Interrupted Time Series Analysis for Single Series and Comparative Designs: Using Administrative Data for Healthcare Impact Assessment

Joseph M. Caswell, Ph.D.

Lead Analyst

Institute for Clinical Evaluative Sciences (ICES) North
and
Epidemiology, Outcomes & Evaluation Research
Health Sciences North Research Institute (HSNRI)
Northeast Cancer Centre

Presentation Overview

- Quasi-experimental research
- Interrupted time series (ITS)
- ITS single series with example
- ITS comparative with example
- Autocorrelation
- Adjusting standard errors (SE) in SAS
- Guide and macro

Quasi-Experimental Research

- Experimental research:
  - "Gold-standard" is randomized controlled trial (RCT)
  - i.e., drug trials (random assignment to treatment and placebo groups)
  - Treatment and control groups balanced on baseline measures
  - Can be time-consuming, expensive, and even unethical (i.e., withholding care)

- Quasi-experimental research:
  - Alternative(s) available when RCT is not an option
  - Quickly implemented, cost-efficient
  - Often times, observational data are already available
  - Various statistical methodologies for observational studies (i.e., propensity scores, ITS, instrumental variables, etc)

Observational Studies

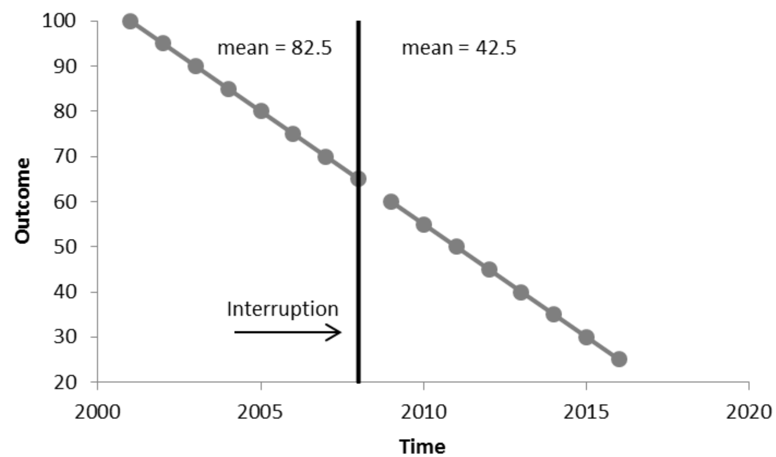
- Administrative data:
  - Routinely collected by hospitals and other healthcare facilities
  - Great source for conducting observational health studies
- ICES data holdings:
  - Examples:
    - Ontario Cancer Registry (OCR)
    - Registered Persons Database (RPDB)
    - Discharge Abstract Database (DAD)
    - Ontario Health Insurance Plan (OHIP)
    - Many, many more
- We can use administrative data to build a study cohort and create various indicator variables

What is ITS Analysis?

- Increasingly popular quasi-experimental alternative
- Analysis of time series data (i.e., an outcome measured over time)
- Comparison before and after an intervention or interruption
- Particularly useful for assessing impact of policy or some other healthcare initiative

Time Series Caveats

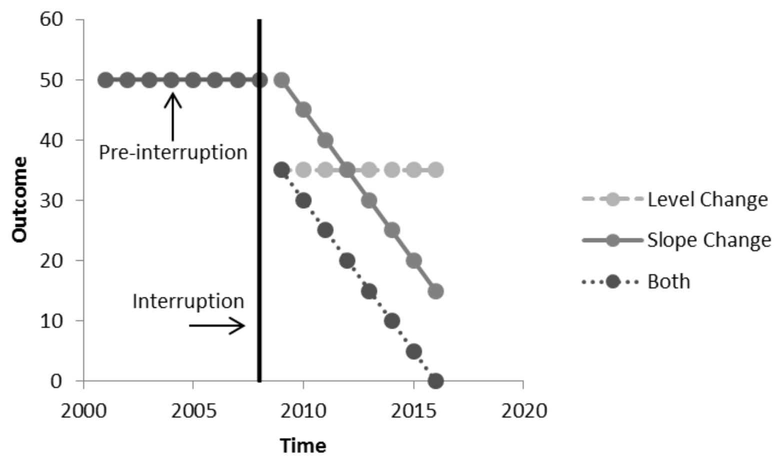
- Example where simple pre- to post- comparison would be misleading



- Must control for pre-interruption trend
- Must also control for any autocorrelation (will get to this later!)

ITS Analysis

- Can identify different effects:
  - Level change (immediate effect)
  - Slope change (sustained effect)
  - Both



Single Series ITS Analysis

- Single time series for outcome variable
  - Example: annual rates of influenza, monthly counts of administered chemotherapy, etc
- Measured before and after some intervention
  - Example: implementing a new hand hygiene regimen, changing policy for use of chemotherapy, etc
- Are there significant changes in level and/or slope following the intervention?

Some Statistical Methodology

- ITS analyses use regression-based techniques
- Added dummy variables for ITS

- Standard linear regression:

  $y = \alpha + \beta x + \varepsilon$

  where α = intercept, β = coefficient, x = independent variable, ε = residual (error)

- Single ITS based on segmented linear regression:

  $y = \alpha + \beta_1 T + \beta_2 X + \beta_3 XT + \varepsilon$

  where T = time,

  X = study phase,

  XT = time after interruption

Year	Rate	*T*	*X*	*XT*
2001	31.67	1	0	0
2002	30.19	2	0	0
2003	32.44	3	0	0
2004	31.50	4	0	0
2005	29.62	5	0	0
2006	30.18	6	0	0
2007	29.76	7	0	0
2008	29.89	8	0	0
2009	25.42	9	1	1
2010	24.26	10	1	2
2011	25.11	11	1	3
2012	24.07	12	1	4
2013	23.95	13	1	5
2014	22.78	14	1	6
2015	21.12	15	1	7

Single ITS Example

- What do the results tell us?

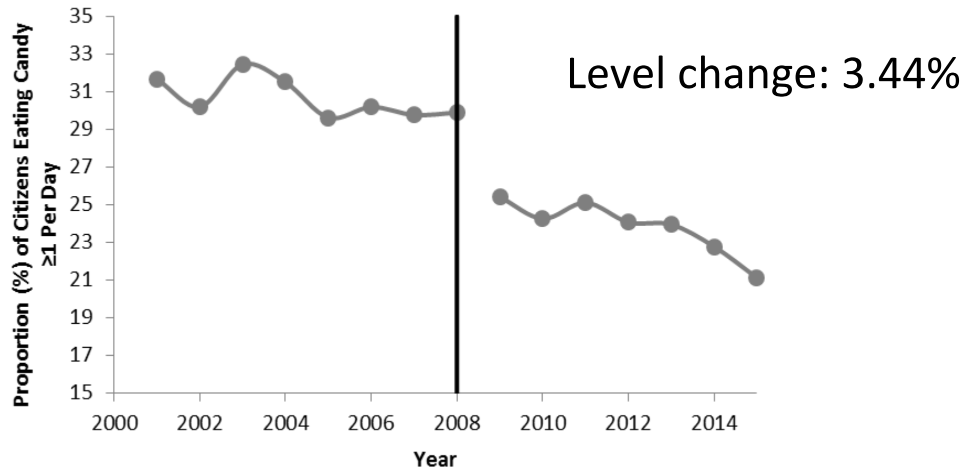
- Fictional example:

  - The town of Squaresville kept records of % population eating candy at least once per day

  - Implemented new candy tax in 2008 (interruption)

  - Data from before and after candy tax analyzed

Single ITS Example



RESULTS

Parameter	Interpretation	Estimate	Standard Error	Probability
β_1	Pre- Trend	-0.27702	0.1242	0.0475
β_2	Post- Level Change	-3.43952	0.8562	0.0020
β_3	Post- Trend Change	-0.33083	0.1964	0.1202
$\beta_1+\beta_3$	Post- Trend	-0.60786	-	<.0001

Comparative Design ITS Analysis

- We can strengthen ITS approach by including comparable "control" series (i.e., no interruption)
- Outcome measured from two sources (treatment and control) during same time period
- Were level and/or slope changes of treatment series significantly different from control series?
- Used far less often compared to single series ITS, even when control series are available

Building on Single Series Method

- Treatment and control time series are appended

- Regression equation is expanded:

  $y = \alpha + \beta_1 T + \beta_2 X + \beta_3 XT + \beta_4 Z + \beta_5 ZT + \beta_6 ZX + \beta_7 ZXT + \varepsilon$

  where Z = treatment or control,

  ZT = time for treatment and 0 for control,

  ZX = study phase for treatment and 0 for control,

  ZXT = time after interruption for treatment and 0 for control

Year	Rate	T	X	TX	Z	ZT	ZX	ZTX
2001	31.67	1	0	0	1	1	0	0
2002	30.19	2	0	0	1	2	0	0
2003	32.44	3	0	0	1	3	0	0
2004	31.50	4	0	0	1	4	0	0
2005	29.62	5	0	0	1	5	0	0
2006	30.18	6	0	0	1	6	0	0
2007	29.76	7	0	0	1	7	0	0
2008	29.89	8	0	0	1	8	0	0
2009	25.42	9	1	1	1	9	1	1
2010	24.26	10	1	2	1	10	1	2
2011	25.11	11	1	3	1	11	1	3
2012	24.07	12	1	4	1	12	1	4
2013	23.95	13	1	5	1	13	1	5
2014	22.78	14	1	6	1	14	1	6
2015	21.12	15	1	7	1	15	1	7
2001	30.81	1	0	0	0	0	0	0
2002	30.96	2	0	0	0	0	0	0
2003	31.23	3	0	0	0	0	0	0
2004	28.65	4	0	0	0	0	0	0
2005	29.33	5	0	0	0	0	0	0
2006	29.10	6	0	0	0	0	0	0
2007	30.27	7	0	0	0	0	0	0
2008	28.64	8	0	0	0	0	0	0
2009	27.95	9	1	1	0	0	0	0
2010	29.55	10	1	2	0	0	0	0
2011	29.14	11	1	3	0	0	0	0
2012	26.87	12	1	4	0	0	0	0
2013	28.72	13	1	5	0	0	0	0
2014	27.89	14	1	6	0	0	0	0
2015	29.82	15	1	7	0	0	0	0

Comparative ITS Example

- What do the results tell us?

- Fictional example:

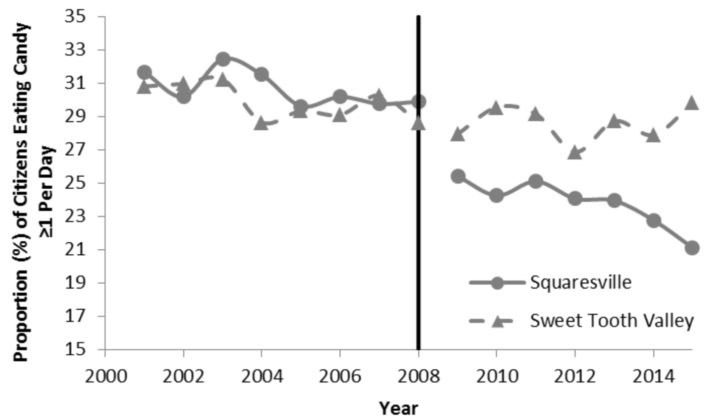
  - The town of Squaresville wanted to compare their results to a control series

  - Another nearby town, Sweet Tooth Valley, also kept records of % population eating candy at least once per day

  - Sweet Tooth Valley did not implement a candy tax in 2008 (no interruption)

  - Data sampled at same rate and during same time period as Squaresville time series

Comparative ITS Example



Post-Intervention
Level difference 2.88%
Δslope difference 0.69%

RESULTS

Parameter	Interpretation	Estimate	Standard Error	Probability
β_1	Control Pre- Trend	-0.28988	0.1398	0.0500
β_2	Control Post- Level Change	-0.56345	0.9632	0.5645
β_3	Control Post- Trend Change	0.356667	0.2210	0.1208
β_4	Treatment/Control Pre- Level Difference	0.724643	0.9981	0.4755
β_5	Treatment/Control Pre- Trend Difference	0.012857	0.1977	0.9487
β_6	Treatment/Control Post- Level Difference	-2.87607	1.3622	0.0463
β_7	Treatment/Control Change in Slope Difference Pre- to Post-	-0.6875	0.3125	0.0386

But Remember…Autocorrelation!

- What is it?
- An outcome measured at some point in time is correlated with past values of itself
- The lag order is "how far back in time" the correlation extends
- Example: monthly data, seasonal cycle (lag order = 12 autocorrelation)

Temperature in January correlated with temperature in January from previous year, etc

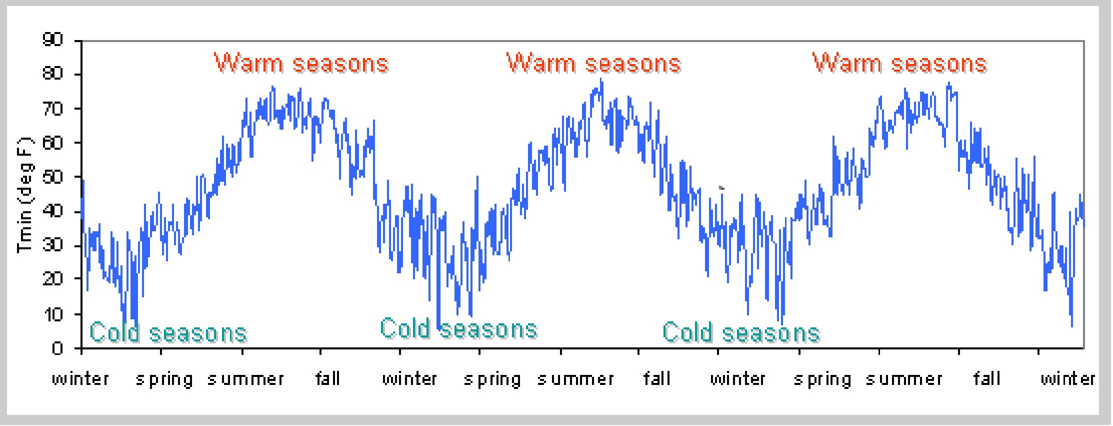
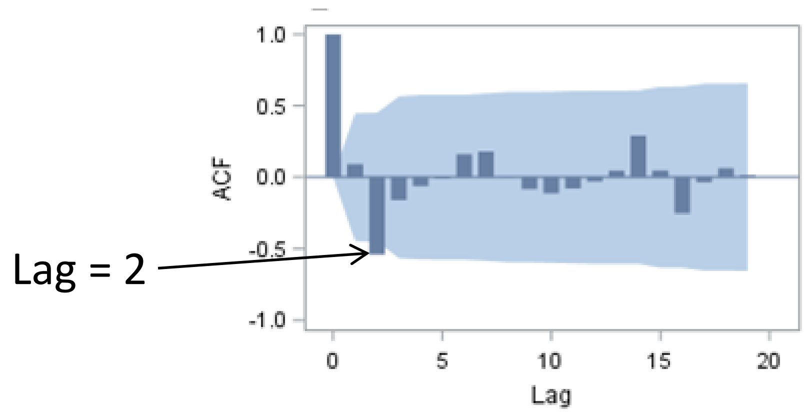


Image source: [NOAA]

Autocorrelation



- How do we test for it?

1. Conduct our ITS regression analysis

2. Obtain residuals (error)

3. Identify residual autocorrelation:

   Compute autocorrelation functions (ACF) up to specified lag
   (i.e., monthly data ~12-24)
   lag 0 always has correlation of 1…series correlated with itself

4. Identify optimal lag order:

   Check partial ACF, use last significant lag before others drop to non-significant

- Note: this is a very general explanation (see Box-Jenkins approach for time series)

- Look for patterns that are guided by theoretical considerations

Adjusting Standard Errors (SE)

- What can we do when autocorrelation is present?

- Use more complicated models
  (i.e., autoregressive or ARIMA models, generalized linear mixed model)

Or

- Use OLS regression as usual, adjust SE

ITS with Adjusted Standard Errors (SE)

- Newey-West autocorrelation adjusted standard errors
- Can do this in SAS with *proc model* after creating ITS dummy variables (*T*, *X*, *TX*):

```
proc model data=DATASET_NAME;
        parms b0 b1 b2 b3;
        OUTCOME_VAR = b0 + (b1*t) + (b2*x) + (b3*tx);
        fit OUTCOME_VAR /gmm kernel=(bart,LAG+1,0)
                vardef=n;
        test b1+b3;
run; quit;
```

Underlined section is for adjusted SE
LAG+1 is a positive integer (lag order + 1)

SAS documentation: http://support.sas.com/kb/40/098.html

SAS Macro and Guide for ITS

- I have written a macro to perform ITS analyses in SAS software
- Based on Stata program by Ariel Linden (2015)
- Can perform single series or comparative ITS analyses
- Will create all necessary dummy variables
- Will adjust for autocorrelation (order needs to be determined before analysis) using Newey-West standard errors
- Will test $\beta_1+\beta_3$ for single series analysis (post- slope)
- Compute residual and predicted values for further model diagnostics
- Produce table of estimates and plot time series
- Companion guide also available

SAS Macro and Guide for ITS

- Guide paper and full macro code:

  https://www.linkedin.com/pulse/interrupted-time-series-analysis-single-comparative-designs-caswell-1/

- Contact me directly:

jcaswell@hsnri.ca

joseph.caswell@ices.on.ca

Thanks for listening!