Time-Trend Analysis With Cross-Sectional Survey Data

Dan Liao and Rachel Harter

RTI International

AAPOR 77th Annual Conference
May 12th, 2022
Chicago, IL

RTI INTERNATIONAL

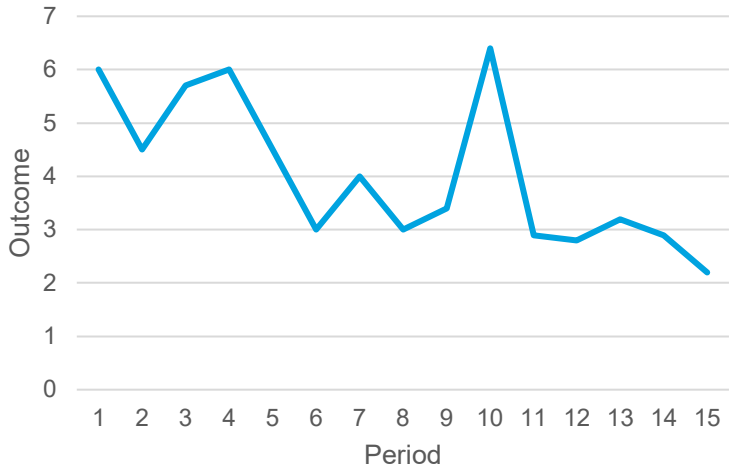
Many Definitions of “Trend”

- an underlying pattern of behavior in a time series (Wikipedia)
- a pattern found in time series datasets (statistica.com)
- the general drift or tendency in a set of data (mathgoodies.com)
- a general direction of **change** (Merriam-Webster)
- the general movement over time of a **statistically detectable change** (Merriam-Webster)

Trend Analysis

…to study changes over time with survey data collected periodically

- Many ways to study
- Many purposes



Review of Trend Testing Methods and Applications Used by U.S. Federal Agencies

Motivation

- the National Survey on Drug Use and Health (NSDUH) 2014 and 2015 partial-redesign

Purpose

- to develop strategies and guidelines for conducting trend testing analyses for cross-sectional surveys collected periodically

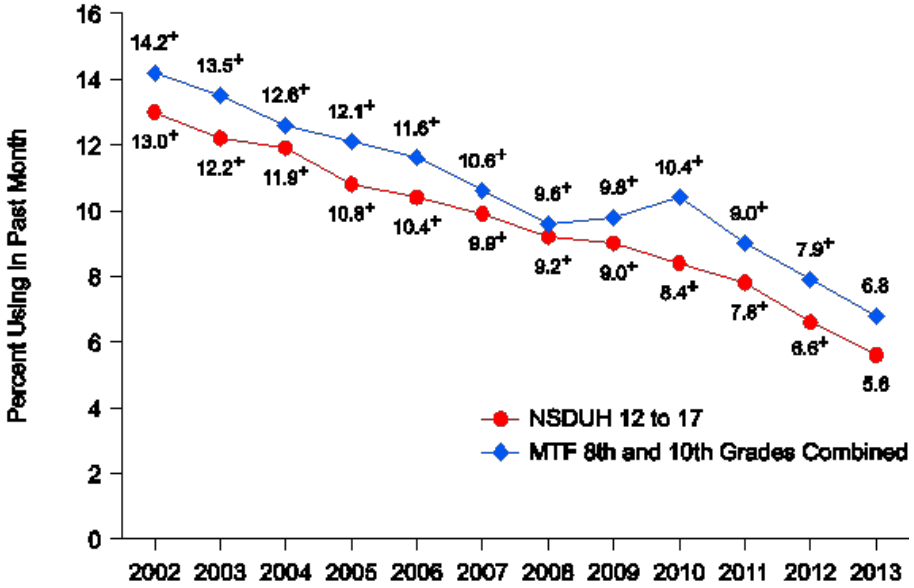
Activities

- conducted a review of trend testing methods and application used by U.S. federal agencies
- contacted federal personnel who graciously contributed their expertise on the trend analysis methods used by their programs
- performed a literature review across scientific journals in different fields on their trend testing innovations
- designed a series of simulation studies to evaluate several popular trend testing methods

See full report for more details https://www.samhsa.gov/data/sites/default/files/NSDUH-TrendTestTask1-2015.pdf

A Time Series Graph with Pairwise Comparisons to Determine the Trend

Past Month Cigarette Use among Youths in NSDUH and MTF: 2002-2013 (Figure 8.2 in the 2013 NSDUH National Findings Report)



MTF = Monitoring the Future;

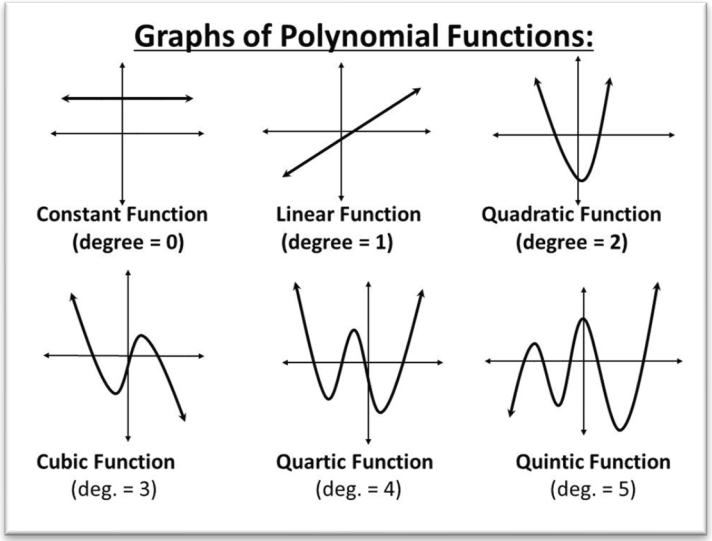
NSDUH = National Survey on Drug Use and Health.

+ Difference between this estimate and the 2013 estimate is statistically significant at the .05 level.

Orthogonal Polynomials for Testing Linear and Quadratic Trends

- Orthogonal polynomials are equations associated with a power of the independent variable over time *T* (e.g., *T*, linear; T^2, quadratic; T^3, cubic, etc.)

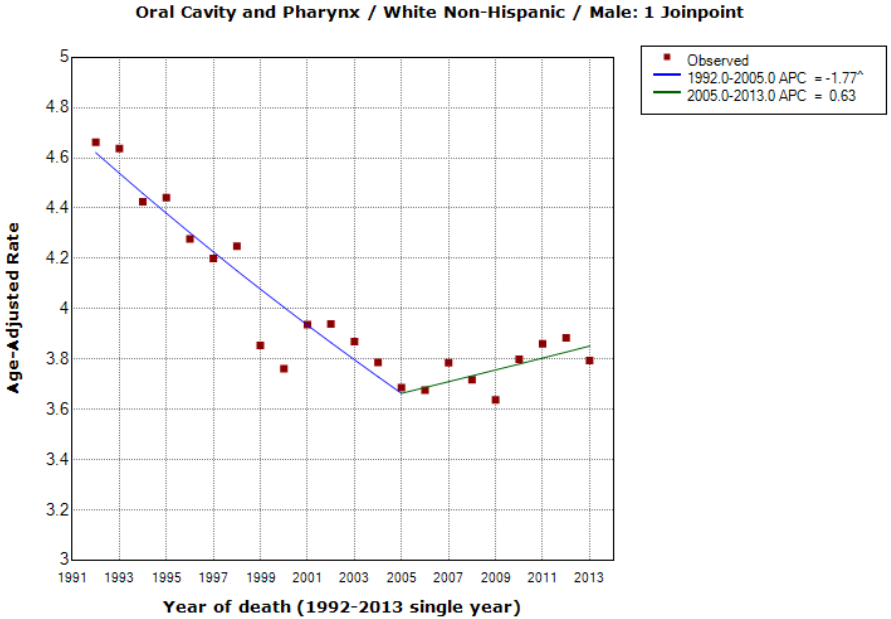
- Can be used in two different ways:
  - *t* Test with Orthogonal Contrast Matrix (Ford, 2013; Carroll, Kit, and Lacher, 2015)
  - Regression with Orthogonal Polynomials (CDC, 2016)



* Graph from: https://medium.com/@olamiotan_/regression-analysis-a2e2ea2f3fb0

Joinpoint Regression Model to Identify Apparent Trend Change(s)

- If the quadratic (or higher order) time variable in the regression model is significant, then Joinpoint Trend Analysis Software can be used to determine the point(s) at which the trends change (Kim, et al, 2000).
- Available at https://surveillance.cancer.gov/joinpoint/



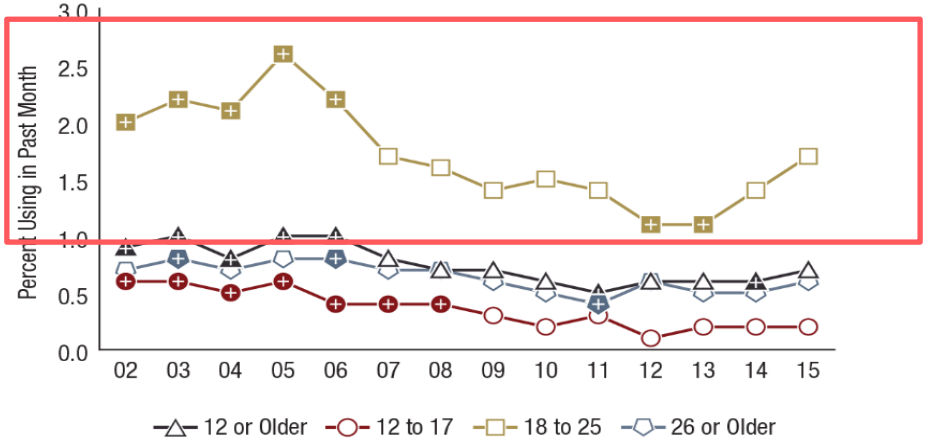
*Graph from: https://surveillance.cancer.gov/help/joinpoint/setting-parameters/advanced-analysis-tools-tab/jump-model-comparability-ratio

Simulation Scenarios

3 artificial scenarios over ten years

Scenario 1: population trend line is “flat”
Scenario 2: population trend line is linearly increasing
Scenario 3: population trend line has an apparent change in the middle (“two-piece”)

Empirical scenario: population trend line is the same as the cocaine use in the past month for ages 18-25 from the 2002-2015 NSDUH

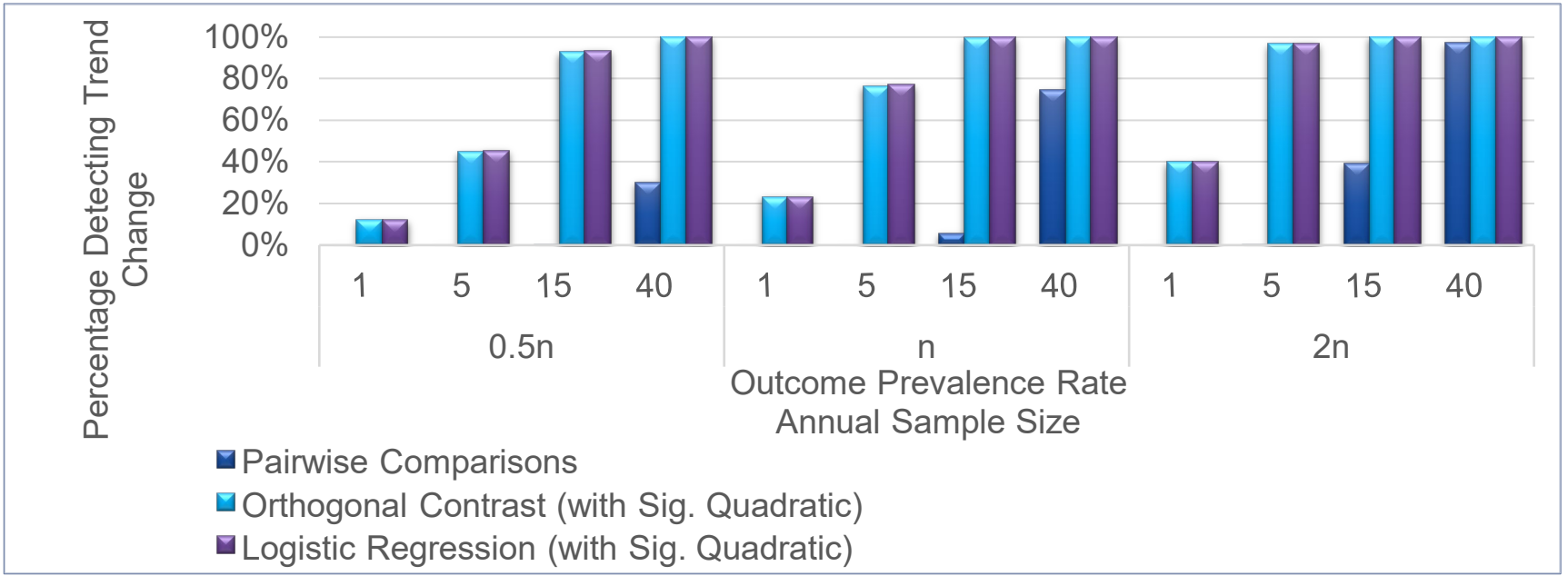


Simulation Conclusion 1: Type I Error Rates (False Positives)

- *When the trend line is flat (scenario 1), how often is a significant trend change detected?*
  - always around 4-to-5%
  - not impacted by sample size or the size of the outcome prevalence
  - not affected by trend testing method

- **Note: Bonferroni correction** should be used to adjust the p-values for the pairwise tests. Otherwise, the type I error rate can be very high.

Simulation Conclusion 2: Power to Detect Trends (Example from Scenario 3 for Detecting a Quadratic Trend)



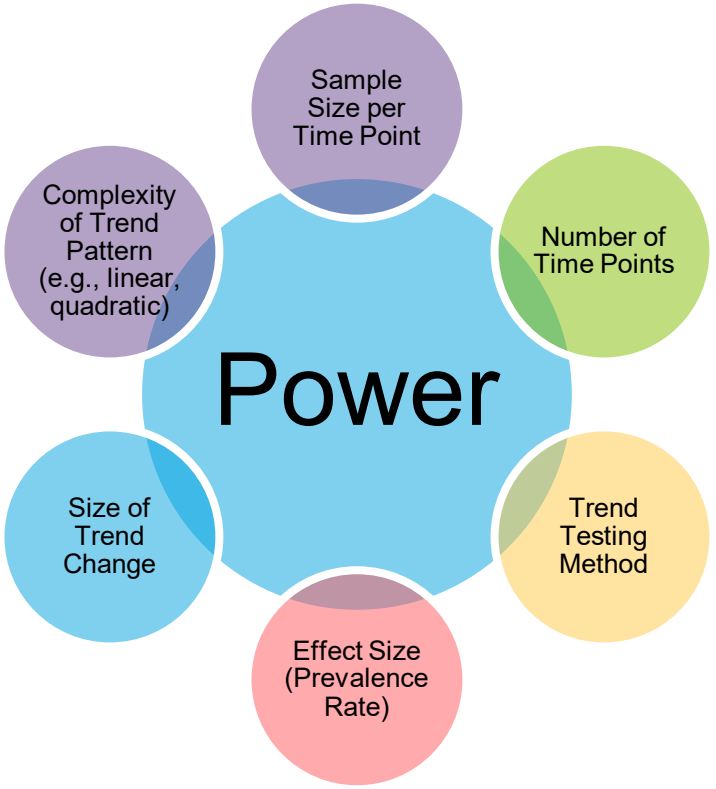
The simulation set the onset of change at Year 7. The annual increase of p_t was set as 2% of p_1 before the change and annual decrease of p_t was set as 2%* p_1 after the change:

- For example, if p_1 in Year 1 is 1%, we have: p_2=1.02%, p_3=1.04%, p_4=1.06%, p_5=1.08%, p_6=1.10%, p_7=1.12%, p_8=1.10%, p_9=1.08% and p_{10}=1.06%.

Simulation Conclusion 2: Power to Detect Trends

How often can a trend be detected?

- **Time series graphs with pairwise comparisons** might detect very few trends.
- **Regression method and orthogonal contrasts** are recommended to test for linear or higher-order trends.
- When prevalence rate is relatively large (close to 50%) and sample size is relatively large, **Joinpoint regression analysis** performs well (with very low Type I and Type II error rates).



Simulation Conclusion 3: Detect the “Joinpoints” (Empirical Study)

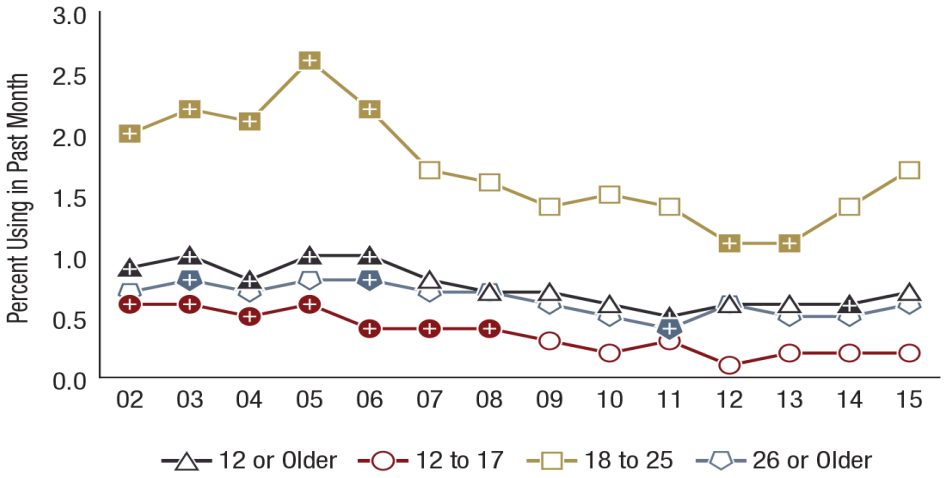
Can the Joinpoint regression method detect a joinpoint at the right time point?

Among the simulations where two joinpoints were detected, a joinpoint was detected:

- at Year 2005 in over 90% of the simulations
- at Year 2012 in around 50% of the simulations
- at Year 2013 in around 25% of the simulations

when the annual sample size was set to equal the NSDUH annual sample size for ages 18-25.

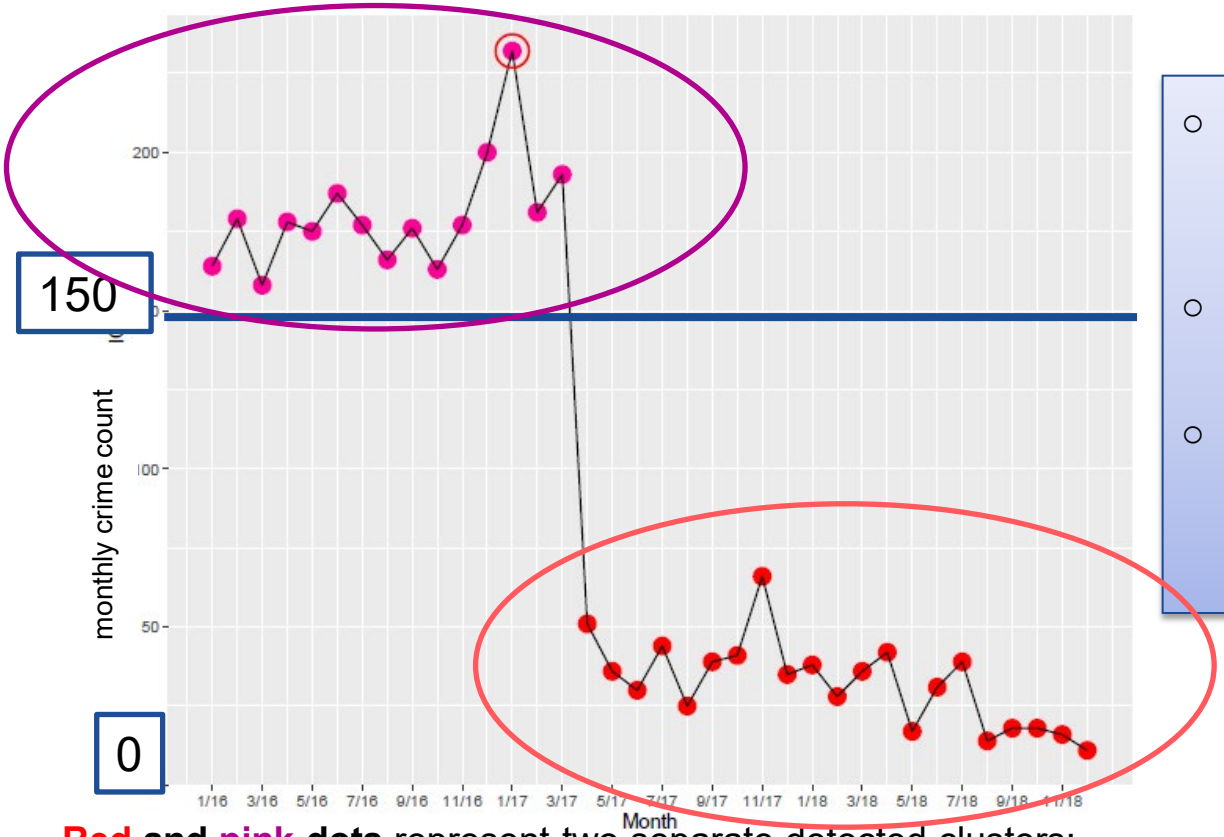
These rates were higher with larger annual sample sizes.



And Now for Something Completely Different…

Some applications or data types may call for a different approach altogether.

Example of Significant Changes in Monthly Crime Counts Within a Law Enforcement Agency



- Detect the **"change in level"** through cluster analysis.
- Detect **outlier(s)** rather than quadratic trend first.
- Within each cluster, after removing the outlier(s), perform the trend analysis.

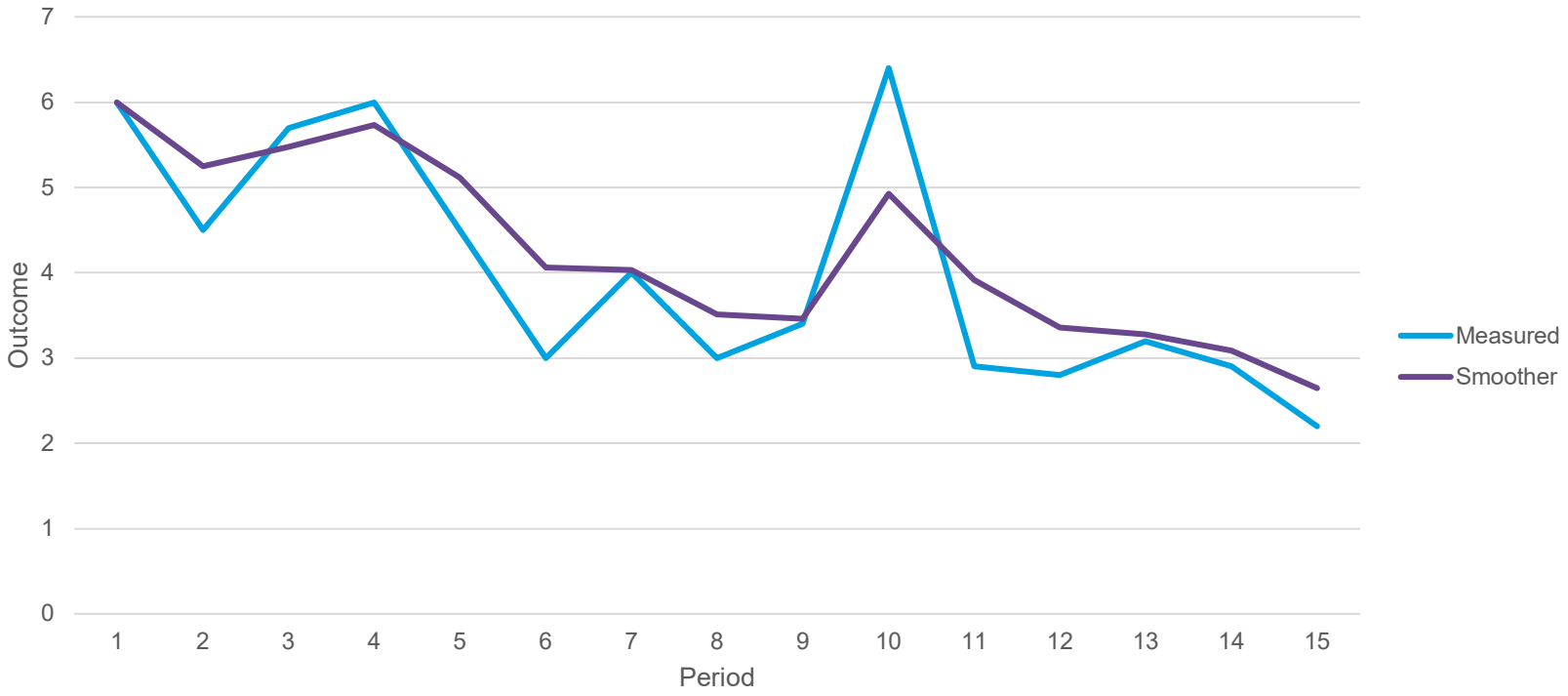
Red and pink dots represent two separate detected clusters;
Circled dot is a detected outlier within the first cluster.

Nielsen Prototype Trend Analysis

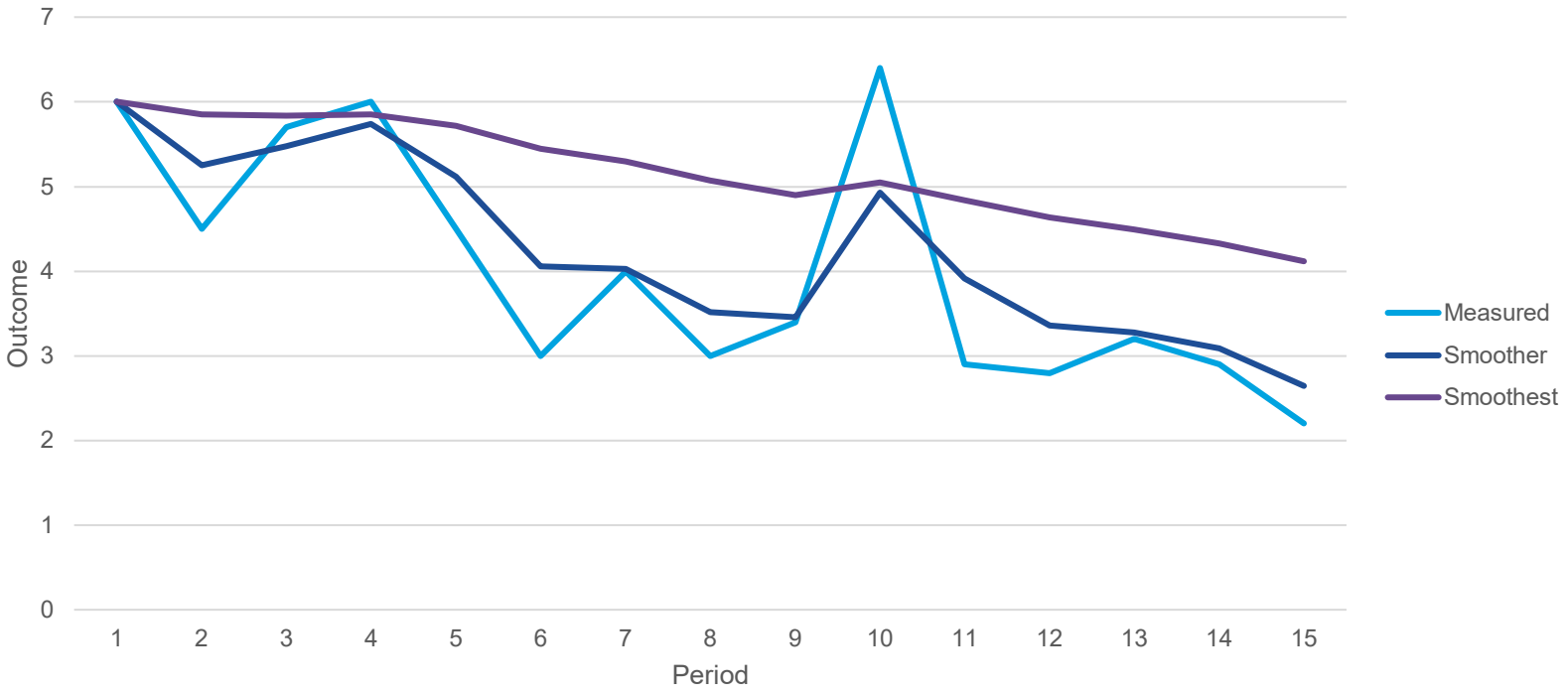
1. Smooth a data series.
2. Identify high and low points to establish trends.
3. Measure the change between trend endpoints.
4. Compare the change to other smoothed changes in the set over the same timeframe.

This process is described in Bayer et al.(1989).

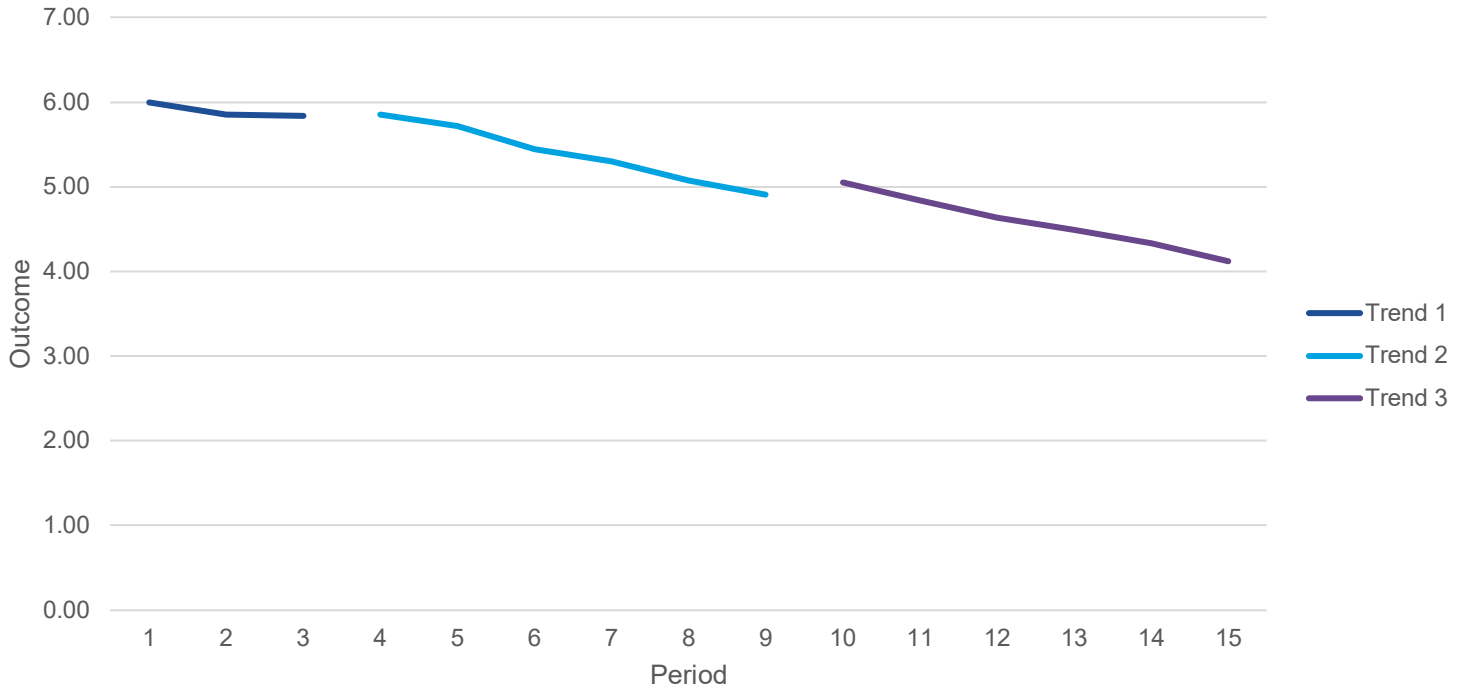
Smooth the Series using a Kalman Filter



Smooth the Series using a Kalman Filter



Identify Endpoints of Trends



Compare Smoothed Changes for the Same Time Period

Brand	P1 (at *t1*)	P2 (at *t2*)	Diff=P2-P1	Diff_Ratio= $\frac{P2-P1}{\frac{1}{2}(P2+P1)}$	Signed Score =Diff*|Diff_Ratio|
A	38	42	4	10%	40
B	21.5	18.5	-3	-15%	-45
C	7	9	2	25%	50
D	6	5	-1	-18%	-18
E	3	2	-1	-40%	-40

The score, an approximation of a measure from information theory, is the product of the absolute and relative changes. (Strobel 1982; Harter 1987)

Summary

The commonly-used methods for testing trends usually have a reasonable **Type I error rate** (less than 5%). When the trend line is flat in the population data, there is over 95% chance that each method won't detect any significant trend.

Power to detect a trend depends on multiple factors including sample size, effect size, size of change, and trend testing method.

The pattern of trend lines can be very complex, which will require innovative methods to tackle (e.g., joinpoint regression analysis, clustering & outlier detection, Kalman filter).

How to interpret the trend changes? This is a separate topic. Although the purpose may be to study "an underlying pattern of behavior in a time series," researchers need to consider how extraneous factors such as a study redesign or emerging data quality issue may contribute to the detected changes.

References

- Bayer, J., Cartwright, P. A., Harris, T., & Harter, R. M. (1989). “A market share trend analysis for an expert system.” Paper presented at the ORSA/TIMS Marketing Science Conference.
- Harter, R. M. (1987). “Exception reporting: Judging what is significant.” *Nielsen Marketing Trends, 1*, 20–23.
- Strobel, D. (1982).” Determining Outliers in Multivariate Surveys by Decomposition of a Measure of Information.” In *Proceedings of the Section on Business and Economic Statistics*, American Statistical Association.
- Kim, H.-J., Fay, M. P., Feuer, E. J., & Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine, 19*, 335–351 (correction: 2001; *20*, 655).
- Ford, E. S. (2013). Trends in predicted 10-year risk of coronary heart disease and cardiovascular disease among U.S. adults from 1999 to 2010. Journal of the American College of Cardiology, 61, 2249–2252. Retrieved from https://doi.org/10.1016/j.jacc.2013.03.023
- Carroll, M. D, Kit, B. K., & Lacher, D. A. (2015). Trends in elevated triglyceride in adults: United States, 2001–2012. NCHS Data Brief, 198. Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/products/databriefs/db198.htm
- Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Adolescent and School Health. (2016, June). Youth Risk Behavior Surveillance System (YRBSS): Conducting trend analyses of YRBS data. Retrieved from https://www.cdc.gov/healthyyouth/data/yrbs/pdf/yrbs_conducting_trend_analyses.pdf

Thank you

Contact: Dan Liao| email: dliao@rti.org; Rachel Harter, rharter@rti.org