



# Comparing Generalized Variance Functions to Direct Variance Estimation for the National Crime Victimization Survey

Bonnie Shook-Sa, Lance Couzens, Rick Williams,  
David Heller, and Marcus Berzofsky

2013 FCSM  
Washington, DC  
November 5, 2013

## Acknowledgement

- This research was funded through a cooperative agreement with the Bureau of Justice Statistics, U.S. Department of Justice.
- The views and ideas expressed here are those of the authors and do not reflect the official ideas or position of the Bureau of Justice Statistics or the U.S. Department of Justice.

# Overview

- National Crime Victimization Survey (NCVS) design and implementation
- Current NCVS variance estimation method
- Direct variance estimation for the NCVS
  - Motivation
  - Evaluation
  - Results
  - Conclusions

## Introduction to the NCVS

- Conducted since 1973
- Sponsored by the Bureau of Justice Statistics (BJS)
- Nationally representative sample of approximately 40,000 households and 75,000 persons per year
- Provides estimates of the frequency and characteristics of criminal victimization in the United States

## NCVS Design

- Target population is the civilian, noninstitutionalized population
- Stratified, four-stage sample design
  - Primary Sampling Units (PSUs) - large metro areas, counties, or groups of counties
  - Segments within PSUs
  - Households within segments
  - All persons 12 and older within households
- Panel survey - each sampled household is interviewed once every 6 months over a 3-year period
- Instruments collect household, person, and crime incident-level data

## NCVS Estimation

- Three analysis files and sets of weights
  - Household-level file
  - Person-level file
  - Incident-level file
- Used for the calculation of key estimates
  - *Victimization totals* - estimated number of criminal victimizations
  - *Victimization rates* - victimizations per 1,000 persons (for personal crimes) or households (for property crimes)

# NCVS Variance Estimation

- Must account for complex sample design
- Must account for person, household, and victimization weights, as appropriate
- Two options
  - Generalized Variance Functions (GVFs)
  - Direct Variance Estimation
    - Taylor Series Linearization (TSL)
    - Balanced Repeated Replication (BRR)

# Generalized Variance Functions (GVFs)

- Current method used for NCVS estimation
- GVFs express variances as functions of survey estimates and pre-specified parameters from models fit to the direct variance estimates for similar survey outcomes
- Allow users to calculate design-consistent variance estimates without knowledge of the sample design
- Provide approximate standard errors for numerous survey estimates



## Why Direct Variance Estimation?

- GVEs can be cumbersome
- GVEs do not allow for complex analyses (e.g. regression modeling)
- Accuracy of GVEs for estimate types not included in developing parameters is unknown
- Direct estimation using existing software packages (e.g. SAS, SUDAAN, SPSS, Stata, and R) would remove these issues

# Challenges to Direct Variance Estimation

- Specifying the NCVS design
  - Full set of sample design variables are not publically available
  - Public Use Files (PUFs) contain “pseudostrata” and “half sample” codes to approximate the design
  - Single- and pooled-year estimates
- Cannot calculate direct variance estimates with the current NCVS file structure

## Direct Variance Estimation Evaluation

- Specified NCVS design based on publicly available data
- Developed BRR weights for single-year estimates
- Prepared NCVS files to accommodate direct variance estimation
- Compared TSL, BRR, and GVF estimates for single years
- Selected a single direct variance method
- Compared selected method to GVFs for pooled-year estimates, cross single-year comparisons, and cross pooled-year comparisons

## Specifying the NCVS design

- Account for redesigns by creating a “year group” stratification variable: 1992-1996, 1997-2005, and 2006-2011
- PSUs within year groups treated as the same PSUs, PSUs across year groups treated as independent
- TSL
  - Strata are year groups and pseudostrata
  - PSUs are half samples
- BRR
  - Created replicate weights for single-year estimation
  - Based on same design as TSL

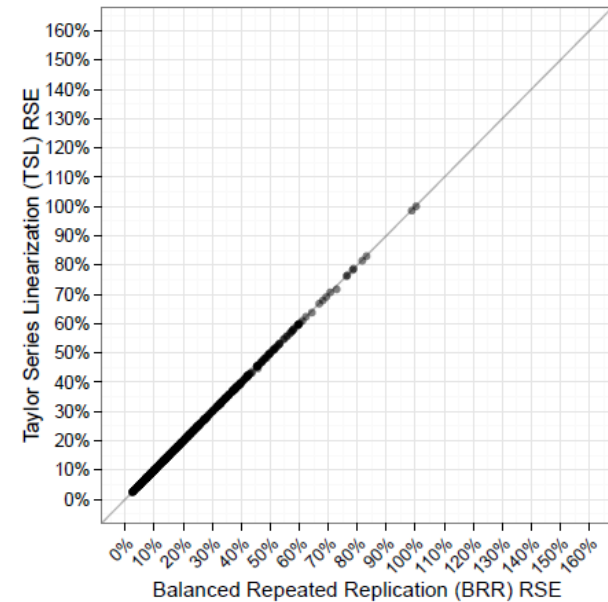
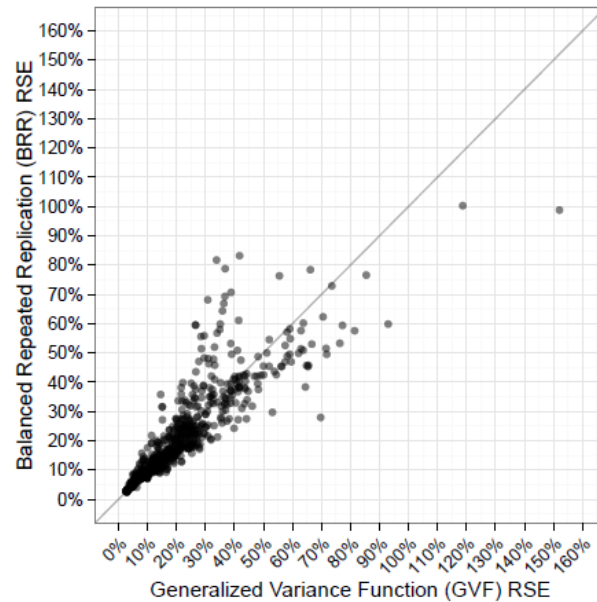
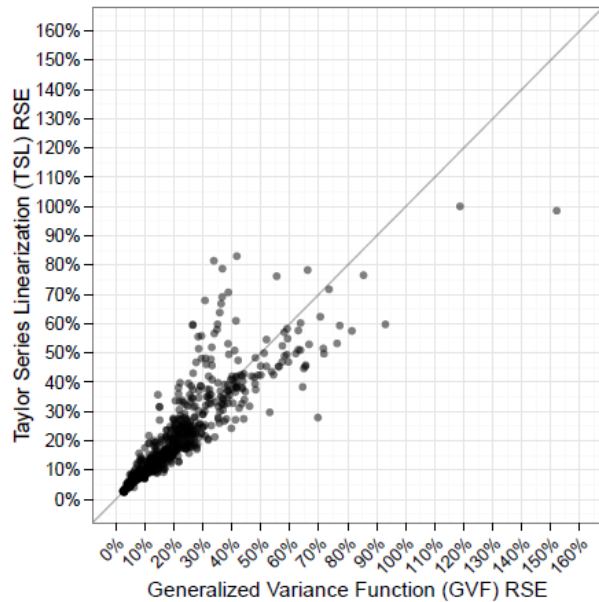
## Preparing NCVS files for direct variance estimation

- Victimization totals are calculated based only on the incident-level file and the victimization weight
  - File only contains data for reported crimes
  - Add “dummy records” for missed PSUs
- Victimization rates are calculated from multiple files using multiple weights
  - Household and incident files for property crimes
  - Person and incident files for personal crimes
  - Move incident counts and victimization weights from the incident file to the household and person files
  - Victimization weights must be parsed into their components and applied to estimates as appropriate

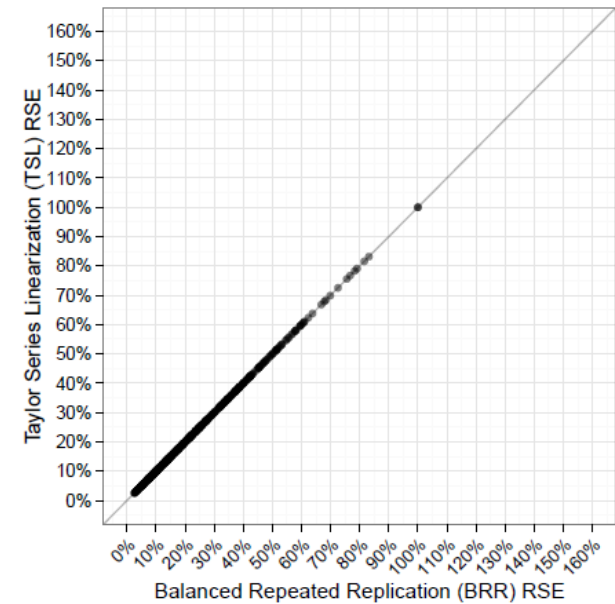
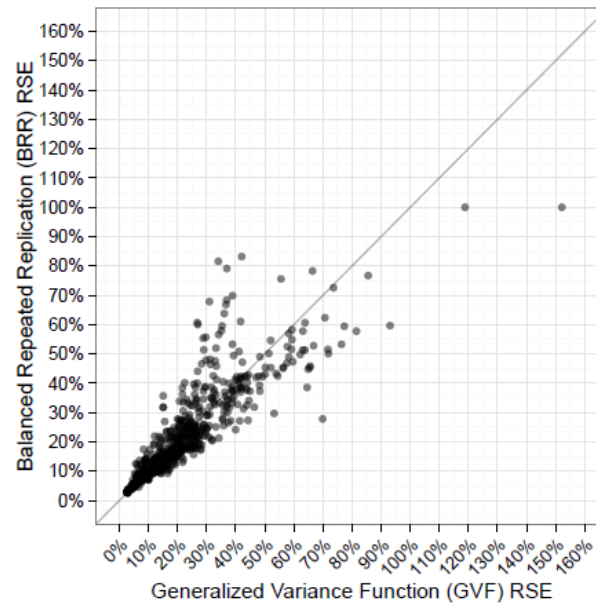
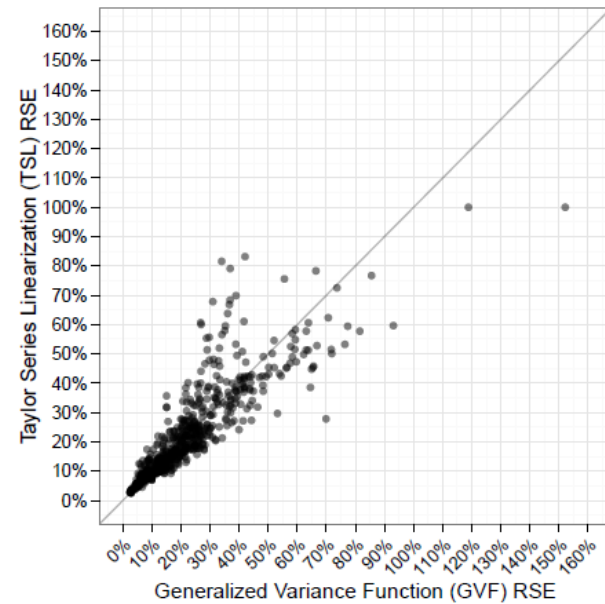
## Compare TSL, BRR, and GVF for single-year estimates

- Calculated crime victimization totals and rates for various crime types and domains of interest
- Yearly estimates for 2008-2011
- Calculated Percent Relative Standard Errors (RSEs) of each estimate
- Compared across estimation methods

# Compare TSL, BRR, and GVF for single-year estimates (Rates)



# Compare TSL, BRR, and GVF for single-year estimates (Totals)





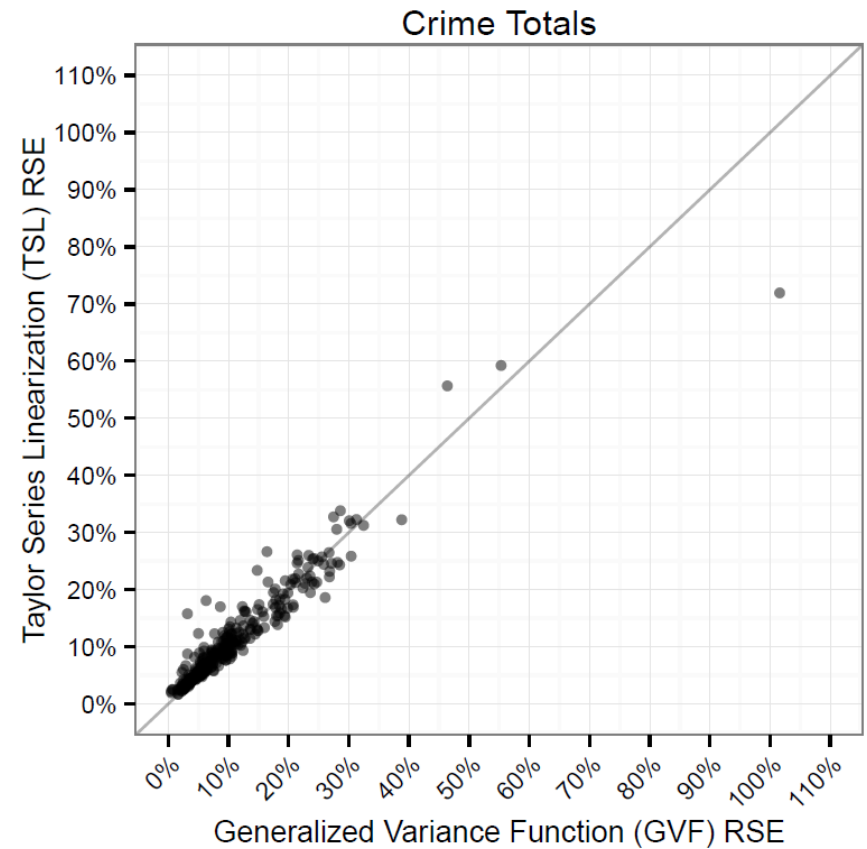
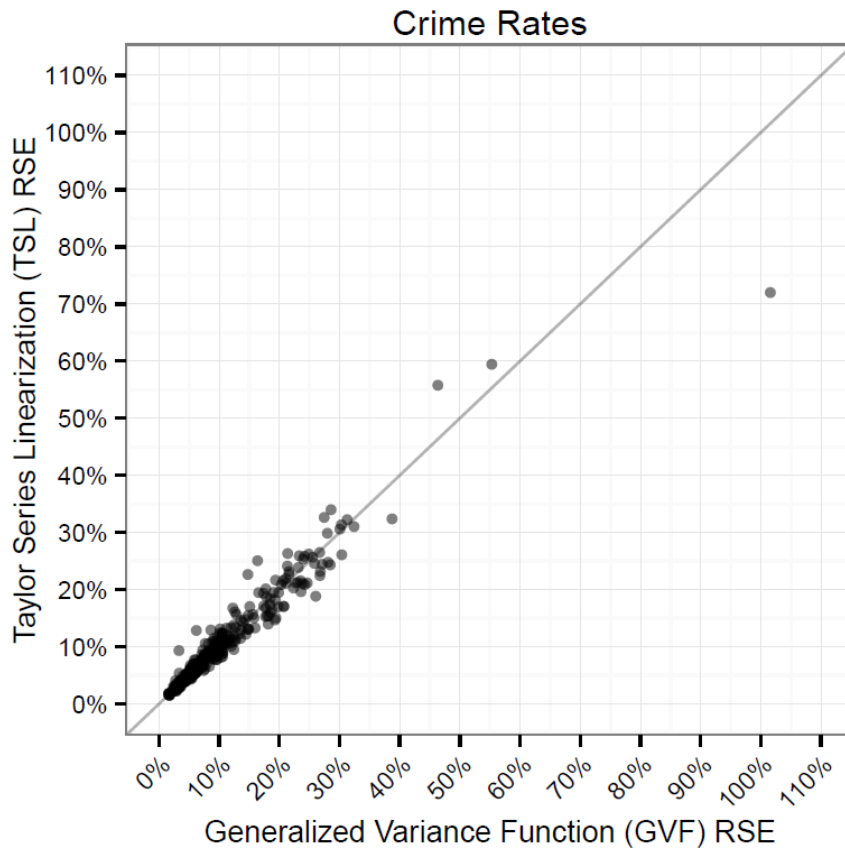
## Conclusions from Single-Year Comparisons

- Both TSL and BRR track very well with GVF estimates
- TSL and BRR are nearly the same, use TSL
  - BRR is more complicated and computationally expensive for pooled-year estimates
  - BRR is not available in the most commonly used software package for NCVS analysts (SPSS)

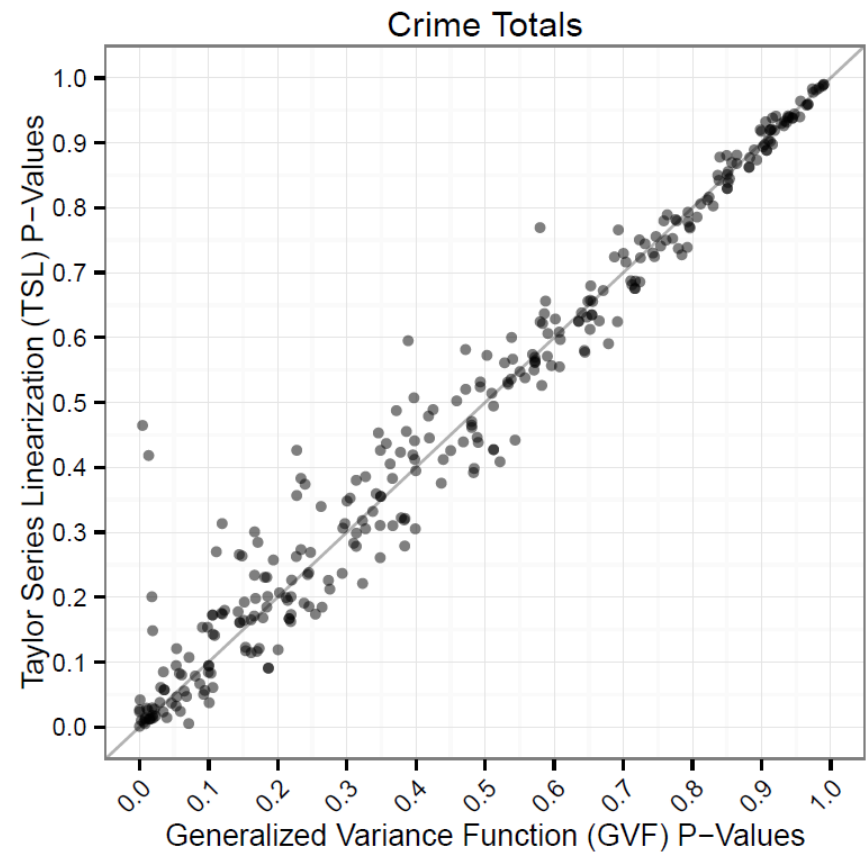
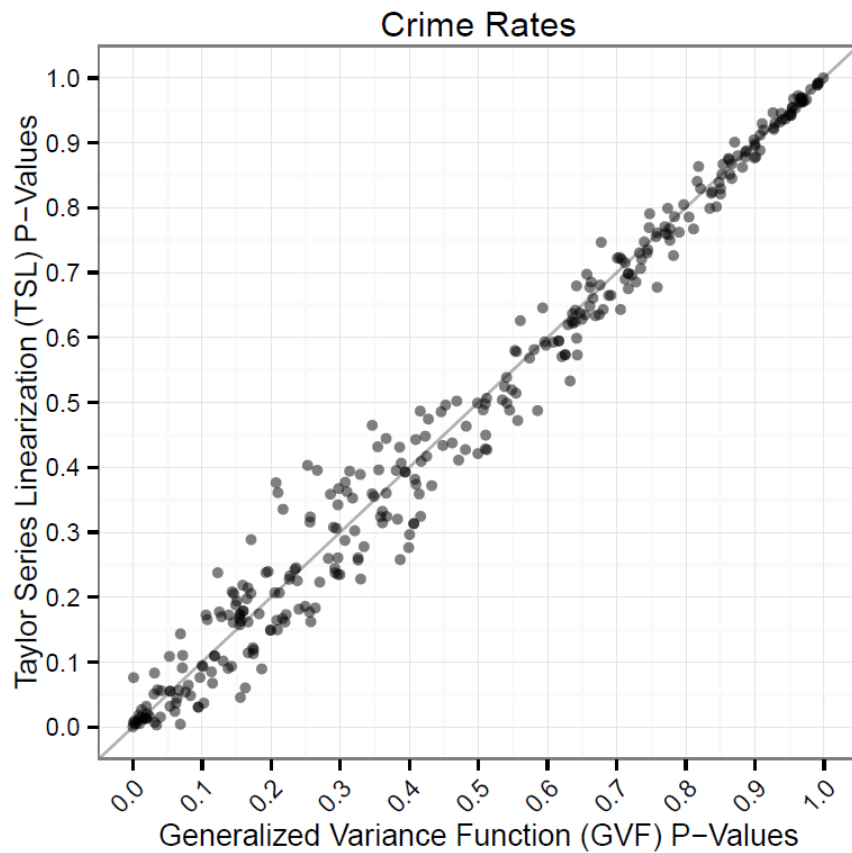
## Pooled-Year Estimates and Cross-Year Comparisons

- Pooled-year estimates
  - 2002-2004 and 2005-2007
  - Percent RSEs for GVF's vs. TSL
- Cross single-year comparisons
  - 2004 vs. 2005 and 2005 vs. 2006
  - P-values for GVF's vs. TSL
- Cross pooled-year comparisons
  - 2002-2004 vs. 2005-2007
  - P-values for GVF's vs. TSL
- Victimization rates and totals

# Comparison of TSL and GVF for pooled year estimates – RSEs

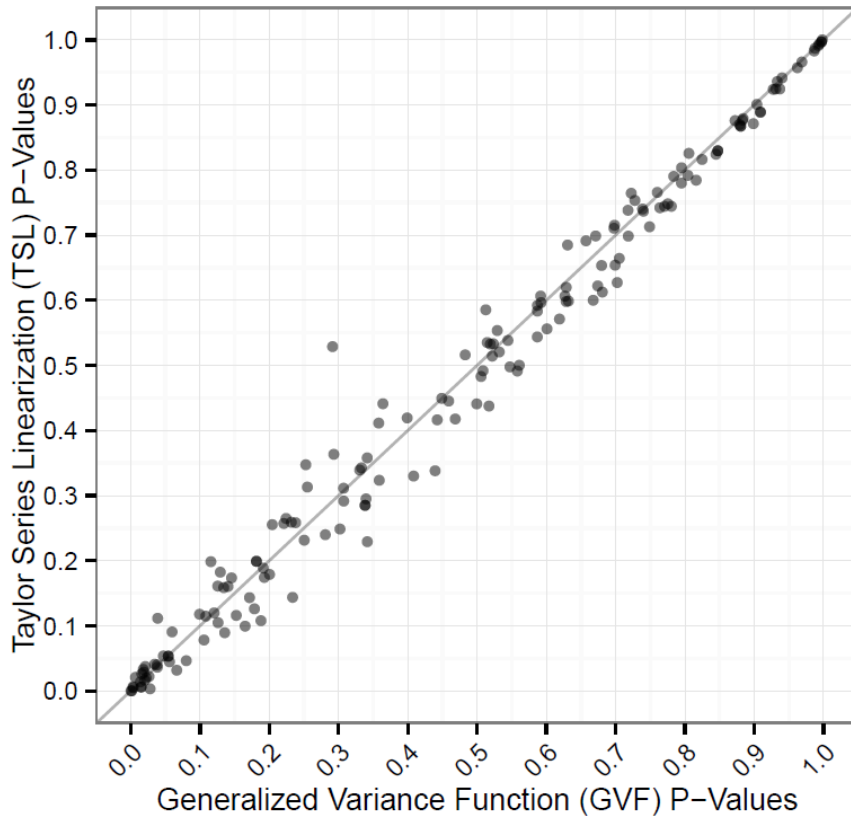


# Comparison of TSL and GVF for cross single-year comparisons – p-values

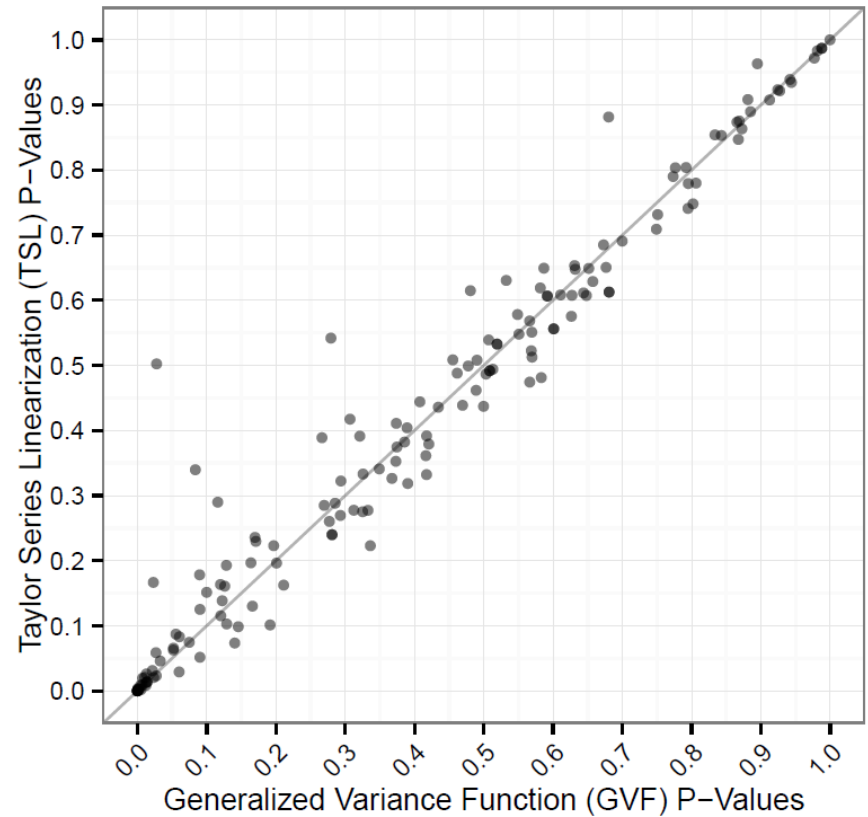


# Comparison of TSL and GVF for cross pooled-year comparisons – p-values

Crime Rates



Crime Totals



## Conclusions

- Direct variance techniques can be utilized for the NCVS based on publicly-available data
- Data manipulation is needed to prepare NCVS files for direct variance estimation
- Direct variance estimates tracked well with GVF estimates for the crime types and domains analyzed
  - Single and pooled year estimates
  - Single and pooled cross-year comparisons

# Contact

**Bonnie Shook-Sa**

*Research Statistician*

*RTI International*

[bshooksa@rti.org](mailto:bshooksa@rti.org)