NATIONAL SURVEY ON DRUG USE AND HEALTH

AN OVERVIEW OF TREND TESTING METHODS AND APPLICATIONS IN NSDUH AND OTHER STUDIES

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland

April 2017

This page intentionally left blank

NATIONAL SURVEY ON DRUG USE AND HEALTH

AN OVERVIEW OF TREND TESTING METHODS AND APPLICATIONS IN NSDUH AND OTHER STUDIES

RTI Project Nos. 0213757.004.107.008.003.007 &
0213405.005.002.002.003.005
Contract No. HHSS283201300001C

RTI Authors:
Dan Liao
Christopher Sroka

SAMHSA Authors:
Matthew Williams
Sarra Hedden

Project Director:
David Hunter

SAMHSA Project Officer:
Peter Tice

For questions about this report, please e-mail Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

April 2017

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2017). *National Survey on Drug Use and Health: An Overview of Trend Testing Methods and Applications in NSDUH and Other Studies.* Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This methodological document was prepared for the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a registered trademark and a trade name of Research Triangle Institute). In addition to the listed authors, significant contributors include Rachel Harter and Ralph Folsom at RTI and Jonaki Bose and Eunice Park-Lee at CBHSQ. Debbie Bond word processed and formatted the report, while Richard Straw copyedited it and Teresa Bass coordinated its web production. The authors thank the following federal personnel who graciously contributed their expertise on the trend analysis methods used by their programs.

Name	**Affiliation**	**Abbreviation**
Shannan Catalano	Bureau of Justice Statistics	BJS
Lynn Langton	Bureau of Justice Statistics	BJS
Joy Sharp	Bureau of Transportation Statistics	BTS
David Raglin	Census Bureau	Census
Nicole Scanniello	Census Bureau	Census
Eileen O'Brien	Energy Information Administration	EIA
Wendy Barboza	National Agricultural Statistics Service	NASS
Nathan Cruze	National Agricultural Statistics Service	NASS
Chris Chapman	National Center for Education Statistics	NCES
Marilyn Seastrom	National Center for Education Statistics	NCES
Jennifer Madans	National Center for Health Statistics	NCHS
Donald Malec	National Center for Health Statistics	NCHS
Cathie Alderks	Substance Abuse and Mental Health Services Administration	SAMHSA
Elizabeth Crane	Substance Abuse and Mental Health Services Administration	SAMHSA
Laura Milazzo-Sayre	Substance Abuse and Mental Health Services Administration	SAMHSA

Table of Contents

Chapter **Page**

List of Figures v

List of Tables v

1. Introduction 1

2. Summary of Trend Testing Methods in Current Practices in NSDUH 3
   - 2.1 Overview 3
   - 2.2 *t* Test for Pairwise Comparison between Two Different Time Points 3
     - 2.2.1 Example of *t* Test for Pairwise Comparison in the 2014 NSDUH Detailed Tables 5
     - 2.2.2 Example of *t* Test for Pairwise Comparison in the 2014 First Findings Reports 5
   - 2.3 *t* Test for Testing Linear and Quadratic Trends across Multiple Time Points 6
     - 2.3.1 SUDAAN Code (Test of Linear Trends with DESCRIPT) 6
     - 2.3.2 SUDAAN Code (Test of Quadratic Trends with DESCRIPT) 7
     - 2.3.3 Example of Linear Trend Testing in the 2013 NSDUH National Findings Report 7
     - 2.3.4 Example of Linear and Quadratic Trend Testing in an Analytic Study 8
   - 2.4 Parametric Regression for Trend Testing 9
     - 2.4.1 Example of Trend Testing in the 2014 Sample Redesign Impact Analysis 10
     - 2.4.2 Example of Trend Testing in a Study of Trends in Cigarette Use, by Serious Psychological Distress Status, in a National U.S. Sample 11
   - 2.5 Testing Estimates at Two Different Time Points for Small Area Estimation Documents 12

3. Summary of Trend Testing Methods Used by Other Agencies 17
   - 3.1 Overview 17
   - 3.2 *t* Test for Pairwise Comparisons and for Testing Linear and Quadratic Trends with Orthogonal Contrast Matrices across Multiple Time Points 17
   - 3.3 Regression Analysis for Trend Testing 18
     - 3.3.1 Logistic Regression and Linear Regression 18
     - 3.3.2 Time Series Analysis 19
     - 3.3.3 Hierarchical and Multilevel Models 19
   - 3.4 Testing a Nonlinear Trend and Identifying an Apparent Trend Change 20
     - 3.4.1 Three-Step Analysis in Youth Risk Behavior Survey (YRBS) 20
     - 3.4.2 Other Studies That Use Orthogonal Polynomial Trend Contrasts 21
     - 3.4.3 Other Studies That Use Joinpoint Regression 21
   - 3.5 Trend Testing without Considering Sample Variation at Each Time Point 22
   - 3.6 Caveats When Doing Trend Testing with Medical Expenditure Panel Survey (MEPS) Data 22

Table of Contents (continued)

Chapter **Page**

4. Literature Review of Selected Topics in Trend Testing 25
- 4.1 Overview 25
- 4.2 Detecting Outliers in Trends and Time Series 25
- 4.3 Nonparametric and Bayesian Methods in Time Series 26
  - 4.3.1 Nonparametric Methods 26
  - 4.3.2 Bayesian Methods 27
- 4.4 Methods Related to Joinpoint Regression 28
- 4.5 Sensitivity of Analyses to Time Window 29
- 4.6 Scanning Statistics 29
- 4.7 Updating Analyses in the Presence of New Data 29

5. Summary and Future Research 31

References 33

Appendix

A Summary of Documents under This Review That Discussed or Used Trend Analysis 43

B Methods Investigated through Literature Review 53

C Methodological Approaches Used by Other Statistical Agencies 59

List of Figures

Figure **Page**

2.1 Past Month Cigarette Use among Youths in NSDUH and MTF: 2002-2013 (Figure 8.2 in the 2013 NSDUH National Findings Report) 8

2.2 Past Year Mental Health Service Use, by Service Type: Percentages, National Surveys on Drug Use and Health, 2008–2012 (Figure 1 in Ringeisen et al., 2016) 9

List of Tables

Table **Page**

A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis 45

B.1 Methods Investigated through Literature Review 55

C.1 Methodological Approaches Used by Other Statistical Agencies 61

This page intentionally left blank

1. Introduction

The National Survey on Drug Use and Health (NSDUH) provides annual data on alcohol use, tobacco use, illicit drug use, substance use disorders, mental health, receipt of services for behavioral health conditions, and other related measures of interest (e.g., risk perceptions related to substance use). When estimates are comparable from year to year, NSDUH supports analyses of trends that may be useful for tracking key indicators.

When analyzing cross-sectional survey data such as NSDUH's, the "trend" depicts the general underlying pattern of change of an outcome variable over time in a finite population. Trend testing examines whether or not the change (i.e., the trend) of an outcome variable is significant over time. A trend can be flat without a significant change over time, or it can be a significant increase or decrease (getting "better" or "worse") or even more complex over time (e.g., getting "better" first, then getting "worse").

A variety of methods for NSDUH trend analysis has been used, such as pairwise testing and statistical regression. However, no formal NSDUH guidelines have been available on how to select an appropriate method in terms of fitness for use under certain constraints. Factors such as limited fiscal and time constraints, targeted research questions, advantages and disadvantages of the testing methods, and the appropriate level of significance and power required should all be considered when choosing a method.

A significant increase or decrease in a trend or a break in a trend may detect real change in a characteristic for the civilian, noninstitutionalized population of the United States. However, changes made to survey measures or procedures could cause the results to deviate from expected trends. Even when no changes have been made to the survey, some changes could simply be the result of random variation because the estimates are based on samples. The redesign impact analysis work in NSDUH's annual Methodological Resource Book (e.g., Center for Behavioral Health Statistics and Quality, 2016b, in press) discusses in detail and explains the factors that cause a trend or changes in trends. This report focuses on statistical details of current trend testing methods.

The report summarizes the current methods used for analyzing trends in NSDUH and the trend testing methods used by other federal agencies, along with a literature review of the topic. The literature review can be used as a reference for researchers wishing to understand the rationale for using each of the methods in that it provides guidance on choosing an appropriate method for analyzing trends and an overview of the language used to interpret NSDUH results.

This report is organized into five chapters and three appendices. Chapter 2 describes the methods for analyzing trends in various NSDUH statistical activities. Chapter 3 presents a summary of the methods used by other federal agencies as revealed by a thorough review of study documents, reports, and other information released by these agencies. After reviewing this information, the authors also directly contacted program staff to learn more about the rationale of using a particular method and their past experience in trend testing.

Chapter 4 summarizes the literature review that addressed specific weaknesses found in NSDUH's current methods and those used in other federal studies. The literature review was mainly conducted using two databases of published articles: (a) the Current Index to Statistics (https://www.statindex.org/) for the latest methodological articles, and (b) the National Library of Medicine's PubMed (https://www.ncbi.nlm.nih.gov/pubmed) for the latest articles describing the application of trend analysis methods to the health sciences.

Chapter 5 summarizes all of the methods discussed in the previous chapters in terms of what research questions they address and how much effort is required for each method. This final chapter also points out several potential areas for future research.

A list of references used in this report and in the literature review also is provided, as are three appendices on trend analysis documentation (Appendix A), the methods investigated through the literature review (Appendix B), and the methodological approaches used by other statistical agencies (Appendix C).

2. Summary of Trend Testing Methods in Current Practices in NSDUH

2.1 Overview

This chapter summarizes the four methods currently used in National Survey on Drug Use and Health (NSDUH) reports for analyzing trends and identifying breaks in trends. In the four sections that follow, each method is first described from its theoretical aspects, then a few examples are presented on how it has been applied and interpreted in NSDUH publications. Section 2.2 describes the *t* test used to compare prevalence estimates from two different time points. This test is used in NSDUH's annual detailed tables to compare the current year's estimates with estimates from individual previous years in order to check if there is any change in a particular measure in the current year relative to the past few years. Section 2.3 describes the *t* test used for testing linear trends across multiple time points that is also used for NSDUH's detailed tables and first findings reports to aid in reporting whether a particular measure has remained stable, increased, or decreased over the entire span of the years of interest. Section 2.4 describes parametric regression models (e.g., linear regression, logistic regression) that are used to test trends and breaks in trends for various NSDUH analytical tasks. Section 2.5 describes the test developed to compare two small area estimates between two time points.

2.2 *t* Test for Pairwise Comparison between Two Different Time Points

Pairwise comparisons have been used widely in NSDUH's detailed tables to compare prevalence estimates from two different time points, which can be based on single years of data (e.g., 2013 and 2014) or more than 1 year of data used for each estimate (e.g., 2011 and 2012 annual averages and 2013 and 2014 annual averages). In the pairwise comparisons, one can test the null hypothesis (no difference between rates) against the alternative hypothesis (there is a difference in prevalence rates) using the standard *t* test with the appropriate degrees of freedom (*df*) for the difference in proportions test, expressed as follows:

$$t_{df} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\mathrm{var}(\hat{p}_1) + \mathrm{var}(\hat{p}_2) - 2\mathrm{cov}(\hat{p}_1, \hat{p}_2)}}, \tag{1}$$

or

$$t_{df} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\mathrm{var}(\hat{p}_1) + \mathrm{var}(\hat{p}_2) - 2\rho(\hat{p}_1, \hat{p}_2)SE(\hat{p}_1)SE(\hat{p}_2)}}, \tag{2}$$

where in both formulas, *df* = the appropriate degrees of freedom, $\hat{p}_1$ = the first prevalence estimate, $\hat{p}_2$ = the second prevalence estimate, $\mathrm{var}(\hat{p}_1)$ = the variance of the first prevalence estimate, and $\mathrm{var}(\hat{p}_2)$ = the variance of the second prevalence estimate. In the first formula, $\mathrm{cov}(\hat{p}_1, \hat{p}_2)$ = covariance between $\hat{p}_1$ and $\hat{p}_2$. In the second formula, the covariance between

$\hat{p}_1$ and $\hat{p}_2$ is displayed as the product of the correlation between $\hat{p}_1$ and $\hat{p}_2$ and the standard errors (SEs) of $\hat{p}_1$ and $\hat{p}_2$, where $\rho(\hat{p}_1, \hat{p}_2)$ = the correlation between $\hat{p}_1$ and $\hat{p}_2$, and $SE(\hat{p}_1)SE(\hat{p}_2)$ = the product of the SEs for $\hat{p}_1$ and $\hat{p}_2$ (i.e., the two formulas are equivalent in that the first formula is defined in terms of the covariance, and the second is defined in terms of the correlations and SEs).

Generally, the correlations between estimates in adjacent years are very small and positive; thus, ignoring the correlation in the second formula will usually result in a slightly more conservative test outcome, which is a test that is less likely to reject the null hypothesis that there is no difference in the two estimates. However, a negative correlation is possible and would result in a liberal test, which means it would be more likely to reject the null hypothesis that there is no difference in the two estimates. Additionally, the second (simplified) formula can be used in the case of two independent (i.e., uncorrelated) samples, as in the case of comparing two nonadjacent year estimates.[1] Note that the first and second prevalence estimates may take the form of prevalence estimates from two different survey years (e.g., 2013 and 2014, respectively), prevalence estimates from sets of combined survey data (e.g., 2011 to 2012 annual averages and 2013 to 2014 annual averages, respectively), or prevalence estimates for populations of interest within a single survey year. Quick tests (where the correlation of 0 is assumed) are great tools for gaining a better understanding of published estimates; however, the results of these quick tests should be confirmed using NSDUH data and appropriate software. Some examples of the quick tests for comparing prevalence estimates between years can be found in the 2014 NSDUH statistical inference report (Center for Behavioral Health Statistics and Quality [CBHSQ], 2016a).

Under the null hypothesis, the test statistic *t* is a random variable that asymptotically follows a *t*-distribution. Therefore, calculated values of *t*, along with the appropriate degrees of freedom, can be used to determine the corresponding probability level (i.e., *p* value). Whether testing for differences between years or from different populations within the same year, the covariance term in the formula for *t* (see formula 1 above) will, in general, not be equal to 0. SUDAAN® is used to compute estimates of *t* along with the associated *p* values such that the covariance term is calculated by taking the sample design into account (RTI International, 2012a, 2012b). A similar procedure and formula for *t* are used for estimated totals. It should be noted, however, that the SE of the total estimates are not directly calculated in SUDAAN for domains forced by the weighting process to match their respective U.S. Census Bureau population estimates; thus, the corresponding test statistics are also not directly computed in SUDAAN. SAS®, SUDAAN, and Stata® examples showing the computational methods for generating *p* values of estimates of *t* and estimated totals can be found in Appendix A in the 2014 NSDUH statistical inference report (CBHSQ, 2016a). Two examples in NSDUH that used the *t* test for pairwise comparison in trend analysis are presented in the following sections.

---

[1] Methods used for NSDUH's detailed tables take into account the correlation induced by multistage sample design, but not serial correlation in these type of data. See Chapters 3 and 4 in this report for a discussion on the methods that account for serial correlation.

2.2.1 Example of *t* Test for Pairwise Comparison in the 2014 NSDUH Detailed Tables

In the 2014 NSDUH detailed tables (CBHSQ, 2015d), the *t* test for pairwise comparison was used to test the 2014 prevalence estimates against their corresponding estimates in the prior years to see whether or not each 2014 estimate was significantly larger or smaller than its counterparts in the prior years. For example, Table 1.1B of the 2014 detailed tables displays the prevalence for lifetime, past year, and past month illicit drug use. Past month marijuana use had a prevalence rate of 7.5 percent in 2013 and 8.4 percent in 2014. The *p* value from the pair comparison between these two estimates was smaller than 0.01; thus, a footnote was added to the 2013 estimate: "Difference between estimate and 2014 estimate is statistically significant at the 0.01 level." When the *p* value from the pairwise comparison was smaller than 0.05 but larger than 0.01, the following footnote was referred to instead: "Difference between estimate and 2014 estimate is statistically significant at the 0.05 level."

2.2.2 Example of *t* Test for Pairwise Comparison in the 2014 First Findings Reports

In the 2014 NSDUH's four first findings reports (CBHSQ, 2015a, 2015b, 2015c, 2015e), similar to other NSDUH reports, trend analyses focused on percentages because the percentages take into account any changes in the size of the total population and facilitate the comparison of estimates across years. Statistical tests were conducted for pairwise comparisons in these reports. Statistically significant differences were described using terms such as "higher," "lower," "increased," or "decreased." Statements used terms such as "similar," "remained steady," or "stable" when a difference was not statistically significant. Analyses of long-term trends in these reports summarize whether the 2014 estimates are different from or similar to estimates in most or all previous years. For example, in the report on *Behavioral Health Trends in the United States: Results from the 2014 National Survey on Drug Use and Health* (CBHSQ, 2015a), Figure 2 presents trend estimates on past month illicit drug use among people aged 12 or older, by age group. The testing results were discussed using the following verbatim text (emphases added here in boldface):

> More than 1 in 5 young adults aged 18 to 25 (22.0 percent) were current users of illicit drugs in 2014 (Figure 2). This percentage corresponds to about 7.7 million young adults in 2014 who were current users of illicit drugs. The percentage of young adults who were current illicit drug users was **stable** between 2009 and 2014. However, the 2014 estimate was **higher** than the estimates from 2002 through 2008.

In this example, pairwise comparisons were conducted to test whether or not the 2014 percentage estimate of young adults aged 18 to 25 who were current users of illicit drug users was significantly different from the prior year estimates. The percentage estimates from 2009 to 2013 were not significantly different from the 2014 estimate, but the percentage estimates from 2002 through 2008 were significantly lower than the 2014 estimate.

2.3 *t* Test for Testing Linear and Quadratic Trends across Multiple Time Points

In addition to comparing 1 year versus another year, it can be useful to test the linear trend for all data points across all years of interest. Linear trend testing can inform users about whether prevalence use has decreased, increased, or remained steady over the entire span of the years of interest or about changes in specific measures. Various methods can be used to test linear trends. Linear trend testing is produced for NSDUH's annual detailed tables as applicable, but it is only used to aid in NSDUH report writing and is not published. An example can be found in Section 2.3.3. These linear trend tests are implemented using the SUDAAN procedure DESCRIPT with CONTRAST statements looking across years to evaluate changes over time.

For linear trend testing within NSDUH's detailed tables, the DESCRIPT procedure is used in the mass production of detailed tables *only* as an aid in report writing regarding whether a particular measure had remained stable, increased, or decreased over time. This method uses the *t* test, similar to the pairwise method used when testing means between years and between demographic levels within the detailed tables. Instead of using PAIRWISE statements, type I errors (incorrectly producing significant differences) are controlled through the use of orthogonal polynomial coefficients in the CONTRAST statement.

In addition, this method can be used not only for linear trend testing, but also for higher order trend testing, such as for quadratic trends (when the trend changes at a certain time point). It has been also applied in a NSDUH analytic study (denoted as "the MH18 task") that examines mental health service use across adolescence (ages 12 to 17), by service type, with 2008 to 2012 NSDUH data (see Section 2.3.4 for more details). With this method, the linear and quadratic patterns across ages 12 to 17 were tested for each service type (Ringeisen et al., 2016). Note that this test is applicable only for equally spaced time points. For instance, if the trend for the prevalence rates of a rare illicit drug use from 2011 to 2014 is of interest while this estimate is missing in 2012, the trend cannot be tested based on the 2011, 2013, and 2014 estimates using this test because the time points among these three estimates are not equally spaced.

Examples of SUDAAN code to test the linear and quadratic trends for alcohol use in the past month (ALCMON), by gender, from 2005 to 2014 are shown in Sections 2.3.1 and 2.3.2, respectively.

2.3.1 SUDAAN Code (Test of Linear Trends with DESCRIPT)

```
PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR ALCMON; *Variable ALCMON is the alcohol use in the past month;
SUBGROUP YEAR IRSEX;
LEVELS 10 2; *Variable YEAR ranges from 2005 to 2014 and Variable IRSEX is gender
(male/female);
TABLES IRSEX; *Test by gender;
CONTRAST YEAR = (-9 -7 -5 -3 -1 1 3 5 7 9) / NAME="LINEAR TREND TEST";
* Orthogonal polynomial coefficients for linear trend testing incorporated in the CONTRAST
```

statement here are derived through a mathematical formula that depends on the number of time points in the test, which is 10 in this example;
PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN / REPLACE STYLE=NCHS;
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN / REPLACE
NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10 TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS";
TITLE "TESTS OF LINEAR TREND OF PAST MONTH ALCOHOL USE FROM 2005 TO 2014 BY GENDER";
RUN;

2.3.2 SUDAAN Code (Test of Quadratic Trends with DESCRIPT)

PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR ALCMON; *Variable ALCMON is the alcohol use in the past month;
SUBGROUP YEAR IRSEX;
LEVELS 10 2; *Variable YEAR ranges from 2005 to 2014 and Variable IRSEX is gender (male/female);
TABLES IRSEX; *Test by gender;
CONTRAST YEAR = (6 2 -1 -3 -4 -4 -3 -1 2 6) */ NAME="QUADRATIC TREND TEST";*
* Orthogonal polynomial coefficients for quadratic trend testing incorporated in the CONTRAST statement here are derived through a mathematical formula that depends on the number of time points in the test, which is 10 in this example;
PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN / REPLACE STYLE=NCHS;
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN / REPLACE
NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10 TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS";
TITLE "TESTS OF QUADRATIC TREND OF PAST MONTH ALCOHOL USE FROM 2005 TO 2014 BY GENDER";
RUN;

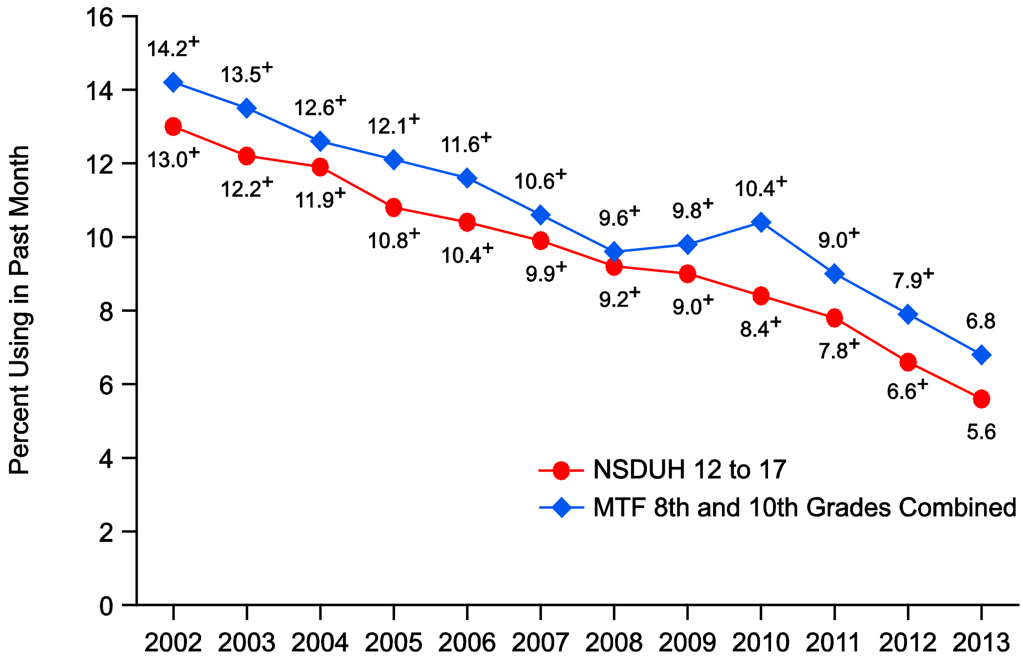
2.3.3 Example of Linear Trend Testing in the 2013 NSDUH National Findings Report

In the 2013 NSDUH national findings report (CBHSQ, 2014, p. 100), a comparison of NSDUH and the Monitoring the Future (MTF) survey estimates among youths for 2002 to 2013 was conducted. The *t* test for linear trend testing was applied to test the trends in both the MTF and NSDUH yearly estimates. The results were interpreted as follows (emphases added here in boldface; note that the Figure 8.2 referred to in the following text is shown here as Figure 2.1):

> The 2012 and 2013 MTF estimates, however, **showed a continuing decline**, **consistent with the NSDUH trend in youth smoking**. Over the long term, the two surveys **showed consistent**

decreases in the prevalence of smoking among youths (Figure 8.2). During the 4-year period from 2010 to 2013, NSDUH showed a 33 percent decline (from 8.4 to 5.6 percent) and MTF showed a 35 percent decline (from 10.4 to 6.8 percent) in current cigarette use.

Figure 2.1 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.

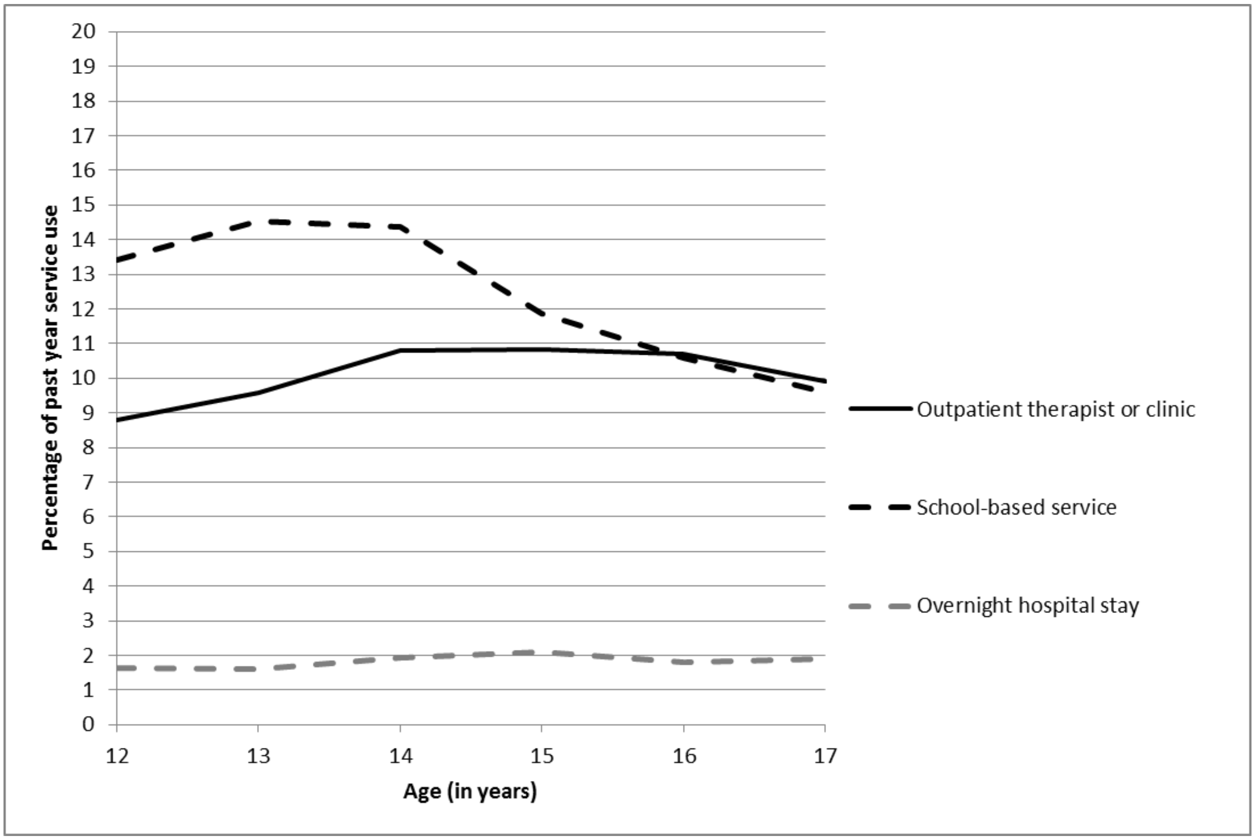
2.3.4 Example of Linear and Quadratic Trend Testing in an Analytic Study

An analytic study (Ringeisen et al., 2016) examined mental health service use by age and by service type among adolescents aged 12 to 17 with 2008 to 2012 NSDUH data. The *t* tests for linear and quadratic trend testing were conducted to compare service use across age groups. The test results are illustrated in the following verbatim text (emphases added here in boldface); note that the Figure 1 referred to in the following text is shown here as Figure 2.2:

> As Figure 1 demonstrates, the use of school-based (t=-4.7, df=900, p<.01) and outpatient therapist or clinic (t=-5.1, df=900, p<.01) services across ages 12–17 **was characterized by a convex quadratic pattern** (i.e., inverted U). By comparison, overnight hospital stay (t=2.1, df=900, p<.01) services **showed a monotone linear pattern** (increasing from 12 to 14 and almost flat from 15 to 17). For all service types, use **increased** from 12 to14 and then either **declined** (therapist/mental health clinic and school-based) or **remained level** from 15 to 17 (overnight hospital stay). The decline was particularly

apparent for school-based services, where service use **decreased** from 14.5% ±0.4 at 13 to 9.6% ±0.3 at 17.

Figure 2.2 Past Year Mental Health Service Use, by Service Type: Percentages, National Surveys on Drug Use and Health, 2008–2012 (Figure 1 in Ringeisen et al., 2016)



2.4 Parametric Regression for Trend Testing

The *t* test methods described in Sections 2.2 and 2.3 can be easily embedded in the SUDAAN DESCRIPT procedure and therefore have been applied widely, despite time constraints, to generate NSDUH's annual detailed tables. However, the *t* test methods have a few limitations: (1) when testing multiple time points, the *t* test with orthogonal contrast matrix is applicable only for equally spaced time points; (2) the *t* test methods focused on testing consistency in value over time, and their ability to identify breaks in trends (i.e., when a trend is changed) is limited; and (3) other covariates (factors) cannot be controlled when analyzing the trends. Parametric regression methods, however, can compensate for these limitations and thus are used for more complex analytical tasks in NSDUH. These methods include linear regression (when the outcome variable is continuous), logistic regression (when the outcome variable is binary), and multinomial regression (when the outcome variable is ordinal or nominal).

In several analytical tasks, linear trend testing was done through a simple regression model that had the outcome variable as the dependent variable and the year as the single independent continuous variable in the model with the intercept term. If the coefficient (i.e., slope) of the year was statistically significant, then the trend was considered statistically significant. In other analytical tasks, more sophisticated models were developed for trend testing. Two examples are given as follows, one from the 2014 NSDUH sample redesign impact assessment and another from an analytical task (SR9) that compared trends in current and heavy cigarette smoking between adults with and without serious psychological distress (SPD) using the Centers for Disease Control and Prevention's (CDC's) National Health Interview Survey (NHIS) data.

2.4.1 Example of Trend Testing in the 2014 Sample Redesign Impact Analysis

The objective of the 2014 sample redesign impact analysis was to assess the impacts of the sample, field operations, and questionnaire changes in the 2014 survey on several key measures in NSDUH by comparing the 2014 estimates with estimates from prior years (CBHSQ, 2016b). In the 2014 sample redesign impact analysis, wherever a potential change in estimates between 2014 and prior years was observed through pairwise *t* tests, a linear trend analysis was conducted to determine whether the 2014 estimate represented a break in trend or if the 2014 estimate was what would be expected given the linear trend in the data. Models to assess linear trends at the national level took the following form:

$$\text{Model (OUTCOME)} = \text{Intercept} + \text{YR14IND} + \text{YEAR7YR}, \tag{3}$$

where YR14IND = 1 if in 2014 and = 2 otherwise, and YEAR7YR = a continuous variable of the year (limited to 2008 to 2014). If YEAR7YR was statistically significant in the model, that indicated that the slope of the linear trend from 2008 to 2014 was significantly different from 0 for the outcome variable modeled. The year 2008 was selected as the beginning of the time period because it conformed with the earliest year that NSDUH mental health estimates became available. The year 2008 also represented a compromise between having sufficient annual time points to reasonably detect a linear trend and at the same time not so many that the linearity of the trend might be compromised (e.g., over longer time periods, curvature or small irregularities in the trend might have occurred), thereby complicating subsequent interpretations. If YR14IND was statistically significant, that indicated that the estimate from 2014 differed significantly from the fitted linear trend. An example of SUDAAN code for this model is shown as follows:

```
PROC RLOGIST DATA=MODEL_DATA DESIGN=WR FILETYPE=SAS DDF=900;
NEST STRATUM PSU;
WEIGHT ANALWT;
REFLEVEL YR14IND=2;
SUBGROUP YR14IND;
LEVELS    2;
MODEL OUTCOME=YR14IND YEAR7YR;
CONDMARG YEAR7YR/ YEAR7YR=(1 2 3 4 5 6 7);
SETENV DECWIDTH=6 COLWIDTH=18;
PRINT beta="Beta" sebeta="Stderr" Deft="Design Effect" t_beta="T:Beta=0" p_beta="p-value"
condmrg="CONDMARG" /cond_mrg=default risk=all tests=default t_betafmt=F8.2
waldchifmt=f6.2 orfmt=f10.2 loworfmt=f10.2 uporfmt=f10.2 dffmt=f7.0 t_prdmrgfmt=f8.2;
OUTPUT /FILENAME=TAB_PRED FILETYPE=SAS COND_MRG=ALL REPLACE;
RUN;
```

Other covariates could also be added to the model in formula (3). For example, to assess linear trends within age groups (CATAG4), the interaction terms between the year variables and the age variables were added in the model:

$$\text{Model (OUTCOME)} = \text{Intercept} + \underline{\text{CATAG4} \times \text{YR14IND}} + \underline{\text{CATAG4} \times \text{YEAR7YR}}. \quad (4)$$

The CATAG4 × YEAR7YR interaction term allows the model the flexibility to fit individual slopes to each of the applicable age groups. If this term was statistically significant, that indicated some differences among the individual fitted slopes. If CATAG4 × YR14IND was statistically significant, that indicated that the 2014 year effect (i.e., the difference between the 2014 estimate and the linear trend) differed across the applicable age groups, which in turn triggered individual tests within each age group to determine if a significant break in the trend occurred within any age group.

2.4.2 Example of Trend Testing in a Study of Trends in Cigarette Use, by Serious Psychological Distress Status, in a National U.S. Sample

An unpublished study used data from the NHIS public use files for the years from 1998 to 2013. The analyses involved two primary dependent variables of interest: current cigarette smoking (current every day or "same day" smoking) and current heavy cigarette smoking (heavy smoking). Average annual trends in current and heavy smoking were examined by the year of the NHIS interview. Two years of data (1997 and 1998, 1999 and 2000, etc.) were combined in analyses to increase the sample size and precision of estimates across smaller subgroups of interest. The time trend variable created and used in the analyses (named as POOL2YR) was a continuous variable that ranged from 1 (survey years 1998 and 1999) to 8 (survey years 2012 and 2013). The main correlates of interest were both time and a measure of psychological distress. The independent variable SPD was a binary variable based on questionnaire responses to specific psychological distress questions.[2] Covariates used as control variables in the regression models included age group (18 to 34, 35 to 54, and 55 or older), education (less than high school, general equivalency diploma, high school graduate, some college, and college graduate), race/ethnicity (not Hispanic or Latino white, not Hispanic or Latino black or African American, Hispanic or Latino, and other), gender (male and female), and region of residence (Northeast, Midwest, South, West).

Logistic regression models adjusted for the aforementioned covariates were fit separately for the dependent variables of current smoking and heavy smoking, with the focus on the main covariate of interest, SPD, and trends over time. Trend lines for each smoking outcome over time were confirmed to be linear (i.e., the effect of the continuous time trend variable was statistically significant in the logistic regression models after controlling for the other covariates), after which a time by SPD interaction term was added to each of the two models to determine whether trends in current smoking or in heavy smoking differed by SPD status over time (i.e., 1998 to 2013). Significant interaction terms ($p < .05$) based on Wald chi-square testing were investigated further via contrasts to determine differences in trends (i.e., the direction and magnitude of each smoking outcome by SPD status). Adjusted odds ratios (AORs) and 95 percent confidence

---

[2] SPD was defined as having a Kessler-6 (K6) psychological distress scale score of 13 or higher (range 0 to 24).

intervals (CIs) were reported for nonreference levels of each variable in the logistic regression models. The odds ratio (OR) for the time trend continuous variable of primary interest showed the percentage increase or decrease in average annual current or heavy smoking prevalence across each 2 adjacent interview years. For example, an OR of 1.03 indicated that the odds of smoking increased by an average of 3 percent for each 2-year time period between 1998 and 2013. The following is a SUDAAN example code of the logistic regression model for the dependent variable of current smoking (CURRSOMK2).

```
PROC RLOGIST DATA =MODEL FILETYPE = SAS DESIGN = WR;
WEIGHT;
NEST STRATUM PSU;
CLASS SEX AGE_3GRP RACE4 EDUCATION2 REGION SPD;
REFLEV SEX = 1 AGE_3GRP = 1 RACE4 = 1 EDUCATION2 = 1 REGION = 1 SPD = 2;
MODEL CURRSOMK2 = SEX AGE_3GRP RACE4 EDUCATION2 REGION POOL2YR SPD
POOL2YR*SPD;
EFFECTS POOL2YR / SPD = 1 exp;
EFFECTS POOL2YR / SPD = 2 exp;
RUN;
```

A verbatim example from the draft report that interprets the trend testing results is as follows (emphases added here in boldface):

> The prevalence of current smoking among those without SPD **steadily decreased over time**, from 23.3% in 1998–1999 to 17.1% in 2012–2013. In contrast, the trend in current smoking prevalence among adults with **SPD did not show a significant decrease or increase**, and exhibited no distinct pattern throughout the period (1998–2013). The prevalence of heavy smoking **decreased** during this time period for those both with and without SPD. However, heavy smoking **declined by about three-quarters** among adults without SPD (from 3.8% in 1998–1999 to 1.0% in 2012–2013) and only by about one-half among those with SPD (from 11.5% in 1998–1999 to 5.3% in 2012–2013).

Logistic regression models have been applied in several NSDUH analytical tasks to estimate data trends over time. Another example is the "PR5a task" in which a logistic regression was run for each substance use and mental health measure to test whether trends over time differed for veterans and nonveterans by including an interaction term for year by veteran status.

2.5 Testing Estimates at Two Different Time Points for Small Area Estimation Documents

In NSDUH's state small area estimation (SAE) documents, state estimates are generated by pooling 2 years of NSDUH data, and the difference between two small area estimates of an outcome of interest (e.g., past year cocaine use) is tested for a particular state (e.g., Alabama) between two time points. If the time points contain a common year (such as 2011–2012 vs. 2012–2013), it is called an "overlapping" year change or a "short-term" trend.[3] If the time points are far apart (such as 2002–2003 vs. 2012–2013), it is called a "nonoverlapping" year change or a "long-term" trend.[4] If a test result is found to be significant, a footnote is added to an SAE table (e.g., "difference between the 2002–2003 estimate and the 2012–2013 estimate is statistically significant at the 0.05 level"). Small area estimates and CIs are produced by fitting logistic mixed models using a survey-weighted hierarchical Bayes (SWHB) methodology.

Based on the SWHB method, a test was developed to estimate the Bayes posterior probability of changes between the 2011–2012 and 2012–2013 small area estimates. To estimate change in state estimates, let $\pi_{sa(1)}$ and $\pi_{sa(2)}$ denote the 2011–2012 and 2012–2013 prevalence rates, respectively, for state-*s* and age group-*a*. The change between $\pi_{sa(1)}$ and $\pi_{sa(2)}$ is defined in terms of the log-odds ratio (lor_{sa}) as opposed to the simple difference because the posterior distribution of the lor_{sa} is closer to Gaussian than the posterior distribution of the simple difference $(\pi_{sa(2)} - \pi_{sa(1)})$. The lor_{sa} is defined as follows:

$$lor_{sa} = \ln\left[\frac{\pi_{sa(2)} / (1 - \pi_{sa(2)})}{\pi_{sa(1)} / (1 - \pi_{sa(1)})}\right],$$

where ln denotes the natural logarithm. The *p* value given is computed to test the null hypothesis of no change (i.e., $\pi_{sa(2)} = \pi_{sa(1)}$ or equivalently $lor_{sa} = 0$). An estimate of lor_{sa} is given by

$$\hat{lor}_{sa} = \ln\left[\frac{p_{sa(2)} / (1 - p_{sa(2)})}{p_{sa(1)} / (1 - p_{sa(1)})}\right],$$

where the $p_{sa(1)}$ are the previously published 2011–2012 state estimates and the $p_{sa(2)}$ are the 2012–2013 state estimates. To compute the variance of $\hat{lor}_{sa}$, that is, $v(\hat{lor}_{sa})$, let $\hat{\theta}_1 = \frac{p_{sa(1)}}{1 - p_{sa(1)}}$ and $\hat{\theta}_2 = \frac{p_{sa(2)}}{1 - p_{sa(2)}}$, then

$$v(\hat{lor}_{sa}) = v[\ln(\hat{\theta}_1)] + v[\ln(\hat{\theta}_2)] - 2\,\mathrm{cov}[\ln(\hat{\theta}_1), \ln(\hat{\theta}_2)],$$

where $\mathrm{cov}[\ln(\hat{\theta}_1), \ln(\hat{\theta}_2)]$ denotes the covariance between $\ln(\hat{\theta}_1)$ and $\ln(\hat{\theta}_2)$. This covariance is defined in terms of the associated correlation as follows:

---

[3] A comparison of 2012–2013 and 2013–2014 population percentages in NSDUH is available at https://www.samhsa.gov/data/sites/default/files/NSDUHsaeShortTermCHG2014/NSDUHsaeShortTermCHG2014.htm

[4] A comparison of 2002–2003 and 2013–2014 population percentages in NSDUH is available at https://www.samhsa.gov/data/sites/default/files/NSDUHsaeLongTermCHG2014/NSDUHsaeLongTermCHG2014.htm

$$\text{cov}[\ln(\hat{\theta}_1), \ln(\hat{\theta}_2)] = \text{correlation } [\ln(\hat{\theta}_1), \ln(\hat{\theta}_2)] \times \sqrt{v[\ln(\hat{\theta}_1)] \times v[\ln(\hat{\theta}_2)]} \,.$$

Note that $v[\ln(\hat{\theta}_1)]$ and $v[\ln(\hat{\theta}_2)]$ used here to calculate $v(\hat{lor}_{sa})$ are the same variances used in calculating the published 2011–2012 Bayesian CIs and the 2012-2013 Bayesian CIs, respectively.

The correlation between $\ln(\hat{\theta}_1)$ and $\ln(\hat{\theta}_2)$ was obtained by simultaneously modeling the 2011, 2012, and 2013 NSDUH data. This simultaneous modeling approach was adopted based on the results of the validation study[5] conducted for measuring change in the 1999–2000 and 2000–2001 state estimates. For this simultaneous model, 4 age groups (12 to 17, 18 to 25, 26 to 34, and 35 or older) by 3 years (2011, 2012, and 2013), that is, 12 subpopulation-specific models, were fitted, each with its own set of fixed and random effects. In this case, the general covariance matrices for the state and within-state random effects were 12 × 12 matrices corresponding to the 12 element (age group × year) vectors of random effects. Note that the survey-weighted, Bernoulli-type log likelihood employed in the SWHB methodology was appropriate for this simultaneous model because the 12 age group × year subpopulations were nonoverlapping. The correlation $[\ln(\hat{\theta}_1), \ln(\hat{\theta}_2)]$ was approximated by the correlation calculated using the posterior distributions of $\ln[\pi_{sa(1)} / (1 - \pi_{sa(1)})]$ and $\ln[\pi_{sa(2)} / (1 - \pi_{sa(2)})]$ from the simultaneous model.

To calculate the *p* value for testing the null hypothesis of no difference $(lor = 0)$, it is assumed that the posterior distribution of lor is normal with $mean = \hat{lor}_{sa}$ and $variance = v(\hat{lor}_{sa})$. With the null value of $(lor = 0)$, the Bayes *p* value or posterior probability of no difference is

$$p \text{ value} = 2 * P[Z \geq abs(z)],$$

where Z is a standard normal random variate, $z = \dfrac{\hat{lor}_{sa}}{\sqrt{v(\hat{lor}_{sa})}}$, and $abs(z)$ denotes the absolute value of z.

This test can also be extended to test nonoverlapping year changes. The correlation between the two multiyear time points will be derived from simultaneous models fit to all of the single-year data involved in the trend contrast. Note that this simultaneous modeling conditions each single-year small area estimate on the data from all of the other years and therefore overestimates the between "pooled year" covariances relative to those that would strictly apply to the independently estimated pooled-year small area estimates. This likely has some tendency to underestimate the variance of the associated log-odds ratios and to therefore reject more than the nominal 5 percent of the tests when there is no change. This simultaneous modeling strategy could be extended to include the intervening years in a longer trend series and thus provide the single-year variances and covariances required to fit efficient linear and possibly quadratic trend

[5] See Appendix E, Section E.2, pp. 152–155, of the Wright (2003) report.

lines to state-by-age group small area estimates. The association between-year correlations would suffer from some of the same overestimation discussed above.

In an issue of *The CBHSQ Report*, Hughes, Lipari, and Williams (2015) used NSDUH data to present national and state (including the District of Columbia) estimates of past month marijuana use and perceptions of great risk from smoking marijuana once a month among youths aged 12 to 17. Reported estimates were annual averages based on combined 2013 and 2014 NSDUH data from 39,600 respondents. Using the testing method described above, the combined 2013–2014 estimates were compared with estimates from combined 2012–2013 data, which were based on responses from 45,000 youths aged 12 to 17. A verbatim example of how the test results were presented in their report (emphases added here in boldface) follows:

> When combined 2012–2013 estimates are compared with combined 2013–2014 estimates, the nation as a whole **did not experience a significant change** in the rate of past month marijuana use among adolescents (7.15 percent in 2012–2013 and 7.22 percent in 2013–2014) (Table 1). On an individual state level, **three states experienced a statistically significant decrease** in the rate of adolescent past month marijuana use (Hawaii: from 9.55 to 7.65 percent, Ohio: from 7.36 to 6.04 percent, and Rhode Island: from 12.95 to 10.69 percent). The remaining 47 states and the District of Columbia experienced no change in past month marijuana use.

This page intentionally left blank

3. Summary of Trend Testing Methods Used by Other Agencies

3.1 Overview

This chapter summarizes the methods used by a few other agencies under the U.S. Department of Health and Human Services (HHS)[6] and other federal statistical agencies for analyzing trends and identifying breaks in trends. For a list of program staff and agencies who were contacted for this report, see the Acknowledgments on p. ii. For a list of the examined documents that discussed or included any performed trend analysis, see Appendix A. Also, for a summary of the methods and approaches used by these other agencies, see Appendices B and C, respectively.

From this investigation, it was discovered that the *t* test and linear and logistic regression models were the most common methods used for trend testing by these agencies. These methods can take complex sample design into account and can be easily implemented via SAS® callable SUDAAN®, Stata®, and survey analysis procedures in SAS. A few studies used more advanced modeling techniques, such as time series analysis, hierarchical and multilevel models, and Joinpoint regression. Time series, hierarchical, and multilevel models can take serial correlations when analyzing data with repeated observations. Hierarchical and multilevel models can also account for random errors at both individual and aggregated levels. However, the implementation of both methods to survey data is intricate in order to consider complex sample design and various weighting treatments (see Sections 3.3.2 and 3.3.3). Joinpoint regression was used in particular when a significant nonlinear trend was found to detect an apparent trend change. An example of how to apply this method to complex survey data can be found in Section 3.4.1.

3.2 *t* Test for Pairwise Comparisons and for Testing Linear and Quadratic Trends with Orthogonal Contrast Matrices across Multiple Time Points

The *t* test can be used to do pairwise comparisons between estimates at two time points and to identify significant trend changes. This method has been used in the production of detailed tables for the National Survey on Drug Use and Health (NSDUH) mainly because of its efficiency and ease of implementation. In addition, the *t* test with orthogonal contrast matrices can be used to test linear, quadratic, and even higher order trends across multiple time points. For more details on this method, see Sections 2.2 and 2.3 in Chapter 2.

The *t* test method is also widely applied in other agencies for trend testing, especially in their trend table productions. For example, the National Center for Health Statistics (NCHS) used the *t* test to do pairwise comparisons between 1988–1994 estimates and 1999–2004 estimates to show trends in oral health status in the United States (Dye et al., 2007). Ford (2013)

---

[6] These include HHS agencies other than the Substance Abuse and Mental Health Services Administration (SAMHSA).

and Carroll, Kit, and Lacher (2015) tested linear survey trends using *t* tests with orthogonal contrast matrices across multiple time points based on the National Health and Nutrition Examination Survey (NHANES) data. The Agency for Healthcare Research and Quality (AHRQ) applies the *t* test for pairwise comparisons frequently in their reports, such as their *Statistical Brief #478* to examine trends in medical expenses and use of medical treatments (Stagnitti, 2015). The Monitoring the Future (MTF) study has also evaluated the trends in youth drug use while employing the *t* test for pairwise comparisons (Miech, Johnston, O'Malley, Bachman, & Schulenberg, 2015).

3.3 Regression Analysis for Trend Testing

3.3.1 Logistic Regression and Linear Regression

Other than the *t* test, logistic regression and linear regression are also commonly used for trend analyses by other federal agencies that were investigated. The complexity of the survey sample design (e.g., stratification and clustering) and sampling weights are often considered in the regression analysis when analyzing complex survey data. Some examples are presented in the following sections.

3.3.1.1 National Center for Health Statistics

Branum and Jones (2015) used data from the 1982, 1988, 1995, 2002, 2006–2010, and 2011–2013 National Survey of Family Growth (NSFG) to examine trends in current long-acting reversible contraceptive (LARC) among women aged 15 to 44. The significance of trends was tested using weighted least-squares regression models to determine whether an apparent change over time was statistically significant. Branum and Lukacs (2008) also used weighted least squares regression to test linear trends in the prevalence and hospitalizations of food allergy among children while using National Health Interview Survey (NHIS) data. The Centers for Disease Control and Prevention (CDC) applied logistic regression on NHIS data to analyze trend changes in national smoking prevalence from 2005 through 2010, overall and by age, race/ethnicity, education, poverty status, and U.S. census region (CDC, 2011).

3.3.1.2 Other Studies

Tong et al. (2013) examined the trends in smoking before, during, and after pregnancy from 40 Pregnancy Risk Assessment Monitoring System (PRAMS) sites from 2000 through 2010 using PRAMS data. Logistic regression with the infant's birth year and other variables as the independent variables was used to assess the linear trends. In order to assess current tobacco use among middle and high school students, the CDC analyzed data from the 2000 and 2011 National Youth Tobacco Survey (NYTS) (CDC, 2012). Logistic regression analysis was also used to analyze trend changes from 2000 to 2011 in tobacco use, by school level, controlling for other factors (including grade, race/ethnicity, and sex). Assessment for linear and quadratic trends were conducted simultaneously. Sequist et al. (2011) applied logistic regression to analyze trends in ambulatory quality of care and physician reports of barriers to quality improvement within the Indian Health Service (IHS).[7]

3.3.2 Time Series Analysis

A few studies also used time series models to take serial correlation[8] into account. Tiesman et al. (2015) enumerated suicides occurring in U.S. workplaces and compared workplace suicide trend estimates with estimates of suicides occurring outside the workplace between 2003 and 2010. They used suicide data from the Census of Fatal Occupational Injury database and the CDC's Web-based Injury Statistics Query and Reporting System (WISQARS). Suicide rates were calculated using denominators from the 2013 Current Population Survey (CPS) and the 2000 U.S. population census. They used autoregressive models with a first-order autoregressive error structure AR(1) to assess the trends of suicide rates while accounting for serial correlation.[9] Both linear and quadratic time variables were tested through modeling to determine the best fit of the suicide data.

3.3.3 Hierarchical and Multilevel Models

The National (Nationwide) Inpatient Sample (NIS) is an annual database of hospital inpatient stays, which is part of the Healthcare Cost and Utilization Project (HCUP) sponsored by AHRQ. Researchers and policymakers use the NIS to analyze national trends in health care utilization and quality. Houchens, Ross, and Elixhauser (2015) enumerated the important revisions to the NIS sample design between 1988 and 2012, suggested ways to manage these changes, and offered advice on statistical methods that may be useful for investigating trends. In their report, they discussed the use of hierarchical or multilevel regressions for trend testing. In NIS trend studies, discharges are nested within years nested within hospitals. Hierarchical models can account for the errors at both the discharge level and the hospital level, the correlation among discharges within hospitals, and serial correlation over time. The sample weights and survey design elements should be incorporated in the modeling. Note that weights at a different level (e.g., hospital and discharges) need to be derived and incorporated separately in hierarchical and multilevel models. If any nonresponse or noncoverage weighting adjustments need to be made, the adjustment should be at a different level as well. This requires additional effort to re-create weights when applying these types of models to NSDUH (see Section 5.5 in CBHSQ, 2016c).

---

[7] Although the IHS published an annual report on *Trends in Indian Health*, they did not perform any trend testing for the trend estimates provided in the report. For the 2014 edition of the report, see https://www.ihs.gov/dps/publications/trends2014/.

[8] An error term is considered to be serially correlated when error terms from different (usually adjacent) time periods are correlated. In time series studies, serial correlation occurs when the errors associated with a given time period carry over into future time periods. For example, if the prevalence of substance use is being predicted, an overestimate in 1 year is likely to lead to overestimates in succeeding years.

[9] This regression did not consider either survey weights or design variables because the data analyzed were not survey data but suicide rates calculated based on other data sources.

3.4 Testing a Nonlinear Trend and Identifying an Apparent Trend Change

Different methods can be used to test quadratic or even higher order trends (e.g., a cubic trend), such as the *t* test with orthogonal contrast matrices, and adding a squared term (i.e., quadratic term) or a higher order term of the time variable to the regression models. If a quadratic (or higher order) trend is found to be significant, researchers in some federal agencies may opt to use time series graphs to identify the time point where the trend change occurs, while others will adopt more sophisticated statistical techniques, such as Joinpoint regression, to identify the trend change.

3.4.1 Three-Step Analysis in Youth Risk Behavior Survey (YRBS)

The Youth Risk Behavior Surveillance System (YRBSS) was developed in 1990 to monitor priority health risk behaviors that contribute markedly to the leading causes of death, disability, and social problems among youths and adults in the United States. The CDC's National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Adolescent and School Health, presents online a method for conducting trend analyses of YRBS data to identify and describe changes in the prevalence of risk behaviors over time (CDC, 2016). This method consists of three steps as described in following three paragraphs.

First, the linear and nonlinear trends are tested via regression models with time variables plus other control variables as the independent variables. To test linear trends, a linear time variable will be included in the model. To test quadratic (or higher order) trends, both linear and quadratic (and higher order) time variables will be included in the model. All of the time variables are recoded into orthogonal coefficients with PROC IML in SAS (see CDC, 2016, for an example of the SAS code).

Second, if the *p* value for the quadratic (or higher order) time variable in the regression model is significant, then there is evidence of a trend change and Joinpoint software (available at https://surveillance.cancer.gov/joinpoint/) will be used to determine the "joinpoints" where the trends change.

Third, after the joinpoint(s) are determined, a regression model with a linear time variable (not a higher order time variable) will be used to test the linearity of the resulting line segments.

In terms of trend result interpretation, CDC (2014) presented some examples on how to interpret a YRBS trend fact sheet.[10] For example, to interpret long-term change, this document included the following verbatim statement:

> The YRBS trend fact sheet column titled "Long-term Change" describes statistically significant linear and quadratic changes over time from 1991 (or the first year in which the data were collected) through 2013 based on logistic regression analyses for each selected behavior. If there is a statistically significant linear trend, then this situation is described as "Increased, 1991–2013" or

[10] YRBS trend fact sheets are available at https://www.cdc.gov/healthyyouth/data/yrbs/results.htm.

"Decreased, 1991–2013." If there is a statistically significant quadratic trend, both parts of the quadratic trend are described, such as "Decreased, 1991–1999 and No change, 1999–2013." If there is no statistically significant linear or quadratic trend, then this situation is described as "No change, 1991–2013."

3.4.2 Other Studies That Use Orthogonal Polynomial Trend Contrasts

The YRBSS is not the only study that uses orthogonal polynomial trend contrasts for testing linear and nonlinear trends in regression analysis. With NHIS data from 2002 to 2012, Clarke, Black, Stussman, Barnes, and Nahin (2015) also used the SAS procedure PROC SURVEYLOGISTIC with orthogonal polynomial trend contrasts to perform weighted linear or quadratic regressions of the annual design-adjusted rates for each outcome variable. Jamal, Dube, and King (2015) analyzed data for aggregated hospital outpatient visits among patients aged 18 years or older from the 2005–2010 National Hospital Ambulatory Medical Care Survey (NHAMCS). Logistic regression with orthogonal polynomials was used to analyze linear trends of tobacco use screening and cessation assistance offered to U.S. adults during their hospital outpatient clinic visits from 2005 through 2010, adjusted for covariates, including sex, age, and race/ethnicity ($\alpha = .05$).

3.4.3 Other Studies That Use Joinpoint Regression

When a quadratic trend is found to be significant, some researchers determine when the trend has changed based on time series graphs or tables. Ford and Dietz (2013), for example, examined trends in energy intake by using linear regression with time specified as the midpoint of the NHANES studies. They also examined nonlinear trends for time by adding a squared term (i.e., a quadratic term) to the models. They calculated unadjusted mean energy intake and mean energy intake adjusted for age, sex, race or ethnicity and other covariates) by using analysis of covariance (ANCOVA). They examined the trends for both unadjusted mean energy intake and adjusted mean energy intake separately. If the quadratic term was found to be significant, the unadjusted or adjusted mean energy intake would be examined over time based on the time series graphs or tables to find the breaks in trends. The following statement that interprets the significant quadratic term is quoted verbatim from their paper:

> Mean energy intake adjusted for age, sex, race or ethnicity, educational status, and BMI increased by 314 kcal (95% CI: 259, 368 kcal) from 1971–1975 through 2003–2004 and then decreased by 74 kcal (95% CI: 21, 126 kcal) in subsequent years (Table 1). For the unadjusted and adjusted means, the signs (both for the period 1971–1975 to 2009–2010 and for 1999–2000 to 2009–2010) for the regression coefficients of the quadratic term for time were negative (data not shown) and the *P* values were significant, suggesting that the upward trend in energy intake before ~2003–2004 had changed course and was decreasing in recent years.

A few studies also used the Joinpoint regression for testing trend changes. Akinbami et al. (2012) and Jackson, Howie, and Akinbami (2013) tested the trends with weighted least squares regression models of the log of each outcome using Joinpoint software to determine whether an apparent change over time was statistically significant.

3.5 Trend Testing without Considering Sample Variation at Each Time Point

In 1997, the Division of Science, Education and Analysis (which became the Division of Research in 1999) within the Maternal and Child Health Bureau (MCHB) published a document on trend analysis and interpretation (Rosenberg, 1997) in which some statistical procedures for trend testing are described. Two statistical methods were mentioned: (a) chi-square test for linear trends and (b) regression analysis. Unlike most trend analyses in NSDUH that use all of the observed values collected from individuals in each survey year, the trend analyses that Rosenberg suggested are mainly for analyzing aggregated data with the observed series of rates or counts. Thus, there is only one value (e.g., the prevalence rate or weighted count) at each time point, and the SE for the rate or count (caused by sample variation) at each time point is ignored.

The MCHB is not the only agency that does trend testing by ignoring sample variation. With the NHANES data, Will, Yuan, and Ford (2014) used the analysis of variance (ANOVA) type of Cochran–Mantel–Haenszel (CMH) test in survey data analysis to test for trends in each study subgroup. This method is similar to a chi-square test.

3.6 Caveats When Doing Trend Testing with Medical Expenditure Panel Survey (MEPS) Data

An online report by AHRQ's Center for Financing, Access, and Cost Trends (2011) includes a section on "Using MEPS Data for Trend Analysis" (p. C-13) in which they discuss some caveats when doing trend testing with Medical Expenditure Panel Survey (MEPS) data. The verbatim text is quoted below with emphases added here in boldface:

> MEPS began in 1996, and the utility of the survey for analyzing health care trends expands with each additional year of data. However, it is important to consider a variety of factors when examining trends over time using MEPS. Statistical significance tests should be conducted to assess the likelihood that observed trends **may be attributable to sampling variation**. The length of time being analyzed should also be considered. In particular, large shifts in survey estimates over short periods of time (e.g. from one year to the next) that are statistically significant should be interpreted with caution, unless they are attributable to known factors such as changes in public policy, economic conditions, or MEPS survey methodology. Looking at changes over longer periods of time can provide a more complete picture of underlying trends. Analysts may wish to consider **using techniques to smooth or stabilize analyses of trends** using MEPS data such as comparing pooled time periods (e.g. 1996-97 versus 2004-05), working with moving averages, or using modeling techniques with several consecutive years of MEPS data to test the fit of specified

patterns over time. Finally, researchers should be aware of **the impact of multiple comparisons on Type I error.** Without making appropriate allowance for multiple comparisons, undertaking numerous statistical significance tests of trends increases the likelihood of concluding that a change has taken place when one has not.

Most of the reports from MEPS use only *t* tests to compare estimates between two time periods, and a few research articles from this study have used regression methods. The sample weights and survey sample design are always considered in their analyses. However, no reports or articles were found that addressed the other caveats, other than the sampling variation (from a complex sample design) mentioned above.

This page intentionally left blank

4. Literature Review of Selected Topics in Trend Testing

4.1 Overview

This chapter provides a literature review of selected topics related to trend testing. It focuses on methods that address specific issues found in the trend testing approaches used by the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), in the National Survey on Drug Use and Health (NSDUH) and those used by some other federal agencies. In most cases, the literature review was limited to articles published within the past 15 years. For some established methods, however, the search was extended further back in time to get a more comprehensive selection of articles. As part of the literature review, several popular textbooks on time series were consulted to assess how well they cover the topics of interest, including textbooks by Hamilton (1994), Brockwell and Davis (2002), Prado and West (2010), and Shumway and Stoffer (2011). With the exception of Bayesian methods and splines, none of the topics of interest is covered in these texts.

As mentioned in Chapter 1, the literature review was mainly conducted using two databases of published articles: (a) the Current Index to Statistics (https://www.statindex.org/) for the latest methodological articles, and (b) the National Library of Medicine's PubMed (https://www.ncbi.nlm.nih.gov/pubmed) for the latest articles describing the application of trend analysis methods to the health sciences. Most of the methods described in this chapter are rarely applied but may have the potential to be adapted in analyses of complex survey data.

A complete listing of the citations reviewed is provided within this report's reference list, following a list of the references cited in other chapters. Literature review citations are organized by topic, then alphabetically by author within each topic. In the sections that follow, key findings from the literature review are summarized. For a complete summary table that lists the methods investigated under the literature review, see Appendix B.

4.2 Detecting Outliers in Trends and Time Series

The literature was searched to find articles related to detecting outliers or extreme values in time series data. This topic is not covered in the textbooks surveyed, so the search was extended back further than 15 years to develop a complete picture of these methods.

Fox (1972) introduced the earliest methods for outlier detection in time series. Fox classified outliers as belonging to one of two types: Type 1 (sometimes called additive outliers) due to error in observation or recording, and Type 2 (sometimes called innovation outliers) due to extreme events. Fox also developed likelihood-based tests to detect each type. Muirhead (1986) developed likelihood ratio and Bayes rules for classifying an outlier when the type is unknown. Tsay (1988) developed a batch-type (i.e., uses the entire dataset) iterative procedure to detect outliers as well as level shifts (drifts) in the trends and changes in the variance over time, a method that is commonly cited in the literature. Balke (1993) showed that

Tsay's method does not work well in the presence of level shifts in the time series and proposed a modification. Peña (1990) developed methods to test for influential observations (those that are not errors but affect estimation) in autoregressive integrated moving average (ARIMA) models and measured their impact on the model parameters. Ljung (1993) connected the dual problems of estimating additive outliers and estimating missing observations and provided a recursive procedure for estimating outliers and smoothing the time series. Choy (2001) proposed an iterative spectrum-based outlier detection algorithm assuming that the underlying outlier-free process follows a linear white noise process. Tsay, Peña, and Pankratz (2000) compared outliers in univariate time series to outliers in multivariate time series and developed an iterative procedure for estimating the latter. Galeano, Peña, and Tsay (2006) developed a method of identifying outliers in multivariate time series by looking for outliers in the projected univariate time series.

For particularly long time series, interest is in identifying sequences of unusual observations that may themselves identify a trend. Keogh, Lonardi, and Yuan-chi Chiu (2002) and Keogh, Lin, and Fu (2005) described methods for finding unusual sequences of values in large time series datasets. Gupta, Gao, Aggarwal, and Han (2014) provided an overview of methods for detecting outliers in high-dimensional data and data streams. Smith (1989) applied extreme value theory to identify periods of high ozone in a time series of pollutant data. Frei and Schär (2001) developed procedures to test for trends in rare climatological events and discriminate these trends from stochastic fluctuations.

"Exception reporting" is a term used to describe a system for monitoring incoming data for inconsistencies with an established pattern in previously collected data. McCabe, Greenhalgh, Gettinby, Holmes, and Cowden (2003) described how methods such as exponentially weighted moving averages, zero-inflated Poisson models, and generalized linear models can be used to monitor for outbreaks in infectious diseases. No other articles were found that were related to this topic, mostly because the term "exception reporting" is used in other contexts (i.e., it also refers to a practice in the United Kingdom in which doctors can select which patients to exclude from performance metrics, a topic of considerable interest to researchers in health policy). The concept appears to be related to online, sequential searches for change points, which is discussed later in this chapter.

4.3 Nonparametric and Bayesian Methods in Time Series

The methods used by other federal agencies tend to be based on frequentist hypothesis tests under the assumption of (asymptotically) normally distributed data. The literature was searched to find nonparametric and Bayesian alternatives to these methods.

4.3.1 Nonparametric Methods

The search for nonparametric methods focused on three techniques.

- The Pettitt (1979) test is a sign test for the hypothesis that a sudden change in mean occurred at time t against the null hypothesis that no change occurred at that time point.
- The Mann-Kendall test uses ranks to test the hypothesis that there is a monotonic trend against the null hypothesis of no trend.

- Sen's slope is an estimator of the slope of a linear trend based on the median of differences between time points.

The search revealed that methods for detecting trends and breaks in trends (particularly these three nonparametric methods) are popular in the environmental sciences for analyzing climate data. Several of the papers found contain useful descriptions and comparisons of nonparametric methods, and for these reasons they may be useful for analyzing NSDUH data. Beharry, Clarke, and Kurmarsingh (2014) provided detailed descriptions of the Pettitt test, Mann-Kendall test, and Sen's slope estimator, and they applied these methods to precipitation data. Gocic and Trajkovic (2013) described the mathematical details for Mann-Kendall, Sen's slope, and the cumulative sum (CUSUM) method for change points and applied the techniques to meteorological data in Serbia. Hess, Iyer, and Malm (2011) compared seven different parametric and nonparametric methods for detecting trends in environmental data, including a version of the Mann-Kendall method for seasonal data. Costa and Soares (2009) described methods used to detect climactic trends from weather data, including the Mann-Kendall and Pettitt tests, as well as other procedures for testing for trends and randomness in the data. Hall and Tajvidi (2000) developed a nonparametric model for identifying trends (linear and nonlinear) in extreme values from a time series and applied the method to temperature data. Yue, Pilon, and Cavadias (2002) demonstrated that the Mann-Kendall test was indistinguishable in practice from Spearman's rank correlation coefficient (i.e., Spearman's rho), another nonparametric method for identifying monotonic trends.

Outside of environmental sciences, Zhou, Zou, Zhang, and Wang (2009) applied the Mann-Kendall test to control charts to monitor for change points in a process. Brodsky and Darkhovsky (2013), in an e-book edition of their 1993 textbook, provided a treatment of nonparametric change point methods.

4.3.2 Bayesian Methods

The time series textbook by Prado and West (2010) that was mentioned in Section 4.1 focuses on Bayesian methods for analyzing time series data. The literature search found additional articles that may be useful for understanding Bayesian approaches to model trends, changes in trends, or unusual time points in NSDUH data. A distinction is often made in the literature between methods that rely on Markov chain Monte Carlo (MCMC) methods and methods that can estimate parameters directly from the posterior distribution. The latter technique avoids the difficulty of ensuring that the MCMC simulation converges.

Abraham and Box (1979) introduced a Bayesian method for detecting both additive and intervention outliers in autoregressive models. McCulloch and Tsay (1993) applied Bayesian methods to conduct statistical inference on changes to the mean and variance of an autoregressive time series. Punskaya, Andrieu, Doucet, and Fitzgerald (2002) used MCMC methods to fit a sequence of piecewise linear trends over a time series, using a prior distribution to identify the number and timing of the change points. Giordani and Kohn (2012) provided a Bayesian method to model shifts in variance and to identify dates of breaks in change point models. Lai and Xing (2011) developed a Bayesian model for multiple change points in a series and demonstrated how the method could be used to solve frequentist problems in change point problems, such as testing for multiple change points versus no change points and checking inferences on the numbers and locations of change points. Bernardo, Moreno, and Casella (2007) described Bayesian methods for detecting multiple change points in sequences of independent but not identically distributed observations. Tartakovsky and Moustakides (2010) provided an overview of the quickest Bayesian method to detect a change point and described its statistical properties, such as the probability of a false detection and the average length of the sequence needed to detect a change. Thum and Bhattacharya (2001) described a model for detecting change and joinpoints in a system when clusters within the system change at differing time points. Assareh, Smith, and Mengersen (2011) applied Bayesian hierarchical modeling to detect change points in a control chart monitoring patient outcomes after cardiac surgery.

4.4 Methods Related to Joinpoint Regression

Joinpoint regression was identified as a method applied to the Youth Risk Behavior Survey (YRBS) to detect trends and breaks in trends. The method is commonly used in cancer studies, and the National Cancer Institute makes software available to implement this technique. Kim, Fay, Feuer, and Midthune (2000) described Joinpoint regression and addressed the issue of adjusting the type I error rate due to testing multiple hypotheses. The literature was searched for information on change point analysis, splines, and piecewise polynomials, techniques that are very similar to the Joinpoint approach.

Several review articles were identified that cover the variety of methods used in change point analysis. Reeves, Chen, Wang, Lund, and Lu (2007) provided a detailed review of several methods for change point detection in climate data. Eckley, Fearnhead, and Killick (2011) provided detailed descriptions of methods to detect change points, both from a frequentist and a Bayesian perspective. Jandhyala, Fotopoulos, MacNeill, and Liu (2013) reviewed frequentist and Bayesian methods of inference for detecting change points, with a focus on advanced computational methods for detecting multiple change points, particularly when the timing of the change points is unknown. Additional change point approaches not covered in Jandhyala et al. (2013) were also found. Wu, Woodroofe, and Mentz (2001) developed a test based on isotonic regression for monotonic trends in short-range dependent sequences. Moskvina and Zhigljavsky (2003) proposed a method to detect change points based on the singular value decomposition of the time series trajectory matrix and compared this approach with methods based on the cumulative sum statistic. Samé, Chamroukhi, Govaert, and Aknin (2011) developed a method based on the expectation-maximization algorithm to detect change points in clustered temporal data. Wang, Wen, and Wu (2007) developed a method for detecting a single change point that is less sensitive than other methods as to where the change point falls in the series. Frick, Munk, and Sieling (2014) used dynamic programming to maximize the probability of correctly identifying the number of change points, estimate change point locations and associated simultaneous confidence intervals (CIs), and compute the values of the sequence at the change points and associated confidence bands. Matteson and James (2014) developed a nonparametric approach to detect multiple change points in multivariate data.

Splines and piecewise polynomials are addressed in the time series textbooks by Prado and West (2010) and Shumway and Stoffer (2011) mentioned in Section 4.1. Some additional literature that may be of interest was also identified. Dominici, McDermott, Zeger, and Samet (2002) applied generalized additive models with smoothing splines to study the time series of air pollution data. Orváth and Kokoszka (2002) detected change points using a smoothed polynomial regression model. Lemire (2007) proposed an approach for smoothing a time series in which the degree of the polynomial in each segment can vary. Dong and Roychowdhury (2003) presented a nonlinear model reduction approach based on piecewise polynomial representations. Huang and Shen (2004) proposed a global smoothing method based on polynomial splines for the estimation of functional coefficient regression models for non-linear time series.

4.5 Sensitivity of Analyses to Time Window

A topic of interest is how the results of a trend analysis may be sensitive to the "time window" over which the data are analyzed (e.g., the time window is from 2002 to 2008 when analyzing trends in the 2002–2008 NSDUH data). The search of terms related to this topic did not result in any articles focusing on this topic. This topic can be explored further when specific techniques for modeling NSDUH data are compared.

4.6 Scanning Statistics

One method for trend detection relies on scanning windows of the series and testing whether each window is consistent with a trend. Glaz, Naus, and Wallenstein (2001) provided a textbook treatment of the subject. A particular challenge with scanning statistics is controlling the type I error rate because multiple correlated hypotheses are being tested. Siegmund, Zhang, and Yaskir (2011) developed methods for estimating and controlling the false discovery rate of scanning statistics.

4.7 Updating Analyses in the Presence of New Data

As new NSDUH data are collected, previous trend analyses need to be updated, and new observations need to be evaluated as potential outliers. No articles on this topic could be found as it relates to small time series collected over spaced-out intervals. However, monitoring streams of data collected at frequent intervals is an area of active research. Methods to detect trends or outliers as the data are collected are described as "online" or "sequential," in contrast to "offline" methods that examine data retrospectively. Articles related to online data monitoring are summarized in this section, although these methods may not be applicable to NSDUH data.

Blazek, Kim, Rozovskii, and Tartakovsky (2001) developed an algorithm to detect change points as soon as possible while controlling the probability of a false positive and applied this method to detecting cyberattacks on a network. Adams and MacKay (2007) introduced a Bayesian approach for inference on the most recent change point from online data. Fearnhead and Liu (2007) proposed a Bayesian algorithm for detecting multiple change points from online data by simulating directly from the posterior distribution (as opposed to using MCMC). Killick, Fearnhead, and Eckley (2012) developed an online change point detection algorithm when the number of likely change points increases as more data are collected. Brown, Grassly, Garnett, and Stanecki (2006) adopted an approach called "level fitting" for trend analysis of HIV prevalence to adjust for expansion of national surveillance systems into lower prevalence sites, which assumes that all surveillance sites in a region follow a similar trend pattern of rise and fall, but that each individual site may be at different prevalence levels.

This page intentionally left blank

5. Summary and Future Research

In this investigation of trend analyses used in the National Survey on Drug Use and Health (NSDUH) and trend analyses used by other federal agencies, it was discovered that the *t* test is used widely to do pairwise comparisons between estimates at two time points. This method is used in the production of various descriptive tables and short reports in NSDUH and other studies due to its efficiency and ease of implementation. The null hypothesis underlying this method is that an estimate did not change from one time point to another. The alternative hypothesis is that an estimate changed significantly from one time point to another. For a detailed description about this method as well as some examples of its applications in NSDUH, see Section 2.2 in this report. For examples of the *t* test's use by other federal agencies, see Section 3.2 and Appendices A and C.

If a long-term trend across multiple time points is of interest, a *t* test with orthogonal contrast matrix can be used to test linear, quadratic, and even higher order trends. Similar to the *t* test for pairwise comparisons, this method can be embedded in the SUDAAN® or SAS® procedures that produce the direct estimates and therefore is easy to implement under given time and cost constraints. When testing linear trends, the underlying null hypothesis can be stated as the slope of a trend line equal to 0; the alternative hypothesis is that the slope is significantly different from 0 and an estimate significantly increased or decreased over time. When testing for curvature in a trend, the underlying null hypothesis is that the trend is linear over time; the alternative hypothesis is that the trend had a significant curvilinear component. For a detailed description about this method and some examples of its applications in NSDUH, see Section 2.3 in this report. For examples of its use by other agencies, see Section 3.2 and Appendices A and C.

There are some limitations in the *t* test with orthogonal contrast matrix when testing linear or higher order trends. For example, when testing multiple time points, this method can be applicable only for equally spaced time points. Statistical regression methods can compensate for this limitation and thus are commonly used for more complex analytical tasks in NSDUH and other studies. When analyzing trends with regression methods, one can control for one or more covariates (factors), detect and control any outlier in the trend, identify breaks in trends, take serial correlation into account (when using time series models), and estimate variance at different levels in hierarchical data. Implementation of regression methods, however, can be much more expensive and time-consuming than the *t* test, depending on the complexity of the regression model and the survey design. For some simple tasks, one may only need to run a linear regression with a single time variable to estimate the slope of a trend line. Implementation of a regression method under this circumstance should require a relatively low cost. Nevertheless, for some more complicated tasks, one may need to conduct a series of model selections to find an appropriate regression model that fits specified analytic purposes, then perform a model assessment to ensure that the underlying model assumptions are not violated with the analyzed data. It can be even more challenging when applying some advanced modeling techniques, such as time series models and multilevel models, to NSDUH and other similar survey data. Sophisticated methods need to be developed to account for complex sample design and various weighting treatments in conjunction with serial correlation and hierarchical structure in the data.

For examples of the trend analysis with regression method, see Sections 2.4 and 3.3 in this report.

When a significant quadratic or higher order change has been detected, Joinpoint regression can be used to determine the location of one joinpoint for a significant quadratic or two joinpoints for a significant cubic trend. See Section 3.4 for more details on Joinpoint regression. Through the literature review, several other techniques were found for change point analysis that are very similar to the joinpoint approach. See Section 4.4 and Appendix C for more details about these techniques.

Although several statistical methods for detecting outliers in trends have been developed over the past several decades, no use of these methods was found for outlier detection in trend analysis at any other federal agency. Section 4.2 provides an overview on these methods. Similarly, various nonparametric and Bayesian methods have been developed for trend analysis (see Section 4.3), but few federal agencies use these methods for trend analysis, probably because they are not readily adaptable to survey data. The methods used by all of the agencies tend to be based on frequentist hypothesis tests under the assumption of (asymptotically) normally distributed data, which can be violated when the data distribution is skewed. These areas can be explored further in future research to run experiments using these methods with NSDUH data.

It has been recognized that the results of a trend analysis may be sensitive to the time window over which the data are analyzed (e.g., see Section 3.6). However, no article focusing on this topic was found through the literature review. Moreover, little relevant literature was found in the areas of scanning statistics and updating analyses in the presence of new data. These areas could also be valuable to explore with NSDUH data to test the sensitivity of the trend results to time windows with different lengths and in the presence of new data.

References

Akinbami, L. J., Moorman, J. E., Bailey, C., Zahran, H. S., King, M., Johnson, C. A., & Liu, X. (2012). Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS Data Brief*, *94*, 1–8. Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/data/databriefs/db94.htm

Branum, A. M., & Jones, J. (2015). Trends in long-acting reversible contraception use among U.S. women aged 15–44. *NCHS Data Brief, 188*. Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/products/databriefs/db188.htm

Branum, A. M., & Lukacs, S. L. (2008). Food allergy among U.S. children: Trends in prevalence and hospitalizations. *NCHS Data Brief, 10*. Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/products/databriefs/db10.htm

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting* (2nd ed.). New Yok, NY: Springer-Verlag.

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

Center for Behavioral Health Statistics and Quality. (2014). *Results from the 2013 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 14-4863, NSDUH Series H-48). Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015a). *Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health* (HHS Publication No. SMA 15-4927, NSDUH Series H-50). Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015b). *Suicidal thoughts and behavior among adults: Results from the 2014 National Survey on Drug Use and Health*. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015c). *Receipt of services for behavioral health problems: Results from the 2014 National Survey on Drug Use and Health*. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015d). *Results from the 2014 National Survey on Drug Use and Health: Detailed tables*. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015e). *Risk and protective factors and initiation of substance use: Results from the 2014 National Survey on Drug Use and Health*. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016a). *2014 National Survey on Drug Use and Health: Methodological resource book (Section 13, statistical inference report)*. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016b). *2014 National Survey on Drug Use and Health: Methodological resource book (Section 15, sample redesign impact assessment, final report)*. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016c). *National Survey on Drug Use and Health: Merging auxiliary data with NSDUH data for mental health research*. Retrieved from https://www.samhsa.gov/data/sites/default/files/NSDUH-N17-Merging%20Auxillary%20Data%20with%20NSDUH-2016.pdf

Center for Behavioral Health Statistics and Quality. (in press). *2015 National Survey on Drug Use and Health: Methodological resource book (Section 14, sample experience report)*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Centers for Disease Control and Prevention. (2011). Vital signs: Current cigarette smoking among adults aged ≥ 18 years—United States, 2005-2010. *Morbidity and Mortality Weekly Report*, *60*, 1207–1212. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6035a5.htm

Centers for Disease Control and Prevention. (2012). Current tobacco use among middle and high school students—United States, 2011. *Morbidity and Mortality Weekly Report*, *61*, 581–585. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6131a1.htm

Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Adolescent and School Health. (2014, June). *Youth Risk Behavior Surveillance System (YRBSS): Interpretation of YRBS trend data*. Retrieved from https://www.cdc.gov/healthyyouth/data/yrbs/pdf/yrbs_trend_interpretation.pdf

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

Center for Financing, Access, and Cost Trends. (2011). *MEPS HC-128: 2009 medical conditions*. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from https://meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h128/h128doc.pdf

Clarke, T. C., Black, L. I., Stussman, B. J., Barnes, P. M., & Nahin, R. L. (2015). Trends in the use of complementary health approaches among adults: United States, 2002–2012. *National Health Statistics Reports, 79*. Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4573565/

Dye, B. A., Tan, S., Smith, V., Lewis, B. G., Barker, L. K., Thornton-Evans, G., . . . Li, C. H. (2007). Trends in oral health status: United States, 1988-1994 and 1999-2004. *Vital and Health Statistics. Series 11, Data from the National Health Survey*, *248*, 1–92. Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/data/series/sr_11/sr11_248.pdf

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

Ford, E. S., & Dietz, W. H. (2013). Trends in energy intake among adults in the United States: Findings from NHANES. *American Journal of Clinical Nutrition*, *97*, 848–853. Retrieved from https://doi.org/10.3945/ajcn.112.052662

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Houchens R. L., Ross D., & Elixhauser A. (2015). Using the HCUP national inpatient sample to estimate trends. *HCUP Methods Series Report, 2006–05*. Rockville, MD: U.S. Agency for Healthcare Research and Quality. Retrieved from https://www.hcup-us.ahrq.gov/reports/methods/methods.jsp

Hughes, A., Lipari, R. N., & Williams, M. (2015). *State estimates of adolescent marijuana use and perceptions of risk of harm from marijuana use: 2013 and 2014*. The CBHSQ Report. Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. Retrieved from https://www.samhsa.gov/data/sites/default/files/report_2121/ShortReport-2121.pdf

Jackson, K. D., Howie, L. D., & Akinbami, L. J. (2013). Trends in allergic conditions among children: United States, 1997–2011. *NCHS Data Brief, 121*. Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/products/databriefs/db121.htm

Jamal, A., Dube, S. R., & King, B. A. (2015). Tobacco use screening and counseling during hospital outpatient visits among US adults, 2005-2010. *Preventing Chronic Disease*, *12*, E132. https://doi.org/10.5888/pcd12.140529

Miech, R. A., Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2015). *Monitoring the Future national survey results on drug use, 1975–2014: Vol. I. Secondary school students*. Ann Arbor, MI: University of Michigan, Institute for Social Research. Retrieved from http://www.monitoringthefuture.org/pubs.html

Prado, R., & West, M. (2010). *Time series: Modeling, computation, and inference*. Boca Raton, FL: Chapman and Hall/CRC.

Ringeisen, H., Miller, S., Munoz, B., Rohloff, H. F., Hedden, S., & Colpe, L. J. (2016). Mental health service use across adolescence: Findings from the National Survey on Drug Use and Health. *Psychiatric Services, 67*, 787–789. https://doi.org/10.1176/appi.ps.201400196

Rosenberg, D. (1997). *Trend analysis and interpretation*. Rockville, MD: Maternal and Child Health Bureau.

RTI International. (2012a). *National Survey on Drug Use and Health: Sample redesign issues and methodological studies* (RTI/0209009.486.001 and 0211838.108.006.005, prepared for the Substance Abuse and Mental Health Services Administration under Contract Nos. 283–2004–00022 and HHSS283200800004C). Research Triangle Park, NC: Author. Retrieved from https://archive.samhsa.gov/data/NSDUH/NSDUHMethodsSIMS2012.pdf

RTI International. (2012b). *SUDAAN®, Release 11.0* [computer software]. Research Triangle Park, NC: Author.

Sequist, T. D., Cullen, T., Bernard, K., Shaykevich, S., Orav, E. J., & Ayanian, J. Z. (2011). Trends in quality of care and barriers to improvement in the Indian Health Service. *Journal of General Internal Medicine*, *26*, 480–486. Retrieved from https://doi.org/10.1007/s11606-010-1594-4

Shumway, R. H., & Stoffer, D. S. (2011). *Time series analysis and its applications, with R examples* (3rd ed.). New York, NY: Springer-Verlag.

Stagnitti, N. M. (2015). Trends in prescribed outpatient opioid use and expenses in the U.S. civilian noninstitutionalized population, 2002–2012. *Statistical Brief #478*. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from https://meps.ahrq.gov/data_files/publications/st478/stat478.shtml

Tiesman, H. M., Konda, S., Hartley, D., Chaumont Menéndez, C., Ridenour, M., & Hendricks, S. (2015). Suicide in U.S. workplaces, 2003-2010: A comparison with non-workplace suicides. *American Journal of Preventive Medicine*, *48*, 674–682. Retrieved from https://doi.org/10.1016/j.amepre.2014.12.011

Tong, V. T., Dietz, P. M., Morrow, B., D'Angelo, D. V., Farr, S. L., Rockhill, K. M., England, L. J., & Centers for Disease Control and Prevention. (2013). Trends in smoking before, during, and after pregnancy—Pregnancy Risk Assessment Monitoring System, United States, 40 sites, 2000-2010. *Morbidity and Mortality Weekly Report*, *62*, 1–19. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/ss6206a1.htm

Will, J. C., Yuan, K., & Ford, E. (2014). National trends in the prevalence and medical history of angina: 1988 to 2012. *Circulation: Cardiovascular Quality and Outcomes*, *7*, 407–413. Retrieved from https://doi.org/10.1161/circoutcomes.113.000779

Wright, D. (2003). *State estimates of substance use from the 2001 National Household Survey on Drug Abuse: Volume II. Individual state tables and technical appendices* (HHS Publication No. SMA 03-3826, NHSDA Series H-20). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

References for Literature Review

Detecting Outliers in Trends and Time Series

Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics*, *11*(1), 81–92. https://doi.org/10.1080/07350015.1993.10509934

Choy, K. (2001). Outlier detection for stationary time series. *Journal of Statistical Planning and Inference*, *99*(2), 111–127. https://doi.org/10.1016/s0378-3758(01)00081-7

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society. Series B. Methodological, 43*, 350–363.

Frei, C., & Schär, C. (2001). Detection probability of trends in rare events: Theory and application to heavy precipitation in the Alpine region. *Journal of Climate*, *14*, 1568–1584. https://doi.org/10.1175/1520-0442(2001)014<1568:dpotir>2.0.co;2

Galeano, P., Peña, D., & Tsay, R. S. (2006). Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association*, *101*, 654–669. https://doi.org/10.1198/016214505000001131

Gupta, M., Gao, J., Aggarwal, & Han, J. (2014). Outlier detection for temporal data. *Synthesis Lectures on Data Mining and Knowledge Discovery*, *5*(1), 1–129. https://doi.org/10.2200/s00573ed1v01y201403dmk008

Keogh, E., Lin, J., & Fu, A. (2005). Hot sax: Efficiently finding the most unusual time series subsequence. In *Proceedings of the Fifth IEEE International Conference on Data Mining*. https://doi.org/10.1109/icdm.2005.79

Keogh, E., Lonardi, S., & Yuan-chi Chiu, B. (2002). Finding surprising patterns in a time series database in linear time and space. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/775047.775128

Ljung, G. M. (1993). On outlier detection in time series. *Journal of the Royal Statistical Society. Series B. Methodological, 55*, 559–567.

McCabe, G. J., Greenhalgh, D., Gettinby, G. , Holmes, E., & Cowden, J. (2003). Prediction of infectious diseases: An exception reporting system. *Journal of Medical Informatics & Technologies*, *5*, 67–74.

Muirhead, C. R. (1986). Distinguishing outlier types in time series. *Journal of the Royal Statistical Society. Series B. Methodological, 48*, 39–47.

Peña, D. (1990). Influential observations in time series. *Journal of Business & Economic Statistics*, *8*(2), 235–241. https://doi.org/10.1080/07350015.1990.10509795

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science, 4*, 367–377. https://doi.org/10.1214/ss/1177012400

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, *7*(1), 1–20. https://doi.org/10.1002/for.3980070102

Tsay, R. S., Peña, D., & Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika*, *87*(4), 789–804. https://doi.org/10.1093/biomet/87.4.789

Nonparametric and Bayesian Methods in Time Series

Nonparametric

Beharry, S. L., Clarke, R. M., and Kurmarsingh, K. (2014). Precipitation trends using in-situ and gridded datasets. *Theoretical and Applied Climatology, 115*, 599–607. https://doi.org/10.1007/s00704-013-0921-1

Brodsky, E., & Darkhovsky, B. S. (2013). *Nonparametric methods in change point problems*. Dordrecht, The Netherlands: Springer Netherlands.

Costa, A. C., & Soares, A. (2009). Homogenization of climate data: Review and new perspectives using geostatistics. *Mathematical Geosciences*, *41*, 291–305. https://doi.org/10.1007/s11004-008-9203-3

Gocic, M., & Trajkovic, S. (2013). Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. *Global and Planetary Change*, *100*, 172–182. https://doi.org/10.1016/j.gloplacha.2012.10.014

Hall, P., & Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science, 15*, 153–167.

Hess, A., Iyer, H., & Malm, W. (2001). Linear trend analysis: A comparison of methods. *Atmospheric Environment*, *35*, 5211–5222. https://doi.org/10.1016/s1352-2310(01)00342-9

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics, 28*, 126–135. https://doi.org/10.2307/2346729

Yue, S., Pilon, P., & Cavadias, G. (2002). Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology (Amsterdam)*, *259*(1), 254–271. https://doi.org/10.1016/s0022-1694(01)00594-7

Zhou, C., Zou, C., Zhang, Y., & Wang, Z. (2009). Nonparametric control chart based on change-point model. *Statistical Papers*, *50*(1), 13–28. https://doi.org/10.1007/s00362-007-0054-7

Bayesian

Abraham, B., & Box, G. E. P. (1979). Bayesian analysis of some outlier problems in time series. *Biometrika*, *66*, 229–236. https://doi.org/10.1093/biomet/66.2.229

Assareh, H., Smith, I., & Mengersen, K. (2011). Bayesian change point detection in monitoring cardiac surgery outcomes. *Quality Management in Health Care*, *20*, 207–222. https://doi.org/10.1097/qmh.0b013e318220897e

Bernardo, J. M., Moreno, E., & Casella, G. (2007). Objective Bayesian analysis of multiple changepoints for linear models. *Bayesian Statistics 8: Proceedings of the Eighth Valencia International Meeting, June 2–6, 2006*. Vol. 8 (pp. 1–27). New York, NY: Oxford University Press.

Giordani, P., & Kohn, R. (2012). Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business & Economic Statistics, 26*(1), 66–77. https://doi.org/10.1198/073500107000000241

Lai, T. L., & Xing, H. (2011). A simple Bayesian approach to multiple change-points. *Statistica Sinica, 21*, 539–569. https://doi.org/10.5705/ss.2011.025a

McCulloch, R. E., & Tsay, R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, *88*, 968–978. https://doi.org/10.1080/01621459.1993.10476364

Punskaya, E., Andrieu, C., Doucet, C., & Fitzgerald, W. J. (2002). Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Transactions on Signal Processing*, *50*, 747–758. https://doi.org/10.1109/78.984776

Tartakovsky, A. G., & Moustakides, G. V. (2010). State-of-the-art in Bayesian changepoint detection. *Sequential Analysis*, *29*(2), 125–145. https://doi.org/10.1080/07474941003740997

Thum, Y. M., & Bhattacharya, S. K. (2001). Detecting a change in school performance: A Bayesian analysis for a multilevel join point problem. *Journal of Educational and Behavioral Statistics*, *26*, 443–468. https://doi.org/10.3102/10769986026004443

Methods Related to Joinpoint Regression

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. https://doi.org/10.1002/(sici)1097-0258(20000215)19:3<335::aid-sim336>3.3.co;2-q

Changepoint

Eckley, I. A., Fearnhead, P., & Killick, R. (2011). Analysis of changepoint models. In *Bayesian time series models*. In D. Barber, A. T. Cemgil, & S. Chiappa (Eds.), *Bayesian time series models* (pp. 205–224). Cambridge, United Kingdom: Cambridge University Press.

Frick, K., Munk, A., & Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *76*, 495–580. https://doi.org/10.1111/rssb.12047

Jandhyala, V., Fotopoulos, S., MacNeill, I., & Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis, 34*, 423–446. https://doi.org/10.1111/jtsa.12035

Matteson, D. S., & James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, *109*, 334–345. https://doi.org/10.1080/01621459.2013.849605

Moskvina, V., & Zhigljavsky, A. (2003). An algorithm based on singular spectrum analysis for change-point detection. *Communications in Statistics. Simulation and Computation*, *32*, 319–352. https://doi.org/10.1081/sac-120017494

Reeves, J., Chen, J., Wang, X. L., Lund, R., & Lu, Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, *46*, 900–915. https://doi.org/10.1175/jam2493.1

Samé, A., Chamroukhi, F., Govaert, G., & Aknin, P. (2011). Model-based clustering and segmentation of time series with changes in regime. *Advances in Data Analysis and Classification*, *5*, 301–321. https://doi.org/10.1007/s11634-011-0096-5

Wang, X. L., Wen, Q. H., & Wu, Y. (2007). Penalized maximal *t* test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology*, *46*, 916–931. https://doi.org/10.1175/jam2504.1

Wu, W., Woodroofe, M., & Mentz, G. (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika, 88(3),* 794–804. https://doi.org/10.1093/biomet/88.3.793

Splines and Piecewise Polynomials

Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, *156*, 193–203. https://doi.org/10.1093/aje/kwf062

Dong, N., & Roychowdhury, J. (2003). Piecewise polynomial nonlinear model reduction. In *Proceedings of the IEEE Design Automation Conference, 2003*, pp. 484–489. https://doi.org/10.1109/dac.2003.1219054

Huang, J. Z., & Shen, H. (2004). Functional coefficient regression models for non-linear time series: A polynomial spline approach. *Scandinavian Journal of Statistics*, *31*, 515–534. https://doi.org/10.1111/j.1467-9469.2004.00404.x

Lemire, D. (2007). A better alternative to piecewise linear time series segmentation. *Siam Data Mining*. Retrieved from https://arxiv.org/pdf/cs/0605103.pdf

Orváth, L., & Kokoszka, P. (2002). Change-point detection with non-parametric regression. *Statistics: A Journal of Theoretical and Applied Statistics, 36*(1), 9–31. https://doi.org/10.1080/02331880210930

Scanning Statistics

Glaz, J., Naus, J., & Wallenstein, S. (2001). *Scan statistics*. New York, NY: Springer-Verlag.

Siegmund, D. O., Zhang, N. R., & Yakir, B. (2011). False discovery rate for scanning statistics. *Biometrika*, *98*, 979–985. https://doi.org/10.1093/biomet/asr057

Updating Analyses in the Presence of New Data

Adams, R. P., & MacKay, D. J. C. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*. Retrieved from https://arxiv.org/pdf/0710.3742v1.pdf

Blazek, R. B., Kim, H., Rozovskii, B., & Tartakovsky, A. (2001). A novel approach to detection of "denial-of-service", attacks via adaptive sequential and batch-sequential change-point detection methods. In *Proceedings of IEEE Systems, Workshop on Information Assurance and Security, United States Military Academy, West Point, NY, 5–6 June 2001* (pp. 220–226).

Brown, T., Grassly, N. C., Garnett, G., & Stanecki, K. (2006). Improving projections at the country level: The UNAIDS Estimation and Projection Package 2005. *Sexually Transmitted Infect, 82(Suppl. 3)*, iii, 34–40. https://doi.org/10.1136/sti.2006.020230

Fearnhead, P., & Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *69*, 589–605. https://doi.org/10.1111/j.1467-9868.2007.00601.x

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107,* 1590–1598. https://doi.org/10.1080/01621459.2012.737745

This page intentionally left blank

Appendix A. Summary of Documents under This Review That Discussed or Used Trend Analysis

This page intentionally left blank

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
National Center for Health Statistics			
National Health Interview Survey (NHIS)	National Health Statistics Report[a]	Clarke et al. (2015)	***Method***: "PROC SURVEYLOGISTIC with orthogonal polynomial trend contrasts was used to perform weighted linear or quadratic regressions of the annual design-adjusted rates for each variable of interest." ***Example***: "There was **a quadratic change** in the overall use of any complementary health approach across the three time points **with a peak of** 35.5% in 2007." "Fish oil use among adults **increased** from 4.8% in 2007 to 7.8% in 2012."
	Data Brief[b]	Akinbami et al. (2012)	***Method***: Weighted least squares regression models of the log of each outcome with the time variable as a continuous independent variable were used with Joinpoint software to determine whether an apparent change over time occurred. ***Example***: "Asthma visits in primary care settings (physician offices and hospital outpatient departments) per 100 persons with asthma **declined** from 2001 to 2009 (Figure 3). Asthma ED visits and hospitalizations per 100 persons with asthma were **stable** from 2001 to 2009."
	Data Brief[b]	Jackson et al. (2013)	***Method***: "The significance of trends was tested using weighted least squares regression models of the log of each outcome and Joinpoint software to determine whether an apparent change over time was statistically significant, taking into account the standard error for each data point. Because there were limited data points over the period, linear regression (zero joinpoints) was specified for all models." ***Example***: "The prevalence of skin allergies **increased** from 7.4% in 1997–1999 to 12.5% in 2009–2011. There **was no significant trend** in respiratory allergies from 1997–1999 to 2009–2011, yet respiratory allergy remained the most common type of allergy among children throughout this period (17.0% in 2009–2011)."
	Data Brief[b]	Branum and Lukacs (2008)	***Method***: "Trend tests were performed to evaluate changes in reported food allergy over time using weighted least squares regression." ***Example***: "From 1997 to 2007, the prevalence of reported food allergy **increased** 18% among children under age 18 years."

See notes at end of table. (continued)

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis (continued)

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
	Morbidity and Mortality Weekly Report	Centers for Disease Control and Prevention (2011)	***Method***: "Using NHIS data, logistic regression was used to analyze temporal changes in national smoking prevalence and cigarettes smoked per day (among daily smokers) during 2005–2010, overall and by age, race/ethnicity, education, poverty status, and U.S. census region. These 6-year linear trend analyses were constructed using 2005 as the baseline to enable comparability with previous national trend estimates; results were adjusted for sex, age, and race/ethnicity, and the Wald test was used to determine statistical significance (defined as $p < 0.05$)." ***Example***: "During 2005–2009, the proportion of U.S. adults who were current cigarette smokers was 20.9% in 2005 and 20.6% in 2009, with **no significant difference** (Figure 1). **No significant changes** in current smoking prevalence for U.S. adults were observed during the 5-year period overall and for each of the four regions: Northeast, Midwest, South, or West ($p \geq 0.05$)."
National Health and Nutrition Examination Survey (NHANES)	Vital and Health Statistics *(similar to NSDUH's national reports)*	Dye et al. (2007)	***Method***: Pairwise comparison using *t* test was used. ***Example***: "Overall, the prevalence of dental caries in primary teeth (dft) **increased** from approximately 40% from 1988–1994 to 42% during 1999–2004." "The prevalence of dental caries in permanent teeth (DMFT) for youths has **decreased significantly** from approximately 25% in 1988–1994 to 21% in 1999–2004."
	Journal Article	Ford (2013)	***Method***: Tests for linear trends were conducted with orthogonal linear contrasts using the *t* test. ***Example***: "The mean predicted 10-year risk for CHD **decreased significantly** from 7.2% during 1999 to 2000 to 6.5% during 2009 to 2010 (Table 1). However, the mean risk for CVD **showed no significant improvement**. When stratified by age, the mean risk for CHD and CVD **decreased significantly** in the 3 oldest age groups." (CHD = coronary heart disease; CVD = cardiovascular disease)
	Data Brief[b]	Margaret et al. (2015)	***Method***: Linear survey trends were tested using orthogonal contrast matrices using the *t* test. ***Example***: "**Decreasing trends were noted** in the percentage of obese men with elevated triglyceride, from 48.0% for 2001–2004 to 38.7% for 2009–2012." "Among adults aged 60 and over from 2001–2004 to 2009–2012, the percentage with elevated triglyceride **declined** from 39.9% to 24.8% for men, and from 43.5% to 30.9% for women." "**Declines** in elevated triglyceride levels were observed in overweight and obese men and women between 2001–2004 and 2009–2012."

See notes at end of table. (continued)

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis (continued)

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
	Journal Article	Ford and Dietz (2013)	***Method***: Trends in energy intake were examined by using linear regression with the time specified as the midpoint of the NHANES series of surveys. They also examined nonlinear trends for time by adding a squared term (i.e., quadratic term) to the models. They examined the trends for both unadjusted mean energy intake and adjusted mean energy intake separately. If the quadratic term was found to be significant, the unadjusted or adjusted mean energy intake would be examined over time based on the time series graphs or tables to find the breaks in trends. ***Example***: "Mean energy intake adjusted for age, sex, race or ethnicity, educational status, and BMI **increased** by 314 kcal (95% CI: 259, 368 kcal) from 1971–1975 through 2003–2004 and then **decreased** by 74 kcal (95% CI: 21, 126 kcal) in subsequent years (Table 1). For the unadjusted and adjusted means, the signs (both for the period 1971–1975 to 2009–2010 and for 1999–2000 to 2009–2010) for the regression coefficients of the quadratic term for time were negative (data not shown) and the P values were significant, suggesting that the upward trend in energy intake before 2003–2004 had changed course and was decreasing in recent years."
	Journal Article	Will et al. (2014)	***Method***: The analysis of variance (ANOVA) type of Cochran–Mantel–Haenszel (CMH) test in Survey Data Analysis was used to test for trends in each study subgroup. (This approach is similar to a chi-square test; the caveat of this type of method is that it ignores the standard error of the rate estimated at each time point.) ***Example***: "**Neither** men nor women in this age category **showed statistically significant declines** in angina symptomatology over time. Prevalence rates for those aged ≥65 years ranged from 3% to 5% for men and 2% to 6% for women. However, among people in this age category, the decline in rates was significant from the first time period (1988–1994) to the most recent time period (2009–2012) for both men and women. The absolute rate for women aged ≥65 years dropped in half during the study period."
National Survey of Family Growth (NSFG)	Data Brief[b]	Branum and Jones (2015)	***Method***: "The significance of trends was tested using weighted least squares regression models to determine whether an apparent change over time was statistically significant, taking into account the standard error for each data point." ***Example***: "Use of long-acting reversible contraceptives (LARCs) **declined** between 1982 and 1988, **remained stable** through 2002, and then **increased nearly five-fold** in the last decade among women aged 15–44, from 1.5% in 2002 to 7.2% in 2011–2013." "**After remaining stable** between 1988 and 2006–2010, LARC use **increased** almost 10-fold among women with no previous births through the 2011–2013 time period."

See notes at end of table.

(continued)

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis (continued)

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
National Hospital Ambulatory Medical Care Survey (NHAMCS)	Journal Article	Jamal et al. (2015)	***Method***: "Multivariate logistic regression with orthogonal polynomials was used to analyze linear trends from 2005 through 2010, adjusted for sex, age, and race/ethnicity (α = .05)." ***Example***: "Current tobacco use **decreased** from 28.9% in 2005 to 22.6% in 2010 among hospital outpatient visits (P for trend <.001)." "From 2005 through 2010, cessation assistance **did not change over time** after adjusting for sex, age, and race/ethnicity (P for trend = .17)."
Pregnancy Risk Assessment Monitoring System (PRAMS)	*Morbidity and Mortality Weekly Report*	Tong et al. (2013)	***Method***: "Statistical linear trends were assessed using logistic regression with smoking as the outcome variable and the infant's birth year as the independent variable;" some other variables were also included in the model as control factors. ***Example***: "The percentage of smokers who quit smoking during pregnancy **decreased significantly** for one site (Louisiana [from 44.2% to 37.1%]) and **increased significantly** for four sites (Illinois [from 40.3% to 56.5%], Massachusetts [from 42.8% to 62.0%], Michigan [from 33.1% to 48.7%], and New Jersey [from 47.9% to 63.6%])."
National Youth Tobacco Survey (NYTS)	*Morbidity and Mortality Weekly Report*	Centers for Disease Control and Prevention (2012)	***Method***: "Logistic regression analysis was used to analyze temporal changes from 2000 to 2011 in current tobacco use, current combustible tobacco use, and current cigarette use, by school level, controlling for grade, race/ethnicity, and sex, and simultaneously assessed for linear and quadratic trends; a p value < 0.05 was considered statistically significant. Statistical software was used for all calculations to account for the complex survey design." ***Example***: "From 2000 to 2011, among middle school students, **significant linear downward trends** were observed for current tobacco use (14.9% to 7.1%), current combustible tobacco use (14.0% to 6.3%), and current cigarette use (10.7% to 4.3%) (Figure 1)."

See notes at end of table. (continued)

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis (continued)

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
Agency for Healthcare Research and Quality (AHRQ)			
Medical Expenditure Panel Survey (MEPS)	Statistical Brief *(Easy-to-read, concise graphical summaries of MEPS data)*	Stagnitti (2015)	***Method***: A pairwise comparison using the *t* test was used. ***Example***: "In 2002, MEPS estimates show that 27.2 million people purchased one or more outpatient prescribed opioids in the U.S. civilian noninstitutionalized population, and in 2012 the total number of people purchasing one or more prescribed outpatient opioid **increased** to 36.7 million (figure 1)." "When comparing 2002 and 2012, MEPS estimates **showed growth** in the total number of outpatient prescription purchases of opioids, **rising from** 85.9 million purchases to 143.9 million purchases, an increase of 67.5 percent (figure 2)." "When comparing 2002 with 2012, total out of pocket expense (in 2012 dollars) for outpatient prescribed opioids **decreased** from $2.3 billion to $1.6 billion, **a 30.4 percentage decrease** (figure 5)."
	Methods Report	Center for Financing, Access, and Cost Trends (2011)	***Method***: Some caveats were provided when doing trend testing with MEPS data. ***Example***: "Statistical significance tests should be conducted to assess the likelihood that observed trends may be attributable to sampling variation." "The length of time being analyzed should also be considered. In particular, large shifts in survey estimates over short periods of time (e.g. from one year to the next) that are statistically significant should be interpreted with caution, unless they are attributable to known factors such as changes in public policy, economic conditions, or MEPS survey methodology." "Looking at changes over longer periods of time can provide a more complete picture of underlying trends. Analysts may wish to consider using techniques to smooth or stabilize analyses of trends using MEPS data such as comparing pooled time periods (e.g. 1996-97 versus 2004-05), working with moving averages, or using modeling techniques with several consecutive years of MEPS data to test the fit of specified patterns over time." "Finally, researchers should be aware of the impact of multiple comparisons on Type I error. Without making appropriate allowance for multiple comparisons, undertaking numerous statistical significance tests of trends increases the likelihood of concluding that a change has taken place when one has not."

See notes at end of table. (continued)

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis (continued)

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
Healthcare Cost and Utilization Project (HCUP) (National Inpatient Sample [NIS])	Methods Report	Houchens et al. (2015)	***Method***: The use of hierarchical or multilevel regressions for trend testing was discussed. ***Example***: "In the context of NIS trend studies, discharges are nested within hospitals. Some hospitals are contained in multiple years of the NIS. Consequently, the nesting structure also could be characterized as discharges nested within years nested within hospitals (repeated measures on the same hospital)." "Hierarchical models account separately for the discharge-level error, the hospital-level error, and the correlation among discharges within hospitals." "Also, these models can account for serial correlation over time." "The sample weights and survey design elements should be incorporated in the modeling." "For example, hospital-level variation should be modeled separately from discharge-level variation, and hospital stratification variables can be included as independent variables for the hospital-level model."
Monitoring the Future Survey (MTF)	Key Findings Report *(similar to NSDUH national findings reports)*	Miech et al. (2015)	***Method***: Pairwise comparison using the *t* test was used. ***Example***: "**Declining use** of a number of licit and illicit substances is a main finding in 2014. Annual prevalence of drug use **declined** for 28 of the 34 drug outcomes reported for the combined pool of 8th, 10th, and 12th graders, shown in Table 2. Annual prevalence of using any illicit drug **decreased slightly, but not significantly,** in all three grades: by 0.6 (ns) percentage points in 8th grade, 2.1 (ns) percentage points in 10th, and 1.5 (ns) percentage points in 12th. For the three grades combined prevalence declined by 1.4 (ns) percentage points." "Overall increases in perceived risk and disapproval appear to have contributed to the downturn in cigarette use. Perceived risk **increased substantially and steadily** in all grades from 1995 through 2004, after which it **leveled** in 8th and 10th grades. However, it **continued rising** in 12th grade until 2006, after which it **leveled and then declined** some in 2008. Disapproval of smoking had been **rising steadily** in all grades since 1996. After 2004, **the rise decelerated** in the lower grades through 2006—again, reflecting a cohort effect in this attitude."

See notes at end of table.

(continued)

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis (continued)

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
Youth Risk Behavior Survey (YRBS)	Methods Report	Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Adolescent and School Health (2014, 2016)	***Method***: A three-step analysis was used: (1) regression analysis when all time variables (linear, quadratic, cubic, etc.) are treated as continuous and are created by coding each year with orthogonal coefficients; (2) using Joinpoint software; and (3) test segments. ***Example***: "The 2013 YRBS fact sheet has a column named as 'Change from 1999 to 2003' that describes statistically significant linear and quadratic changes over time from 1991 (or the first year in which the data were collected) through 2013 based on logistic regression analyses for each selected behavior. If there is a statistically significant linear trend, then this situation is described as 'Increased, 1991–2013' or 'Decreased, 1991-2013.' If there is a statistically significant quadratic trend, both parts of the quadratic trend are described. For example, 'Decreased, 1991–1999 and No change, 1999–2013.' If there is no statistically significant linear or quadratic trend then this situation is described as 'No change, 1991–2013.'"
Indian Health Service (HIS)	Journal Article	Sequist et al. (2011)	***Method***: "We analyzed trends in clinical performance measures within the IHS by fitting logistic regression models with the performance of the target measure as the dependent variable and year as the primary independent variable, after adjusting the standard errors for clustering of patients by health center using generalized estimating equations." ***Example***: "Clinical performance **improved significantly** from 2002 to 2006 for 10 of the 12 performance measures, including preventive services and chronic disease management (Table 2)." "Five-year trends in breast cancer and diabetic retinopathy screening rates **remained relatively flat** within Medicare, Medicaid, and the IHS."

See notes at end of table. (continued)

Table A.1 Summary of Documents under This Review That Discussed or Used Trend Analysis (continued)

Agency/Study	Product Type	Source	Methods and Verbatim Examples of Method Description and Result Interpretation
National Institute for Occupational Safety and Health (NIOSH)	Journal Article	Tiesman et al. (2015)	***Method***: The authors used autoregressive models with a first-order autoregressive error structure AR(1) to assess the trends of suicide rates while accounting for serial correlation. Both linear and quadratic time variables were tested through modeling to determine the best fit of the suicide data. ***Example***: "Between 2003 and 2010, **a significant quadratic trend** in workplace suicides was observed (p=0.035). Workplace suicides **decreased** between 2003 and 2007 and then **sharply increased** (Figure 1)."
Maternal and Child Health Bureau (MCHB)	Methods Report	Rosenberg (1997)	***Method***: The chi-square test for linear trend and regression analysis was used. Both methods are mainly for analyzing aggregated data with the observed series of rates or counts; thus, there is only one value at each time point, and the standard error of the rate or count at each time point is ignored. ***Example***: Chi-square test for linear trend: "The linear trend in these data is **not statistically significant according to this test** since the p-value is 0.2, not less than the customary cutoff of 0.05." Regression analysis: "Notice that modeling the 16 annual log transformed rates, the average annual % change is negative, **indicating a decreasing infant mortality rate**."

[a] National Health Statistics Reports released from the National Center for Health Statistics (NCHS) provide annual data summaries, present analyses of health topics, or present new information on methods or measurement issues.

[b] Data Briefs released from NCHS are statistical publications that provide information about current public health topics in a straightforward format. Each report takes a complex data subject and summarizes it in text and graphics that provide readers with easily comprehensible information in a compact publication.

Appendix B. Methods Investigated through Literature Review

This page intentionally left blank

Table B.1 Methods Investigated through Literature Review

Method	Research Questions Addressed by Method	Citation	Number of Observations	Type of Data
Detecting outliers in trends and time series	1. Is an observation in a time series an outlier? 2. How can one correctly identify level shifts from outliers? 3. How can one detect an unusual sequence of events in a time series? 4. How can one monitor a series in real time to detect an unusual measurement (exception reporting)?	Balke (1993)	100; 180	Environmental; Injury
		Choy (2001)	45; 100; 200	Engineering; Economic; Simulated
		Fox (1972)	100	Simulated
		Frei and Schär (2001)	34,310	Climatological
		Galeano, Peña, and Tsay (2006)	50–500; 57	Simulated; Economic
		Keogh, Lin, and Fu (2005)	5,000; 15,000	Engineering; Medical;
		Keogh, Lonardi, and Yuan-chi Chiu (2002)	800; 35,000	Simulated; Engineering
		Ljung (1993)	50–200	Simulated
		McCabe, Greenhalgh, Gettinby, Holmes, and Cowden (2003)	572	Medical
		Muirhead (1986)	110	Manufacturing
		Peña (1990)	39	Ecological
		Smith (1989)	119, 905	Environmental
		Tsay (1988)	216; 369; 274	Economic; Economic; Economic
		Tsay, Peña, and Pankratz (2000)	296; 184	Engineering; Economic
Nonparametric methods in time series	Making minimal assumptions about the distribution of the data: 1. Is there a sudden change in the mean at time t? 2. Is there a monotonic trend in the data? 3. What is the slope of the linear trend in the data?	Beharry, Clarke, and Kurmarsingh (2014)	110	Climatological
		Costa and Soares (2009)	21	Climatological
		Gocic and Trajkovic (2013)	360	Climatological
		Hall and Tajvidi (2000)	45; 84	Climatological
		Hess, Iyer, and Malm (2001)	1,148	Environmental
		Pettitt (1979)	27	Engineering
		Yue, Pilon, and Cavadias (2002)	22 to 84	Environmental
		Zhou, Zou, Zhang, and Wang (2009)	28	Simulated

(continued)

Table B.1 Methods Investigated through Literature Review (continued)

Method	Research Questions Addressed by Method	Citation	Number of Observations	Type of Data
Bayesian methods in time series	Using the Bayesian paradigm of inference: 1. Is an observation in a time series an outlier? 2. Is there a change in the mean or variance of a time series? 3. Is there one or more changes in the slope of the series (change point), and when did they occur?	Abraham and Box (1979)	70	Simulated
		Assareh, Smith, and Mengersen (2011)	1,971	Clinical
		Bernardo, Moreno, and Casella (2007)	18; 20; 100	Simulated; Simulated; Ecological
		Giordani and Kohn (2012)	212	Economic
		Lai and Xing (2011)	1,000; 113	Simulated; Injury
		McCulloch and Tsay (1993)	158; 250	Economic; Simulated
		Punskaya, Andrieu, Doucet, and Fitzgerald (2002)	500; 3,500	Simulated; Signal data
		Thum and Bhattacharya (2001)	9	Longitudinal educational data
Methods related to Joinpoint regression	Using various computation techniques (Joinpoint regression, smoothing splines, piecewise polynomials), these methods address the questions: 1. What is the long-term trend in the data? 2. Are there regular variations (seasonality) in the data? 3. Are there one or more changes in the slope of the series (change point), and when did they occur?	Kim, Fay, Feuer, and Midthune (2000)	27; 23	Simulated; Medical
		Eckley, Fearnhead, and Killick (2011)	200; 2,000	Simulated
		Frick, Munk, and Sieling (2014)	500; 200; 600	Simulated; Genomic; Engineering
		Matteson and James (2014)	150–900; 2,500; 262	Simulated; Genomic; Economic
		Moskvina and Zhigljavsky (2003)	144	Economic
		Reeves, Chen, Wang, Lund, and Lu (2007)	90–100	Climatological
		Samé, Chamroukhi, Govaert, and Aknin (2011)	60; 60	Simulated; Engineering
		Wang, Wen, and Wu (2007)	6–500; 336	Simulated; Climatological
		Dominici, McDermott, Zeger, and Samet (2002)	2,920	Environmental
		Dong and Roychowdhury (2003)	100	Engineering
		Huang and Shen (2004)	400; 177; 926	Simulated; Economic; Economic
		Lemire (2007)	600; 200	Medical; Economic
Scanning statistics	This method identifies whether there is a trend in the data using scanning statistics and controlling the false discovery rate.	Siegmund, Zhang, and Yakir (2011)	50,000; 16,500	Simulated; Genomic

(continued)

Table B.1 Methods Investigated through Literature Review (continued)

Method	Research Questions Addressed by Method	Citation	Number of Observations	Type of Data
Updating analyses in the presence of new data	Does a new set of observations indicate a change in the slope of the series (change point)?	Adams and MacKay (2007)	4,050; 780; 112	Engineering; Economic; Injury
		Fearnhead and Liu (2007)	500; 40,000	Simulated; Genomic
		Killick, Fearnhead, and Eckley (2012)	75,000	Environmental

This page intentionally left blank

Appendix C. Methodological Approaches Used by Other Statistical Agencies

This page intentionally left blank

Table C.1 Methodological Approaches Used by Other Statistical Agencies

CBHSQ	**Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services** (excluding NSDUH)		
Study	**Notes**		**Data Type/Miscellaneous Comments**

DAWN (Drug Abuse Warning Network)	• Comparisons of estimates for first year versus last year (i.e., 2004 vs. 2011)
	• Comparisons of estimate for recent years (2009 vs. 2011 and 2010 vs. 2011)
	• Statistical significance if difference had *p* value less than 0.05
	• Sometimes compared rates over time, but not in standard Excel files released on the website
		• Methods (cross-year comparisons, see Section 5.8): https://www.samhsa.gov/data/sites/default/files/DAWN2k11ED/DAWN2k11ED/rpts/DAWN2k11-Methods-Report.htm
		• Related weighted least squares for trends in emergency department (ED) visits:
		  ○ National Hospital Ambulatory Medical Care Survey (NHAMCS), National Center for Health Statistics (NCHS)
		  ○ https://www.cdc.gov/nchs/data/databriefs/db194.htm
			Emergency Department; Survey Sample

Mental health treatment service providers (examples from other programs)	• Nationwide Inpatient Sample (NIS) of the Healthcare Cost and Utilization Project (HCUP):
	  ○ conducted by the Agency for Healthcare Research and Quality (AHRQ)
	  ○ chi-square, pairwise comparisons, and regression tests
	  ○ https://jamanetwork.com/journals/jamapsychiatry/fullarticle/209961
		• National Institute of Mental Health (NIMH)/SAMHSA Patient Sample Surveys, National Reporting Program:
		  ○ Pairwise comparisons and regression (for controls)
		  ○ https://doi.org/10.1093/schbul/21.1.75
			Patient Surveys

TEDS (Treatment Episode Data Set) N-SSATS (National Survey of Substance Abuse Treatment Services)	• There is a variety of variables for 10-year span (i.e., 2004-2014). • Reports from DASIS (Drug & Alcohol Services Information System): ○ https://wwwdasis.samhsa.gov/dasis2/index.htm • TEDS data show that certain substance admission counts have changed throughout the years:	○ through a series of maps (charts) ○ initial year as base year and subsequent years indexed ○ Chapter 3 from the TEDS state reports	Administrative Records

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

BJS	**Bureau of Justice Statistics, U.S. Department of Justice**

Victimization Statistics Unit (VSU) National Crime Victimization Survey (NCVS)

- Crime is relatively rare, so oftentimes 2 or 3-year rolling averages are used.
- Comparisons are generally made between most recent year, 5 years prior, and 10 years prior.
- Concerns about reporting percent change over time and assessing substantive versus statistical significance:
  - Reports that discuss change over time generally focus on percent decline
  - Recently, Bureau of Justice Statistics (BJS) has stopped presenting percent change in the annual *Criminal Victimization Bulletin*. Due to the statistical rarity of crime, the base rate of victimization is also relatively low. This leads to small absolute changes in point estimates from 1 year to the next that translate into large percentage changes.
  - Percent change from year to year is currently addressed on a case-by-case basis.
- No current documentation or standards for guidance are available on how to examine trends.
- The NCVS Historical Trends project (NHT) is a multiyear project that integrates historical National Crime Survey (NCS) from 1973 to 1992 with the contemporary NCSV into a single data file. As part of this project, agency guidance on trend analysis is being planned as part of the project deliverables. This would include minimum number of time points for using the *t* test and regression.

Method	**Notes**	**Verbatim Examples**	**Source**
Pairwise comparison using the *t* test	• Generalized Variance Functions (GVF) are used variance estimation. • For trends, a *t* test is used to compare estimates for first and last year. • The *t* test may be between two end points or between an end point and all other points. • Statistical difference if *p* value is less than .05 are described as higher, lower, increased, or decreased. • *P* values of .10 are described as slightly, somewhat, or as some indication of differences, although the VSU has recently moved away from .01. • When no statistical difference is observed, terms such as "similar" and "stable" are used. • *P* values not an indication of data quality, so the BJS flags estimates with coefficient of variances greater than 50 per cent or that are based on 10 or fewer in the sample. • In the report discussion, there is some variation from one author to the next.	• Rates per 1,000: "The rate of serious intimate partner violence against females declined by 72% from 5.9 victimizations per 1,000 females age 12 or older in 1994 to 1.6 per 1,000 in 2011." • Counts: "The number of children living in households that experienced violent crime was about 6 million fewer in 2010 than in 1993." • Percentages: "No significant change was found in the percentage of violent crime reported to police from 2013 to 2014."	*Intimate Partner Violence: Attributes of Victimization, 1993-2011*: https://www.bjs.gov/index.cfm?ty=pbdetail&iid=4801 *Prevalence of Violent Crime among Households with Children, 1993-2010*: https://www.bjs.gov/index.cfm?ty=pbdetail&iid=4472 *Criminal Victimization, 2014*: https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5366 Additional examples: ○ https://www.bjs.gov/content/pub/pdf/fv9311.pdf ○ https://www.bjs.gov/content/pub/pdf/vcay9410.pdf ○ https://www.bjs.gov/content/pub/pdf/ndv0312.pdf

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

BJS	**Bureau of Justice Statistics, U.S. Department of Justice**		
Method	**Notes**	**Examples**	**Source**
Joinpoint Analysis (JPA)	• JPA has been examined as an analytical tool for the NHT project. • Using GVFs for 41+ years is not efficient and prone to error. • Direct estimation is not an option for the NHT due to absence of survey design variables prior to 1992. • The VSU is considering a hybrid approach that combines JPA and GVF. JPA will be used to identify underlying pattern and joinpoints in a trend. Pairwise comparison will then be used to test resulting line segments.		https://surveillance.cancer.gov/joinpoint

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

Corrections Statistical Unit (CU)

- The CU fields eight annual administrative data collections and multiple periodic administrative and in-person surveys to describe the populations, facilities, and operational systems of corrections in the United States, generally binned into four separate groups: probation, local jail, state or federal prison, and parole.
- Although the number of persons involved in the correctional system in any given year is small (about 6.9 million or 2.8 percent of all adults aged 18 or older in the United States in 2014), the number of persons cycling through the correctional system each year is very high (~2 million entries and exits to probation annually, ~12 million admissions and releases from local jails, ~600,000 admissions and releases from prisons, and ~460,000 entries and exits to parole). In addition, collateral consequences of involvement in the correctional system affect persons once they are released back into the community.
- Comparisons are generally made between most recent year and 10 years prior:
  - percent change from previous year to current year;
  - average annual change over 10 years;
  - testing for statistical significance generally only done on measures made periodically, or to measures on the general population; and
  - moving averages across multiple years to show very rare events.
- No current documentation or standards are available for guidance on how to examine trends.

Method	Notes	Examples	Source
The *t* test for pairwise comparison between two points in time or between populations		https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5219 https://www.bjs.gov/content/pub/pdf/urhuspj1112.pdf https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5500 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5479	- Survey of Inmates in State and Federal Adult Correctional Facilities - SILJ (Survey of Inmates in Local Jails) - NIS (National Inmate Survey) - ASJ (Annual Survey of Jails)
Average annual change	With the exception of the ASJ, these are complete enumerations of the correctional populations.	https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5519 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5415 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5414 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5387 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5299 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5602	- National Prisoner Statistics Program - Annual Survey of Jails - National Corrections Reporting Program - Annual Survey of Probation - Annual Survey of Parole - Census of Local Jail Facilities - Survey of Jails in Indian Country

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

Corrections Statistical Unit (CU)			
Method	**Notes**	**Examples**	**Source**
Percent change	With the exception of ASJ, these are complete enumerations of the correctional populations	https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5519 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5415 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5414 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5387 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5299 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5602 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5341	- National Prisoner Statistics program - Annual Survey of Jails - National Corrections Reporting Program - Annual Survey of Probation - Annual Survey of Parole - Census of Local Jail Facilities - Survey of Jails in Indian Country - Deaths in Custody Reporting Program
Visual presentation of trends (graphs, charts, tables of longitudinal data)	With the exception of ASJ, these are complete enumerations of the correctional populations	https://www.bjs.gov/content/pub/pdf/cp13st.pdf https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5519 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5415 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5414 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5387 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5299 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5602 https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5341	- Capital Punishment - National Prisoner Statistics Program - Annual Survey of Jails - National Corrections Reporting Program - Annual Survey of Probation - Annual Survey of Parole - Census of Local Jail Facilities - Survey of Jails in Indian Country - Deaths in Custody Reporting Program
Moving averages displaying 2- to 3- year trends	Moving averages calculated on cause of death.	https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5341	- Deaths in Custody Reporting Program

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

BTS	**Bureau of Transportation Statistics, U.S. Department of Transportation**		
Study	**Notes**		**Data Type**

CFS (Commodity Flow Survey)	• Maintaining comparability/tracking trends across years is difficult.
	• Survey only conducted every 5 years:
	  ○ Classifications and geographies get updated.
	  ○ Continuous process improvement.
	• Methods documentation educates and cautions users (19-22) about comparability of estimates.
		○ https://www.census.gov/econ/cfs/2012/2012%20CFS%20survey%20methodology%20(Feb%202016).pdf
		○ Geographic area changes, industry changes, mode changes (water-borne ships) and mileage calculations, routing software changes, commodity coding changes
		○ Application of noise infusion
		○ Sampling variability and nonresponse
			Economic/Business Surveys

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

Census	**U.S. Census Bureau, U.S. Department of Commerce**	
Study	**Notes**	**Data Type**

ACS (American Community Survey)	New year of data is compared with previous year's data. Analysts from substantive areas perform bulk of data review on the following:
	• SAS programs (compare year to year);
	• QC for data reasonableness;
	• outlier detection;
	• trend determination and evaluation of real trend versus problem with data;
	• education, income, etc. (subject matter experts); and
	• data review (methodologists).
	
	There are required review steps for subject matter analysts:
	• ACS has 16 different subject matter offices.
	• Analysts have review requirements.
	• There are standard ways to quickly review.
	• More formal accountability is required.
	
	Use automated tools:
	• web-based tools run SAS in background,
	• full distribution changes, and
	• year-to-year $\chi 2$ tests and z tests.
	
	Know subject area and geography:
	• 3.5 million addresses,
	• organized by geographic levels,
	• nation and state (set thresholds for significance $\chi 2 > 95\%$),
	• county and place (set tighter thresholds for significance $\chi 2 > 99\%$),
	• same done for point estimates, and derived estimates such as %, ratios, or medians:
	  ○ nation/state - $\lvert z \rvert > 1.96$ and
	  ○ county - $\lvert z \rvert > 2.58$.
	
	Visual tool for mapping include
	• drill down to lower levels of geography (i.e., county and place);
	• how to get geographic context;
	• unweighted sample sizes;
	• mode information, such as mail computer-assisted telephone interviewing (CATI), computer-assisted personal interviewing (CAPI), the Internet; and
	• group-quarters information (composition).
	
	Trend information incudes
	• compare current and previous 4 years with some statistics and
	• sometimes bounce up and down.
	
	Require reviewers (analysts) to look at similar Census Bureau estimates
	• at least at national and state level, and
	• some also look at other federal surveys.
	
	When there are large weights,
	• subject matter staff are not responsible for the weights;
	• look at unweighted and weighted data; and
	• if the weighted data seem off (i.e., somehow a large weight fell through the process):
	  ○ analysts can send data back to the weighting staff;
	  ○ also look at sample size and group quarters population; and
	  ○ see if methodological changes are found.
		Household/Residential Survey

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

EIA	**Energy Information Administration, U.S. Department of Energy**		
	Notes		**Data Type**

	• EIA would like to do "on demand" data trends and is working towards it.
	• They publish some time series trends:
	  ○ Residential Energy Consumption Survey (RECS) and
	  ○ Commercial Buildings Energy Consumption Survey (CBECS).
	• Key items are tracked and published in the Annual Energy Review (AER):
		○ challenge in establishing a baseline year.
		• EIA receives interest from press office for historical context for journalists.
		• EIA is concerned that measurement process changes may confound trends.
			Energy Production and Consumption

NASS	**National Agricultural Statistics Service, U.S. Department of Agriculture**		
	Notes		**Data Type**

	Track record publication:
	https://www.nass.usda.gov/Publications/Track_Records/
	
	Focus is on retrospective assessment:
	• different from trends and
	• something like performance assessment of past estimates (board process).
	
	Root-mean-square-deviation (RMSE) is used for production forecasts during season to final estimates:
	• rolling 20-year window at national level only;
	• reliability measures on pp. 47–48 of August 2015 report at https://www.usda.gov/nass/PUBS/TODAYRPT/crop0815.pdf;
	• model-based yield forecasts are made;
	• linear trend term in synthetic regression portion;
		• used for forecast rather than for interpreting slope parameter to assess historic trend;
		• had issues in drought year (2012);
		• results internal, but method presented at conferences: https://www.amstat.org/meetings/ices/2012/AbstractDetails.cfm?AbstractID=302190.
		
		Board Process:
		• Plot survey indications over time.
		  ○ If large jump, look for evidence to support decrease/increase.
		  ○ If not, typically smooth it down/up.
		• New data (survey, census, administrative data):
		  ○ reevaluate (revision process) and
		  ○ typically administrative data (USDA Farm Service Agency [FAS], slaughter, cotton ginnings, etc.).
			Establishment/Economic Surveys; Agricultural Production

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

NCES	**National Center for Education Statistics, U.S. Department of Education**		
	Notes		**Data Type**
	Studies and collections that focus on trends: • Condition of Education. • Focus on trends: a. National Household Education Survey (NHES), b. National Assessment of Educational Progress (NAEP), and c. Schools and Staffing Survey (SASS) / National Teacher and Principal Survey (NTPS). • Comparison across cohorts: B&B, BPS, and ECLS: a. High School and Beyond (HS&B), b. National Education Longitudinal Study of 1988 (NELS), c. Education Longitudinal Study of 2002 (ELS), d. High School Longitudinal Study of 2009 (HSLS), e. Baccalaureate and Beyond Longitudinal Study (B&B), f. Beginning Postsecondary Students Longitudinal Study (BPS), and g. Early Childhood Longitudinal Study (ECLS). Analysis: • Most reports are descriptive in nature for analysis of data over time. • Regression analysis is used to support an assertion of trends: a. support a statement about an overall pattern of increase or decrease. • The *t* test is used to compare two specific time points: a. may be between a two end points	b. may be between an end point and all others. i. For example, NAEP compares the most recent point to each prior point in time. Reports that typically show key trends: • Condition of Education and Digest of Education Statistics. • Other regularly release compendia reports: a. Indicators of School Crime and Safety, b. NAEP (National Assessment of Educational Progress) release reports, and c. Contributions to OECD (Organisation for Economic Co-operation and Development) reports. • Trend estimates in study-specific reports: a. Search "Trend" on publication site: https://nces.ed.gov/pubsearch/. • Online analysis tools: a. https://nces.ed.gov/nationsreportcard/naepdata b. https://nces.ed.gov/surveys/international/ide/ Statistical Approaches: • Cross-sectional repeating surveys a. Use *t* test to compare point to point if not many time points. b. Regression is used to study average slope direction and magnitude: i. if significant changes (data collection, weighting, etc.), ii. categorical variable to identify variance associated with that change, and iii. analyses most often done with Current Population Survey (CPS) data because they have a lot of time points.	Variety of Surveys (longitudinal and cross-sectional)

(continued)

Table C.1 Methodological Approaches Used by Other Statistical Agencies (continued)

NCHS	**National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services**		
	Notes		**Data Type**
Working Group	No survey specific guidelines are available on trend analysis. Major issues include: • choosing the time period to include in a trend analysis and providing the rationale, • using all time points or just the beginning and ending time points to assess a trend, • pooling data across years or cycles, • conducting time trend analyses of survey data, • assessing a trend when there are three time points, • assessing a trend when there are four or more time points, • trend analyses of binary outcome variables, • trend analyses with covariates, • locating joinpoints at observed time points or between them, • trend analysis using Joinpoint trend analysis software, and • using the Cochran-Mantel-Haenszel test of trends. Comments: • Relying on three common approaches: pairwise comparisons, polynomial regression, and Joinpoint:	○ suggest pairwise limited to comparisons based on only three time points and ○ use trend lines (polynomial or joinpoint) for longer • Distinction between vital statistics and survey data for types of analysis. • Not provided formal advice for multiple testing: ○ suggest mention issues in reports, but not correct for it, and ○ lack of software to implement corrections is an issue. • Detailed section for where to start and how many points to include, which is based on being defensible and transparent rather than on formal statistical properties: ○ rationale needed for choice of beginning and end point: ▪ end point is simply most recent point, ○ data availability, ○ data comparability, ○ external events, ○ prior research, ○ recent or long-term trend, and ○ sensitivity of starting time point (may need to try multiple points).	Survey data and vital statistics