you scored NOT DONE for either of the above criteria in items a) or b), the study should not be included in an EPOC review. Some studies may have only 2 points before and after and these studies should be entered into the Excluded Studies Table.

Quality criteria for ITS designs

The following seven standard criteria should be used to assess the methodological quality of ITS designs included in EPOC reviews. Each criterion is scored **DONE, NOT CLEAR or NOT DONE**. The results of the quality assessment for each study are reported in the Table of Included Studies in RevMan. Examples can be obtained from the EPOC Group Co-ordinator.

a) **Protection against secular changes**:

i) The intervention is independent of other changes

Score DONE if the intervention occurred independent of other changes over time

Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors)

Score NOT DONE if reported that intervention was not independent of other changes in time

ii) There are sufficient data points to enable reliable statistical inference

Score DONE
(a) If at least twenty points are recorded before the intervention **AND** the authors have done a traditional time series analysis (ARIMA model)
OR
(b) If at least 3 points are recorded pre and post intervention **AND** the authors have done a repeated measures analysis
OR
(c) If at least 3 points are recorded pre and post intervention **AND** the authors have used ANOVA or multiple t-tests **AND** there are at least 30 observations per data point.

Score NOT CLEAR if not specified in paper e.g. number of discrete data points not mentioned in text or tables (will be treated as NOT DONE if information cannot be obtained from the authors)

Score NOT DONE if any of the above conditions are unmet

iii) Formal test for trend. Complete this section if authors have used ANOVA modelling.

Score DONE if formal test for change in trend using appropriate method is reported (e.g. see Cook & Campbell 1979)

Score NOT CLEAR if not specified in the paper (will be treated as NOT DONE if information cannot be obtained from the authors)

Score NOT DONE if formal test for change in trend has not been done.

b) Protection against detection bias

i) Intervention unlikely to affect data collection

Score DONE if the investigators report that the intervention itself was unlikely to affect data collection (for example, sources and methods of data collection were the same before and after the intervention)

Score NOT CLEAR if not reported (will be treated as NOT DONE if information cannot be obtained from the authors)

Score NOT DONE if the intervention itself was likely to affect data collection (for example, any change in source or method of data collection reported)

ii) Blinded assessment of primary outcome(s)*

Score DONE if the authors state explicitly that the primary outcome variables were assessed blindly OR the outcome variables are objective e.g. length of hospital stay, drug levels as assessed by a standardised test

Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors)

Score NOT DONE if the outcomes were not assessed blindly

***** Primary outcome(s) are those variables that correspond to the primary hypothesis or question as defined by the authors. In the event that some of the primary outcome

variables were assessed in a blind fashion and others were not, score each separately.

c) Completeness of data set

Score DONE if data set covers 80 - 100% of the total number of participants or episodes of care in the study

Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors)

Score NOT DONE if data set covers less than 80% of the total number of participants or episodes of care in the study

d) Reliable primary outcome measure(s)*

Score DONE if two or more raters with at least 90% agreement or kappa greater than or equal to 0.8
OR the outcome is obtained from some automated system e.g. length of hospital stay, drug levels as assessed by a standardised test

Score NOT CLEAR if reliability is not reported for outcome measures that are obtained by chart extraction or collected by an individual (will be treated as NOT DONE if information cannot be obtained from the authors)

Score NOT DONE if agreement is less than 90% or kappa is less than 0.8

* In the event that some outcome variables were assessed in a reliable fashion and others were not, score each separately.

Example from an EPOC review[1]*

Grilli and colleagues (Grilli et al 1997) conducted a systematic review of the impact of mass media campaigns on health services utilisation. They included 22 papers published between 1979 and 1995 reporting 17 time series. Most campaigns were aimed at promoting of the use of specific health services (either cancer screening or immunisation programmes, or emergency services for patients with suspected myocardial infarction). Using the criteria already described, the authors assessed the quality of the studies. They found that most studies only described the time series data without any statistical analysis, or based their interpretation of results on a comparison of means before and after the intervention.

When information about individual observations over time was reported only graphically in the original paper, the authors derived the data set by computer scanning the figures. They had used this approach previously and found it to be reliable (Grilli R et al 1993). Consistency between the data collected with this approach and those explicitly reported on papers (when this information was provided) was reported to be good and discrepancies were never greater than 1%.

Data were analysed using an auto regressive integrated moving average (ARIMA) model to isolate the effect of the intervention from existing time trends (Cook & Campbell 1979). The authors then estimated a regression coefficient (with its standard error) that described the effect of the campaign. The direction of effect (e.g. positive or negative) was standardised so that a negative coefficient described an improvement in outcome attributable to the intervention.

The authors then pooled the results from individual studies using the random effects model described by DerSimonian & Laird (DerSimonian & Laird 1986). They chose this method

[1] Thanks to Roberto Grilli and colleagues for allowing us to use their review as an example.

because it does not assume a single underlying (fixed) treatment effect, but takes observed differences that cannot be explained by chance into account in the pooled estimate and its precision. When the impact of the intervention was assessed in individual studies on more than one outcome measure, the outcome that best reflected the targeted intervention was selected for pooling. Where there were multiple appropriate outcomes the median effect was selected. Where there were only two, the more conservative result was selected.

References

Cook TD, Campbell DT. Quasi-experiments: interrupted time-series designs in Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin Company, Boston, MA 1979.

DerSimonian R, Laird N. Meta analysis in clinical trials. Controlled Clinical Trials 1986;7:177-188.

Grilli R, Freemantle N, Minozzi S, Domenighetti G, Finer D. The impact of mass media on health services utilisation [Protocol]. In: Bero L, Grilli R, Grimshaw J, Oxman A (eds.) Effective Practice and Organisation of Care Module of The Cochrane Database of Systematic Reviews, [updated 01 December 1997]. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 1. Oxford: Update Software; 1998. Updated quarterly.

Grilli R, Oxman A.D., Julian J.A. Chemotherapy for advanced non-small-cell lung cancer. How much benefit is enough? J Clin Oncol ,11:1866-1872,1993.

McCain LJ. McCleary R. The statistical analysis of the simple interrupted time series quasi-experiment. in Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin Company, Boston, MA 1979.