

Youth Risk Behavior Surveillance System (YRBSS)



# Conducting Trend Analyses of YRBS Data

August 2013

---

Where can I get more information? Visit [www.cdc.gov/yrbss](http://www.cdc.gov/yrbss) or call 800-CDC-INFO (800-232-4636).

---



National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention  
Division of Adolescent and School Health



### Introduction

**Purpose** The purpose of this document is to describe a method for conducting trend analyses of Youth Risk Behavior Survey (YRBS) data to identify and describe changes in the prevalence of risk behaviors over time.

---

**Contents** Introduction  
Purpose  
Contents  
Background  
Data Requirements  
Software Requirements

Analysis: 3 steps  
Step 1: Conduct regression analysis  
Step 2: Use Joinpoint software to determine critical values  
Step 3: Test segments  
Example

References

---

**Background** National, state, territorial, tribal, and local YRBS's are periodically conducted using similar sampling methods, survey administration procedures, and questionnaires to produce a series of independently selected cross-sectional data sets which can be compared across years. Trend analyses using YRBS data allow analysts to identify changes in the prevalence of risk behaviors over time, and to describe that change. This document assumes the reader has a working knowledge of YRBS data as well as a thorough knowledge of complex survey data and analysis, statistical software programming, regression analysis, and trend analysis using Joinpoint software from the National Cancer Institute (NCI). This document is not intended to teach statistical methodology.

Readers unfamiliar with YRBS data should review:

- the latest [national data documentation](#),
  - “[Methodology of the Youth Risk Behavior Surveillance System—2013](#)”,
  - “[Software for Analysis of YRBS Data](#)”,
  - “[Combining YRBS Data Across Years and Sites](#)”, and the
  - [Frequently Asked Questions](#) page on the YRBS website.
-

---

## Conducting Trend Analyses of YRBS Data

---

**Data Requirements** The first step in conducting this type of trend analysis is to assemble a trend data set by combining multiple YRBS data sets. In general, a trend data set should be created by combining (concatenating) data sets from surveys for the time period of interest. For example, a trend analysis to determine progress in achieving Healthy People 2010 objectives might include national YRBS data from the years 1999 through 2009. In addition, the dataset must include a variable (e.g., YEAR) that identifies the survey year for each observation. Finally, the dependent variables of interest (i.e., the risk behaviors to be tested for trend) must be based on questions and response options consistently worded across all the survey years included in the trend analysis. Important guidance in combining YRBS data sets is available in the document “[How to Combine YRBS Data Across Years and Sites.](#)”

---

**Software Requirements** The Division of Adolescent and School Health (DASH) conducts all complex survey analyses in [SAS-callable SUDAAN](#) (Research Triangle Institute, Research Triangle Park, NC). [While other statistical software designed to analyze complex survey data will suffice](#), this document describes trend analysis in the context of SAS-callable SUDAAN. In addition, [Joinpoint software](#), a free trend analysis program offered through NCI, is used when there are significant non-linear (quadratic, cubic, etc.) trends to identify the year(s) where the trend changes. More information about Joinpoint, as well as the software download information, can be found on the [NCI website](#).

---

### 3-Step Analysis

#### **Step 1: Conduct Regression Analysis**

Trend analyses for dichotomous risk behaviors, such as current smoking, are conducted using logistic regression. Similar analyses of continuous risk behaviors are conducted using linear regression. The logistic regression model used by DASH regresses the risk behavior (dependent variable) on continuous linear and non-linear time variables. In addition, models typically control for sex, race/ethnicity, and grade in school using categorical variables. Testing for linear and non-linear trends is accomplished using the following method:

- 1) Test for linear trends using a model that contains only a linear time variable (plus variables controlling for sex, race/ethnicity, and grade).
- 2) Test for quadratic trends by re-running the model with both linear and quadratic time variables.
- 3) Test for cubic trends by re-running the model a third time and include linear, quadratic, and cubic time variables.

In each case, only the highest-order time variable in the model is valid and can be accurately interpreted. All time variables (linear, quadratic, cubic, etc.) are treated as continuous and are created by coding each year with orthogonal coefficients calculated using PROC IML in SAS (see example SAS code at the end of this document).

If the p-value for the linear time variable is less than the *a priori* significance level (DASH typically uses  $\alpha=0.05$ ), then there is evidence of a linear change. If the associated beta for the significant linear time variable is negative (i.e., less than 0), there is evidence of a linear decrease. Similarly, if the associated beta is positive (i.e., greater than 0), there is evidence of a linear increase.

If the p-value for the quadratic time variable is less than the *a priori* significance level (DASH typically uses  $\alpha=0.05$ ), then there is evidence of a quadratic change. When quadratic changes are detected, the next step is to calculate the adjusted (e.g., for sex, race/ethnicity, and grade) prevalence and standard error by year, and then export these values into Joinpoint software to determine the critical year(s) or “joinpoints” where the non-linear trends change.

---

---

## Conducting Trend Analyses of YRBS Data

---

**Step 2: Use Joinpoint software to determine critical values (for significant quadratic or cubic time components)**

When a significant quadratic change has been detected, the next step is to calculate the adjusted prevalence (predicted marginal) and associated standard error (SE) for each year in the analysis. In SUDAAN, this is easily done by adding the “PREDMARG” statement to the RLOGIST procedure code. Sample SUDAAN code for obtaining adjusted prevalence can be found in Bieler, et al.,<sup>1</sup> a paper on estimating model-adjusted risks and risk ratios in complex survey data. The year, adjusted prevalence estimate, and SE can be exported to a tab-delimited file and then imported into Joinpoint software.

Once values have been imported into Joinpoint, use Joinpoint to determine the location of 1 joinpoint for a significant quadratic or 2 joinpoints for a significant cubic trend. The last analytical step is to return to SUDAAN to test each line segment (i.e., before and after each joinpoint) for linearity.

---

**Step 3: Test segments**

After determining the joinpoint(s) for significant non-linear trends (e.g., quadratic, cubic, etc.) the resulting line segments are tested for linearity in SUDAAN. To test both sides of a significant quadratic trend with a single joinpoint, two data sets are created – the first includes all years from the first available year up to and including the joinpoint year; the second includes all years beginning with the joinpoint year up to the last available year. For example, if a trend analysis covers years 1991 to 2011 and the Joinpoint is determined to be 1999, the first data set would contain 1991, 1993, 1995, 1997, and 1999 and the second data set would contain 1999, 2001, 2003, 2005, 2007, 2009, and 2011.

With each data set, run a logistic regression analysis containing a linear, but not a higher order (e.g., quadratic or cubic) time component. If the linear component is significant at the *a priori* level, this indicates either a significant decrease or a significant increase for that time period. If the associated beta for the significant linear time component is negative (i.e., less than 0), there is evidence of a linear decrease. Similarly, if the associated beta is positive (i.e., greater than 0), there is evidence of a linear increase.

---

---

## Conducting Trend Analyses of YRBS Data

---

### Example

Testing line segments for linearity to the left and right of a joinpoint (1999) in the event of a significant quadratic trend (QN29: Ever smoked a cigarette)

Year	1991	1993	1995	1997	1999	2001	2003	2005	2007	2009	2011
Prevalence	70.1%	69.5%	71.3%	70.2%	70.4%	63.9%	58.4%	54.3%	50.3%	46.3%	44.7%

/\*Produce orthogonal coefficients for linear trend testing to the left and right of the joinpoint year 1999\*/

```
PROC IML; X={1991 1993 1995 1997 1999}; XP=ORPOL(X,1); PRINT XP; RUN; QUIT;
PROC IML; X={1999 2001 2003 2005 2007 2009 2011}; XP=ORPOL(X,1); PRINT XP; RUN; QUIT;
```

```
DATA VARSET; SET DATA1.TRENDWK; /* Testing for linearity to the left of the joinpoint. */
IF (YEAR=1991 OR YEAR=1993 OR YEAR=1995 OR YEAR=1997 OR YEAR=1999);
```

```
IF YEAR=1991 THEN T5L_L=-0.632456;
ELSE IF YEAR=1993 THEN T5L_L=-0.316228;
ELSE IF YEAR=1995 THEN T5L_L=8.571E-21;
ELSE IF YEAR=1997 THEN T5L_L=0.3162278;
ELSE IF YEAR=1999 THEN T5L_L=0.6324555;
```

```
PROC RLOGIST DATA=WORK.VARSET DESIGN=WR FILETYPE=SAS;
NEST SURVYEAR STRATUM PSU/PSULEV=3 MISSUNIT;
WEIGHT WEIGHT; SUBPOPN TOTALPOP=1;
CLASS SEX RACE GRADE YEAR SURVYEAR;
MODEL QL29 = SEX RACE GRADE T5L_L;
PRINT/BETAFMT=F8.5 SEBETAFMT=F8.5 P_BETAFMT=F8.5;
RUN;
```

**In this example, T5L\_L had a p-value=0.52261 and beta=0.03704. Therefore, there was “no significant change in the prevalence of ever smoking a cigarette during 1991-1999.”**

```
DATA VARSET; SET DATA1.TRENDWK; /* Testing for linearity to the right of the joinpoint. */
IF (YEAR=1999 OR YEAR=2001 OR YEAR=2003 OR YEAR=2005 OR YEAR=2007 OR
YEAR=2009 OR YEAR=2011);
```

```
IF YEAR=1999 THEN T7L_R=-0.566947;
ELSE IF YEAR=2001 THEN T7L_R=-0.377964;
ELSE IF YEAR=2003 THEN T7L_R=-0.188982;
ELSE IF YEAR=2005 THEN T7L_R=0;
ELSE IF YEAR=2007 THEN T7L_R=0.1889822;
ELSE IF YEAR=2009 THEN T7L_R=0.3779645;
ELSE IF YEAR=2011 THEN T7L_R=0.5669467;
```

```
PROC RLOGIST DATA=WORK.VARSET DESIGN=WR FILETYPE=SAS;
NEST SURVYEAR STRATUM PSU/PSULEV=3 MISSUNIT;
WEIGHT WEIGHT; SUBPOPN TOTALPOP=1;
CLASS SEX RACE GRADE YEAR SURVYEAR;
MODEL QL29 = SEX RACE GRADE T7L_R;
PRINT/BETAFMT=F8.5 SEBETAFMT=F8.5 P_BETAFMT=F8.5;
RUN;
```

**In this example, T7L\_R had a p-value<0.0001 and beta=-0.99165. Therefore, there was a “significant linear decrease in the prevalence of ever smoking a cigarette during 1999-2011.”**

### References

- 1) Bieler GS, Brown GG, Williams RL, Brogan DJ. Estimating model-adjusted risks, risk differences, and risk ratios from complex survey data. *American Journal of Epidemiology* 2010;171(5):618-23.
-